جامعة حلوان
كلية التجارة وإدارة الأعمال

الإحصاء التطبيقي

الجزء الثاني

الاستدلال الإحصائي

الطبعة الثالثة

الدكتورة
عفاف على حسن الدش
أستاذ الإحصاء
ووكيل كلية التجارة وإدارة الأعمال
للدراسات العليا والبحوث - جامعة حلوان

الناشر
جهاز نشر وتوزيع الكتاب الجامعى
جامعة حلوان

٢٠٠٦

كلية التجارة وإدارة الأعمال

الإحصاء التطبيقي

الجزء الثاني

الاستدلال الإحصائي

الطبعة الثالثة

الدكتورة

عفاف على حسن الدش

أستاذ الإحصاء

ووكيل كلية التجارة للدراسات العليا والبحوث

جامعة حلوان

الناشر

جهاز نشر وتوزيع الكتاب الجامعي

جامعة حلوان

٢٠٠٦

الإحصاء التطبيقي

الجزء الثاني

الاستدلال الإحصائي

الطبعة الثالثة

الدكتورة

عفاف على حسن الدش

أستاذ الإحصاء

ووكيل كلية التجارة للدراسات العليا والبحوث جامعة حلوان

الطبعة الأولى :	جماد ثاني سنة ١٤٢٢ هـ - ٢٠٠١م رقم الإيداع :٢٠٠١/١٣٦٧٢م الترقيم الدولي : **I.S.B.N.977-32-6** الناشر : جهاز نشر وتوزيع الجامعي
الطبعة الثانية :	رجب سنة ١٤٢٥ هـ - سبتمبر رقم الإيداع : ٢٠٠٤/١٥٦١٢ الترقيم الدولي : **I.S.B.N.977-776** الناشر : جهاز نشر وتوزيع الجامعي
الطبعة الثالثة :	سنة ١٤٢٧هـ -سبتمبر رقم الإيداع : الترقيم الدولي : الناشر : جهاز نشر وتوزيع الجامعي

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

"وقل ربي زدنى علماً"

صدق الله العظيم

الفهرس

الموضوع	الصفحة

الجزء الأول

الاستدلال الإحصائي لمعلمات المجتمع

Statistical Inference of Population′s Parameters

الباب الأول
الاستدلال الإحصائي وصناعة القرارات
Statistical Inference and Decisions Making

(١-١) دور الإحصاء في صناعة القرارات

The Statistical Role in Decisions Making

(١-٢) الاستدلال الإحصائي

Statistical Inference

(١-٣) بعض أساليب الاستدلال الإحصائي

Some Statistical Inference Techniques

(١-٤) استخدام الحزم الإحصائية

Using The Statistical Software Package

(١-١) دور الإحصاء في صناعة القرارات

The Statistical Role in Decisions Making

لقد أصبحت علوم الإحصاء (الإحصاء التطبيقي ، الإحصاء الرياضي ، الإحصاء السكاني ، الإحصاء الفيزيائية ، .... ، الخ) بمفاهيمها الحديثة المرتبطة ارتباط وثيق بتكنولوجيا المعلومات والحاسبات والاتصالات ضرورة لا غنى عنها في جميع فروع المعرفة : الاقتصادية ، الطبية ، السياسية ، الإدارية ،.... الخ.

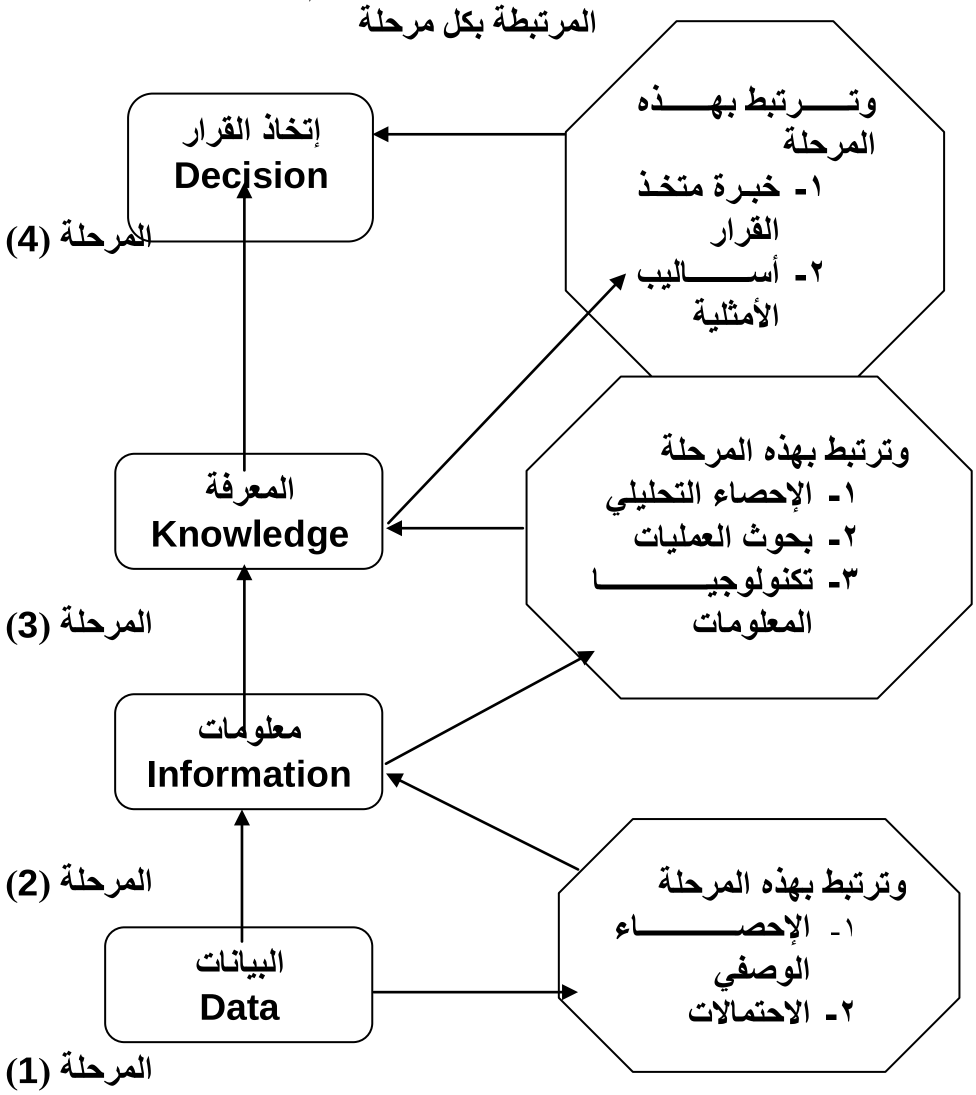
وتمثل علوم الإحصاء بمفاهيمها الحالية مجموعة النظريات والأساليب المستخدمة لجمع Collecting وعرض Presentation وتحليل Analysis البيانات Data للوصول إلى أفضل القرارات Best Decision بالنسبة للمشاكل موضع الدراسة أو التخطيط للأنظمة القائمة أو المستخدمة.

ولقد أدى التطور العظيم في علوم الحاسبات والاتصالات في هذا القرن إلى تطوير علوم الإحصاء وسهولة تناولها وتطبيقها على نطاق واسع في جميع المجالات (الزراعية ، الصناعية ، التجارية ، الخدمية ،.... الخ).

وكما ذكرنا سابقاً في الباب الأول من الجزء الأول من هذا المرجع * إن عملية صناعة القرار تمر بأربعة مراحل متتالية على النحو الموضح في الشكل (١-١).
ومن الشكل يتضح أن الانتقال من المرحلة الثانية (استخلاص المعلومات) إلى المرحلة (استخلاص المعرفة) يتطلب استخدام أساليب الاستدلال الإحصائي.

* أ.د عفاف الدش (٢٠٠٥): الإحصاء التطبيقي – الجزء الأول – الطبعة السادسة – جهاز نشر الكتاب الجامعي – جامعة حلوان

مراحل صناعة القرار والعلوم المرتبطة بكل مرحلة



شكل (١-١)

ومن الشكل يتضح أن البيانات هي المادة الرئيسية في صناعة القرار فبقدر صحتها وتكاملها واتساقها تتوقف دقة النتائج المستخلصة وبالتالي تتوقف صحة القرارات المبنية على هذه النتائج.

كذلك نجد أن مراحل صناعة القرار مراحل متتالية وترتيبها ترتيب تصاعدي بمعنى أن مرحلة جمع البيانات وعرض البيانات لابد أن تسبق مرحلة استخلاص المعلومات ومرحلة استخلاص المعلومات لابد أن تسبق مرحلة استخلاص المعرفة ، وهكذا.

ومن الشكل يتضح أيضاً أن تحويل البيانات **Data** إلى معلومات **Information** يتطلب استخدام أساليب الإحصاء الوصفي والاحتمالات . فباستخدام أساليب الإحصاء الوصفي والاحتمالات يمكن الحصول على مؤشرات تصف خصائص الظواهر محل الاعتبار كذلك وصف السلوك الاحتمالي لتلك الظواهر وبذلك يتم تحويل البيانات إلى معلومات – وقد تم تقديم ذلك في الجزء الأول من الكتاب.

ومن الشكل يتضح أن المرحلة الثالثة في صناعة القرار هو المعرفة **Knowledge** وتحويل المعلومات التي تم الحصول عليها في المرحلة الثانية على معرفة في المرحلة الثالثة يتطلب استخدام أساليب الاستدلال الإحصائي (أو التحليل الإحصائي) وهو ما يتم تناوله بالتفصيل في هذا الجزء من الكتاب.

وفيما يلي سوف نوضح المراحل المتتالية لصناعة القرار ودور أساليب الإحصاء الوصفي والاستدلال الإحصائي في كل مرحلة باختصار من خلال المثال التالي:-

مثال(١-١)

ترغب إحدى شركات إنتاج المثلجات في تقدير الحجم الأمثل من الكميات التي يتم عرضها في أحد الأسواق من منتجاتها في العام القادم حيث تنتج الشركة ثلاثة أحجام (ربع لتر – نصف لتر - لتر) من الزجاجات لمشروب معين ومن الدراسات التسويقية السابقة للشركة أثبتت أن الكميات المعروضة بسعر مناسب تتوقف فقط على الكميات المطلوبة عند هذا السعر.

ولتقدير حجم الطلب المتوقع من كل حجم من الأحجام الثلاثة في السوق في العام القادم أجريت دراسة ميدانية من خلال عينة عشوائية مناسبة من المستهلكين حيث:-

أولاً: تم تصميم استمارة مناسبة لجمع البيانات عن الكميات الشهرية المستهلكة من كل حجم ، ونسبة المستهلك من كل حجم في العينة حيث يمكن باستخدام الإحصاء الوصفي:-

- تصميم استمارة جمع بيانات مناسبة.
- تفريغ البيانات في جداول إحصائية مناسبة.
- تحديد أقل عدد وأكبر عدد من عدد الزجاجات المستهلكة لكل حجم.

وهذه المرحلة تمثل المرحلة الأولى في صناعة القرار وهي عملية جمع البيانات وعرضها.

ثانياً: كذلك باستخدام أساليب الإحصاء الوصفي فإنه من البيانات التي يتم الحصول عليها في أولاً يتم استخراج المعلومات التالية:-

- متوسط عدد الزجاجات المستهلكة من كل حجم خلال الشهر في العينة.
- نسبة الزجاجات المستهلكة من كل حجم خلال الشهر في العينة.

وهذه المؤشرات تمثل معلومات أي المرحلة الثانية في عملية صناعة القرار.

ويلاحظ أن استخلاص المعلومات في ثانياً من البيانات في أولاً يعتمد كلياً على الأساليب الإحصائية.

ثالثاً: المؤشرات التي تم الحصول عليها في ثانياً من بيانات العينة تمثل الظاهرة في العينة وبالتالي تختلف من عينة لأخرى.

وباستخدام المؤشرات المحسوبة من العينة محل الدراسة وأساليب الاستدلال الإحصائي يمكن الحصول على:-

- **تقدير لمتوسط عدد الزجاجات المطلوبة من كل حجم خلال شهر في العام القادم في السوق محل الدراسة (مجتمع الدراسة).**
- **تقدير نسبة عدد الزجاجات المطلوبة من كل حجم خلال شهر في العام القادم في السوق محل الدراسة (مجتمع الدراسة).**

كذلك باستخدام الدراسات التسويقية السابقة يمكن وصف السلوك الإحصائي للظاهرة (الطلب على المنتجات). ويتضح مما أعلاه أنه باستخدام أساليب الاستدلال الإحصائي فإنه يمكن تحويل المؤشرات التي تصف الظاهرة في العينة إلى تقديرات لهذه المؤشرات في المجتمع محل الدراسة أي أمكن تحويل المعلومات إلى معرفة عن المجتمع محل الدراسة.

رابعاً: وباستخدام التقديرات للأعداد المطلوبة من كل حجم في العام القادم التي تم الحصول عليها في ثالثاً وتطبيق أساليب الأمثلية* بالإضافة إلى خبرة متخذ القرار فإنه يمكن تحديد حجم العرض الأمثل من كل نوع وكذلك تحديد المخاطرة المترتبة على ذلك.

ويهدف هذا الجزء من الكتاب إلى:-

أ- تقديم أهم المفاهيم والأساليب للاستدلال الإحصائي التي يمكن باستخدامها تحويل المعلومات إلى معرفة والتي تشكل ضرورة لعملية صناعة القرار.

ب- تقديم العديد من التطبيقات التي توضح كيفية استخدام هذه الأساليب وأهمية تطبيقها.

(١-٢) الاستدلال الإحصائي Statistical Inference

* أ.د عفاف الدش (١٩٨٧): بحوث العمليات واتخاذ القرارات – مكتبة عين شمس – جمهورية مصر العربية

وكما سبق تعريف علم الإحصاء بأنه العلم [**] **الذي يهتم بطرق وأساليب جمع وعرض وتحليل البيانات كذلك بطرق وأساليب الأمثلية في حالة وجود معرفة غير كاملة أو ظروف عدم التأكد بالنسبة للظاهرة محل الدراسة.**

ووفقاً لهذا التعريف لعلم الإحصاء فإننا يمكننا تقسيم العلم إلى ثلاثة فروع رئيسية على النحو التالي:-

١- الإحصاء الوصفي **Descriptive Statistics**

وهو الفرع الذي يهتم بتقديم وتطوير وتطبيق طرق وأساليب جمع وعرض البيانات وبعبارة أخرى الفرع الذي يمكن باستخدامه تحويل البيانات إلى معلومات وتم تناول هذا الفرع في الجزء الأول من هذا الكتاب.

[**] أ.د عفاف الدش (٢٠٠٥):الإحصاء التطبيقي – الجزء الأول – الطبعة السادسة – جهاز نشر الكتاب
الجامعي – جامعة حلوان

٢ـ الاستدلال الإحصائي Statistical Inference

كذلك يسمى بالتحليل الإحصائي وهو الفرع الذي يهتم بتقديم وتطوير وتطبيق طرق وأساليب التقدير Estimation والتنبؤ * Forecasting للظواهر محل الدراسة وأسبابها وعلاقتها مع الظواهر الأخرى ، أي الطرق والأساليب التي تحول المعلومات Information المستخلصة من العينة (أو العينات) إلى معرفة Knowledge لخصائص وسلوك وعلاقات الظواهر محل الدراسة في المجتمع.**

وفي كثير من الكتابات يتم تقسيم أساليب الاستدلال الإحصائي إلى فئتين هما:ـ

- **الفئة الأولى: أساليب التقدير والتنبؤ.**
- **الفئة الثانية: أساليب اختبارات الفروض الإحصائية.**

وسوف نتناول بالدراسة في هذا الكتاب أهم أساليب الاستدلال الإحصائي بصفة عامة وبصفة خاصة من الجانب التطبيقي.

٣ـ صناعة القرارات في بيئة غير يقينية

Decision Making in Uncertain Environment

ويختص هذا الفرع بتقديم وتطوير وتطبيق أساليب اتخاذ القرارات المثلى في ظل وجود معرفة غير كاملة أو في ظل ظروف عدم التأكد Uncertain * Environment.**

(١-٣) بعض أساليب الاستدلال الإحصائي

* Paul ,N.and Other (2003) : " Statistics for Business and Economics" Prentice Hall., New Jersey

** Prem , S (1995) : "Statistics for Business and Economics" John Wiley and Sons,Inc,Newyork.

*** أ.د عفاف الدش (١٩٨٧): بحوث العمليات واتخاذ القرارات – مكتبة عين شمس – جمهورية مصر العربية

Some Techniques of Statistical Inference

وكما ذكرنا أعلاه أنه يمكن تقسيم أساليب الاستدلال الإحصائي إلى فئتين:-

- **الفئة الأولى: أساليب التقدير والتنبؤ.**
- **الفئة الثانية: أساليب الاختبارات الإحصائية.**

أولاً : أساليب التقدير والتنبؤ

ويمكن تقسيم أساليب التقدير والتنبؤ إلى قسمين على النحو التالي:-

القسم الأول: ويتناول أساليب لتقدير بعض معلمات المجتمع محل الدراسة سواء كانت تقديرات بنقطة Point Estimators أو تقديرات بفترة Interval Estimators وسوف نتناول في الباب الرابع بعض أساليب التقدير لمتوسط المجتمع ، والنسبة في المجتمع ، والتباين في المجتمع باستخدام المؤشرات المحسوبه من العينة العشوائية.

القسم الثاني: ويتناول أساليب لتقدير بعض العلاقات بين المتغيرات في المجتمع محل الدراسة وسوف نتناول في الباب السابع والعاشر بعض هذه الأساليب.

ثانياً : أساليب الاختبارات الإحصائية

ويمكن تقسيم أساليب إجراء الاختبارات الإحصائية إلى قسمين أيضاً على النحو التالي:-

القسم الأول: ويتناول أساليب الاختبارات الإحصائية عن معلمات المجتمع محل الدراسة مثل متوسط المجتمع أو النسب في المجتمع ،.....الخ. وتسمى هذه الاختبارات بالاختبارات المعلمية أو اللامعلمية ، وسوف نتناول بعض هذه الأساليب بالتفصيل في البابين الخامس والسادس.

القسم الثاني: ويتناول الاختبارات الإحصائية عن العلاقة بين متغيرين أو أكثر في المجتمع محل الدراسة.وتسمى هذه الاختبارات باختبارات العلاقات أو اختبارات جودة التوفيق ن وسوف نتناول بعض هذه الاختبارات في الباب الثامن والتاسع والعاشر.

(١-٤) استخدام الحزم الإحصائية

Using the Statistical Software Package

وكما ذكرنا في الجزء الأول من هذا المرجع * أهمية وضرورة استخدام الحاسب الآلي Computer في تطبيق الطرق والأساليب الإحصائية على نطاق واسع.

وقد صممت برامج حاسب خاصة بالطرق والأساليب الإحصائية بصفة عامة تسمى بالحزم الإحصائية وبصفة خاصة بأساليب الاستدلال الإحصائي، وتعتبر الحزمة الإحصائية المسماه بـ SPSS Windows من أهم الحزم الإحصائية لتطبيق أساليب الاستدلال الإحصائي.

ومما هو جدير بالذكر أن جميع أساليب الاستدلال الإحصائي المقدمة في هذا الكتاب اعتباراً من الباب الخامس حتى الباب العاشر لها برامج الحاسب في الحزمة الإحصائية SPSS Windows التي يمكن استخدامها مباشرة في الحصول على النتائج المطلوبة

---

* عفاف الدش (٢٠٠٥) : الإحصاء التطبيقي – الجزء الأول - الطبعة السادسة – جهاز نشر وتوزيع الكتاب الجامعي – جامعة حلوان

الباب الثاني
العينات الإحصائية
Statistical Samples

(٢-١) أنواع العينات
Types Samples
(٢-٢) العينة العشوائية البسيطة
Simple Random Sample
(٢-٣) العينة العشوائية المنتظمة
Systematic Random Sample
(٢-٤) العينة العشوائية الطبقية
Stratified Random Sample
(٢-٥) العينة العشوائية العنقودية
Cluster Random Sample
(٢-٦) العينات غير العشوائية
Non – Random Sample
(٢-٧) تطبيقات **Applications**
(٣-٨) تمرينات **Exercises**

(٢-١) أنواع العينات Types Samples

في الجزء الأول من هذا الكتاب تناولنا أساليب الدراسة الإحصائية وطرق جمع البيانات الإحصائية وانتهينا إلي أنه يوجد أسلوبين للدراسة الإحصائية هما:

الأسلوب الأول: هو ما يسمي بأسلوب الحصر الشامل حيث يتم جمع البيانات المطلوبة من جميع مفردات المجتمع محل الدراسة.

الأسلوب الثاني: وهو ما يسمي بأسلوب العينة حيث يتم جمع البيانات المطلوبة من بعض مفردات المجتمع وتسمي هذه المفردات بالعينة. وبالتالي فإن العينة ما هي الا فئة جزئية من المجتمع محل الدراسة بحيث تمثل هذه الفئة المجتمع كما ذكرنا سابقاً.

وتوجد أنواع متعددة من العينات يرجع الاختلاف بينها إلى الاختلاف في الأساليب التي يتم بها سحب (اختيار) مفردات كل منهم.

ومن الأهمية أن تكون مفردات العينة ممثلة تمثيلاً دقيقاً للمجتمع، لذلك تتعدد أساليب اختيار مفردات العينة نظراً لعدم تجانس مفردات المجتمع محل الدراسة (حيث يمكن قياس تجانس مفردات المجتمع بالنسبة للظاهرة أو المتغير محل الدراسة باستخدام التباين والانحراف المعياري أو باستخدام أحد مؤشرات الاختلاف الأخرى (انظر الباب الرابع من الجزء الأول من هذا الكتاب)

وفي هذا الباب سوف نتناول بالدراسة النظرية والتطبيقية الأساليب الإحصائية المختلفة لاختيار مفردات العينة من حيث طرق اختيارها (سحب مفرداتها) والخصائص المميزة لكل عينة عن العينات الأخرى بالإضافة إلى مجموعة من الأمثلة التي توضح طريقة السحب (الاختيار) وإبراز أهم الخصائص.

وعادة يتم تقسيم أنواع العينات وفقاً لأساليب سحب مفردات العينة من المجتمع إلى نوعين سوف نتناولهما فيما يلي:

النوع الأول: العينات العشوائية أو الاحتمالية

Random (Probabilistic) Samples

وهذا النوع من العينات يعتمد أسلوب اختيار مفردات العينة فيه على العنصر العشوائي أو الاحتمالي. ويوجد عديد من العينات التي تنتمي إلى هذا النوع. سوف نتناول منها: العينة العشوائية البسيطة، والعينة العشوائية المنتظمة، والعينة العشوائية الطبقية، والعينة العشوائية العنقودية في الفصول (٢-٢) - (٢-٥).

النوع الثاني: العينات غير العشوائية (أو التحكمية)

Non-Random (Judgment) Samples

وهذا النوع من العينات يعتمد أسلوب اختيار مفردات كل عينة من هذا النوع أساساً على الباحث. أي يكون تحديد مفردات العينة تحكمي Judgment، أي يرجع إلى التقدير الشخصي للباحث ولا يخضع لأي معيار موضوعي. ويوجد عديد من العينات التي تنتمي إلى هذا النوع سوف نتناول منها: العينة العمدية، والعينة الحصصية في الفصل (٢-٦).

ومما هو جدير بالذكر فإنه قبل تحديد اختيار العينة لابد أولاً من قياس مدي تجانس مفردات المجتمع بالنسبة للظاهرة (أو الظواهر) محل الدراسة. ويمكن التعرف مبدئياً على التجانس بين مفردات المجتمع من خلال عينة استطلاعية، حتى يمكن تحديد الأسلوب الملائم لاختيار العينة لكي تمثل المجتمع تمثيلاً دقيقاً. لأنه في حالة عدم الاختيار الأمثل لمفردات العينة سوف يؤدي إلى استخراج نتائج (معلومات) لا يمكن استخراج معرفة صحيحة للمجتمع المسحوب منه العينة وبالتالي إلى قرارات غير صحيحة.

(٢-٢) العينة العشوائية البسيطة

Simple Random Sample

تكون العينة عينة عشوائية بسيطة إذا توفر فيها الشرطين التاليين:

١- إذا كان للعينة نفس الفرصة Chance في الظهور كما لأي عينة أخري من العينات الممكن سحبها من المجتمع محل الدراسة. أو بعبارة أخري احتمال سحب أي عينة من العينات المتماثلة الممكن سحبها متساوي بالنسبة لجميع العينات الممكنة. فإذا فرضنا أن حجم المجتمع المراد سحب عينة منه يساوي N، وحجم العينة المراد سحبها يساوي n (سحب بدون إرجاع)، فإن عدد العينات المتماثلة الممكن سحبها يساوي C_n^N.

وبالتالي يتحقق هذا الشرط إذا كان احتمال سحب العينة يساوي $\frac{1}{C_n^N}$، أما إذا اختلف احتمال سحب العينة عن $\frac{1}{C_n^N}$ فإن العينة في هذه الحالة تكون عينة عشوائية غير بسيطة.

٢- إذا كان لكل مفردة من مفردات المجتمع نفس الفرصة في الظهور في العينة كما لأي مفردة أخري بالمجتمع. بمعني أن عدد العينات التي تظهر فيها مفردة معينة من مفردات المجتمع يساوي نفس عدد العينات التي تظهر فيها أي مفردة أخري. أو بعبارة أخري جميع مفردات المجتمع لكل منها نفس احتمال الظهور في العينة. وبالتالي يصبح عدد العينات الممكنة التي يمكن أن تظهر فيها إحدى المفردات* يساوي C_{n-1}^{N-1}. وبالتالي يتحقق هذا الشرط إذا كان ~~احتمال ظهور أي مفردة من مفردات المجتمع في العينات الممكنة~~

* Heinz Kohler (1994): " Statistics for Business and Economics" Third Edition. Harper Collins College Publishers, U.S.A.

يساوي $\frac{C_{n-1}^{N-1}}{C_{n}^{N}}$ (عدد الحالات المواتية ÷عدد الحالات الممكنة) أنظر تعريف الاحتمال بملحق رقم (١).

والمثال التالي سوف يوضح كيفية توافر هذين الشرطين في العينة العشوائية البسيطة.

مثال (٢-١)

إذا فرضنا أن لدينا مجتمع محدود مكون من 4مفردات فقط A,B,C,D وترغب في سحب (سحب بدون إرجاع) عينة عشوائية بسيطة مكونة من 3مفردات فقط.

فإننا نجد أن عدد العينات الممكن سحبها يساوي C_3^4 حيث :

$$C_3^4 = \frac{4!}{3!\times 1!} = 4 \text{ عينات}$$

أي أن عدد العينات الممكنة يساوي 4 عينات، ويمكن سردها على النحو التالي:

1) (A,B,C) 2) (A,B,D)
3) (B,C,D) 4) (A,C,D)

أي أن احتمال سحب أي عينة من العينات الأربعة المذكورة أعلاه يساوي $\frac{1}{4}$ حيث

$$\frac{1}{C_n^N} = \frac{1}{4}$$

أي لكل عينة نفس الفرصة للسحب مثل العينات الأخرى (أي يتحقق الشرط الأول لكون أي عينة من العينات الأربعة السابقة تكون عينة عشوائية بسيطة).

كذلك نجد أن عدد العينات الممكنة التي يمكن أن تظهر فيها المفردة A يساوي 3 عينات. بالمثل كل من المفردات B,C,D نجد أن لكل مفردة منهم يمكن أن تظهر في 3عينات فقط، أي كل عدد العينات الممكنة التي تظهر بها كل مفردة يساوي

$$C_{n-1}^{N-1} = C_{3-1}^{4-1} = C_2^3 = 3 \text{ عينات}$$

وبالتالي احتمال ظهور أي مفردة **A,B,C,D** يساوي

$$\frac{C_2^3}{C_3^4} = \frac{3}{4}$$

وهذا يعني أن

$$P_r(\text{ظهور}A) = P_r(\text{ظهور}B) = P_r(\text{ظهور}C) = P_r(\text{ظهور}D) = \frac{3}{4}$$

وبذلك يتحقق الشرط الثاني لتكون أي عينة من العينات الأربعة السابقة عينة عشوائية بسيطة.

مثال (٢-٢)

بنفس الأسلوب إذا كان المجتمع يتكون من خمسة مفردات هي **(A), (B), (C), (D), (E)** ونرغب في سحب (سحب بدون ارجاع) عينة مكونة من 3مفردات فإننا نجد أن عدد العينات الممكنة في هذه الحالة يساوي $C_3^5 = 10$ عينات ممكنة على النحو التالي:

1) (A,B,C)	**2) (A,B,D)**	**3) (A,B,E)**
4) (A,C,D)	**5) (A,C,E)**	**6) (A,D,E)**
7) (B,C,D)	**8) (B,C,E)**	**9) (E,C,A)**
10) (C,D,E)		

واحتمال سحب أي عينة من هذه العينات يساوي $\frac{1}{10} = 0.1$ وهذا يعني تحقق الشرط الأول.

كذلك نجد أن عدد العينات الممكنة التي يمكن أن تظهر فيها المفردة **A** يساوي 6عينات وهو نفس العدد لظهور المفردة **B** أو **C** أو **D** أو **E** واحتمال ظهور أي مفردة في العينة يساوي:

$$\frac{C_{3-1}^{5-1}}{C_3^5} = \frac{(5-1)!}{(3-1)!\times(5-3)!} \div \frac{5!}{3!\times(5-3)!} = \frac{4!}{2!\times 2!} \div \frac{5!}{3!\times 2!}$$

$$= \frac{4\times 3\times 2\times 1}{2\times 1\times 2\times 1} \div \frac{5\times 4\times 3\times 2\times 1}{3\times 2\times 1\times 2\times 1} = \frac{6}{1} \div \frac{10}{1} = \frac{6}{10}$$

مثال (٢-٣)

في إحدى الإعلانات عن وظائف خالية وجود إعلان عن وظيفتين لمخططي برامج حاسب في أحد البنوك التجارية من خريجي كليات الحاسبات وخبرة لا تقل عن عشرة سنوات ولا يزيد السن عن 40 سنة. فتقدم لهاتين الوظيفتين 5 أشخاص A,B,C,D,E تنطبق على كل منهم الشروط المعلنة.

١- حدد العينات الممكنة لاختيار أثنين من المتقدمين.

٢- أثبت أن العينات التي تم تحديدها في (١) تعتبر عينات عشوائية بسيطة.

الحل:

١- بما أن عدد الأشخاص المتقدمين N حيث N=5 تنطبق عليهم جميع الشروط محل الاعتبار. وبالتالي فإن مفردات المجتمع تعتبر متجانسة، وكذلك نجد أن حجم العينة n حيث n=2 في هذه الحالة، وبالتالي فأن عدد العينات الممكنة حيث: $C_n^N = C_2^5$

$$C_n^N = C_2^5 = \frac{5!}{2!\times(5-2)!} = \frac{5\times 4}{2} = 10$$ **عينات**

1- (A,B) , 2- (A,C) , 3- (A,D) , 4- (A,E)
5- (B,C) , 6- (B,D) , 7- (B,E) , 8- (C,D)
9- (C,E) , 10- (D,E)

٢- وبالتالي فإن احتمال ظهور أي عينة يساوي $\frac{1}{C_3^5}$ أي:

$$\frac{1}{C_3^5} = \frac{1}{10} \qquad \textbf{(2.1)}$$

أي تحقق الشرط الأول لكي تكون العينة عشوائية. وإذا أشرنا إلى احتمال ظهور العينة التي تتضمن **A** بالرمز $P_r(A)$ ، بالمثل احتمال ظهور العينة التي تتضمن **B** بالرمز $P_r(B)$، ... الخ. من (١) نجد أن:

$$P_r(A) = P_r(B) = P_r(C) = P_r(D) = P_r(E) = \frac{4}{10}$$

وبما أن

$$\frac{C_{n-1}^{N-1}}{C_n^N} = \frac{C_1^4}{C_2^5} = \frac{4}{10} \qquad \textbf{(2.2)}$$

من (2.1) ، (2.2) نجد أن أي عينة من العينات السابق سردها في (١) تعتبر عينة عشوائية بسيطة.

طرق سحب العينة العشوائية البسيطة

توجد طريقتين لسحب مفردات العينة العشوائية البسيطة وهما:

- طريقة البطاقات المتماثلة.
- طريقة جداول الأعداد العشوائية.

أولاً: طريقة البطاقات المتماثلة

إذا كان حجم مجتمع الدراسة N مفردة ويراد سحب عينة عشوائية بسيطة منه حجمها nمفردة. حيث أن حجم المجتمع N صغيراً (نعتبر حجم المجتمع صغير إذا كان عدد المفردات أقل من 100مفردة) فإننا نتبع الخطوات التالية لسحب عينة عشوائية بسيطة باستخدام طريقة البطاقات:

١- نوجد عدد N من البطاقات (الكروت) المتماثلة تماماً من حيث الحجم، اللون، الشكل، ... الخ.

٢- ترقم مفردات المجتمع على النحو1,2,3,....,N كذلك نرقم البطاقات من 1 إلى N. أي تخصص بطاقات لكل مفردة من مفردات المجتمع.

٣- تخلط جميع البطاقات المرقمة وتوضع في الصندوق (أو الكيس) وتخلط جيداً ثم يتم سحب عدد n من البطاقات من الصندوق (أو الكيس).

وهذه الطريقة توفر الشرطين السابق ذكرهما لكي تكون العينة المسحوبة عينة عشوائية بسيطة. فهي توفر أن تكون جميع العينات المسحوبة من الصندوق (أو الكيس) وكل عينة بحجم n لها نفس احتمال السحب، كذلك توفر احتمال ظهور أي مفردة من مفردات المجتمع في العينة مساوي لاحتمال ظهور أي مفردة أخري من مفردات المجتمع، وبالتالي يتوافر في هذه الطريقة تحقيق الشرطين السابق ذكرهما.

ورغم بساطة هذه الطريقة في اختيار مفردات العينة العشوائية البسيطة إذا كان عدد مفردات المجتمع محل الدراسة صغير (100 مفردة أو أقل)، فإنه يصبح من الصعب استخدام هذه الطريقة في حالة إذا كان عدد مفردات المجتمع كبير جداً مثلاً 200000 مفردة أو أكثر فإنه من الصعب إيجاد 200000 بطاقة أو أكثر متماثلة تماماً وبالتالي يصعب استخدام هذه الطريقة.

ونظراً لتوافر وإمكانية استخدام الحاسب الآلي (الكمبيوتر) فإنه يمكن في حالة إذا كان حجم المجتمع كبير (أو صغير أيضاً) استخدام طريقة جداول الأعداد العشوائية لاختيار مفردات عينة عشوائية بسيطة (أي يتوافر فيها الشرطين السابقين) كما سوف نوضح فيما يلي

ثانياً: طريقة جداول الأعداد العشوائية

وفي حالة إذا كان حجم المجتمع N المطلوب سحب عينة عشوائية بسيطة منه كبيراً فإن طريقة استخدام البطاقات السابقة تصبح طريقة غير عملية لسحب العينة. لذلك صمم الإحصائيون جداول تسمي جداول الأعداد العشوائية Random Numbers Tables يتكون كل عدد فيها من الأرقام من صفر إلى تسعة [0-9] بحيث أنه في أي سحبة تكون فرصة ظهور أي عدد منها مساويا لفرصة ظهور أي عدد أخر (يعتمد تصميم جداول الأعداد العشوائية على النظرية القائلة* أن دالة التوزيع التراكمية لأي متغير عشوائي (بغض النظر عن التوزيع الاحتمالي الذي يتبعه) تمثل متغير عشوائي يتبع التوزيع الاحتمالي المنتظم في الفترة [1 - 0] وملحق رقم (٣) يعطي جزء من جداول الأعداد العشوائية (الأعداد من 00000 إلى 99999 حيث أن أي عدد في هذا الملحق له نفس الفرصة للظهور في السحب) وسحب أي عينة باستخدام جداول الأعداد العشوائية يتوفر فيها الشرطين السابق ذكرهما بالنسبة للعينة العشوائية البسيطة. وسوف نوضح كيفية استخدام جدول الأعداد العشوائية في سحب عينة عشوائية بسيطة من خلال الأمثلة التالية:

* أ.د/ عفاف الدش (١٩٨٩): " بحوث العمليات وصناعة القرارات " – مكتبة عين شمس – القاهرة.

مثال (٢-٤)

في سنة ١٩٩٢ وجد أن 100 شركة (متماثلة من حيث الهيكل المالي) من شركات القطاع الخاص بجمهورية مصر العربية حققت مستوي ربح يزيد 50% عن ربحها في السنة السابقة ١٩٩١. ووجد أحد بيوت الخبرة أهمية دراسة الوضع المالي لهذه الشركات من خلال عينة عشوائية بسيطة حجمها 10 شركات تمثل هذه الشركات.

لسحب العينة المكونة من 10 شركات من الـ 100 شركة عن طريق جداول الأعداد العشوائية، سوف نتبع الخطوات التالية:

١- نقوم بإعطاء عدد لكل شركة من الشركات محل الدراسة (مجتمع الدراسة وعددهم 100 شركة) كما هو موضح بجدول (٢-١).

٢- نختار أي عمود (أو أي صف) في جدول الأعداد العشوائية بملحق رقم (٣) وليكن العمود رقم 4 ونأخذ منه 10 أعداد متتالية (حجم العينة) من أعلي العمود (أو من بداية الصف) على النحو التالي:

(85690) , (31891) , (15549) , (11447) , (08883) , (91572) ,

(1960) , (46617) , (6117) , (35384) , (80997).

ملحوظة: عند اختيار الأعداد من جدول الأعداد العشرية لابد أن تكون الأعداد متتالية وذلك من خصائص تصميم الجداول العشوائية حتى يكون احتمال ظهور أي عدد متساوي مع احتمال ظهور أي عدد متساوي مع احتمال ظهور أي عدد أخر.

جدول (٢-١)* يعطي بيان بأسماء الشركات محل الدراسة والأعداد المناظرة لكل منها.

الشركة	الرقم	الشركة	الرقم
شركة التغليف	50	الشركة المتحدة للأغذية	00
شركة الصبغات المتحدة	51	الشركة الشرقية للمنسوجات	01
شركة الملابس المصرية	52	شركة الحاج محمد على	02
شركة فهمي ماجد	53	الشركة المصرية لصناعة الحرير	03
شركة الأزياء الحديثة	54	شركة الاتحاد العربي	04
شركة داليا للألبان	55	شركة الأخوان	05
شركة الأعلاف الصناعية	56	شركة ميخائيل جرجس	06
شركة الآلات الحديثة	57	شركة عثمان محمد لمواد البناء	07
شركة الملاحة البحرية	58	شركة الأباب المصرية	08
شركة الاقطان العربية	59	شركة الكيماويات	09
شركة الأقطان المصرية	60	الشركة المصرية للسياحة	10
شركة مستحضرات التجميل المتحدة	61	شركة المقاولون المتحدة	11
شركة الأغذية المصرية	62	الشركة الصناعية للالكترونات	12
شركة موريس كريم	63	الشركة العربية للنسيج	13
شركة محمود فاروق	64	الشركة العربية للنقل	14
شركة كمال مجلي	65	شركة منصور أحمد	15
شركة الجلود المتحدة	66	شركة السيارات المتحدة	16

* جميع الأسماء للشركات أسماء افتراضية غير واقعية.

الشركة	الرقم	الشركة	الرقم
شركة الأصواف المحلية	67	شركة الأسمدة الحديثة	17
شركة عصام لقطع الغيار	68	شركة الأدوية المصرية	18
شركة الغازات الصناعية	69	شـــركة الخـــزف والصـــينى المصرية	19
شركة الصباغة العامة	70	شركة الأسمدة المصرية	20
شركة المحاريث	71	شركة المطاط المصرية	21
شركة المستحضرات الطبية	72	شركة الأسمدة الشرقية	22
شركة محمود وحسين	73	شركة خوفو للسياحة	23
شركة ماجد غبور	74	شركة الحديد والصلب	24
شركة النصر للسيارات	75	شركة مصر للأغذية	25
شركة الأخوان العرب	76	شركة الكيماويات	26
شركة الورق المصرية	77	شركة المياه المقطرة	27
شركة الجبنة الشرقية	78	شركة الأفلام الحديثة	28
شركة الخيوط المتحدة	79	شركة المثلجات الحديثة	29
شركة الصناعات اليدوية	80	شركة المطبوعات العامة	30
شركة الخدمات الأمنية	81	شركة الهندسة الحديثة	31
شركة الخدمات العامة	82	شركة عثمان حامد	32
شركة الصيانة الكهربية	83	شركة وود للأخشاب	33
شركة الصرافة العربية	84	الشركة المصرية للأخشاب	34
شركة الطباعة القومية	85	شركة الموبيليات الحديثة	35
شركة المحاريث القومية	86	شركة محسن للآلات	36
شركة المنظفات الصناعية	87	شركة الكمبيوتر المصرية	37

الشركة	الرقم	الشركة	الرقم
شركة البسكويت المصرية	88	شركة ايجبت	38
شـــركة منتجـــات الأليـــاف القومية	89	الشركة الشرقية للورق	39
شركة الجلود المصرية	90	شـــركة أحمـــد محمـــود للتليفزيون	40
شـــركة أحمـــد محمـــود للكيماويات	91	شركة الحديد العامة	41
الشركة الفنية للصيانة	92	شـــركة ســـامي معـــوض للبصريات	42
شركة الراديو المصرية	93	شركة المعادن المتحدة	43
شركة عاشور للرخام	94	شركة العطور العامة	44
شركة الجلود الحديثة	95	شركة الاتحاد العامة	45
شركة الأحذية العربية	96	شركة الصرافة	46
شركة المطبوعات العربية	97	شركة الشاي العامة	47
شركة السفن الحديثة	98	شركة الحلويات الشرقية	48
الشركة الهندسية العامة	99	الشركة الهندسية للمعادن	49

٣- وبما أن عدد مفردات المجتمع محل الدراسة تتراوح من 00 إلى 99، إذن يتم أخذ الرقم الأول والثاني فقط من الأعداد المختارة من جدول الأعداد العشوائية كما هو موضح بجدول (٢-٢). وإذا كان عدد مفردات المجتمع يتراوح من 000 إلى 999 فإنه يتم أخذ الرقم الأول والثاني والثالث من كل عدد من الأعداد العشوائية المختارة، بالمثل عندما يكون عدد مفردات المجتمع أكثر من ذلك. وبالتالي تكون العينة المختارة هي:

(90, 91, 49, 47, 83, 72, 3, 17, 84, 97)

٤- بعد تحديد الرقمين الأول والثاني في كل عدد عشوائي يتم تحديد الشركات التي لها نفس الأعداد المختارة، كما هو موضح بالجدول (٢-٢).

جدول (٢-٢): يوضح العينة العشوائية البسيطة التي تم اختيارها باستخدام جداول الأعداد العشوائية

أسم الشركة المناظر	العدد العشوائي
شركة الجلود المصرية	85690
شركة أحمد محمود للكيماويات	31891
شركة الحلويات الشرقية	15549
شركة الصرافة	11447
شركة الصيانة الكهربائية	08883
شركة المستحضرات الطبية	91572
الشركة المصرية لصناعة الحرير	19603
السيارات المتحدة	46617
شركة الصرافة العربية	35384
شركة المطبوعات العربية	80997

ومما هو جدير بالذكر في هذا المثال أنه قد يتكرر الرقمين الأول والثاني من العدد العشوائي المختار في أكثر من عدد، وفي هذه الحالة يأخذ العدد مرة واحدة فقط ثم يضاف الأعداد الأخرى التالية في سلسلة الأعداد المتتالية في جدول الأعداد العشوائية. ويمكن تعميم ذلك في حالة الثلاث أرقام أو أكثر. ففي هذا المثال نجد أن الرقم الأول والثاني من كل عدد من الأعداد (41591) , (65947) , (86172) , (61572) في العمود رقم (4) سبق تكرارها كما هو موضح بجدول (٢-٢) لذلك تم حذفها.

مثال (٢-٥)

إذا كان لدينا مجتمع مكون من 10 أفراد على النحو التالي:

الأسم	الرقم	الأسم	الرقم
عزت مرقص	5	أحمد حسن	0
أشرف صفوت	6	محمود محمد	1
عماد عبد الحليم	7	نادية كامل	2
فاطمة على	8	ياسمين حافظ	3
إيمان عبد الله	9	حسين كامل	4

والمطلوب: سحب عينة عشوائية بسيطة باستخدام جداول الأعداد العشوائية مكونة من 7 أفراد.

الحل:

باستخدام العمود رقم (2) بملحق (٣) نجد أن:

أشرف صفوت	84346
ياسمين حافظ	01373
(تحذف لأنها تكررت)	17686
فاطمة على	59688
(تحذف لأنها تكررت)	95016
محمود محمد	65207
أحمد حسن	34510

فنجد أن الرقم الأول في كل عدد من الأعداد التي تمثل العينة هو 7,6,8,6,3,6 وبالتالي نجد أن الرقم 6 تكرر للمرة الثانية والثالثة في العددين 17686 , 95016 لذا يجب حذفهما لتكرار الرقم الأول في كل منهما لذلك سوف نقوم باختيار عددين أخرين في التسلسل بالعمود الثاني أيضاً فيكونا:

إيمان عبد الله	56299
حسين كامل	75444

(تم استبعاد الأعداد 02113 , 87979 , 53577 , 56740 نظراً لتكرار الرقم الأول في كل عدد من هذه الأعداد مع الرقم الأول في الأعداد التي تم اختيارها في العينة لذا تم استبعادهم). وبالتالي تصبح العينة العشوائية البسيطة على النحو التالي:

أشرف صفوت، ياسمين حافظ، فاطمة على، محمود محمد، أحمد حسن، إيمان عبد الله، حسين كامل.

(٢-٣) العينة العشوائية المنتظمة

Systematic Random Sample

في هذا الفصل نتناول تعريف وطريقة سحب العينة العشوائية المنتظمة. وكذلك أهم خصائص هذه العينة، وسوف نوضح ذلك من خلال المثال التالي:

في أحدي الكليات ذات الأعداد الكبيرة (كلية التجارة مثلاً) ترغب الكلية في دراسة قدرة الطالب أو الطالبة في الفرقة الأولي في الكلية على الاستيعاب من الأساتذة المحاضرين في المدرجات وذلك من خلال أخذ عينة عشوائية بسيطة حجمها 50 طالب وطالبة. فإذا كان طلاب الفرقة الأولي يتلقون المحاضرات بمدرج يستوعب 2000 طالب وطالبة، ومرقمة أماكن الجلوس في المدرج وفقاً لقربها من المنصة التي يحاضر عليها الأستاذ، فالمقعد رقم (0001) أقرب ما يمكن من المنصة والمقعد رقم (2000) أبعد ما يمكن من المنصة. ويجلس الطلاب في المقاعد عشوائياً (أي ليس مخصص لكل طالب مقعد معين). وبما أن حجم العينة التي ترغب الكلية استخدامها في الدراسة يساوي n حيث n=50 أي تمثل $\frac{1}{40}$ من حجم مجتمع الدراسة $\left(\frac{50}{2000}=\frac{1}{40}\right)$ فإنه يمكن اختيار هذه العينة باستخدام طريقة الأعداد العشوائية كما وضحنا في الفصل السابق.

وقد تكون العينة العشوائية البسيطة التي يتم سحبها مكونة من الطلاب التي تشغل المقاعد من (0001) إلى (0050) وبالتالي تمثل هذه العينة الطلاب الأقرب إلى المنصة، أي تمثل الطلاب الأقدر على المتابعة والاستيعاب من المحاضر. كذلك قد تكون العينة التي يتم سحبها من الطلاب التي تشغل المقاعد من (1951) إلى (2000) وبالتالي تمثل الطلاب الأبعد ما يمكن من المنصة أي الطلاب الأقل قدرة على المتابعة والاستيعاب. وبالتالي تكون العينة في هذه الحالة لا تمثل المجتمع من حيث المتغير محل الدراسة. وفي هذه الحالة والحالات المشابهة لا تصلح العينة العشوائية البسيطة لتمثل جميع مفردات المجتمع.

أما إذا كان المدرج المكون من 50 بنش Banch، كل بنش مكون من 40 مقعداً مثلاً فتكون المقاعد في البنش الأول من (0001) إلى (0040) وفي البنش الثاني من

(0041) إلى (0080) ... الخ. وتكون المقاعد في البنش الأخير من (1961) إلى (2000)، وتكون العينة ممثلة لجميع الطلاب بالمدرج إذا تم اختيار طالب أو طالبة عشوائياً من كل بنش (أنظر الفصل السابق باستخدام طريقة البطاقات أو باستخدام جدول الأعداد العشوائية).

ولتسهيل هذه العملية يمكن أن نختار المفردة من البنش الأول عشوائياً ولتكن المفردة رقم (0036) ثم يضاف 40 فنحصل على رقم المفردة الثانية (36+40=0076) بالمثل يضاف 40 فنحصل على المفردة الثالثة (76+40=0116) ... وهكذا فتكون المفردة الأخيرة رقم (1996). وفي هذه الحالة تسمي العينة المختارة بالعينة العشوائية المنتظمة.

وبصفة عامة في حالة إذا أمكن تقسيم المجتمع بطريقة منتظمة وفقاً للظاهرة محل الدراسة كما في المثال السابق فلكي تمثل العينة جميع مفردات المجتمع يفضل استخدام العينة المنتظمة. ويمكن تلخيص خطوات سحب العينة العشوائية المنتظمة على النحو التالي:

١- إذا كان حجم المجتمع N فإنه يتم تقسيمه إلى فئات جزئية Classes (وفقاً لطبيعة المتغير محل الدراسة) عددها k وحجم كل فئة جزئية r أي أن:

$$N = kr \qquad (2.3)$$

حيث كل من r , k أعداد صحيحة موجبة كذلك يكون احتمال سحب أي عينة منتظمة يساوي $\frac{1}{k}$ أي نسبة حجم العينة بالنسبة لحجم المجتمع.

$$\frac{r}{N} = \frac{r}{kr} = \frac{1}{k} \qquad (2.4)$$

٢- يتم اختيار مفردة من مفردات الفئة الأولي عشوائياً ولتكن المفرد رقم (i) حيث I = 1 , 2 , 3 , …, r

٣- وبالتالي تكون المفردة الثانية هي المفردة رقم [i+(2-1)r]، والمفردة الثالثة هي المفردة رقم [i+(3-1)r]، وبالمثل المفردة الرابعة، ... ، وتكون المفردة الأخيرة في العينة هي المفردة رقم [i+(k-1)r]*.

ومما سبق يمكن تلخيص أهم خصائص العينة العشوائية المنتظمة على النحو التالي:

١- إذا أمكن تقسيم المجتمع إلى فئات جزئية منتظمة وفقاً لطبيعة المتغير محل الدراسة فإن العينة العشوائية المنتظمة سوف تمثل جميع مفردات المجتمع وهذا لا يتوافر بالنسبة للعينة العشوائية البسيطة في هذه الحالة.

٢- طريقة اختيار مفردات العينة العشوائية المنتظمة تعتبر بسيطة وسهلة وبصفة خاصة بالنسبة لغير المتخصصين خاصة بالنسبة للعينات ذات الحجم الكبير كما هو متبع في الدراسات السكانية والاجتماعية.

٣- تعتبر العينة العشوائية المنتظمة تقريب ممتاز للعينة العشوائية البسيطة وتحقق نتائج مماثلة للعينة العشوائية البسيطة في حالة توافر الشروط التالية:

أ- التجانس لجميع مفردات المجتمع محل الدراسة من حيث المتغيرات محل الدراسة.

ب- عشوائية الترتيب لمفردات المجتمع (بمعني أن الرقم المعطي للمفردة ليس له أي دلالة على خصائص المفردة) أي أن صفة الدورية غير موجودة وذلك لأنه لو وجد دلالة بين الرقم المعطي وخصائص المفردة فإن ذلك سوف يؤدي

* حيث تمثل الأعداد التي تمثل مفردات العينة متوالية عددية أساسها يساوي r.

اختيار جميع مفردات العينة المنتظمة التي لها نفس خصائص أول مفردة تم اختيارها عشوائياً من الفئة الأولي. فمثلاً إذا تم اختيار أول مفردة في الفئة الأولي في بداية الفئة فإن ذلك سوف يؤدي إلى أن جميع مفردات العينة سوف تكون في بداية الفئات. وبالتالي تكون العينة عينة متحيزة وغير ممثلة للمجتمع محل الدراسة.

وإذا لم يتحقق الشرطين السابقين (أ) ، (ب) في العينة العشوائية المنتظمة فإن النتائج التي سوف تستخلص من هذه العينة تكون نتائج متحيزة (كما سوف نوضح ذلك في الباب الثالث الخاص بالتقديرات الإحصائية).

مثال (٢-٦)

إذا فرضنا أن تخصص الطلاب بكلية التجارة يبدأ في بداية الصف الثالث، وتجري دراسة عن أسباب تفضيل الطلاب للدراسة بقسم المحاسبة في إحدي السنوات الدراسية، بإحدى الجامعات. فإذا وجد أن عدد الطلاب في قسم الحاسبة بالصفيين الثالث والرابع 1200 طالب وطالبة منها 700 بالصف الثالث، 500 بالصف الرابع. ويراد سحب عينة عشوائية منتظمة مكونه من 12 طالب وطالبة من الصفين الثالث والرابع تمثل الطلاب محل الدراسة (مجتمع الدراسة).

الحل:

١- يرقم طلاب الصف الثالث من (0001) إلى (0700) وطلاب الصف الرابع من (0701) إلى (1200).

٢- بما أن k=100 فإن:

$$\frac{1}{k}=\frac{r}{N}=\frac{12}{1200}=\frac{1}{100}$$

وبما أن k=100 وبالتالي يتم تقسيم طلاب قسم المحاسبة بالصف الثالث 12 فئات جزئية وكل فئة جزئية بها 100 طالب وطالبة.

٣- بطريقة البطاقات (أو الأعداد العشوائية) يتم اختيار المفردة الأولي عشوائياً ولتكن رقم (82) أي i=82 من المائة مفردة في الفئة الأولي ثم بإضافة k=100 على التوالي إلى الرقم (82) نحصل على مفردات العينة على النحو التالي:

(82), (182), (282), (382), (482), (582), (682), (782), (882), (982), (1082), (1182)

وبالتالي وفقاً للعينة العشوائية المنتظمة يتم اختيار 7 مفردات من طلاب الفصل الثالث (82, 182, 282, 482, 582, 682)، 5 مفردات من طلاب الصف الرابع (782, 882, 982, 1082, 1182).

(٢-٤) العينـــة العشـــوائية الطبقيـــة Stratified Random

إذا كان المطلوب دراسة مستوي أجر المشتغلين بإحدى الشركات باستخدام أسلوب العينة. فنجد أن أجر العامل الفني يختلف عن أجر العامل غير الفني، كذلك يختلف أجر العامل عن أجر الإداري من الدرجة الأولي عن أجر الإداري من الدرجة الثانية ... الخ. وبالتالي فإن المتغير محل الدراسة وهو أجر المشتغل بالشركة يختلف وفقاً لاختلاف العمل الذي يقوم به المشتغل. وهنا نجد أن البيانات المطلوبة عن المتغير محل الدراسة (أجر المشتغل بالشركة) غير متجانسة. فمجتمع الدراسة هنا من حيث الأجر يمكن تقسيمه إلى فئات متجانسة من حيث الأجر الشهري الذي يمثل المتغير محل الدراسة على النحو التالي:

- فئة (مجموعة) العاملين غير الفنيين ولنرمز لعدد مفرداتها بالرمز N_1.
- فئة (مجموعة) العاملين الفنيين ولنرمز لعدد مفرداتها بالرمز N_2.
- فئة (مجموعة) العاملين الإداريين من الدرجة الأولي وعدد مفرداتها N_3.
- فئة (مجموعة) العاملين الإداريين من الدرجة الثانية وعدد مفرداتها N_4.

فإذا فرضنا أننا نرغب في سحب عينة حجمها **n** من المشتغلين بالشركة ممثل فيها جميع المشتغلين فلابد من سحب عينة عشوائية من كل فئة (أو طبقة)، ولتكن n_1 حجم العينة من الطبقة الأولي، و n_2, n_3, n_4 أحجام العينات العشوائية من الطبقات الثانية والثالثة والرابعة على الترتيب. وتكون حجم العينة المطلوبة **n** بحيث:

$$n = n_1 + n_2 + n_3 + n_4 \qquad (2.5)$$

وفي هذه الحالة تسمي هذه العينة بالعينة العشوائية الطبقية **Stratified Random Sample** وتكون هذه العينة ممثلة لجميع فئات (أو طبقات) المجتمع محل الدراسة حيث تسمي كل فئة (أو مجموعة) بالطبقة **Strata (Subgroup)** وتتميز العينة العشوائية الطبقية بأنها تمثل جميع الطبقات في المجتمع محل الدراسة. كذلك استخدام العينة الطبقية يؤدي إلى توافر بيانات عن كل طبقة تمكن من دراسة كل طبقة على حدة. ويتم تقسيم المجتمع إلى طبقات وفقاً لطبيعة وخصائص المتغير محل الدراسة.

مثال (٢-٧)

إذا كان الهدف من الدراسة هو معرفة كيفية قضاء وقت الفراغ في الأجازات الصيفية بالنسبة لطلاب المرحلة الثانوية في مدينة القاهرة عن طريق استخدام عينة من هؤلاء الطلاب . فهنا نجد أن مجتمع الدراسة هو جميع طلاب المرحلة الثانوية بمدينة القاهرة.

ونجد أن قضاء وقت الفراغ في الأجازة الصيفية مرتبط بالمستوي الاقتصادي والاجتماعي والثقافي للأسرة التي ينتمي إليها الطالب والذي غالباً ما ينعكس في مكان سكن الطالب. وبالتالي يمكن تقسيم مدينة القاهرة إلى أحياء مثل حي المعادي، حي السيدة زينب، حي العباسية، ... الخ. وبالتالي يمكن تمثيل عدد الطلاب في الثانوية العامة في كل حي من أحياء القاهرة بطبقة من طبقات المجتمع محل الدراسة.

وبصفة عامة إذا كان لدينا مجتمع حجمه **N** مكون من عدد **k** من الطبقات حيث:

$$N = \sum_{j=1}^{k} N_j \qquad \textbf{(2.6)}$$

حيث N_j تشير إلى عدد المفردات في الطبقة رقم **j**. وبالتالي فإن العينة العشوائية الطبقية التي حجمها **n** هي عبارة عن مجموع العينات العشوائية البسيطة المسحوبة من كل طبقة، أو بعبارة أخري:

$$n = \sum_{j=1}^{k} n_j \qquad \textbf{(2.7)}$$

حيث n_j تمثل حجم العينة العشوائية البسيطة المسحوبة من الطبقة رقم **j**.

وتوجد ثلاثة طرق لتوزيع حجم العينة **n** على الطبقات المختلفة، أو بعبارة أخري تحديد عدد مفردات العينة في كل طبقة، أي تحديد n_j ، **j=1,2,3,…,k** . وسوف نستعرض هذه الطرق الثلاثة باختصار فيما يلي:

أولاً: طريقة التوزيع المتساوي Equal Allocation Method

وتعتمد هذه الطريقة على توزيع مفردات العينة على الطبقات المختلفة بالتساوي، أي عدد المفردات في الطبقة الواحدة يساوي عدد مفردات العينة على عدد الطبقات أو بعبارة أخري:

$$\boxed{n_j = \frac{n}{k} \qquad , \ j = 1,2,\ldots\ldots,k \qquad \textbf{(2.8)}}$$

وبالتالي:

$$n_1 = n_2 = n_3 = \ldots\ldots\ldots = n_k$$

مثال (٢-٨)

إذا كان لدينا مجتمع مكون من ثلاثة طبقات بحيث

$$N_1 = 150 \quad , \quad N_2 = 250 \quad , \quad N_3 = 600$$

أي أن:

$$N = N_1 + N_2 + N_3 = 150 + 250 + 600 = 1000 \text{ مفردة}$$

فإذا كان المطلوب سحب عينة عشوائية طبقية حجمها 75 مفردة، ففي هذه الحالة يكون حجم الطبقة **j** على النحو التالي:

$$n_j = \frac{n}{k} = \frac{75}{3} = 25 \text{ مفردة} \qquad \textbf{J=1,2,3.}$$

وبالتالي حجم كل عينة من العينات العشوائية البسيطة التي يجب سحبها من كل طبقة يساوي 25 مفردة.

ونلاحظ أن توزيع العينة بهذه الطريقة لم يأخذ في الاعتبار حجم كل طبقة، فمثلاً نجد أن حجم الطبقة الثالثة $N_3 = 600$ يمثل 4 أضعاف حجم الطبقة الأولي $N_1 = 150$. ورغم ذلك استخدام هذه الطريقة أدي إلى أن نسبة تمثيل الطبقة الولي في العينة يساوي نسبة تمثيل الطبقة الثالثة في العينة.

ورغم بساطة طريقة التوزيع المتساوي في تحديد حجم العينة في كل طبقة إلا أنها تعتبر طريقة جيدة فقط عندما يكون حجم الطبقات المختلفة متساوي تقريباً، كذلك تقارب الاختلاف في قيم المشاهدات في الطبقات المختلفة. لذلك لا تصلح هذه الطريقة في وجود حالة واحدة على الأقل من الحالتين التاليتين:

١- اختلاف حجم الطبقات في المجتمع محل الدراسة.

٢- اختلاف التشتت (الاختلاف) في القيم في الطبقات المختلفة.

وفي حالة وجود واحدة على الأقل من الحالتين المذكورتين أعلاه فإنه لا يصلح استخدام طريقة التوزيع المتساوي لاختيار العينة ولكن يمكن استخدام إحدى الطريقتين التاليتين

ثانياً: طريقة التوزيع المتناسب

Proportional Allocation Method

وتتميز هذه الطريقة عن طريقة التوزيع المتساوي في أنها تأخذ حجم الطبقات المختلفة في الاعتبار (الحالة الأولي)، حيث يتم تحديد حجم العينة العشوائية البسيطة من الطبقة رقم **j** على النحو التالي:

$$n_j = n\left(\frac{N_j}{N}\right) \qquad j=1,2,\ldots,k \qquad \textbf{(2.9)}$$

مثال (٢-٩)

في المثال السابق إذا استخدمنا طريقة التوزيع المتناسب نجد أن:

$$n_1 = 75\left(\frac{150}{1000}\right) = 11.25 \approx 11 \text{ مفردة}$$

$$n_2 = 75\left(\frac{250}{1000}\right) = 18.7 \approx 19 \text{ مفردة}$$

$$n_3 = 75\left(\frac{600}{1000}\right) = 45 \text{ مفردة}$$

وبهذه الطريقة نجد أن حجم العينة المسحوبة من كل طبقة متناسب مع حجم الطبقة في المجتمع محل الدراسة.

ثالثا: طريقة التوزيع الأمثل

Optimum Distribution′s Method

وتعتبر هذه الطريقة هي الطريقة المثلي لتحديد حجم العينة العشوائية البسيطة المأخوذة من أي طبقة من الطبقات المختلفة. حيث تأخذ هذه الطريقة في الاعتبار حجم كل طبقة في المجتمع. كذلك تأخذ في الاعتبار تشتت المتغير محل الدراسة داخل كل طبقة (الحالة الثانية). ويتم تحديد حجم العينة في الطبقة j على النحو التالي:

$$n_j = n\left[\frac{\sigma_j N_j}{\sigma_j \sum_{j=1}^{k} N_j}\right] \qquad j = 1,2,\ldots\ldots,k \qquad \textbf{(2.10)}$$

حيث σ_j يمثل الانحراف المعياري للمتغير محل الدراسة في الطبقة رقم j.

ملحوظة: عندما يتساوي الانحراف المعياري في الطبقات المختلفة أي عندما تتساوي قيم σ_j، j=1,2,...,k . فإن حجم العينة في الطبقة j باستخدام طريقة التوزيع الأمثل يساوي حجم العينة في الطبقة j باستخدام طريقة التوزيع المتناسب حيث نجد أن:

$$n_j = n\left[\frac{\sigma_j N_j}{\sum_{j=1}^{k} \sigma_j N_j}\right] = n\left[\frac{\sigma_j N_j}{\sigma_j \sum_{j=1}^{k} N_j}\right] = n\left[\frac{N_j}{\sum_{j=1}^{k} N_j}\right] = n\left[\frac{N_j}{N}\right]$$

(2.11)

مثال (٢-١٠)

إذا كان لدينا مجتمع حجمه 10000 مفردة مقسم إلى 5 طبقات، أي j=1,2,3,4,5 بحيث:

$$N_1 = 1500 \ , \ N_2 = 2000 \ , \ N_3 = 500 \ ,$$
$$N_4 = 3000 \ , \ N_5 = 3000$$
$$\sigma_1 = 2 \ , \ \sigma_2 = 1 \ , \ \sigma_3 = 5 \ , \ \sigma_4 = 56 \ , \ \sigma_5 = 2$$

ويراد سحب عينة عشوائية طبقية حجمها 200 مفردة. حدد حجم العينة المسحوبة من كل طبقة باستخدام:

1. **طريقة التوزيع المتساوي.**
2. **طريقة التوزيع المتناسب.**
3. **طريقة التوزيع الأمثل.**
4. **عقب على النتائج التي تم الحصول عليها في (١) ، (٢) ، (٣).**

الحل:

١- باستخدام طريقة التوزيع المتساوي. حجم العينة في الطبقة **j** يساوي n_j حيث:

$$n_j = \frac{n}{k} = \frac{200}{5} = 40 \text{ مفردة} \qquad J=1,2,3,4,5 \qquad (2.12)$$

٢- باستخدام طريقة التوزيع المتناسب نجد أن:

$$n_j = n\left(\frac{N_j}{N}\right) \qquad j=1,2,3,4,5$$

$$n_1 = 200\left[\frac{1500}{10000}\right] = 30 \text{ مفردة} \qquad (2.13)$$

$$n_2 = 200\left[\frac{2000}{10000}\right] = 40 \text{ مفردة} \qquad (2.14)$$

$$n_3 = 200\left[\frac{500}{10000}\right] = 10 \text{ مفردات} \qquad (2.15)$$

$$n_4 = 200\left[\frac{3000}{10000}\right] = 200 \text{ مفردة} \qquad (2.16)$$

$$n_5 = 200\left[\frac{3000}{10000}\right] = 200 \text{ مفردة} \qquad (2.17)$$

٣- باستخدام طريقة التوزيع الأمثل نجد أن:

$$\boxed{n_j = n\left[\frac{\sigma_j N_j}{\sum_{j=1}^{5}\sigma_j N_j}\right] \qquad j=1,2,3,4,5 \qquad (2.18)}$$

وبما أن:

$$\sum_{j=1}^{5}\sigma_j N_j = 2(1500)+1(2000)+5(500)+6(3000)+$$
$$2(3000)=31500$$

$$n_1 = 200\left[\frac{2(1500)}{31500}\right] \approx 19 \text{ مفردة} \qquad (2.19)$$

$$n_2 = 200\left[\frac{1(2000)}{31500}\right] \approx 13 \text{ مفردة} \qquad (2.20)$$

$$n_3 = 200\left[\frac{5(500)}{31500}\right] \approx 16 \text{ مفردة} \qquad (2.21)$$

$$n_4 = 200\left[\frac{6(3000)}{31500}\right] \approx 114 \text{ مفردة} \qquad (2.22)$$

$$n_5 = 200\left[\frac{2(3000)}{31500}\right] \approx 38 \text{ مفردة} \qquad (2.23)$$

٤- مما سبق يتضح ما يلي:

أ- طريقة التوزيع المتساوي أدت إلى سحب 40 مفردة من كل طبقة ولم يراعي اختلاف حجم الطبقات. كذلك لم يراعي مدي الاختلاف داخل كل طبقة الممثل في قيمة الانحراف المعياري، فنجد مثلاً أن حجم الطبقة الثالثة $N_3 = 500$ أقل من حجم الطبقة الخامسة $N_5 = 3000$. ورغم ذلك باستخدام التوزيع المتساوي نجد أن حجم العينة المأخوذة من الطبقة الثالثة مساوي لحجم العينة المأخوذة من الطبقة الخامسة.

ب- من العلاقات (2.13)-(2.17) نجد ان حجم العينة المأخوذة من كل طبقة باستخدام طريقة التوزيع المتناسب يتناسب مع حجم كل طبقة. فنجد أن

$n_4 = n_5 = 60$ مفـردة، حيـث أن حجـم الطبقـة الرابعـة والخامسـة متسـاوي. أي أن $N_4 = N_5 = 3000$ مفـردة. ولكـن هـذه الطريقـة لـم تأخذ في الاعتبار مدي تجانس قيم المتغير محل الدراسة في الطبقات المختلفة. فنجد أن $\sigma_4 = 6$ ، $\sigma_5 = 2$ أي أن قيم المتغير في الطبقة الرابعة أكثر تشتت (أقل تجانس) عن قيم المتغير في الطبقة الخامسة.

ج- ومـن العلاقـات (2.19)-(2.23) نجد أن توزيـع مفـردات العينـة علـى الطبقـات المختلفة باستخدام طريقة التوزيع الأمثل أخذ في الاعتبار حجم الطبقة كذلك أخذ مدي التجانس (أو الاختلاف) للمتغير داخل كل طبقة. فمثلاً نجد أن حجم الطبقة الرابعة مساوي لحجم الطبقة الخامسة حيث $N_4 = N_5 = 3000$ ولكن التشتت في الطبقة الرابعة أكبر من التشتت في الطبقة الخامسة حيث $\sigma_4 = 6$ ، $\sigma_5 = 2$ وانعكس ذلك في حجم العينة المأخوذة من الطبقة الرابعة والخامسة. فنجد أن $n_4 = 114$ مفردة، $n_5 = 38$ مفردة

(٢-٥) العينة العشوائية العنقودية

Cluster Random Sample

في الفصول السابقة أتضح أن اختيار عينة عشوائية بسيطة أو عينة عشوائية منتظمة أو عينة عشوائية طبقية يتطلب تحديد إطار لجميع مفردات المجتمع محل الدراسة (تحديد مفردات المجتمع). وفي كثير من الحالات قد يكون من الصعب أو من المستحيل تحديد هذا الإطار.

فمثلاً إذا كانت الدراسة عن إصابة الأفراد في المجتمع المصري بمرض البلهارسيا في سنة ما (السنة محل الدراسة). فهنا نجد أن مجتمع البحث هو جميع المصابين بالمرض في جميع المحافظات من ريف وحضر في هذه السنة. وهنا يكون تحديد مفردات مجتمع الدراسة صعب جداً، فذلك التحديد يتطلب إجراء الفحوص على جميع السكان وهذا مكلف جداً وقد يكون غير ممكن مما يترتب عليه تعذر تحديد جميع المصابين بالبلهارسيا في جميع محافظات الجمهورية (ريف وحضر).

كذلك قد يكون تحديد إطار مفردات المجتمع محل الدراسة ممكن ولكن ذلك يتطلب تكاليف باهظة ووقت طويل وجهد كبير جداً قد يتعذر توافرهم.
فمثلاً إذا كان الهدف من الدراسة تحديد مستوي تعليم الأم في الأسرة المصرية، ففي هذه الحالة تحديد إطار مفردات المجتمع يتطلب تحديد مستوي تعليم الأم في جمهورية مصر العربية وهذا ممكن لكنه يتطلب جهد كبير ووقت لإجراء حصر لجميع الأسر بالإضافة إلي التكاليف الباهظة نظراً لتناثر الأسر في جميع المحافظات المختلفة.

وللتغلب على المشاكل المذكورة أعلاه فيما يختص بإيجاد إطار يحدد مفردات المجتمع فإننا يمكن استخدام عينة عشوائية عنقودية لا يتطلب تحديد مفردات المجتمع. وسوف نوضح تعريف وأسلوب اختيار مفردات العينة العشوائية العنقودية من خلال المثال التالي:

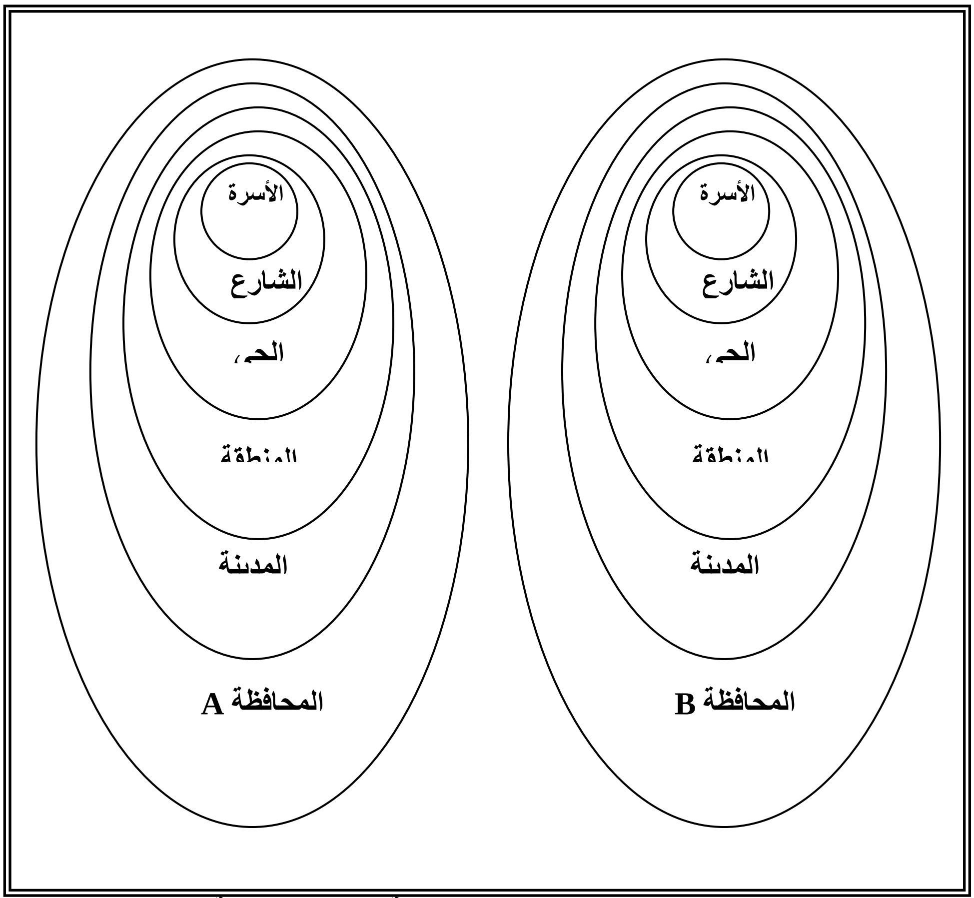
مثال (٢-١١)

في إحدى الدراسات عن تحديد مستوي تعليم الأم في الأسرة في مجتمع الحضر في جمهورية مصر العربية فإننا يمكن تقسيم الجمهورية إلى محافظات وبالتالي يمكن اختيار عينة عشوائية بسيطة (أو منتظمة أو طبقية) من المحافظات وعن طريق عينة المحافظات المختارة يتم تحديد مستوي تعليم الأم في جميع الأسر بالمحافظات المختارة في العينة ولم يتطلب هذا تحديد جميع الأسر في المجتمع وتسمي العينة في هذه الحالة بالعينة العشوائية العنقودية ذات المرحلة الواحدة Single-Stage Cluster Sample.

ومن الممكن تقسيم المحافظات التي تم اختيارها في عينة المحافظات السابقة إلى مدن ثم اختيار عينة عشوائية بسيطة (أو منتظمة أو طبقية) من المدن في كل محافظة من المحافظات المختارة في العينة، ويتم تحديد مستوي تعليم الأم في الأسر في المدن المختارة من الحافظات المختارة. وفي هذه الحالة تسمي العينة بالعينة العشوائية العنقودية ذات المرحلتين Tow-Stage Cluster Sample.

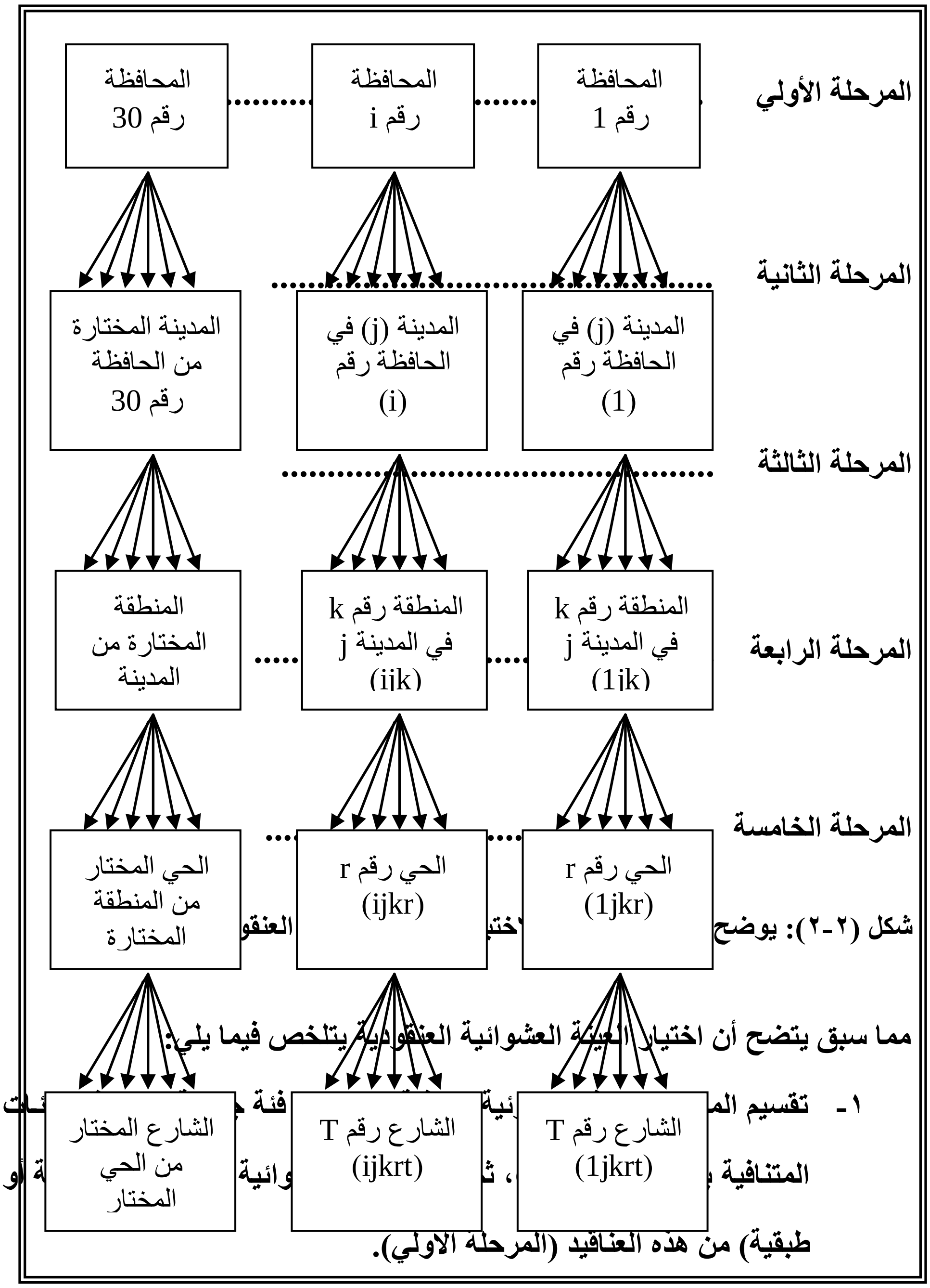
كذلك من الممكن أن تكون العينة العشوائية على ثلاث مراحل إذا تم تقسيم المدن إلى مناطق، أو أربعة مراحل إذا تم تقسيم المناطق إلى أحياء، أو خمسة مراحل إذا تم تقسيم الأحياء إلى شوارع.

ويوضح شكل (٢-١) أن الأسرة تمثل عناصر في فئة الشارع المختار، كذلك الشوارع المختارة تمثل فئة جزئية من الحي المختار، والأحياء المختارة تمثل فئة جزئية من المنطقة المختارة، والمناطق المختارة تمثل فئة جزئية من المدينة، والمدن تمثل فئة جزئية من المحافظة المختارة.



شكل (٢-١): يوضح الفئات الجزئية التي تشكل العينة

وشكل (٢-٢) يوضح المراحل المختلفة لتحديد العينة العنقودية متعددة المراحل(خمس مراحل في هذا المثال).



شكل (٢-٢): يوضح ... العنقو...

مما سبق يتضح أن اختيار العينة العشوائية العنقودية يتلخص فيما يلي:

١- تقسيم الم... فئة ج... ات المتنافية ب... ، ث... وائية ... ة أو طبقية) من هذه العناقيد (المرحلة الاولي).

٢- من العنقود أو العناقيد المختارة في (١) يتم تقسيم كل عنقود إلى فئات جزئية متنافية ثم يتم اختيار عينة عشوائية بسيطة (أو منتظمة أو طبقية) من هذه الفئات الجزئية (المرحلة الثانية).

٣- نستمر في تقسيم كل فئة جزئية في العنقود إلى أن نصل إلى المفردة محل البحث بعد عدد مناسب من المراحل كما هو موضح في شكل (٢-٢).

وتتميز العينة العنقودية بأنها تمثل المجتمع محل الدراسة بالإضافة إلى أنها لا تتطلب تحديد جميع مفردات المجتمع.

(٦-٢) العينات غير العشوائية Non-Random Samples

وهذا النوع من العينات غير العشوائية (غير الاحتمالية) تكون عينات متحيزة ولا تمثل المجتمع محل الدراسة وبالتالي أي نتائج مستخلصة من هذه العينات لا يمكن تعميمها على المجتمع حيث يحدد الباحث مفردات العينة، أي يكون تحديد مفردات العينة تحكمي Judgment لا يخضع للاختيار العشوائي ويرجع إلى التقدير الشخصي للباحث ولا يخضع لأي معيار موضوعي، ومثال ذلك العينات المأخوذة من مرضى معينين لإجراء اختبار أحد العقاقير عليهم، أو العينات المأخوذة من طلاب في مستوي معين لإجراء دراسات عليهم. وبالتالي فإن النتائج المستخلصة من هذه العينات تكون متحيزة ومن أمثلة هذه العينات العينة العمدية والعينة الحصصية.

العينة العمدية Purposive Sample

ويتم تحديد مفردات العينة العمدية بواسطة الباحث (أي يختار متعمداً مفردات معينة من المجتمع) لذا سميت بالعينة العمدية.

فمثلاً عند اختيار عينة مكونة من 5 طلاب من أحد الفصول الدراسية بالمرحلة الإعدادية بمدرسة معينة لتحديد متوسط مستوي الطالب بهذا الفصل. فإذا قام مدرس الفصل بإختيار الطلاب الخمسة فإن هذه العينة تكون عينة عمدية حيث اختيار هؤلاء الطلاب يخضع لتقدير خبرة المدرس الشخصية فقط، في حين إذا قام مدرس أخر بالمدرسة باختيار 5 طلاب بنفس الفصل فإنه سوف يختار 5 طلاب أخرين. لذا فإن النتائج المستخلصة من هذا النوع من العينات تكو متحيزة ولا يمكن أن تمثل مجتمع البحث.

العينة الحصصية Quota Sample

ويتم تحديد مفردات العينة الحصصية بتقسيم المجتمع محل الدراسة إلى طبقات وفقا لخاصية أو مجموعة من الخصائص المعينة، ثم يتم توزيع حجم العينة على الطبقات المختلفة توزيعاً يتناسب مع حجم كل طبقة من الطبقات ثم تحديد مفردات العينة في كل طبقة بطريقة تحكمية (وليست عشوائية) تعتمد على الباحث.

وبالتالي تختلف العينة الحصصية عن العينة العشوائية الطبقية (باستخدام طريقة التوزيع المناسب) في أسلوب اختيار مفردات العينة في كل طبقة فقط. ففي العينة الحصصية يقوم الباحث باختيار المفردات أي تحديد المفردات بطريقة تحكمية (تعتمد على اتجاهه وخبرته وقدراته) أما في العينة العشوائية الطبقية فتحديد مفردات العينة في كل طبقة يتم بطريقة عشوائية (احتمالية).

مما سبق يتضح أن النتائج المستخلصة من العينات غير العشوائية (التحكمية) لا يمكن أن تستخدم في التعميم على المجتمع محل الدراسة. ولكن يستخدم هذا النوع من العينات غير العشوائية في استطلاعات الرأي والاتجاهات في ظواهر أو مشاكل معينة مثل مشاكل التفاوض مع صندوق النقد في مشاكل الديون الخارجية أو بعبارة أخري في المشاكل التي يلعب فيها الباحث دوراً هاما في تحديد مفردات العينة، وبالتالي في النتائج المستخلصة ومثال ذلك أيضا التجارب المعملية.

(٢-٧) تطبيقات Applications

تطبيق (٢-١): حدد أي نوع من العينات التي يجب سحبها في كل حالة من الحالات التالية:

١- دراسة كمية الزيوت النباتية والحيوانية المستهلكة شهرياً في الأسرة المصرية بمدينة القاهرة.

٢- دراسة متوسط دخل الأسرة في جمهورية مصر العربية.

٣- دراسة متوسط إنتاجية الفدان من القطن في سنة ما في جمهورية مصر العربية.

٤- الإصابة بأمراض سوء ونقص الأغذية بإحدى المدارس الابتدائية التابعة للتأمين الصحي.

٥- دراسة مستوي العلاج بالتأمين الصحي للأطفال التابعين للتأمين الصحي بالمرحلة الابتدائية.

الحل:

١- بما أن الهدف من سحب العينة دراسة كمية الزيوت النباتية والحيوانية المستهلكة شهرياً في الأسرة المصرية بمدينة القاهرة. بالتالي فإن مجتمع الدراسة هو جميع الأسر بمدينة القاهرة حيث تمثل الأسرة المفردة محل الدراسة.

وبما أن المتغير محل الدراسة هو كمية الزيوت المستهلكة شهرياً في الأسرة المصرية بمدينة القاهرة وهذا المتغير مرتبط بمستوي معيشة الأسرة الاقتصادي والاجتماعي. لذا فإن تقسيم مدينة القاهرة إلى أحياء وفقاً لمستوي

معيشة الأسرة (من الناحية الاقتصادية والاجتماعية معاً) يصبح ضرورة نظراً لتفاوت مستوي المعيشة في الأحياء المختلفة، وبالتالي فإن تقسيم القاهرة إلى أحياء مثل حي المعادي، حي مصر الجديدة، حي المنيل، ... الخ يعكس الاختلاف في مستوي المعيشة للأسرة وبالتالي لاستهلاكها للزيوت الحيوانية والنباتية.

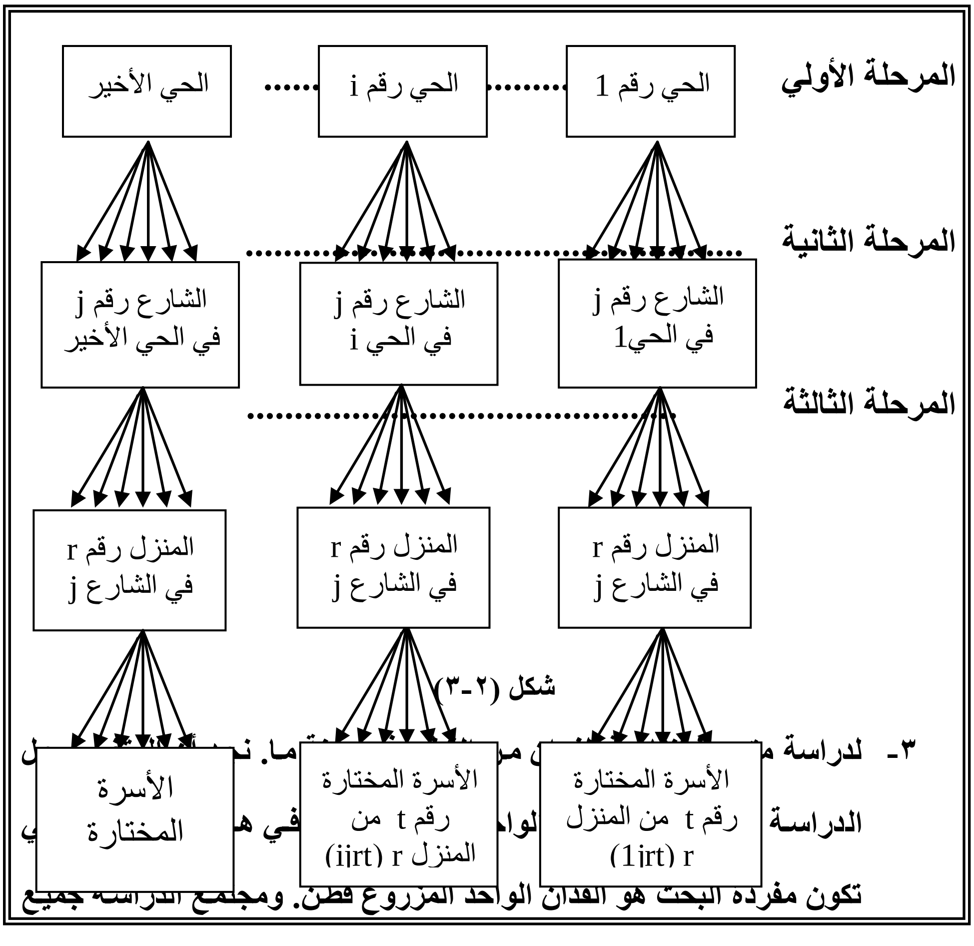
وبالتالي تصبح العينة العشوائية الطبقية باستخدام طريقة التوزيع الأمثل هي أنسب أنواع العينات التي يمكن استخدامها في هذه الدراسة.

وفي حالة تعذر وضع إطار للمجتمع محل الدراسة لاستخدام العينة الطبقية المقترحة لأسباب تتعلق بالنفقات والوقت المطلوب، لذا يمكن استخدام العينة العشوائية العنقودية على ثلاث مراحل مثلاً في المرحلة الأولي يتم تقسيم القاهرة إلى أحياء، والمرحلة الثانية يتم تقسيم الحي إلى شوارع، والمرحلة الثالثة يتم تقسيم الشوارع إلى منازل ويوضح الشكل (٢-٣) هذه الحالة.

٢- بما أن الهدف من سحب العينة دراسة متوسط دخل الأسرة في جمهورية مصر العربية. فإن المفردة في المجتمع محل الدراسة هي الأسرة ومجتمع الدراسة جميع الأسر المصرية. وبما أنه من الصعب تحديد جميع مفردات مجتمع الدراسة. لذا تستخدم العينة العشوائية العنقودية متعددة المراحل. وبما أن المتغير محل الدراسة هو متوسط دخل الأسرة وهذا المتغير مرتبط بالأحياء التي تعيش فيها الأسرة. لذا يمكن تقسيم الجمهورية إلى محافظات (المرحلة الأولي) ثم تقسيم كل محافظة إلى مناطق (المرحلة الثانية) ثم تقسيم كل منطقة إلى مدن وقري (المرحلة الثالثة) ثم تقسيم المدن والقرى إلى أحياء (

المرحلة الرابعة) ثم تقسيم الأحياء إلى شوارع (المرحلة الخامسة) ثم تقسيم الشوارع إلى منازل (المرحلة السادسة).

وبالتالي فإن العينة العشوائية العنقودية ذات المراحل الست (ممكن أن تكون المراحل أقل أو أكثر) هي أنسب أنواع العينات التي يمكن استخدامها في هذه الدراسة.



شكل (٢-٣)

٣- لدراسة ... ما. ...

الدراسة ... في ...

تكون مفردة البحث هو الفدان الواحد المزروع قطن. ومجتمع الدراسة جميع الأفدنة المزروعة قطن في هذه السنة. وبما أنه يتعذر تحديد جميع الأفدنة المزروعة قطن كذلك اختلاف إنتاجية الفدان الواحد وفقاً للمحافظة المزروع

بهـا ووفقـاً للحـوض لاخـتلاف خصـوبة الأرض. لـذلك يكـون مـن المناسـب استخدام العينة العشوائية العنقودية ذات المراحل المتعددة على النحو التالي:

المرحلة الأولي: تقسيم الجمهورية في هذه السنة وفقاً لزراعة القطن إلى محافظات.

المرحلة الثانية: تقسيم كل محافظة إلى مراكز.

المرحلة الثالثة: تقسيم كل مركز إلى قرى.

المرحلة الرابعة: تقسيم كل قرية إلى أحواض.

٤- فـي هـذه الحالـة تكـون أنسـب عينـة هـي العينـة العمديـة التـي يقـوم باختيارهـا الطبيب المقيم بالمدرسة.

٥- في هذه الحالـة يكـون مجتمـع الدراسـة هـو جميـع الأطفـال بالمرحلـة الابتدائيـة الخاضعين لنظام التأمين الصحي ويكون الطفل مـن هـؤلاء الأطفـال هـو مفـردة البحـث. ونظـراً لاخـتلاف مسـتوي الخدمـة بالتـأمين الصـحي فـي المحافظـات المختلفة لذا تكون العينـة الطبقيـة المسـحوبة باسـتخدام التوزيـع الأمثـل أنسـب عينة وفي حالة تفاوت مستوي الخدمة بالتامين الصحي في المحافظة الواحدة (وفقاً للحي مثلاً) فإنه يفضـل فـي هـذه الحالـة اسـتخدام العينـة العنقوديـة علـى ثلاث مراحل:

المرحلة الأولي: يتم تقسيم الأطفال وفقاً للمحافظات.

المرحلة الثانية: يتم تقسيم المحافظات إلى مدن وقرى.

المرحلة الثالثة: يتم تقسيم المدينة أو القرية إلى أحياء.

تطبيق (٢-٢): أعلنت إحدى الشركات الاستثمارية عن حاجتها إلى 5 أفراد تخصص تصميم برامج حاسب. فتقدم لهذه الوظائف 15 فرد تنطبق عليهم الشروط المطلوبة. والجدول التالي يوضح المتقدمين وفقاً لسنوات الخبرة، الجنس، الإقامة. والمطلوب:

١- أوجد عدد العينات العشوائية البسيطة الممكن سحبها.

٢- أسحب عينة عشوائية منتظمة من المتقدمين.

٣- اسحب عينة عشوائية طبقية وفقاً للجنس باستخدام طريقة التوزيع المتناسب.

جدول (٢-٣)

الرقم	الاسم	الجنس	محل الإقامة	سنوات الخبرة
1	أحمد صفوت محمود	ذكر	القاهرة	10 سنوات
2	ميخائيل سلامة	ذكر	الإسكندرية	7 سنوات
3	فاتن على حسن	أنثي	القاهرة	8 سنوات
4	حسام الدين عزت	ذكر	القاهرة	9 سنوات
5	داليا حافظ حسني	أنثي	الجيزة	9 سنوات
6	عصام الدين محمود	ذكر	السويس	5 سنوات
7	أحمد عادل فوزي	ذكر	القاهرة	8 سنوات
8	إيمان حسين على	أنثي	بور سعيد	7 سنوات
9	نادية كامل حسام	أنثي	القاهرة	4 سنوات
10	نادر حسن سعد	ذكر	الجيزة	10 سنوات
11	علاء الدين واصف	ذكر	السويس	10 سنوات
12	كمال فؤاد عزت	ذكر	الإسكندرية	10 سنوات
13	محمد يوسف أحمد	ذكر	القاهرة	10 سنوات

الرقم	الاسم	الجنس	محل الإقامة	سنوات الخبرة
14	حسين محمود سامي	ذكر	القاهرة	10 سنوات
15	خالد محمود خالد	ذكر	الجيزة	10 سنوات

الحل:

١- بما أن حجم المجتمع N=15 ، وحجم العينة n=5.

$\therefore$ عدد العينات العشوائية البسيطة الممكنة يساوي C_n^N حيث:

$$C_n^N = C_5^{15} = \frac{5!}{5!\times 10!} = 3003$$ عينة

٢- بما أن حجم المجتمع N=15، وحجم العينة المراد سحبها n=5 مفردات

$\therefore$ يمكن تقسيم المجتمع إلى 5 مجموعات بحيث تحتوي كل مجموعة على

$$\frac{N}{n} = \frac{15}{5} = 3$$

المجموعة الأولي من [1-3]، والمجموعة الثانية [4-6]، والمجموعة الثالثة [7-9]، والمجموعة الرابعة [10-12]، والمجموعة الخامسة [13-15]. فإذا أخذنا مفردة واحدة عشوائياً باستخدام الكروت مثلاً من المجموعة الأولي ولتكن المفردة رقم (2) وبالتالي تكون المفردة الثانية هي المفردة (5=3+2)، والمفردة الثالثة هي المفردة رقم (8=3+5)، والمفردة الرابعة هي المفردة رقم (11=3+8)، والمفردة الخامسة هي المفردة رقم (14=3+11). ومن ثم تصبح مفردة العينة هي:

ميخائيل سلامة	2
داليا حافظ حسن	5
إيمان حسين على	8
علاء الدين واصف	11
حسين محمود سامي	14

٣- إذا تم تقسيم المتقدمين وفقاً للجنس إلى طبقتين: الطبقة الأولي تمثل الذكور، والطبقة الثانية تمثل الإناث. وبما أن عدد الذكور المتقدمين يساوي $N_1 = 11$ ، وعدد الإناث يساوي $N_2 = 4$ ، وحجم العينة **n=5**. وبالتالي نفرض أن حجم العينة من الطبقة الأولي والثانية هما n_2, n_1 على الترتيب

$$n_1 = n\left(\frac{N_1}{N}\right) = 5\left(\frac{11}{15}\right) = 3.67 \approx 4$$ مفردات

$$n_2 = n\left(\frac{N_2}{N}\right) = 5\left(\frac{4}{15}\right) = 1.33 \approx 1$$ مفردة

وباستخدام طريقة الكروت نقوم بسحب عينة عشوائية بسيطة من الذكور عدد مفرداتها 4 مفردات. ولتكن:

عصام الدين محمود	6	أحمد صفوت	1
خالد محمود خالد	15	كمال فؤاد عزت	12

وباستخدام الكروت أيضاً نقوم بسحب مفردة واحدة من طبقة الإناث ولتكن المفردة

فاتن على حسن 3

تطبيق (٢-٣): في إحدى مؤسسات الطباعة يوجد بها 50 عامل طباعة، فإذا قامت المؤسسة باستيراد ماكينات طباعة حديثة وترغب في تدريب عينة من العاملين على هذه الماكينات مكونه من 5عمال. والمطلوب:

١- حدد عدد العينات العشوائية الممكن سحبها.

٢- حدد الخطوات التي يجب إتباعها لاختيار عينة عشوائية بسيطة باستخدام جداول الأعداد العشوائية.

الحل:

١- بما أن في هذه الحالة يكون السحب بدون إرجاع حيث حجم المجتمع N=50، وحجم العينة n=5. وبالتالي فإن:

عدد العينات الممكنة = $C_n^N = C_5^{50} = 2114760$ طريقة

٢- لسحب عينة عشوائية بسيطة بطريقة جداول العداد العشوائية. تقوم بإعطاء أرقام لجميع العاملين بالمطبعة [00-99] ثم نختار أي عمود في جدول العداد العشوائية وليكون العمود الأول ونأخذ منه 5 أعداد متتالية (حجم العينة) من أعلي العمود فتكون على النحو التالي:

(71794) , (64017) , (43747) , (59688) , (95016) , (65207) , (34510)

ونظراً لأن العددين (94 , 88) خارج مجتمع الدراسة. لذا تم استبعاد العددين (71794) , (59688) وتم استبدالهما بالعددين (25207) , (34510).

وبما أن عدد مفردات المجتمع محل الدراسة تتراوح بين [00-49].

∴ يتم أخذ الرقم الأول والثاني فقط من الأعداد على النحو التالي:

17, 47, 16, 7, 10

تطبيق (٢-٤): تقوم إحدى الشركات بإنتاج ثلاثة أنواع من اللمبات الكهربائية A,B,C وفقاً للحجم. فإذا كان إنتاج الشركة اليومي من الأنواع الثلاثة على النحو التالي 12000, 8000, 10000 على الترتيب. والمراد اختبار الإنتاج اليومي باستخدام عينة مكونه من 150 لمبة. حرر أي نوع من العينات يكون مناسباً لاستخدامه في هذه الحالة مع التعليل.

الحل:

بما أن إنتاج الشركة من كل نوع متماثل تماماً فإننا في هذه الحالة يمكننا استخدام العينة العشوائية الطبقية وتم اختيارها باستخدام أسلوب التوزيع المتناسب على النحو التالي:

$$\because N_1 = 10000, \quad N_2 = 8000, \quad N_3 = 12000, \quad n = 150$$

$$\because n_j = n\left(\frac{N_j}{\sum_{j=1}^{3} N_j}\right)$$

$$\therefore n_1 = 150\left(\frac{10000}{30000}\right) = 50$$ لمبة من النوع **A**

$$\therefore n_2 = 150\left(\frac{8000}{30000}\right) = 40$$ لمبة من النوع **B**

$$\therefore n_3 = 150\left(\frac{12000}{30000}\right) = 60$$ لمبة من النوع **C**

تطبيق (٢-٥): قامت إحدى الجامعات المصرية بإصدار مجلة علمية في علوم الإحصاء وتطبيقاتها، ومن الأهمية معرفة رأى المتخصصين المصريين من القرار في هذه المجلة من خلال عينة حجمها **100** مفردة من المتخصصين.

والمطلوب: حدد أي نوع من العينات يكون أنسب في هذه الحالة

الحل:

بما أن المتخصصين المصريين في الإحصاء قد يكونوا مقيمين داخل الوطن أو خارجة، كذلك يتوقف رأي المتخصص على درجته العلمية (أستاذ، أستاذ مساعد، مدرس، مدرس مساعد، معيد) كذلك على جهة العمل.

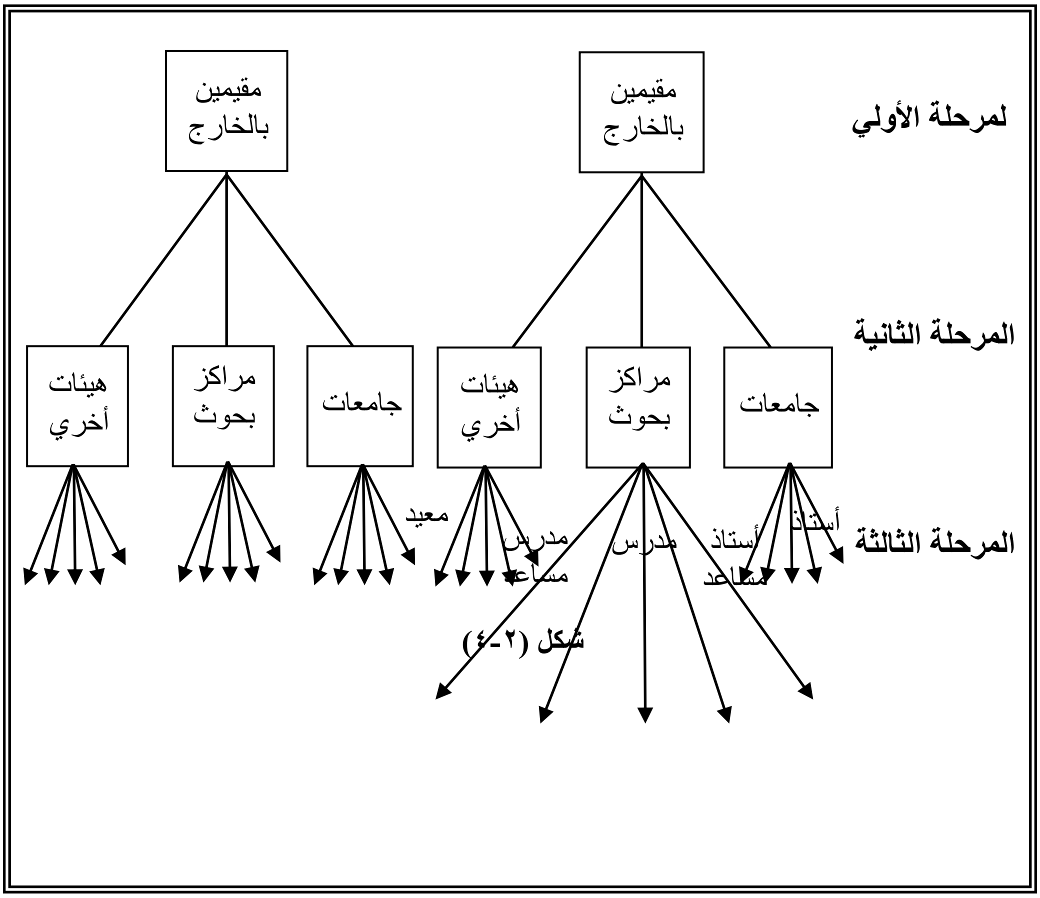
في هذه الحالة يمكن استخدام العينة العشوائية العنقودية على 3 مراحل على النحو التالي:

المرحلة الأولي: يتم تقسيم المتخصصين إلى مقيمين بأرض الوطن وفي الخارج.

المرحلة الثانية: تقسيم المتخصصين وفقاً لجهة العمل مثل الجامعات، مراكز البحوث، هيئات أخري (مثل البنوك، الشركات، ...).

المرحلة الثالثة: تقسيم المتخصصين وفقاً للدرجة العلمية.

ويوضح الشكل التالي هذه الحالة:



شكل (٢-٤)

تطبيق (٢-٦): حدد أي نوع من العينات يجب سحبها لدراسة كل حالة من الحالات التالية:

١- مدي تتبع المشاهدين للتليفزيون من سكان القاهرة لإحدى البرامج الاقتصادية الأسبوعية التي تقدم الساعة العاشرة مساءاً.

٢- الاستهلاك اليومي للمياه الغازية بالقرى السياحية بالساحل الشمالي بجمهورية مصر العربية.

٣- عدد الليالي السياحية التي يقضيها السائح الاجنبي في مدينة الأقصر في شهر ديسمبر.

الحل:

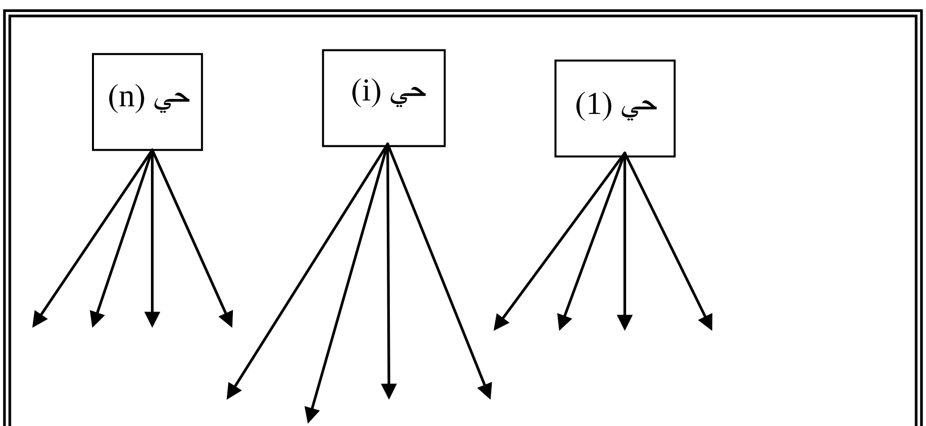
١- ∴ متابعة البرامج الاقتصادية التليفزيونية تختلف وفقاً للمستوي الثقافي والاجتماعي للمشاهد، كذلك التعبئة العمرية للمشاهد بالإضافة إلى جودة وأهمية البرنامج الاقتصادي، كذلك جنس المشاهد (ذكر أو أنثي) لذا في هذه الحالة يمكن سحب عينة عشوائية عنقودية على ثلاثة مراحل حيث:

المرحلة الأولي: يتم تقسيم القاهرة إلى أحياء سكنية بحيث تعكس المستوي الاقتصادي والاجتماعي والثقافي للمشاهدين مثل حي مصر الجديدة، حي السيدة زينب، ....

المرحلة الثانية: تقسيم مشاهدي التليفزيون في كل حي إلى فئات عمرية مثل من -15، -20، -25، -30، ...، 65 فأكثر.

المرحلة الثالثة: تقسيم المشاهدين في كل فئة عمرية إلى ذكور وإناث.

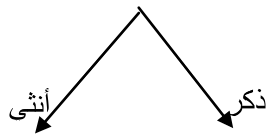
والشكل التالي يوضح ذلك:



المرحلة الأولي ....

المرحلة الثانية **15- 20- .... 65 فأكثر**

المرحلة الثالثة



شكل (٢-٥)

٢- في هذه الحالة تكون المفردة في العينـة هي مراكز توزيـع الميـاه الغازيـة فـي القـرى السـياحية حيـث يمكـن أن تضـيف هـذه المراكـز إلـى 3 أنـواع وفقـاً للكمية التي تقوم بتوزيعها يومياً على النحو التالي:

أ) مراكز صغيرة (تقوم بتوزيع أقل من 50 زجاجة يومياً).

ب) مراكز متوسطة (تقوم بتوزيع 200-50 زجاجة يومياً).

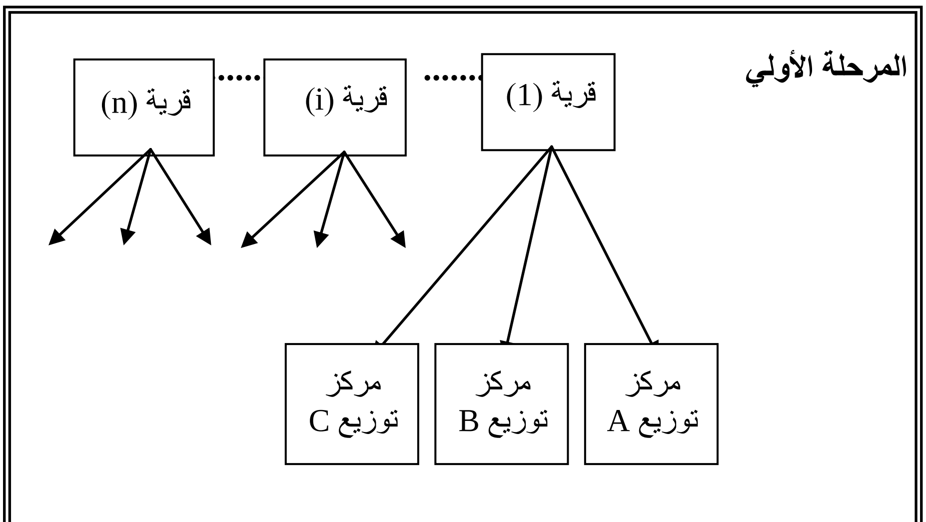
ج) مراكز كبيرة (تقوم بتوزيع أكثر من 200 زجاجة يومياً).

وفـي هـذه الحالـة يمكـن اسـتخدام العينـة العشـوائية العنقوديـة علـى مـرحلتين علـى النحو التالي:

المرحلة الأولي: يتم تقسيم القرى السياحية وفقاً لكثافة النزلاء في كل قرية.

المرحلـة الثانيـة: يـتم تقسـيم مراكـز التوزيـع إلـى الأنـواع A,B,C المـذكورة أعلاه.

كما هو موضح بالشكل التالي:



المرحلة الثانية

شكل (٢-٦)

كذلك يمكن استخدام العينة الطبقية وفي هذه الحالة سوف تكون مطابقة تقريباً للعينة العنقودية السابقة حيث سوف تعتبر الطبقات هي مراكز التوزيع وفقاً للكمية التي يقوم المركز بتوزيعها.

فالطبقة الأولي: هي مراكز التوزيع الصغيرة.

الطبقة الثانية: هي مراكز التوزيع المتوسطة.

الطبقة الثالثة: هي مراكز التوزيع الكبيرة.

٣- تتوقف عدد الليالي السياحية التي يقضيها السائح الأجنبي في مدينة الأقصر في شهر ديسمبر على (يعتبر السائح هو مفردة البحث):

١) مستوي الفندق (5 نجوم، 4 نجوم، 3 نجوم، نجمتان).

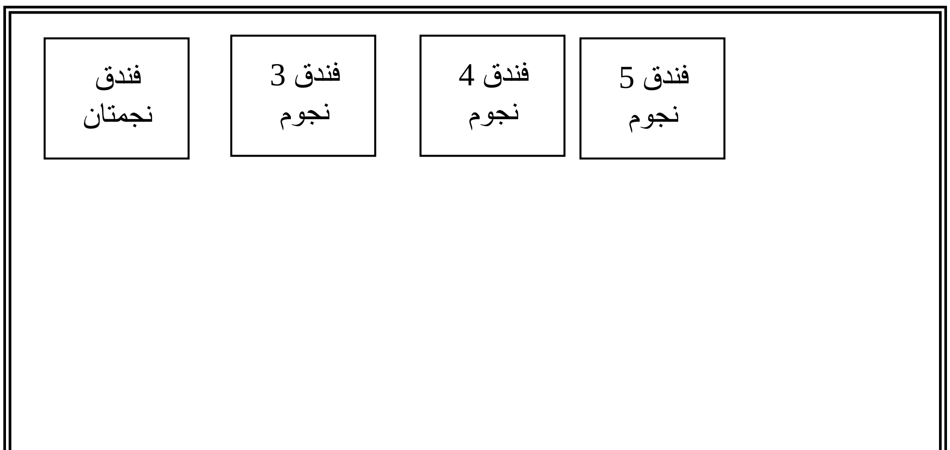
٢) جنسية السائح (أمريكي، إنجليزي، عربي، أسباني، ...).

ففي هذه الحالة يمكن استخدام العينة العشوائية العنقودية على مرحلتين:

المرحلة الأولي: تقسيم الفنادق وفقاً لمستواها.

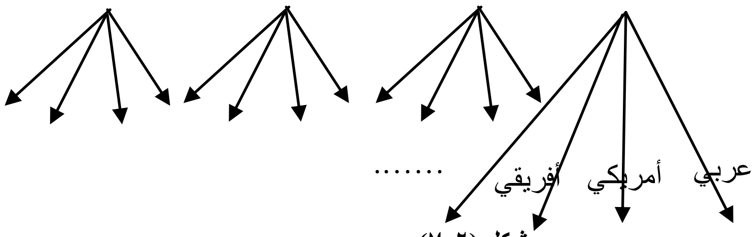
المرحلة الثانية: تقسيم السائحين داخل كل فندق وفقاً لجنسياتهم (أو القارة

الموفد منها).



المرحلة الأولي

المرحلة الثانية



شكل (٢-٧)

(٢-٨) تمرينات Exercises

(٢-١) تكلم باختصار عن أهم خصائص كل عينة من العينات التالية:

أ- العينة العشوائية البسيطة.

ب- العينة العشوائية المنتظمة.

ج- العينة العشوائية الطبقية.

د- العين الحصصية.

(٢-٢) قارن بين كل من العينة العشوائية الطبقية والعينة الحصصية. ثم أذكر بعض الأمثلة التي يمكن أن تستخدم كل عينة منهما في الدراسة.

(٢-٣) عند سحب عينة عشوائية طبقية أو عينة عشوائية عنقودية نقوم بتقسيم مجتمع الدراسة إلى مجموعات. أذكر باختصار أوجه التشابه والاختلاف بين العينة العشوائية الطبقية والعينة العشوائية العنقودية.

(٢-٤) حدد أي نوع من العينات التي يجب سحبها لدراسة كل حالة من الحالات التالية:

١- مدي تتبع المشاهدين للتليفزيون من سكان القاهرة لأحد البرامج الاقتصادية الأسبوعية التي تقدم الساعة العاشرة مساءاً.

٢- الاستهلاك اليومي للمياه الغازية بالقرى السياحية بالساحل الشمالي بجمهورية مصر العربية.

٣- عدد الليالي السياحية التي يقضيها السائح الأجنبي في مدينة الأقصر في شهر ديسمبر

(٢-٥) في إحدى مؤسسات الطباعة يوجد بها 500 عامل طباعة، فإذا قامت المؤسسة باستيراد ماكينات طباعة حديثة وترغب في تدريب عينة من العاملين على هذه الماكينات مكونه من 50 عامل. والمطلوب:

١- حدد عدد العينات العشوائية الممكن سحبها.

٢- حدد الخطوات التي يجب إتباعها لاختيار عينة عشوائية بسيطة باستخدام جدول الأعداد العشوائية.

٣- حدد الخطوات التي يجب إتباعها لاختيار عينة عشوائية منتظمة.

(٢-٦) تقوم إحدى الشركات بإنتاج اللمبات الكهربائية ثلاثة أنواع A,B,C وفقاً للحجم. فإذا كان إنتاج الشركة اليومي من الأنواع الثلاثة على النحو 10000,8000,12000 على الترتيب، والمراد اختيار الإنتاج اليومي باستخدام عينة مكونه من 150 لمبة. والمطلوب:

حدد أي نوع من العينات يكون مناسباً لاستخدامه في هذه الحالة مع التعليل.

(٢-٧) قامت إحدى الجامعات المصرية بإصدار مجلة علمية في العلوم الإحصائية وتطبيقاتها. ومن الأهمية معرفة رأي المتخصصين المصريين من القراء في هذه المجلة من خلال عينة حجمها 100 مفردة من المتخصصين. والمطلوب:

١- حدد أي نوع من العينات يكون أنسب في هذه الحالة.

٢- حدد الخطوات التي جب إتباعها لسحب العينة التي يتم تحديدها في (١).

(٢-٨) حدد أي نوع من العينات التي يجب سحبها في الحالات التالية، مع توضيح خطوات سحب العينة المختارة.

١- دراسة متوسط استهلاك الأسرة الشهري للكهرباء في مدينة القاهرة عن طريق عينة حجمها 500 مفردة.
٢- دراسة خصوبة التربة الزراعية في الوجه القبلي بجمهورية مصر العربية عن طريق عينة مكونه من 50فدان.
٣- دراسة جودة الإنتاج اليومي المكون من 100 تليفزيون ٢٠بوصة، 500 تليفزيون ١٧بوصة من إنتاج شركة النصر للتليفزيونات عن طريق عينة حجمها 10تليفزيونات.

(٢-٩) استخدم جداول الأعداد العشوائية لاختيار عينة عشوائية بسيطة حجمها 10سيارات من 500 سيارة متماثلة.

(٢-١٠)استخدم جداول الأعداد العشوائية في اختيار عينة عشوائية منتظمة حجمها 200مفردة من 10000 شكوى مقدمة من الطلاب وأولياء الأمور لوزارة التربية والتعليم.

(٢-١١) يعمل 200 عامل فني في إحدى الشركات العقارية. كيف يمكن أخذ عينة عشوائية منتظمة حجمها 10% من هؤلاء العمال.

(٢-١٢) يرغب متخذ القرارات في إحدى الشركات للمنتجات الغذائية فتح سوق جديد لمنتجات الشركة بإحدى المدن الجديدة. لذا أراد استطلاع رأي سكان هذه المدينة بالنسبة لمنتجات الشركة عن طريق عينة حجمها 100 فرد من سكان هذه المدينة. المطلوب:
حدد نوع وخصائص العينة التي جب سحبها.

الباب الثالث
التوزيعات الاحتمالية لبعض مؤشرات العينة
(توزيعات المعاينة)
Sampling Distributions

(٣-١) مؤشرات العينة **Sampling Measures**

(٣-٢) التوزيع الاحتمالي للوسط الحسابي(متوسط) في العينة
Probability Distribution of Sample's Mean

(٣-٣) التوزيع الاحتمالي للتباين في العينة
Probability Distribution of Sample's Variance

(٣-٤) التوزيع الاحتمالي للنسبة في العينة
Probability Distribution of Sample's Proportion

(٣-٥) تطبيقات **Applications**

(٣-٦) تمرينات **Exercises**

(٣-١) مؤشرات العينة Sample′s Measures

إذا كان الأجر اليومي لخمسة عاملين A ,B ,C ,D بالقطاع الخاص بالجنيه على النحو التالي : 70, 40 , 60 , 80 , 50 على الترتيب .فإذا تم اختيار عينة عشوائية بسيطة من ثلاثة عاملين من هؤلاء العاملين فنجد أن عدد العينات الممكنة في هذه الحالة يساوى $C_3^5 = 10$ عينات كما هو موضح بجدول (٣-١) .فإذا قمنا بحساب الوسط الحسابي والتباين للأجر في كل عينة نجد أنهما يختلفان من عينة لأخرى كما هو موضح بالجدول.

جدول (٣-١) : يوضح الوسط الحسابي والتباين للأجر اليومي لكل عينة من العينات

رقم العينة	العينة	الوسط الحسابي للأجر في العينة $(\bar{X}_j)$	تباين الأجر اليومي في العينة (S_j^2)
1	(A,B)	(50+80)/2=65	$[(50-65)^2+(80-65)^2]/2=225$
2	(A,C)	(50+60)/2=55	$[(50-55)^2+(80-55)^2]/2=25$
3	(A,D)	(50+40)/2=45	$[(50-45)^2+(80-45)^2]/2=25$
4	(A,E)	(50+70)/2=60	$[(50-60)^2+(80-60)^2]/2=100$
5	(B,C)	(80+60)/2=70	$[(80-70)^2+(80-70)^2]/2=100$
6	(B,D)	(80+40)/2=60	$[(80-60)^2+(80-60)^2]/2=400$
7	(B,E)	(80+70)/2=75	$[(80-75)^2+(80-75)^2]/2=25$
8	(C,D)	(60+40)/2=50	$[(60-50)^2+(80-50)^2/2=100$
9	(C,E)	(60+70)/2=65	$[(60-65)^2+(80-65)^2]/2=25$
10	(D,E)	(70+70)/2=70	$[(40-55)^2+(80-55)^2]/2=225$

ويتضح من الجدول أن احتمال سحب أي عينة يساوي $\frac{1}{10} = \frac{1}{C_5^3}$ كذلك يتضح أن الوسط الحسابي في العينة رقم **j** يساوي $(\overline{X}_j)$ والتباين يساوي (S_j^2) يختلفان من عينة إلى أخرى ، وبالتالي فإن الوسط الحسابي في العينة $(\overline{X}_j)$ يمثل متغير عشوائي ، كذلك التباين (S_j^2) يمثل متغير عشوائي أيضاً .

ومن هنا يتضح أن أي مؤشر مثل الوسط الحسابي ، أو التباين ، أو المنوال ، أو الوسيط ،...الخ يتم حسابه من بيانات العينة يمثل متغير عشوائي قيمته تختلف عينة من عينة إلى أخرى ومما سبق يتضح أن أي مؤشر من مؤشرات العينة المسحوبة من مجتمع الدراسة يعتبر متغير عشوائي يختلف من عينة إلى أخرى لنفس المجتمع.

وعادة تتم دراسة المتغيرات والظواهر محل الدراسة في المجتمع بهدف اتخاذ قرارات مثلى **Optimum decisions** للمشاكل الموجودة أو التخطيط لرسم سياسات صحيحة للأنظمة القائمة أو المستحدثة .

وكما ذكرنا في الباب السابق نظراً لصعوبة أو استحالة استخدام أسلوب الحصر الشامل في دراسة خصائص المجتمع في معظم الحالات ، لذلك تتم دراسة خصائص المجتمع محل الدراسة باستخدام المؤشرات التي تم الحصول عليها من العينة . فبدراسة التوزيعات الاحتمالية للمؤشرات في العينة وخصائص هذه المؤشرات التي تم حسابها من بيانات العينة المحسوبة من مجتمع الدراسة يمكن الوصول إلى معرفة رقمية كاملة عن خصائص المتغيرات محل الدراسة في المجتمع كما سوف يتضح في هذا الباب والأبواب التالية.

لذلك سوف نتناول في هذا الباب التوزيعات الاحتمالية لبعض المؤشرات الهامة التي يتم حسابها من العينة العشوائية وعلاقة هذه المؤشرات بالمؤشرات المناظرة لها للمتغيرات في المجتمع محل الدراسة.

لذلك يعتبر هذا الباب الخطوة الأولى في تحويل المعلومات المتمثلة في المؤشرات المحسوبة من بيانات العينة إلى معرفة متعلقة بالمجتمع محل الدراسة.

(٣-٢) التوزيع الاحتمالي للوسط الحسابي في العينة

Probability Distribution of Sample's Mean

إذا فرضنا أن المشاهدات $X_1, X_2, \ldots, X_n$ هي قيم المتغير محل الدراسة في عينة حجمها n مفردة ، فإن الوسط الحسابي لقيمة هذا المتغير في هذه العينة وسوف نشير إليه بالرمز $\bar{X}$ حيث :-

$$\bar{X} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (3.1)$$

نظرية (٣-١)

إذا كان $\bar{X}$ هو الوسط الحسابي (المتوسط) للمتغير في العينة العشوائية البسيطة التي حجمها n و(السحب بدون إرجاع) من مجتمع حجمه N والوسط الحسابي للمتغير في المجتمع يساوي μ وتباين المتغير في المجتمع يساوي σ^2 فإن :-

القيمة المتوقعة للمتغير $\bar{X} = \mu(\bar{X}) = \mu$ **(3.2)**

$$\text{تباين }(\bar{X}) = \sigma^2(\bar{X}) = \frac{\sigma^2}{n}\left[\frac{N-n}{N-1}\right] \qquad (3.3)$$

نظرية(٣-٢)

عندما يكون حجم المجتمع كبيراً جداً ، أي $N \to \infty$ ، أو عندما يكون السحب بدون إرجاع فإن :-

$$\text{القيمة المتوقعة للمتغير } \bar{X} = \mu(\bar{X}) = \mu \qquad (3.4)$$

$$\text{تباين }(\bar{X}) = \sigma^2(\bar{X}) = \frac{\sigma^2}{n} \qquad (3.5)$$

مثال (٣-١)

إذا كان لدينا مجتمع مكون من 5 مفردات ، فإذا كان المتغير X يمثل عدد ساعات العمل اليومية لكل مفردة حيث $X_1 = 2, X_2 = 3, X_3 = 6, X_4 = 8, X_5 = 11$ فإذا سحبت عينة عشوائي مكونة من مفردتين من هذا المجتمع.

المطلوب

١- **احسب الوسط الحسابي (المتوسط) μ والتباين σ^2 لعدد ساعات العمل اليومية في المجتمع محل الدراسة.**

٢- **حدد العينات العشوائية الممكن سحبها بحيث يكون حجم كل عينة مفردتين إذا كان السحب مع الإرجاع With replacing .**

٣- **حدد العينات العشوائية الممكن سحبها إذا كان السحب بدون إرجاع Without replacing.**

٤- أوجد توقع الوسط الحسابي والتباين لساعات العمل اليومية في كل عينة من العينـات المسـحوبة فـي (٢) ، (٣) مـع المقارنـة بالمتوسـط والتبـاين فـي المجتمع.

الحل

١- إذا كان متوسط عدد ساعات العمل في المجتمع يساوي μ وتباينه يساوي σ^2 فإن:-

$$\mu = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{2+3+6+8+11}{5} = 6 \text{ ساعات} \tag{3.6}$$

$$\sigma^2 = \frac{\sum_{i=1}^{5}(x_i-\mu)^2}{5}$$

$$= \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}$$

$$= \frac{16+9+0+4+25}{5} = \frac{54}{5} = 10.8 \tag{3.7}$$

٢- إذا كـان

السحب مع الإرجاع فإن عدد العينـات العشوائية المختلفـة الممكن سـحبها من المجتمع يساوي عدد المفردات في المجتمع مرفوع لأس يساوي عدد المفردات في العينـة أي يساوي N^n أي $25 = (5)^2$ والجدول التالي يوضح هذه العينات.

جدول (٣-٢)

رقم العينة	العينة	الوسط الحسابي في العينة j : $\bar{X}_j$	$(\bar{X}_j - \mu)$	$(\bar{X}_j - \mu)^2$
1	(2,2)	2	-4	4
2	(3,2)	2.5	-3.5	12.25
3	(6,2)	4	-3	4
4	(8,2)	5	-1	1
5	(11,2)	6.5	0.5	0.25
6	(2,3)	2.5	3.5	12.25
7	(3,3)	3	-3	9
8	(6,3)	4.5	-1.5	2.25
9	(8,3)	5.50	-0.5	0.25
10	(11,3)	7	1	1
11	(2,6)	4	-2	4
12	(3,6)	4.5	-1.5	2.25
13	(6,6)	6	0	0
14	(8,6)	7	1	1
15	(11,6)	8.5	2.5	6.25
16	(2,8)	5	-11	1
17	(3,8)	5.50	-0.5	0.25
18	(6,8)	7	1	1

رقم العينة	العينة	الوسط الحسابي في العينة j : $\overline{X}_j$	$(\overline{X}_j - \mu)$	$(\overline{X}_j - \mu)^2$
19	(8,8)	8	2	4
20	(11,8)	9.5	3.5	12.25
21	(2,11)	6.5	0.5	0.25
22	(3,11)	7	1	1
23	(6,11)	8.5	2.5	6.25
24	(8,11)	9.5	3.5	12.25
25	(11,11)	11	5	25
		150		135

ونظراً لأن العينات المسحوبة عينات عشوائية فإن احتمال سحب أي عينة متساوي مع احتمال سحب أي عينة أخرى . ومن الجدول نجد أن احتمال سحب أي عينة يساوي $\frac{1}{25}$ وبالتالي فإن القيمة المتوقعة للوسط الحسابي في العينة $\mu(\overline{X})$ تصبح على النحو:-

$$\mu(\overline{X}) = \sum_{j=1}^{25} \overline{X}_j P_r(\overline{X})$$

$$= \sum_{j=1}^{25} \frac{1}{25} \overline{X}_j = \frac{1}{25}(150) = 6 \qquad (3.8)$$

من المعادلتين (3.8) , (3.6) نجد أن القيمة المتوقعة للمتغير $\overline{X}$ الذي يمثل الوسط الحسابي أو (المتوسط) في العينة $\mu(\overline{X})$ يساوي الوسط الحسابي (المتوسط) في المجتمع، أي أن:-

$$\mu(\overline{X}) = \mu$$

ومن الجدول السابق أيضاً نجد أن تباين المتغير $\overline{X}_j$ يساوي $\sigma^2(\overline{X})$ حيث :-

$$\sigma^2(\overline{X}) = \frac{1}{25}[\sum_{j=1}^{25}(\overline{X}_j - \mu)^2]$$

(3.9)

$$= \frac{1}{25}(135) = 5.4$$

ومن المعادلتين **(3.9) , (3.7)** نجد أن :-

$$\sigma^2(\overline{X}) = 4.5$$

٣- إذا تم سحب مفردات العينة بدون إرجاع ، فإن عدد العينات العشوائية الممكن سحبها في هذه الحالة يساوي C_2^5 من العينات الممكنة حيث :-

$$C_2^5 = \frac{5!}{2!(5-2)!} = 10 \text{ عينة}$$

والجدول التالي يوضح هذه العينات

جدول (٣-٣)

رقم العينة	العينة	متوسط العينة j $\overline{X}_j$	$(\overline{X}_j - \mu)$	$(\overline{X}_j - \mu)^2$
1	**(2 , 3)**	**2.5**	**-3.5**	**12.25**
2	**(2 , 6)**	**4**	**-2**	**4**
3	**(2 , 8)**	**5**	**-1**	**1**
4	**(2 , 11)**	**6.5**	**0.5**	**0.25**
5	**(3 , 6)**	**4.5**	**-1.5**	**2.25**
6	**(3 , 8)**	**5.5**	**-0.5**	**0.25**
7	**(3 , 11)**	**7**	**1**	**1**
8	**(6 , 8)**	**7**	**1**	**1**

رقم العينة	العينة	متوسط العينة j $\bar{X}_j$	$(\bar{X}_j - \mu)$	$(\bar{X}_j - \mu)^2$
9	**(6 , 11)**	**8.5**	**2.5**	**6.25**
10	**(8 , 11)**	**9.5**	**3.5**	**12.25**
$\sum$		**60**		**40.5**

ومن الجدول يتضح أن احتمال سحب أي عينة يساوي $\frac{1}{10}$ **كذلك نجد أن :-**

$$\mu(\bar{X}) = \sum_{j=1}^{10} \bar{X}_j P_r(\bar{X}_j)$$

$$= \frac{1}{10}(\sum_{j=1}^{10} \bar{X}_j) = \frac{1}{10}(60) = 6$$

أي أن: $\mu(\bar{X}) = \mu$ **(3.10)**

كذلك نجد أن

$$\sigma^2(\bar{X}) = \sum_{j=1}^{10} (\bar{X}_j - \mu)^2 P_r(\bar{X}_j)$$

(3.11)

$$\sigma^2(\bar{X}_j) = \frac{1}{10}[\sum_{j=1}^{10} (\bar{X}_j - \mu)^2] = \frac{1}{10}(40.5$$

من المعادلتين (3.11) , (3.7) نجد أن :-

$$\sigma(\bar{X}_j) = 40.5$$

$$= \frac{(5-2)}{(5-1)}(\frac{10.8}{2})$$

$$= (\frac{\sigma^2}{n})(\frac{N-n}{N-1})$$

نظرية(٣-٣)

إذا كان المتغير محل الدراسة x يتبع التوزيع المعتاد بتوقع μ وتباين σ^2 وكان $\bar{X}$ هو الوسط الحسابي (المتوسط) للمتغير في عينة عشوائية مسحوبة من هذا المجتمع حجمها n فإن المتغير $\bar{X}$ يتبع التوزيع المعتاد بتوقع يساوي $\mu(\bar{X})$ وتباين $\sigma^2(\bar{X})$ حيث:-

$$\mu(\bar{X}) = \mu \qquad (3.12)$$

$$\sigma^2(\bar{X}) = (\frac{\sigma^2}{n})(\frac{N-n}{N-1}) \qquad (3.13)$$

إذا كان السحب بدون إرجاع ، كذلك

$$\mu(\bar{X}) = \mu \qquad (3.14)$$

$$\sigma^2(\bar{X}) = (\frac{\sigma^2}{n}) \qquad (3.15)$$

إذا كان السحب مع الإرجاع

فإذا فرضنا أن Y متغير بحيث:-

$$Y=\frac{\bar{X}-\mu(\bar{X})}{\sigma(\bar{X})}=\frac{\bar{X}-\mu}{\sigma}$$
$$=\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \qquad (3.16)$$

فإن Y متغير يتبع التوزيع المعتاد القياسي إذا كان السحب مع الإرجاع (أو سحب بدون إرجاع وحجم المجتمع كبير جداً)، ولكن:

$$Y=\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}}$$
$$=\frac{(\bar{X}-\mu)(\sqrt{n(N-1)})}{\sigma(\sqrt{N-n})} \qquad (3.17)$$

إذا كان السحب بدون إرجاع وحجم المجتمع صغير.

مثال(٣-٢)

تقوم إحدى شركات صناعة الأحذية بتصنيع 10000 حذاء يومياً . حيث فترة صلاحية الحذاء من هذا المنتج يعتبر متغير يتبع التوزيع المعتاد بتوقع سنة واحدة وتباين 0.563 فإذا سحبت عينة حجمها 25 حذاء من الإنتاج اليومي لهذه الشركة ، أوجد ما يلي:-

١- احتمال أن يكون متوسط عمر الحذاء في العينة اكبر من 1.4 سنة.

٢- احتمال أن يكون متوسط عمر الحذاء في العينة أقل من 3 شهور.

الحل

إذا أشرنا لفترة صلاحية الحذاء من إنتاج هذا المصنع بالرمز **X** ، كذلك لمتوسط فترة صلاحية الحذاء في العينة المسحوبة من هذا المصنع بالرمز $\overline{X}$ نجد أن $\overline{X}$ يتبع التوزيع المعتاد. حيث:-

$$n = 25 \qquad , \qquad \mu(\overline{X}) = 1$$

$$\sigma^2(\overline{X}) = \frac{\sigma^2}{\sqrt{n}} = \frac{0.563}{\sqrt{25}} = 0.0225$$

$$\sigma\ (\overline{X}) = 0.15$$

فإذا فرضنا أن :

$$Y = \frac{\overline{X} - \mu}{\sigma(\overline{X})} = \frac{\overline{X} - 1}{0.15}$$

فإن **Y** متغير يتبع التوزيع المعتاد القياسي وبالتالي فإن :-

١- احتمال أن يكون متوسط عمر الحذاء في العينة أكبر من **1.4** سنة يساوي $P_r(\overline{X} \geq 1.4)$ بحيث:

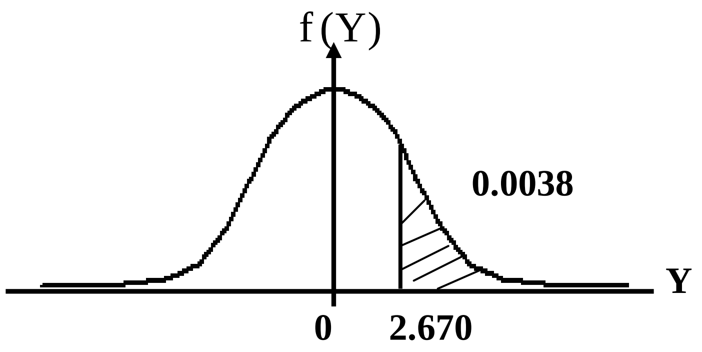
$$P_r(\overline{X} \geq 1.4) = P_r(\frac{1.4 - 1}{0.15})$$

$$= P_r(Y > 2.67) = 0.5 - P_r(0 < Y < 2.67)$$

وباستخدام جدول التوزيع المعتاد القياسي بملحق ٤ :

$$P_r(\overline{X} \geq 1.4) = 0.5 - 0.4962 = 0.0038$$

كما هو موضح بالشكل التالي:-



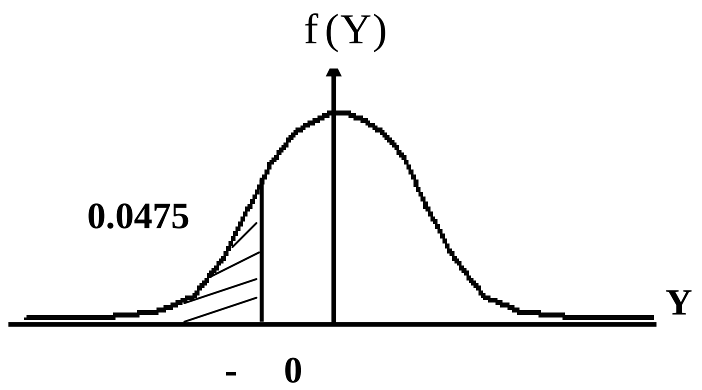
شكل (٣-١)

وبالمثل

٢- احتمال أن يكون متوسط عمر الحذاء في العينة أقل من 3 شهور

$$P_r(\overline{X} < 0.75) = P_r(Y < \frac{0.75-1}{0.15})$$
$$= P_r(Y < -1.67) = 0.5 - 0.4525$$
$$= 0.0475$$

كما هو موضح بالشكل التالي:-



شكل (٣-٢)

نظرية (٣-٤)

إذا كان المتغير محل الدراسة في مجتمع يتبع أي توزيع احتمالي غير التوزيع المعتاد، وكان $\overline{X}$ هو الوسط الحسابي للمتغير في عينة عشوائية مسحوبة من هذا المجتمع حجمها n فإن التوزيع الاحتمالي للمتغير $\overline{X}$ يؤول Approaches إلى التوزيع المعتاد بتوقع $\mu(\overline{X})$ وتباين $\sigma^2(\overline{X})$ عندما يكون حجم العينة n كبير (وعادة يكون حجم العينة كبير عندما تكون أكبر من أو تساوي 30 مفردة أي أن $n \geq 30$).

وتسمى هذه النظرية بنظرية النزعة المركزيةCentral Limit Theorem، وتعتبر هذه النظرية من أهم النظريات في دراسة توزيعات المعاينة وبصفة خاصة من الجانب التطبيقي ويرجع ذلك إلى أنه عند استخدام أسلوب العينة في دراسة خصائص المجتمع عادة يكون التوزيع الاحتمالي للمتغير محل الدراسة في المجتمع غير متاحUnavailable.

مثال (٣-٣)

تقوم إحدى شركات صناعة المستحضرات الطبية بإنتاج نوع معين من المستحضرات . وبدراسة فترة صلاحية هذا المستحضر وجد أنها تتبع التوزيع الأسي (أنظر ملحق(٢)) ، فإذا كانت الشركة تنتج 15000 عبوة في اليوم من هذا المستحضر ، فإذا أخذت عينات عشوائية بسيطة حجم كل منها n=100 مفردة من إنتاج هذه الشركة في يوم معين ، وبفحصها وجد أن القيمة المتوقعة لمتوسط فترة الصلاحية في العينة تساوي $\mu(\overline{X}) = 10$ أيام وانحراف معياري $\sigma(\overline{X}) = 5$ حيث $\overline{X}$ الوسط الحسابي لفترة الصلاحية في العينة. أوجد:

١- التوقع μ والتباين σ^2 لفترة الصلاحية في المجتمع المسحوب منه العينات.

٢- احتمال أن يكون متوسط فترة الصلاحية في العينة أكبر من 15 يوم.

٣- احتمال أن يكون متوسط فترة الصلاحية في العينة أقل من 3 أيام.

الحل

١- من نظرية (٣-١) نجد أن :-

$$\mu = \mu(\overline{X}) = 10 \quad \text{أيام}$$

$$\sigma^2 = n\sigma^2(\overline{X}) = 100(5) = 500$$

٢- وفقاً لنظرية النزعة المركزية (نظرية (٣-٤)) نجد أن المتغير Y حيث :-

$$Y = \frac{\overline{X} - \mu(\overline{X})}{\sigma(\overline{X})}$$ متغير يتبع التوزيع المعتاد القياسي.

احتمال أن يكون متوسط فترة الصلاحية في العينة أكبر من 15 يوم يساوي

$$\begin{aligned} P_r(\overline{X} > 15) &= P_r(Y > \frac{15-10}{5}) \\ &= P_r(Y > 1) = 0.5 - P_r(0 < Y < 1) \\ &= 0.5 - 0.3413 \\ &= 0.1578 \cong 0.16 \end{aligned}$$

وهذا يعني أن 16% من العينات الممكن سحبها من الإنتاج اليومي للشركة يكون في كل منها متوسط فترة الصلاحية في العينة أكبر من 15 يوم .

٣- كذلك احتمال أن يكون متوسط فترة الصلاحية أقل من 3 أيام يساوي:-

$$P_r(\overline{X} < 3) = P_r(Y < \frac{3-10}{10})$$
$$= P_r(Y < -0.7) = 0.5 - P_r(0 < Y < 0.7)$$
$$= 0.5 - 0.258$$
$$= 0.242 \cong 0.24$$

وهذا يعني أن 24% من العينات الممكن سحبها من الإنتاج اليومي للشركة يكون في كل منها متوسط فترة الصلاحية للعبوة من هذا المستحق أقل من 3 أيام.

(٣-٣) التوزيع الاحتمالي لتباين العينة

Probability Distribution of Sample's Variance

إذا كانت $x_1, x_2, ..., x_n$ عينة عشوائية بسيطة حجمها n فإن أشرنا إلى تباين هذه العينة بالرمز S^2 بحيث:-

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})}{n-1} \qquad (3.17)$$

حيث $\overline{X}$ هو الوسط الحسابي للعينة ، ويستخدم المقياس S^2 كتقدير لتباين المتغير x في المجتمع محل الدراسة والذي نرمز له بالرمز σ^2 الذي سحبت منه العينة وفي الفصل الأول من هذا الباب وجدنا أن قيمة التباين في العينة S^2 تختلف من عينة إلى أخرى كما هو موضح في جدول (٣-١) وبالتالي فإن تباين العينة S^2 يمثل متغير عشوائي يأخذ قيم مختلفة في العينات المختلفة المسحوبة من نفس المجتمع وبنفس الحجم.

نظرية (٣-٥)

إذا كانت $x_1, x_2, \ldots, x_n$ **عينة عشوائية مسحوبة من مجتمع معتاد توقعه** μ **وتباينه** σ^2 **فإن المتغير Y حيث:-**

$$Y = \frac{n-1}{\sigma^2} S^2 \qquad (3.18)$$

متغير يتبع توزيع χ^2 **بدرجات حرية (n-1) وبالتالي فإن المتغير** S^2 **مضروب في المقدار الثابت** $(\frac{\sigma^2}{n-1})$ **يتبع توزيع** χ^2_{n-1} **كذلك توقع** S^2 **يساوي** σ^2 **أي أن:**

$$\mu(S^2) = \frac{\sigma^2}{n-1}(n-1) = \sigma^2$$

حيث (n-1) يساوي توقع المتغير الذي يتبع توزيع χ^2 **بدرجات حرية (n-1)**

"أنظر ملحق رقم ٢".

مثال (٣-٤)

إذا سحبت عينة عشوائية حجمها 5 مفردة من مجتمع يتبع التوزيع المعتاد بتوقع $\mu = 10$ **وتباينه** $\sigma^2 = 4$

المطلوب

١- **أوجد احتمال أن يكون التباين في العينة أقل من 2.**

٢- **أوجد احتمال أن يكون التباين في العينة أكبر من 3.**

٣- **أوجد القيمة المتوقعة للتباين في العينة كذلك القيمة المتوقعة للانحراف المعياري.**

الحل

إذا فرضنا أن التباين في العينة يساوي $S^2, n = 5, \sigma^2 = 4$ وبالتالي فإن المتغير Y حيث:-

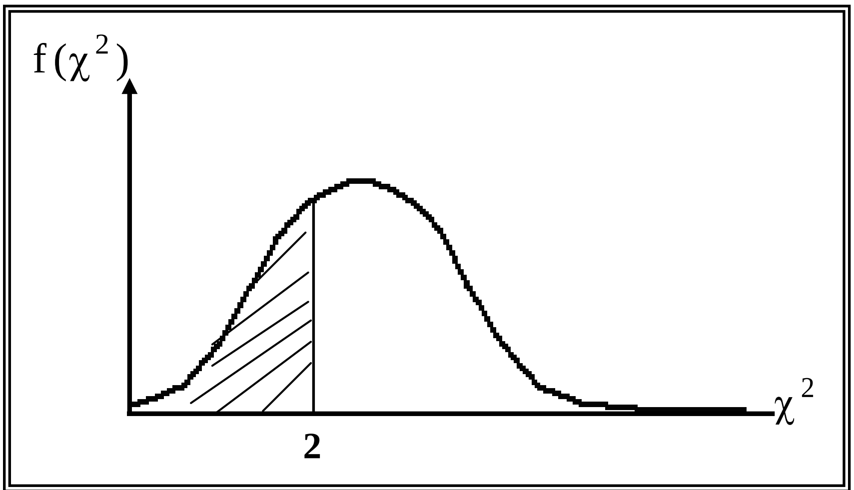
$$Y = \frac{n-1}{\sigma^2} S^2 \approx \chi^2_{n-1} \qquad \text{(نظرية ٣-٥)}$$

$$1) P_r(S^2 < 2) = P_r\left(\left(\frac{n-1}{\sigma^2}\right) S^2 < 2\left(\frac{n-1}{\sigma^2}\right)\right)$$

وباستخدام ملحق رقم (٥) نجد أن :-

$$P_r\left(\chi^2_{n-1} < 2\left(\frac{5-1}{4}\right)\right) = P_r(\chi^2 < 2) = 0.3$$

وشكل (٣-٣) يوضح احتمال أن يكون التباين في العينة أقل من 2:-

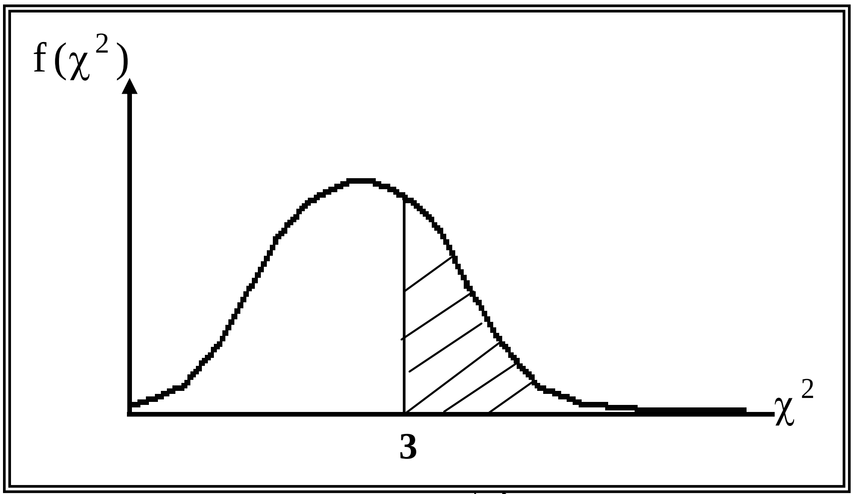


شكل (٣-٣)

$$2) P_r(S^2 > 3) = P_r\left(\left(\frac{n-1}{\sigma^2}\right) S^2 > 3\left(\frac{n-1}{\sigma^2}\right)\right)$$

وباستخدام ملحق رقم (٥) نجد أن: $P_r(\chi^2 > 3) = 0.5$

وشكل (٣-٤) يوضح احتمال أن يكون التباين في العينة أكبر من 3:-



شكل (٣-٤)

$$3)\ \mu(S^2) = \sigma^2 = 4$$
$$\mu(S) = \sqrt{4} = 2$$

(٣-٤) التوزيع الاحتمالي للنسبة في العينة

Probability Distribution of Sample's Proportion

عادة ما يكون من الأهمية معرفة نسبة مفردات المجتمع الإحصائي Statistical Population التي تتصف بخاصية معينة ، فعلى سبيل المثال نسبة المصابين بالبلهارسيا في جمهورية مصر العربية . كذلك نسبة الوحدات المعيبة من منتج معين . وعادة ما يكون غير ممكن حساب النسبة في المجتمع أو يكون الحساب ممكن ولكنه مكلف جداً أو يتطلب وقت غير متاح ، وبالتالي فإننا نستخدم أسلوب العينة في الحصول على معرفة دقيقة عن النسبة في المجتمع كما سوف يوضح فيما يلي:-

فإذا فرضنا أن $x_1, x_2, \ldots, x_n$ تمثل مفردات مجتمع حجمه N وهذه المفردات أما أن تملك الخاصية محل الاهتمام وأما لا تملكها ، فإذا فرضنا أن المفردة التي تملك

هذه الخاصية تعطي الرقم (واحد) والتي لا تملك هذه الخاصية تعطى الرقم (صفر) أو بعبارة أخرى

$$X_i = \begin{cases} 1 & \text{إذا توافرت الخاصية في المفردة} \\ 0 & \text{إذا لم تتوافر الخاصية في المفردة} \end{cases} \quad ,i = 1,2,3,...,N \qquad (3.19)$$

أي يتم تحويل البيانات من بيانات وصفية إلى بيانات كمية وهذا يحدث في الكثير من الدراسات الاجتماعية ، الإدارية ، السياسية ، الطبية ، ...الخ وبصفة خاصة في دراسات استطلاع الرأي في صورة نعم أو لا فإذا أعطينا الشخص الذي أجاب بنعم الرقم (واحد) وبالتالي سوف نعطي الرقم (صفر) للشخص الذي أجاب بـ لا ، كذلك في الدراسات الخاصة بمراقبة جودة الإنتاج ، فإذا كانت الوحدة المنتجة مطابقة للمواصفات تعطي الرقم (واحد) وبالتالي الوحدة غير المطابقة تعطي الرقم (صفر).

تعريف(٣-١)

إذا كان N حجم المجتمع محل الدراسة بحيث N_1 من مفرداته تتصف بخاصية معينة وبالتالي $(N - N_1)$ من المفردات لا تتصف بهذه الخاصية ، فإذا فرضنا أن θ هي نسبة المفردات التي لها هذه الخاصية في المجتمع ، فإن:-

$$\theta = \frac{\text{عدد المفردات التي بها الخاصية}}{\text{عدد مفردات المجتمع}} = \frac{N_1}{N} \qquad (3.20)$$

$$0 \le \theta \le 1 \qquad (3.21)$$

تعريف (٣-٢)

إذا سحبت عينة عشوائية بسيطة حجمها n من مجتمع حجمه N ووجد أن عدد N_1 مفردات العينة له نفس الخاصية المعينة وبالتالي العدد $(N - N_1)$ من مفردات العينة لا تمتلك هذه الخاصية ، فإذا كانت $\hat{\theta}$ تمثل النسبة في العينة فإن:-

$$\hat{\theta} = \frac{\text{عدد المفردات في العينة الذين يمتلكون الخاصية}}{\text{عدد مفردات العينة}} = \frac{N_1}{n} \qquad (3.22)$$

$$0 \le \hat{\theta} \le 1 \qquad (3.23)$$

وكما في حالة الوسط الحسابي (المتوسط) ، وتباين في العينة حيث وجدنا في الفصول السابقة أن كل من الوسط الحسابي $\overline{X}$ والتباين σ^2 في العينة يمثلان متغيرات عشوائية ، بالمثل النسبة في العينة $\hat{\theta}$ تختلف من عينة لأخرى من العينات المسحوبة من نفس المجتمع وبالتالي فهي تمثل متغير عشوائياً أيضاً.

نظرية(٣-٦)

إذا كانت $\theta, \hat{\theta}$ هما النسبة فى المجتمع و النسبة في العينة العشوائية المسحوبة من نفس المجتمع على الترتيب ، حيث حجم المجتمع غير محدود infinite (أو حجم المجتمع محدودfinite ولكن السحب مع الإرجاع with replacement) فإن:

$$\mu(\hat{\theta}) = \theta \quad , \quad \sigma^2(\hat{\theta}) = \frac{\theta(1-\theta)}{n} \qquad (3.24)$$

حيث n حجم العينة

نظرية (٣-٧)

إذا كـان حجـم المجتمـع محـدد finite والسـحب بـدون إرجـاع without replacement فان:-

$$\mu(\hat{\theta}) = \theta \quad , \quad \sigma^2(\hat{\theta}) = \left(\frac{N-n}{N-1}\right)\left(\frac{\theta(1-\theta)}{n}\right) \qquad (3.25)$$

نظرية (٣-٨)

إذا كان حجم العينـة $n \geq 30$ فان التوزيع الاحتمـالي للنسبة في العينـة $\hat{\theta}$ يؤول إلى التوزيع المعتاد بتوقع $\mu(\hat{\theta})$ وتباين $\sigma^2(\hat{\theta})$ حيث أن المتغير محل الدراسة في المجتمع المسحوب منه العينـة يعتبر متغير يتبع توزيع ذات الحدين (أنظر ملحق رقم ٢).

مثال (٣-٥)

إذا كان الإنتاج اليومي لأحد خطوط الإنتـاج للتلفزيونـات الملونـة هو 4 وحدات يوميـة . فـإذا كـان 25% مـن هـذه الوحـدات غيـر مطابقـة للمواصـفات ، 75% وحـدات مطابقـة للمواصفات فإذا تم سحب عينـة عشوائية بسيطة مكونـة من وحدتين من إنتاج يوم معين

المطلوب:

<u>أولاً: إذا كان السحب مع الإرجاع :-</u>

١- **أوجد عدد العينات العشوائية البسيطة الممكن سحبها ، ثم أوجد نسبة المعيب في كل عينـة من هذه العينات.**

٢- **أوجد التوزيع الاحتمالي للنسبة في العينـة ووضح ذلك بيانياً.**

٣- **أثبت أن:-**

$$\mu(\hat{\theta}) = \theta$$

$$\sigma^2(\hat{\theta}) = \frac{\theta(1-\theta)}{n}$$

حيث $\hat{\theta}, \theta$ هما نسبة المعيب في المجتمع والعينة على الترتيب.

<u>ثانياً: إذا كان السحب بدون إرجاع:-</u>

٤- أوجد عدد العينات العشوائية البسيطة التي يمكن سحبها نمن الإنتاج اليومي بحيث تتكون كل عينة من وحدتين.

٥- احسب نسبة الوحدات المعيبة في العينات الممكنة ثم أوجد التوزيع الاحتمالي لنسبة المعيب في العينات الممكنة ووضح ذلك بيانياً.

٦- أثبت أن:-

$$\mu(\hat{\theta}) = \theta$$

$$\sigma^2(\hat{\theta}) = (\frac{N-n}{N-1})(\frac{\theta(1-\theta)}{n})$$

الحل:-

بما أن $N = 4, n = 2, \theta = 0.25$. وبالتالي فإن عدد الوحدات السليمة يساوي 3 وحدات ، فإذا كانت x_1, x_2, x_3 تشير إلى الوحدات السليمة ، x_4 تشير إلى الوحدة المعيبة في المجتمع فإن:-

<u>أولاً:</u> في حالة السحب مع الإرجاع ، فإن عدد العينات الممكن سحبها يساوي:-

$(N)^n = (4)^2 = 16$ عينة

والجدول التالي (٣-٤) يوضح العينات الممكنة ونسبة المعيب في كل عينة $\hat{\theta}$.

جدول (٣-٤)

رقم العينة	العينة	النسبة في العينة $\hat{\theta}_j$	$P_r(\hat{\theta})$
1	(x_1, x_1)	**0/2=0**	**1/16**
2	(x_1, x_2)	**0/2=0**	**1/16**

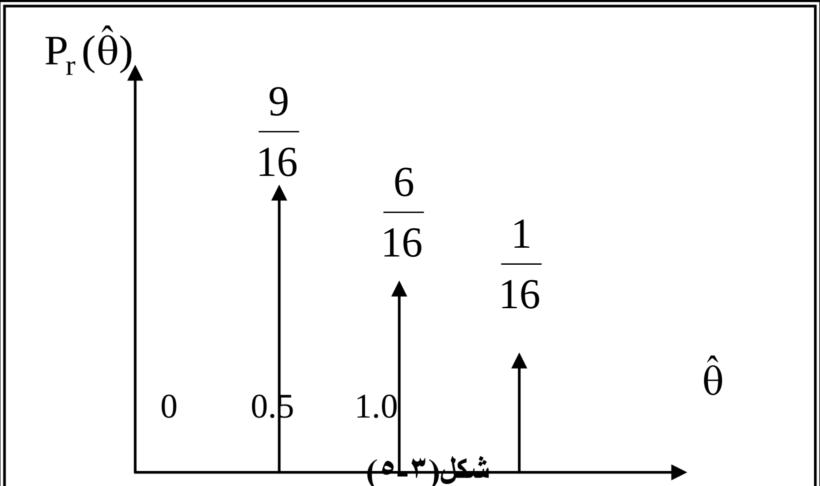
3	(x_1, x_3)	**0/2=0**	**1/16**
4	(x_1, x_4)	**1/2=0.5**	**1/16**
5	(x_2, x_2)	**0/2=0**	**1/16**
6	(x_2, x_3)	**0/2=0**	**1/16**
7	(x_3, x_4)	**1/2=0.5**	**1/16**
8	(x_2, x_4)	**1/2=0.5**	**1/16**
9	(x_2, x_3)	**0/2=0**	**1/16**
10	(x_2, x_2)	**0/2=0**	**1/16**
11	(x_4, x_4)	**2/2=1**	**1/16**
12	(x_4, x_1)	**1/2=0.5**	**1/16**
13	(x_4, x_2)	**1/2=0.5**	**1/16**
14	(x_4, x_3)	**1/2=0.5**	**1/16**
15	(x_3, x_2)	**0/2=0**	**1/16**
16	(x_3, x_1)	**0/2=0**	**1/16**
$\sum$			**16/16=1**

(٢) الجدول التالي يوضح التوزيع الاحتمالي للمتغير $\hat{\theta}$ في حالة السحب مع الإرجاع.

جدول (٣-٥)

النسبة في العينة	0	0.5	1.0	$\sum$
$P_r(\hat{\theta})$	**9/16**	**6/16**	**1/16**	**1**

والشكل التالي يوضح التوزيع الاحتمالي في الجدول السابق:-



شكل(٣-٥)

ويتضح من الشكل أن التوزيع ملتوي جهة اليمين (أي يوجد التواء موجب).

(٣) ومن الجدول نجد أن توقع النسبة في العينة $\mu(\hat{\theta})$ حيث:-

$$\mu(\hat{\theta}) = \sum_{j=1}^{3} \hat{\theta}(P(\hat{\theta}))$$

$$= 0(\frac{9}{16}) + 0.5(\frac{6}{16}) + 1(\frac{1}{16}) = \frac{4}{16} = 0.25 = \theta$$

أي أن توقع النسبة في العينة يساوي النسبة في المجتمع.

كذلك نجد أن تباين النسبة في العينة $\sigma^2(\hat{\theta})$ حيث:-

$$\sigma^2(\hat{\theta}) = [\sum_{i=1}^{3}(\theta^2 P_r(\hat{\theta}))] - [\mu(\hat{\theta})]^2$$

$$= [0(\frac{9}{16}) + (0.5)^2(\frac{6}{16}) + 1(\frac{1}{16})] - [0.25]^2$$

$$= \frac{25}{16} - (\frac{4}{16})^2 = 0.094 \qquad (3.26)$$

وبما أن

$$\frac{\theta(1-\theta)}{n} = \frac{0.25(1-0.25)}{2} = 0.094 \qquad (3.27)$$

ثانياً : في حالة السحب بدون إرجاع

في هذه الحالة يكون عدد العينات الممكنة C_n^N حيث:

$C_2^4 = 6$ عينات

والجدول التالي يوضح العينات الممكنة ونسبة المعيب في كل عينة:-

جدول (٣-٦)

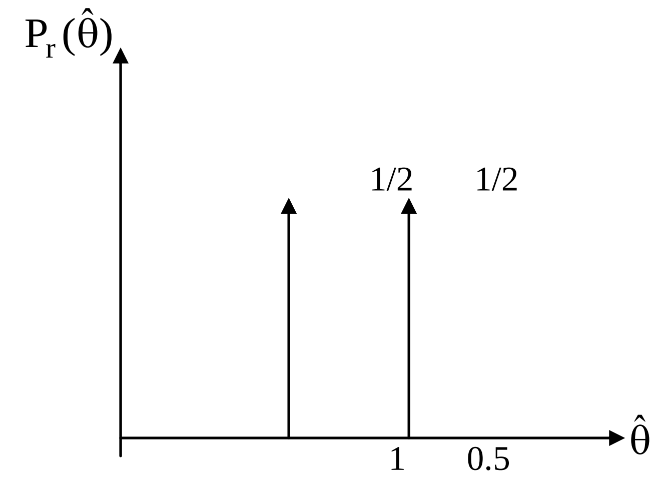
رقم العينة	العينة	النسبة في العينة $\hat{\theta}_j$	$P_r(\hat{\theta}_j)$
1	(x_1, x_2)	**0/2=0**	**1/6**
2	(x_1, x_3)	**0/2=0**	**1/6**
3	(x_1, x_4)	**1/2=0.5**	**1/6**
4	(x_2, x_3)	**0/2=0**	**1/6**
5	(x_2, x_4)	**1/2=0.5**	**1/6**
6	(x_3, x_4)	**1/2=0.5**	**1/6**
$\sum$			**1**

(٥) والجدول التالي يوضح التوزيع الاحتمالي لنسبة المعيب في العينة $\hat{\theta}$.

جدول (٣-٧)

$\hat{\theta}$	0	0.5	$\sum$
$P_r(\hat{\theta})$	3/6	3/6	1

والشكل التالي يوضح التوزيع الاحتمالي في الجدول السابق:-



شكل (٣-٦)

(٦) ومن جدول (٣-٧) نجد أن:-

$$\mu(\hat{\theta}) = 0(\frac{3}{6}) + 0.5(\frac{3}{6}) = 0.25 = \theta \tag{3.28}$$

$$\sigma^2(\hat{\theta}) = [\sum_{i=1}^{2} \hat{\theta}(P_r(\hat{\theta})] - [\mu(\hat{\theta})]^2$$

$$= [0(\frac{3}{6}) + 0.5(\frac{3}{6})] - (0.25)^2$$

$$= 0.125 - 0.0625 = 0.0625 \tag{3.29}$$

وبما أن:-

$$\frac{\theta(1-\theta)}{n}(\frac{N-n}{N-1}) = \frac{0.25(0.75)}{2}(\frac{4-2}{4-1}) = 0.0625 \tag{3.30}$$

ومن المعادلة (2.28) نجد أن توقع النسبة في العينة يساوي النسبة في المجتمع.

ومن المعادلتين (3.30) , (3.29) نجد أن التباين للنسبة في العينة يساوي :-

$$\frac{\theta(1-\theta)}{n}\left(\frac{N-n}{N-1}\right)$$

(٣-٥) تطبيقات Application

تطبيق (٣-١): الجدول التالي يعطي قيمة الإيراد في إحدى السنوات لخمسة من شركات الأدوية بالألف جنيه:-

جدول (٣-٨)

رقم الشركة	1	2	3	4	5	$\sum$
قيمة الإيراد بالألف جنيه	250	500	170	280	300	1500

فإذا تم سحب عينة عشوائية بسيطة حجمها شركتين من هذه الشركات:-

١- أوجد عدد العينات الممكن سحبها كذلك متوسط قيمة الإيراد في كل عينة.

٢- أوجد التوزيع الاحتمالي للإيراد المتوقع في العينة ووضح ذلك بيانياً.

٣- أوجد التوقع والتباين لمتوسط الإيراد بالعينة.

الحل

$$\Theta\ N = 5 \quad , \mathbf{n} = \mathbf{2}$$

$\therefore$ عدد العينات الممكن سحبها $= C_2^5 = \dfrac{5!}{2!\,3!} = 10$ عينات

كما هو موضح بالجدول التالي حيث أن:-

$$\mu = \frac{\sum_{i=1}^{5} x_i}{N} = \frac{250+500+170+280+300}{5} = 300$$

جدول (٣-٩)

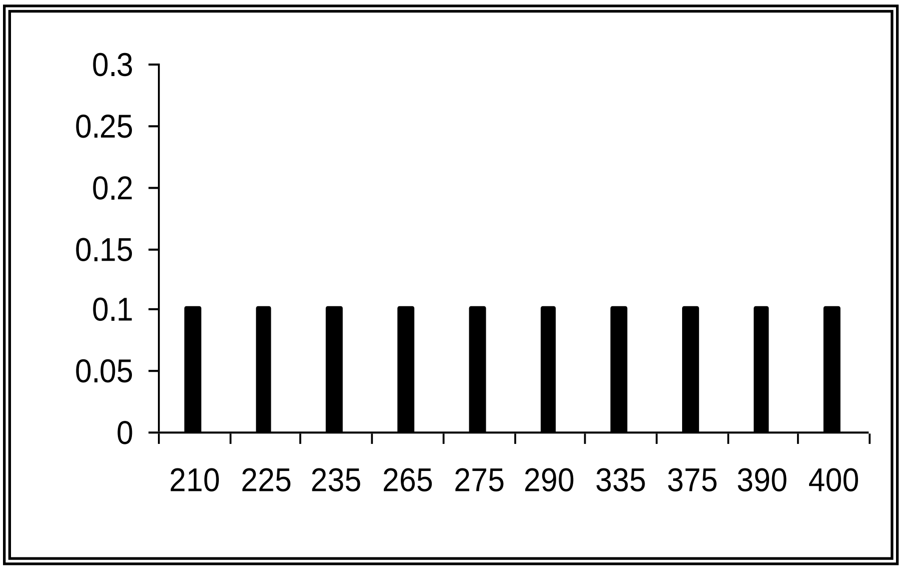
رقم العينة	العينة	متوسط العينة $\bar{X}_j$	$(\bar{X}_j - \mu)^2$
1	**(1,2)**	$\frac{250+500}{2} = 375$	$(375-300)^2 = (75)^2 = 5625$
2	**(1,3)**	$\frac{250+170}{2} = 210$	$(210-300)^2 = (-90)^2 = 8100$
3	**(1,4)**	$\frac{250+280}{2} = 265$	$(265-300)^2 = (-35)^2 = 1225$
4	**(1,5)**	$\frac{250+300}{2} = 275$	$(275-300)^2 = (-25)^2 = 625$
5	**(2,3)**	$\frac{500+170}{2} = 335$	$(335-300)^2 = (35)^2 = 1225$
6	**(2,4)**	$\frac{500+280}{2} = 390$	$(390-300)^2 = (90)^2 = 8100$
7	**(2,5)**	$\frac{500+300}{2} = 400$	$(400-300)^2 = (100)^2 = 10000$
8	**(3,4)**	$\frac{170+280}{2} = 225$	$(225-300)^2 = (-75)^2 = 5625$
9	**(3,5)**	$\frac{170+300}{2} = 235$	$(235-300)^2 = (-65)^2 = 4225$
10	**(3,6)**	$\frac{280+300}{2} = 290$	$(290-300)^2 = (-10)^2 = 100$
$\sum$		**3000**	**44850**

٢- ومن الجدول السابق نجد أن التوزيع الاحتمالي لمتوسط العينة على النحو الموضح بالجدول التالي:-

جدول (٣-١٠)

قيمة المتوسط $\bar{X}_j$	210	225	235	265	275	290	335	375	390	400	$\sum$
$P_r(\bar{X}_j)$	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**1**

والشكل التالي يوضح التوزيع الاحتمالي لمتوسط العينة $\bar{X}_j$:-



شكل (٣-٧)

ومن الجدول السابق نجد أن:-

توقع متوسط العينة $= \mu(\bar{X}) = \sum \bar{X} P_r(\bar{X})$

$= \frac{1}{10}(3000) = 300$ **ألف جنية**

كذلك

تباين متوسط العينة

$$= \sigma^2(\overline{X}) = \frac{\sum_{j=1}^{n}(\overline{X}_r - \mu)^2}{N}$$

$$= \frac{44850}{10} = 4485$$

وبالتالي الانحراف المعياري لمتوسط العينة $\sigma(\overline{X})$ حيث:-

$$\sigma(\overline{X}) = \sqrt{\sigma^2(\overline{X})} = \sqrt{4485} = 67$$ ألف جنية

تطبيق (٣-٢): في التطبيق السابق أوجد تباين الإيراد في كل عينة من العينات الممكنة السابقة ثم أوجد احتمال أن يكون الانحراف المعياري في العينة 40 ألف جنيه فأكثر.

الحل

جدول (٣- ١١)

رقم العينة	العينة	تباين العينة $S_j^2 = \frac{\sum_{j=1}^{n}(x - \overline{X}_j)^2}{2}$	الانحراف المعياري S_j
1	**(1,2)**	$\frac{(250-375)^2 - (500-375)^2}{2} = 15625$	**125**
2	**(1,3)**	$\frac{(250-210)^2 + (170-210)^2}{2} = 1600$	**40**
3	**(1,4)**	$\frac{(250-265)^2 - (280-265)^2}{2} = 225$	**15**
4	**(1,5)**	$\frac{(250-275)^2 + (300-275)^2}{2} = 625$	**25**

الانحراف المعياري S_j	تباين العينة $S_j^2 = \frac{\sum_{j=1}^{n}(x-\overline{X}_j)^2}{2}$	العينة	رقم العينة
165	$\frac{(500-335)^2+(170-335)^2}{2}=19662$	**(2,3)**	**5**
110	$\frac{(500-390)^2+(280-390)^2}{2}=12100$	**(2,4)**	**6**
100	$\frac{(500-400)^2+(300-400)^2}{2}=10000$	**(2,5)**	**7**
55	$\frac{(170-225)^2+(280-225)^2}{2}=3025$	**(3,4)**	**8**
65	$\frac{(170-235)^2+(300-235)^2}{2}=4225$	**(3,5)**	**9**
10	$\frac{(280-290)^2+(300-290)^2}{2}=100$	**(4,5)**	**10**

والعمودين الأخيرين من الجدول السابق يوضحان التباين والانحراف المعياري في كل عينة S_j, S_j^2 على الترتيب.

والجدول التالي يعطي التوزيع الاحتمالي للانحراف المعياري في العينة:-

جدول (٣-١٢)

S_j	**10**	**15**	**25**	**40**	**55**	**65**	**100**	**110**	**125**	**165**	$\sum$
$P_r(S_j)$	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**0.1**	**1**

ومن الجدول السابق يتضح أن احتمال أن يكون الانحراف المعياري للإيراد في العينة 40 ألف جنيه فأكثر يساوي $P_r(S_j > 40)$ حيث:-

$$P_r(S_j > 40) = P_r(S_j = 55) + P_r(S_j = 65) + P_r(S_j = 100)$$
$$+ P_r(S_j = 110) + P_r(S_j = 125) + P_r(S_j = 165)$$
$$= 0.7$$

تطبيق (٣-٣): **قامت إحدى بيوت الخبرة بالإعلان عن وظيفتي مخططي برامج كمبيوتر ، حيث تعتبر سنوات العمل في تخطيط البرامج (سنوات الخبرة) هو أهم معيار للتفضيل بين المتقدمين فإذا تقدم لهاتين الوظيفتين 10 أفراد ، والجدول التالي يوضح سنوات الخبرة لكل متقدم:-**

جدول (٣-١٣)

رقم المتقدم	1	2	3	4	5	6	7	8	9	10
عدد سنوات الخبرة	6	1	2	9	5	8	4	3	10	7

المطلوب

١- **أوجد متوسط سنوات الخبرة للمتقدمين** μ **، كذلك التباين لسنوات الخبرة** σ^2.

٢- **أوجد عدد البدائل (العينات)الممكنة لاختيار 2 من المتقدمين ، ثم أوجد الوسط الحسابي لسنوات الخبرة في كل عينة.**

٣- **أوجد التوزيع الاحتمالي لمتوسط سنوات الخبرة في العينة ، ثم أوجد توقعه** $\mu(\bar{X})$.

٤- **أوجد احتمال أن يكون متوسط سنوات الخبرة في العينة أكبر من 8 سنوات.**

الحل

$$1)\quad \mu = \frac{\sum_{i=1}^{10} x_i}{N} = \frac{6+1+2+9+5+8+4+3+10+7}{10}$$

$$= \frac{55}{10} = 5.5$$

$$\sigma^2 = \frac{\sum_{i=1}^{10} (x_i - \mu)^2}{N}$$

$$= \frac{1}{10}[(6-5.5)^2 + (1-5.5)^2 + ... + (7-5.5)^2 = 4.85$$

٣- عدد البدائل الممكنة يساوي C_2^{10} حيث

$C_2^{10} = 45$ بديل

والجدول التالي يوضح البدائل الممكنة ومتوسط سنوات الخبرة في كل بديل (عينة):-

جدول(٣-١٤)

متوسط سنوات الخبرة $\overline{X}_j$	العينة	رقم العينة	متوسط سنوات الخبرة $\overline{X}_j$	العينة	رقم العينة
6.5	**(3,10)**	**24**	**1.5**	**(1,2)**	**1**
4.5	**(4,5)**	**25**	**2**	**(1,3)**	**2**
5	**(4,6)**	**26**	**2.5**	**(1,4)**	**3**
5.5	**(4,7)**	**27**	**3**	**(12,5)**	**4**
6	**(4,8)**	**28**	**3.5**	**(1,6)**	**5**

6	(1,7)	4	29	(4,9)	6.5
7	(1,8)	4.5	30	(4,10)	7
8	(1,9)	5	31	(5,6)	5.5
9	(1,10)	5.5	32	(5,7)	6
10	(2,3)	2.5	33	(5,8)	6.5
11	(2,4)	3	34	(5,9)	7
12	(2,5)	3.5	35	(5,10)	7.5
13	(2,6)	4	36	(6,7)	6.5
14	(2,7)	4.5	37	(6,8)	7
15	(2,8)	5	38	(6,9)	7.5
16	(2,9)	5.5	39	(6,10)	8
17	(2,10)	6	40	(7,8)	7.5
18	(3,4)	3.5	41	(7,9)	8
19	(3,5)	4	42	(7,10)	8.5
20	(3,6)	4.5	43	(8,9)	8.5
21	(3,7)	5	44	(8,10)	9
22	(3,8)	5.5	45	(9,10)	9.5
23	(3,9)	6			

والجدول التالي يعطي التوزيع الاحتمالي للمتغير $\overline{X}_j$ ، كذلك باستخدامه يمكن حساب $\mu(\overline{X}), \sigma^2(\overline{X})$.

جدول (٣-١٥)

$\overline{X}_j$	$P_r(\overline{X})$	$\overline{X}_j P_r(\overline{X}_j)$
1	**0**	**0**
1.5	**1/45**	**1.5/45**
2	**1/45**	**2/45**
2.5	**2/45**	**5/45**
3	**2/45**	**6/45**
3.5	**3/45**	**10.5/45**
4	**3/45**	**12/45**
4.5	**4/45**	**18/45**
5	**4/45**	**20/45**
5.5	**5/45**	**27.5/45**
6	**4/45**	**24/45**
6.5	**4/45**	**26/45**
7	**3/45**	**21/45**
7.5	**3/45**	**22.5/45**
8	**2/45**	**16/45**
8.5	**2/45**	**17/45**
9	**1/45**	**9/45**
9.5	**1/45**	**9.5/45**
1	**1**	**5.5**

ومن الجدول يتضح أن: $\mu(\overline{X}) = 5.5$

أي تساوي التوقع في المجتمع أي أن:-

$$\mu(\overline{X}) = \mu$$

$$4)\ \Theta\ \sigma^2(\overline{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = 4.31$$

ومن نظرية النزعة المركزية نظرية (٣-٤) نجد أن المتغير $Y = \dfrac{\overline{X} - \mu(\overline{X})}{\sigma(\overline{X})}$

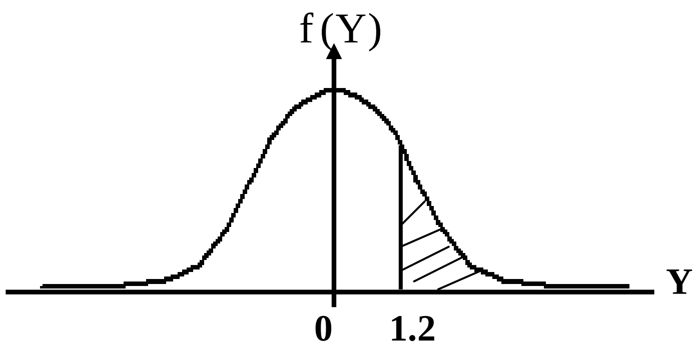
يتبع التوزيع المعتاد القياسي وبالتالي فإن:-

احتمال أن يكون متوسط سنوات الخبرة في العينة أكبر من 8.5 سنوات يساوي

$$P_r(x > 8) = P_r(Y > \frac{5.5 - 8.5}{2.08})$$

$$= P_r(Y > 1.2) = 0.5 - P_r(0 < Y < 1.2)$$

$$= 0.5 - 0.3849 \approx 0.12$$



شكل (٣-٨)

تطبيق (٣-٤): سحبت عينة عشوائية بسيطة من مجتمع معتاد فوجد أن توقع وتباين الوسط الحسابي في العينة على النحو التالي:-

$$\mu(x) = 100 \qquad , \sigma^2(x)225$$

احسب الاحتمالات التالية:-

1- احتمال أن يكون المتوسط في العينة أقل من 118 .
2- احتمال أن يكون المتوسط في العينة لا يقل عن 119 ولا يزيد عن 129 .
3- احتمال أن يكون المتوسط أكبر من 125 .

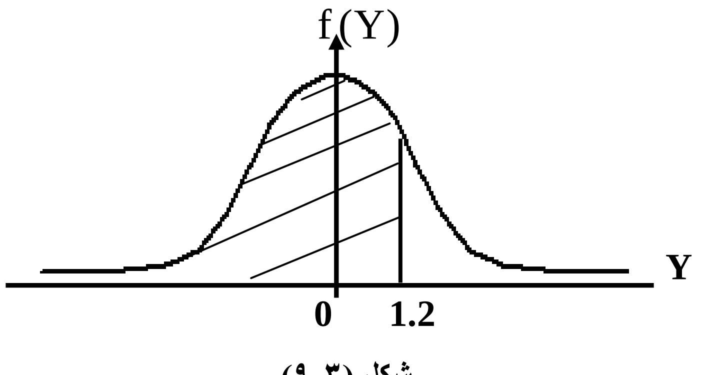
الحل

∵ المتوسط في العينة $\overline{X}$ متغير يتبع التوزيع المعتاد بتوقع $\mu = 100$ وتباين

$$\sigma^2(x) = 225$$

∴ المتغير Y حيث $Y = \dfrac{\overline{X} - \mu}{\sigma(\overline{X})}$ يتبع التوزيع المعتاد القياسي .

$$1)\ P_r(x > 118) = P_r(Y < \frac{118-100}{15})$$
$$= P_r(Y < 1.2) = 0.5 + P_r(0 < Y < 1.2)$$
$$= 0.5 + 0.3849 = 0.8849$$

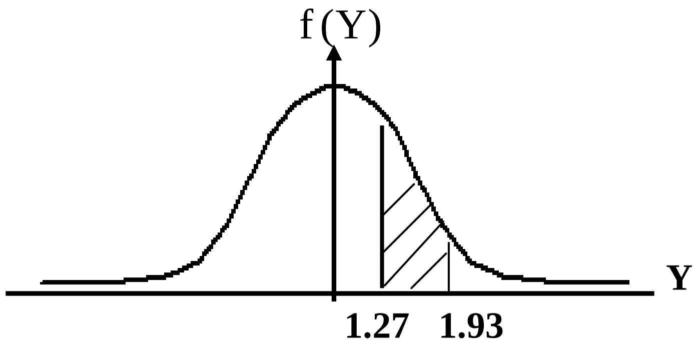


شكل (٣-٩)

$$2)\ P_r(119 < x < 129) = P_r(\frac{119-100}{15} < Y < \frac{118-100}{15})$$

$$= P_r(1.27 < Y < 1.93) = P_r(0 < Y < 1.93) - P_r(0 < Y < 1.27)$$

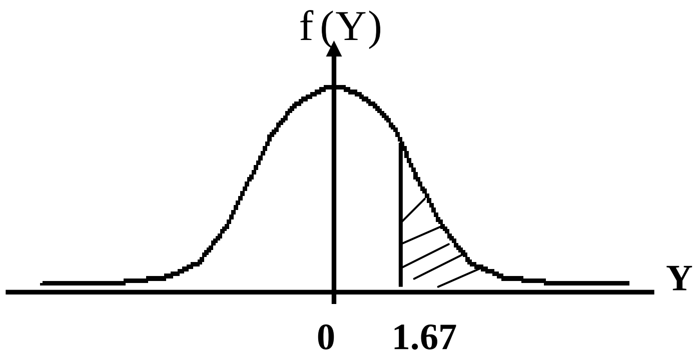
$$= 0.3238 - 0.1443 = 0.1795$$



شكل (٣-١٠)

$$3)\ P_r(x > 125) = P_r(Y > \frac{125-100}{15})$$

$$= P_r(Y > 1.67) = 0.5 - 0.0025 = 0.475$$



شكل (٣-١١)

تطبيق (٣-٥): يقوم أحد البنوك بتحمل نفقات علاج العاملين به فإذا كان عدد العاملين بالبنك 1500 عامل ، وفي أحد السنوات رصد البنك لكل عامل في المتوسط

25.75 جنيه للعلاج بانحراف معياري 5.25 جنيه . فإذا أخذت عينة مكونة من 100 عامل واختير أحد العاملين عشوائياً.

المطلوب

1. **احسب الانحراف المعياري لمتوسط المنفق على العامل في العينة.**
2. **احسب احتمال أن يكون متوسط الإنفاق على العلاج للعامل في العينة لا يزيد عن 27 جنيه ولا يقل عن 25 جنيه في هذه السنة.**
3. **احسب احتمال أن يكون متوسط الإنفاق على علاج العامل في العينة أكبر من 26 جنيه.**

الحل

$\because$ حجم المجتمع المسحوب منه العينة N=1500 عامل وحجم العينة n=100 ومن نظرية (٣-١) نجد أن :-

1. **تباين متوسط العينة $\sigma^2(\overline{X})$ حيث أن :-**

$$\sigma^2(\overline{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

$$= \frac{5.25}{100}\left(\frac{1500-100}{1500-1}\right) = 0.257$$

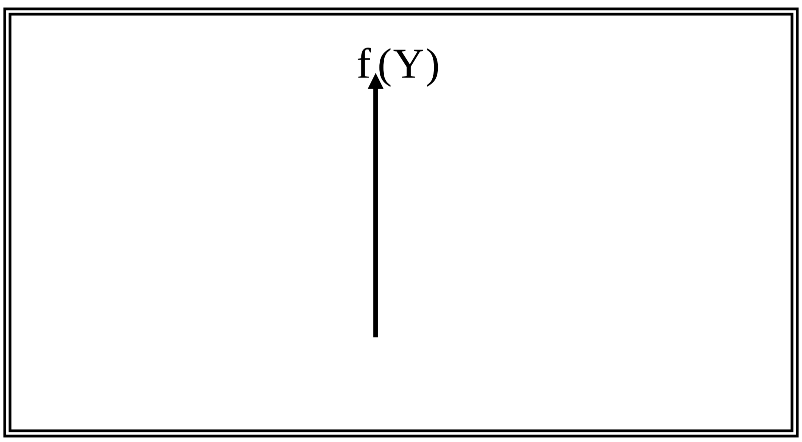
$$\sigma(\overline{X}) = \sqrt{0.257} \approx 0.507$$

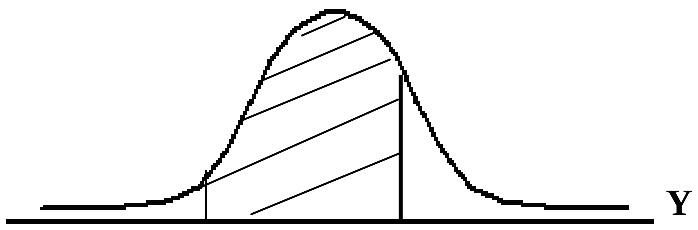
2. **من نظرية (٣-٣) نجد أن المتغير Y حيث :-**

$Y = \dfrac{\overline{X}-\mu(\overline{X})}{\sigma(\overline{X})}$ يتبع التوزيع المعتاد القياسي

$$\therefore P_r(25 < x < 27) = P_r(-1.47 < Y < 2.45)$$

$$= 0.4292 + 0.4929 = 0.9221$$





شكل (٣-١٢)

$$3)\ P_r(x > 26) = P_r(Y > \frac{26 - 25.75}{0.51})$$
$$= P(Y > 0.19) = 0.5 - 0.1879$$
$$= 0.3121$$

تطبيق (٣-٦): في إحدى محافظات الجمهورية معروف أن 60% من المرشحين لانتخابات مجلس الشعب من المستقلين . فإذا اختيرت عينة من المرشحين في هذه المحافظة مكونة من 150 مرشح أوجد ما يلي:-

١- النسبة المتوقعة للمرشحين المستقلين في العينة من هذه المحافظة.

٢- احتمال أن تكون نسبة المرشحين المستقلين في هذه العينة لا تزيد عن 50% ولا تقل عن 70% .

الحل:

إذا كانت $\overline{Y}, Y$ **هما النسبة في المجتمع والعينة على التوالي:-**

$$\because Y = 0.60 \quad , \quad n = 150$$

ومن نظرية (٣-٣) نجد أن :-

$$\mu(\overline{Y}) = Y \rightarrow \mu(\overline{Y}) = 0.6$$
$$\sigma^2(\overline{Y}) = \frac{Y(1-Y)}{n}(\frac{N-n}{N-1})$$
$$\frac{0.6(10.6)}{150}(\frac{1500-150}{1500-1}) = 0.00144$$
$$\sigma(\overline{Y}) = 0.03796$$

وباستخدام نظرية (٣-٨) نجد أن المتغير Z حيث:-

$$Z = \frac{\overline{Y} - \mu(\overline{Y})}{\sigma(\overline{Y})}$$

يتبع التوزيع المعتاد القياسي

$$\therefore P_r(0.5 < x < 0.7) = P_r(\frac{0.5-0.6}{0.04} < Z < \frac{0.7-0.6}{0.04})$$
$$= P_r(-2.5 < Z < 2.5) = 2P_r(0 < Z < 2.5)$$
$$= 2(0.4952) = 0.9904$$

تطبيق (٣-٧): قامت إحدى المؤسسات التي تصدر صحيفة يومية بدراسة الطلب اليومي على الجريدة في إحدى المحافظات بالوجه البحري فوجد أن 60% من سكان هذه المحافظة يقومون بشراء الجريدة . فإذا سحبت عينة من٠ سكان المحافظة مكونة من50 فرداً.

المطلوب

١- أوجد نسبة الأفراد المتوقعة الذين يقومون بشراء الجريدة في العينة.

٢- أوجد احتمال أن تكون نسبة الأفراد في العينة لا تزيد عن 10% .

الحل

1) $\Theta\ \theta = 0.60$

$$\therefore \mu(\hat{\theta}) = 0.60$$

$$\sigma^2(\theta) = \sqrt{\frac{\theta(1-\theta)}{n}}$$

$$= \sqrt{\frac{(0.60)(0.40)}{50}} = \sqrt{\frac{0.24}{50}} = \sqrt{0.0048} = 0.07$$

٢- احتمـــال أن تكـــون نســـبة الأفـــراد فـــي العينـــة الـــذين يشـــترون الجريـــدة

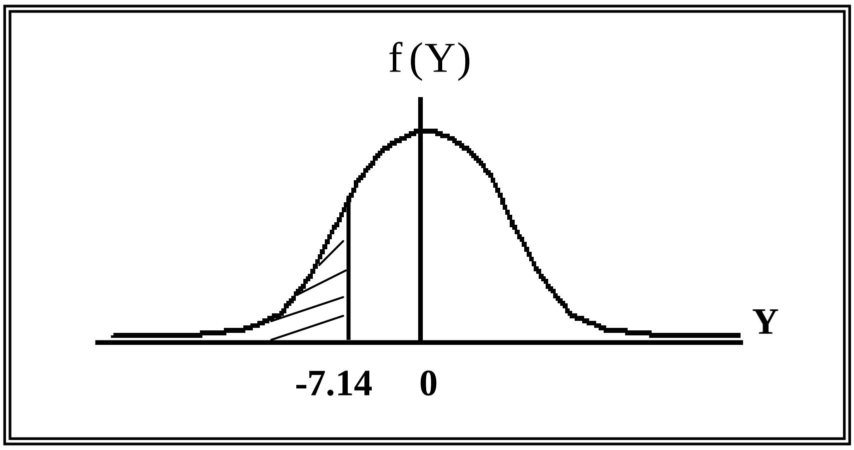
$$P_r(\theta < 0.10) =$$

$$\Theta\ \frac{\hat{\theta} - \mu(\hat{\theta})}{\sqrt{\sigma(\theta)}}$$

متغير يتبع التوزيع القياسي فإن:-

$$P_r(\theta < 0.10) = P_r(Y < \frac{0.10 - 0.60}{0.07})$$

$$= P_r(Y < -7.14) = 0.5 - 0.5 = 0$$



شكل (٣-١٣)

تطبيق (٣-٨): إذا أجريت دراسة لاستطلاع أراء 5 أفراد بشأن استمرار أحد برامج التلفزيون . فإذا كانت الإجابة " نعم " تعني الموافقة على استمرار البرنامج ، وإذا كانت الإجابة " لا " فهذا يعني عدم الموافقة على استمرار البرنامج . فإذا كانت البيانات على النحو التالي :-

" نعم ، لا ، نعم ، لا ، لا"

المطلوب

١- أوجد نسبة الأفراد الراغبين في استمرار البرنامج.

٢- إذا سحبت عينة مكونة من 3 أفراد ، أوجد التوزيع الاحتمالي لنسبة الراغبين في العينة.

٣- احسب احتمال أن تكون نسبة الموافقة على استمرار البرنامج في العينة لا تقل عن 30% ولا تزيد عن 80% .

الحل

١- إذا فرضنا أن $\overline{Y}, Y$ تشير إلى النسبة في المجتمع والعينة على الترتيب:-

$$Y = \frac{2}{5} = 0.4$$

٢- إذا فرضنا أن x تشير إلى الإجابة المقررة رقم j في المجتمع حيث j=(1,2,3,4,5) وفي هذه الحالة يكون عدد العينات الممكن سحبها يساوي $C_2^5 = 10$ عينات كما هو موضح بالجدول التالي:-

جدول (٣-١٦)

رقم العينة	عينة الأفراد	عينة الآراء	النسبة في العينة $\overline{Y}$
1	X_1, X_2, X_3	نعم ، لا ، نعم	2/3
2	X_1, X_2, X_4	نعم ، لا ، لا	1/3
3	X_1, X_2, X_5	نعم ، لا ، لا	1/3
4	X_2, X_3, X_4	لا ، نعم ، لا	1/3
5	X_2, X_3, X_5	لا ، نعم ، لا	1/3
6	X_2, X_4, X_5	لا ، لا ، لا	0
7	X_1, X_2, X_4	نعم، نعم ، لا	2/3
8	X_1, X_2, X_5	نعم، نعم ، لا	2/3
9	X_1, X_4, X_5	نعم ، لا ، لا	1/3
10	X_3, X_4, X_5	نعم ، لا ، لا	1/3

ومن الجدول السابق يمكن تكزين جدول التوزيع الاحتمالي للنسبة في العينة $\overline{Y}$ على النحو التالي:-

جدول (٣-١٧)

$\overline{Y}$	0	1/3 = 0.33	2/3 = 0.67	$\sum$
$P_r(\overline{Y})$	0.1	0.6	0.3	1

ومن جدول التوزيع الاحتمالي السابق للنسبة في العينة $\overline{Y}$ نجد أن

$$P_r(0.3 \leq Y \leq 0.8)$$

$$= P_r(Y = 0.33) + P_r(Y = 0.67) = 0.6 + 0.3 = 0.9$$

(٣-٦) تمرينــات
Exercises

(٣-١): الجدول التالي يعطي الربح السنوي في إحدى السنوات لخمسة شركات من شركات الأدوية بالألف جنيه:-

جدول (٣-١٨)

رقم الشركة	1	2	3	4	5
الربح السنوي بالألف جنيه	250	500	170	280	300

فإذا تم سحب عينة عشوائية بسيطة حجمها شركتين من هذه الشركات

المطلوب

١- أوجد عدد العينات الممكن سحبها كذلك أوجد الوسط الحسابي (المتوسط) والانحراف المعياري للربح في هذه السنة في كل عينة من العينات الممكن سحبها .

٢- أوجد التوزيع الاحتمالي لمتوسط الربح في العينة ووضح ذلك بيانيا.

٣- أوجد التوقع والتباين لمتوسط الربح بالعينة.

(٣-٢): في التمرين السابق أوجد توقع التباين للربح في العينة ثم أوجد احتمال أن يكون تباين العينية 40 ألف جنية فأكثر .

(٣-٣): سحبت عينة عشوائية بسيطة من مجتمع معتاد فوجد أن توقع وتباين الوسط الحسابي في العينة علي النحو التالي

$$\mu(\bar{x}) = 100, \sigma^2(\bar{x}) = 15$$

احسب الاحتمالات التالية:-

١- احتمال أن يكون المتوسط في العينة أقل من 118.

٢- احتمال أن يكون المتوسط في العينة لا يقل عن 119 ولا يزيد عن 129.

٣- احتمال أن يكون المتوسط في العينة أكبر من 125.

(٣-٤): مجتمع يتكون من 10000 مفردة ويتوزع المتغير محل الدراسة توزيعاً منتظماً . فإذا سحبت عينة حجمها 20 مفردة ، ما هو احتمال أن يكون قيمة متوسط

المتغير في العينة لا تزيد عن **40** ولا تقل عن **30** إذا كانت القيمة المتوقعة لمتوسط المتغير في العينة وتباينه هما $\mu(x) = 30, \sigma^2(x) = 0.25$. حدد الاختلاف في قيمة الاحتمال المطلوب إذا كان حجم العينة **100** مفردة مع ذكر السبب.

(٣-٥): قامت إحدى شركات إنتاج محركات الطائرات بإجراء دراسة عن فترة صلاحية المحرك فأثبتت الدراسة أن فترة صلاحية المحرك تتبع التوزيع المعتاد بتوقع 2000 ساعة طيران وانحراف معياري $\sigma = 324$ ساعة طيران ، فإذا سحبت عينة عشوائية بسيطة مكونة من 40 محرك

المطلوب

١- حدد التوزيع الاحتمالي لمتوسط فترة الصلاحية للمحرك في العينة $\overline{X}$ ، ووضح ذلك بيانياً.

٢- أوجد احتمال أن تكون فترة صلاحية المحرك المتوقعة في العينة $\mu(\overline{X})$ أكبر من **2500** ساعة.

٣- أوجد احتمال أن يكون متوسط فترة صلاحية المحرك في العينة أقل من **1800** ساعة.

(٣-٦): قامت إحدى المؤسسات التي تصدر صحيفة يومية بدراسة الطلب اليومي على الجريدة في إحدى المحافظات بالوجه البحري فوجد أن 60% من سكان هذه المحافظة يقومون بشراء الجريدة، فإذا سحبت عينة من سكان المحافظة مكونة من 50 فرد

المطلوب

١- أوجد نسبة الأفراد المتوقعة الذين يقومون بشراء الجريدة في العينة.

٢- أوجد احتمال أن تكون نسبة الأفراد في العينة لا تزيد عن **10%** .

(٣-٧): إذا كانت درجات طلاب في مادة الإحصاء (النهاية العظمى للدرجة 100 درجة) هي :-
.50,20,80,30,70

المطلوب

١- أوجد الدرجة المتوقعة للطالب وكذلك الانحراف المعياري.

٢- إذا تم سحب عينة عشوائية بسيطة مكونة من درجات 3 طلاب وكان السحب مع الإرجاع . أوجد التوزيع الاحتمالي لمتوسط العينة وحدد العلاقة بين متوسط العينة ومتوسط الدرجة في المجتمع.

٣- إذا تم السحب بدون إرجاع أوجد التوزيع الاحتمالي لمتوسط العينة ثم أوجد التوقع والتباين ومتوسط الدرجة في العينة

٤- يعتبر الطالب ناجح في المادة إذا حصل على 50 درجة فأكثر

أ- أوجد نسبة الناجحين في العينات الممكنة في حالة السحب بدون إرجاع.

ب- أوجد احتمال أن تكون نسبة الناجحين في العينة 50% فأكثر.

جـ أوجد احتمال أن تكون نسبة الناجحين في العينة 0.

(٣-٨): في التمرين السابق أوجد التوزيع الاحتمالي لتباين العينة في حالة السحب بدون إرجاع ثم حدد العلاقة بين التباين في المجتمع والتباين في العينة.

(٣-٩): إذا أجرى بحث على 5 أفراد لمعرفة حاملي ميكروب معين فكانت النتائج على النحو التالي:-

غير حامل للميكروب ، حامل للميكروب ، غير حامل للميكروب ، حامل للميكروب ، غير حامل للميكروب.

المطلوب

١- اوجد نسبة الحاملين للميكروب في المجتمع.

٢- إذا تم سحب عينة عشوائية بسيطة من 3 أفراد وبفحص الأفراد فيها كانت النتائج على النحو التالي:-

حامل للميكروب ، غير حامل للميكروب ، حامل للميكروب . أوجد نسبة الحاملين للميكروب في العينة ثم أوجد التوقع والتباين لهذه النسبة.

٣- إذا تم سحب عينة عشوائية بسيطة من فردين ، أوجد التوزيع الاحتمالي للنسبة في العينة.

٤- أوجد العلاقة بين النسبة في المجتمع والنسبة في العينة.

(٣-١٠): إذا أجريت دراسة لاستطلاع أراء 10 أفراد بشأن استمرار أحد برامج التلفزيون . فإذا كانت الإجابة " نعم " تعني الموافقة على استمرار البرنامج ، وإذا كانت الإجابة " لا " فهذا يعني عدم الموافقة على استمرار البرنامج . فإذا كانت البيانات على النحو التالي :-

" نعم ، لا ، نعم ، لا ، لا ، لا ، نعم ، لا ، نعم ، نعم "

المطلوب

١- أوجد نسبة الأفراد الراغبين في استمرار البرنامج.

٢- إذا سحبت عينة مكونة من 3 أفراد ، أوجد التوزيع الاحتمالي للنسبة في العينة ، ثم حدد العلاقة بين النسبة في العينة والنسبة في المجتمع.

٣- احسب احتمال أن تكون نسبة الموافقون على استمرار البرنامج في العينة لا تقل عن 30% ولا تزيد عن 80% .

الباب الرابع
التقدير الإحصائي
Statistical Estimation

(٤-١) التقـــدير الإحصــــائي **Statistical Estimation**

(٤-٢) خصائص التقدير الجيد
Properties of Good Estimators

(٤-٣) تقدير توقع المجتمع
The Estimation of Population's Mean

(٤-٤) تقدير تباين المجتمع
The Estimation of Population's Variance

(٤-٥) تقدير النسبة في المجتمع
The Estimation of Population's Proportion

(٤-٦) خطأ المعاينة والدقة
Sampling Error and Precision

(٤-٧) حجم العينة الأمثل لتقدير متوسط المجتمع
Optimum Sample Size for Estimating Population's Mean

(٤-٨) حجم العينة الأمثل لتقدير النسبة في المجتمع
Optimum Sample Size for Estimating Population's Proportion

(٤-٩) تطبيقات
Applications

(٤-١٠) تمرينات
Exercises

(٤-١) التقدير الإحصائي Statistical Estimation

إذا كان الأسلوب المتبع في جمع البيانات للمتغيرات محل الدراسة هو أسلوب الحصر الشامل * فإن القيم الحقيقية لمعالم المجتمع **Population′s Parameters** (معالم المجتمع جميع المؤشرات التي تعكس خصائص المتغير أو المتغيرات محل الدراسة في المجتمع محل الدراسة) يتم تحديدها بالضبط من البيانات المجمعة عن المتغير (أو المتغيرات) موضع الدراسة. ولكن نظراً للعيوب التي تنشأ نتيجة إتباع أسلوب الحصر الشامل أو استحالة استخدامه في عدة حالات فعادة يفضل استخدام أسلوب العينة في جمع البيانات. ومن بيانات العينة يتم حساب معلمات العينة **Sample′s Parameters** التي باستخدامها يتم تقدير معلمات المجتمع محل الدراسة.

ومن خلال دراستنا في الباب السابق، يتضح أن معلمات (مؤشرات) العينات المسحوبة من مجتمع معين تمثل متغيرات عشوائية ذات توزيعات احتمالية .

ونظراً لأننا نستخدم معلمات العينة كتقديرات **Estimators** لمعلمات المجتمع المناظرة لها، لذلك سوف نحاول في هذا الباب الإجابة على التساؤلات التالية :

١- هل التقدير المحسوب من بيانات العينة تقدير جيد أم لا ؟

٢- هل يتم حساب التقدير من عينة واحدة أم من عدة عينات ؟ وما هو الحجم الأمثل لهذه العينة (أو العينات) ؟

٣- كلمة تقدير تعني أننا نحصل على قيمة تقريبية لمعلمة المجتمع باستخدام بيانات العينة. لذلك من الأهمية معرفة كمية الخطأ نتيجة عملية التقدير، وهل هذه الكمية من الخطأ مقبولة أم لا .

وسوف نتناول الإجابة على هذه الأسئلة من خلال فصول هذا الباب.

* أ.د/ عفاف الدش (٢٠٠٥): الإحصاء التطبيقي – الجزء الأول – جهاز نشر وتوزيع الكتاب الجامعي – جامعة حلوان – القاهرة.

(٤-٢) خصائص التقدير الجيد
Properties of Good Estimators

يعتبر التقدير المحسوب من بيانات العينة كتقدير لأحد معالم المجتمع تقدير جيد إذا كان التوزيع الاحتمالي لهذا التقدير يتركز حول المعلمة المجهولة المراد تقديرها. فعلي سبيل المثال إذا سحبت عينة عشوائية من مجتمع معتاد (أي السلوك الاحتمالي للمتغير محل الدراسة في المجتمع يمثل متغير يتبع التوزيع المعتاد) بمتوسط μ وكان المطلوب تقدير قيمة μ عن طريق الوسط الحسابي (المتوسط) للعينة $\overline{X}$. ويعتبر التقدير $\overline{X}$ تقدير جيد للمعلمة μ حيث أن التوزيع الاحتمالي للمتغير ($\overline{X}$) يتركز حول قيمة المعلمة (μ).

وفيما يلي سوف تقدم خصائص التقدير الجيد أو بعبارة أخري الخصائص التي يمكن باستخدامها معرفة هل التوزيع الاحتمالي للتقدير يتركز حول المعلمة المراد تقديرها أم لا.

أولاً: عدم التحيز Unbiasedness

نظرية (٤-١)

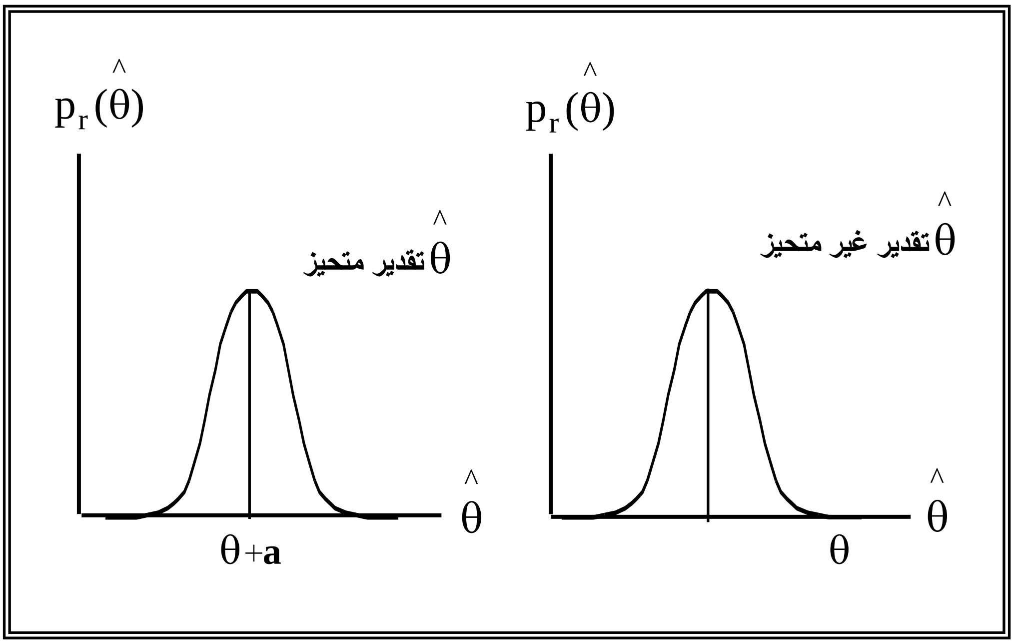
إذا كانت $\hat{\theta}$ هي تقدير لمعلمة المجتمع θ، حيث تم حساب $\hat{\theta}$ من بيانات عينة عشوائية، فإن $\hat{\theta}$ تعتبر تقدير غير متحيز للمعلمة θ إذا كان:

$$\mu(\hat{\theta}) = \theta \qquad (4.1)$$

وتعتبر $\hat{\theta}$ تقدير متحيز Biased إذا كان:

$$\mu(\hat{\theta}) = \theta + a \qquad (4.2)$$

حيث a مقدار ثابت Constant يسمي بحد التحيز Bias Term. والشكل التالي يوضح التوزيع الاحتمالي للتقدير $\hat{\theta}$ إذا كان غير متحيز أو متحيز.



شكل (٤-١)

مثال (٤-١)

إذا سحبت عينة عشوائية بسيطة حجمها n من مجتمع معتاد توقعه μ وتباينه σ^2.

١- أثبت أن الوسط الحسابي في العينة $\overline{X}$ تقدير غير متحيز.

٢- إذا فرضنا أن S^2 تشر إلى تباين المتغير في العينة، فإذا كان:

$$S_1^2 = \frac{\sum (x-\overline{X})^2}{n-1} \quad , \quad S_2^2 = \frac{\sum (x-\overline{X})^2}{n}$$

أثبت أن S_1^2 تقدير غير متحيز لتباين المجتمع σ^2، وان S_2^2 تقدير متحيز لتباين المجتمع σ^2.

الحل:

١- بما أن

نظرية (٣-٤)　　$\mu(\overline{X}) = \mu$

فإن $\overline{X}$ يعتبر تقدير غير متحيز للمعلمة μ (أنظر نظرية (٣-١))

٢- من نظرية (٣-٥) نجد أن:

$$\mu(S_1^2) = \sigma^2$$

فإن التقدير S_1^2 يعتبر تقدير غير متحيز للمعلمة σ^2.

كذلك بما أن:

$$\mu(S_2^2) = \mu\left(\frac{n-1}{n} - S_1^2\right) = \left(\frac{n-1}{n}\right) + \mu(S_1^2) = \left(\frac{n-1}{n}\right) + \sigma^2 \qquad (4.3)$$

من المعادلة (4.3) نجد أن:

$$\mu(S_2^2) = \sigma^2 + \left(\frac{n-1}{n}\right) \qquad (4.4)$$

من المعادلة (4.4) نجد أن $\mu(S_2^2) \neq \sigma^2$. وبالتالي فإن S_2^2 يعتبر تقدير متحيز للمعلمة σ^2. حيث حد التحيز في هذه الحالة **a** بحيث:

$$a = \left(\frac{n-1}{n}\right)$$

ثانياً: الاتساق　　Consistency

يقال أن التقدير $\hat{\theta}$ تقدير متسق **Consistent** للمعلمة θ إذا كان $\hat{\theta}$ تقترب من θ كلما ذاد حجم العينة العشوائية البسيطة **n** (أي $\hat{\theta} \to \theta$ عندما $n \to \infty$)

ويمكن أثبات أن التقدير $\hat{\theta}$ يكون تقدير متسق للمعلمة θ إذا كان $\hat{\theta}$ تقدير غير متحيز للمعلمة θ حيث أن:

$$\lim_{n\to\infty} \sigma^2(\hat{\theta}) = 0 \qquad \textbf{(4.5)}$$

مثال (٤-٢)

إذا كان $\overline{X}$ متوسط العينة العشوائية المسحوبة من مجتمع معتاد توقعه μ وتباينه σ^2 فإن:

$$\mu(\overline{X}) = \mu \qquad \textbf{(4.6)}$$

$$\sigma^2(\overline{X}) = \frac{\sigma^2}{n} \qquad \textbf{(4.7)}$$

من (4.6) نجد أن $\overline{X}$ تقدير غير متحيز لـ μ، ومن (4.7) نجد أن:

$$\lim_{n\to\infty} \sigma^2(\overline{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0 \qquad \textbf{(4.8)}$$

يتضح من المعادلتين (4.6) ، (4.8) أن الوسط الحسابي في العينة ($\overline{X}$) تقدير غير متحيز ومتسق لتوقع المجتمع (μ).

ثالثاً: الكفاءة Efficiency

إذا كان يوجد تقديران $\hat{\theta}_1, \hat{\theta}_2$ للمعلمة θ، فإننا نقول أن $\hat{\theta}_1$ تقدير للمعلمة θ أكثر كفاءة من $\hat{\theta}_2$ إذا كان:

$$\sigma^2(\hat{\theta}_1) < \sigma^2(\hat{\theta}_2) \qquad \textbf{(4.9)}$$

مثال (٤-٣)

إذا كان $\overline{X}$ هو متوسط عينة حجمها n تم سحبها مع الإرجاع من مجتمع حجمه N وتباينه σ^2 فإن:

$$\sigma^2(\overline{X}) = \frac{\sigma^2}{n} \qquad \textbf{(4.10)}$$

أما إذا كان السحب بدون إرجاع فإن:

$$\sigma^2(\overline{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) \qquad \textbf{(4.11)}$$

وبما أن المقدار $\left(\frac{N-n}{N-1}\right)$ بحيث:

$$0 < \left(\frac{N-n}{N-1}\right) < 1$$

أي أن المقدار $\left(\frac{N-n}{N-1}\right)$ قيمة كسرية.

إذن ($\overline{X}$) في حالة السحب بدون إرجاع كما هو واضح في (4.11) أقل من تباين ($\overline{X}$) في حالة السحب مع الإرجاع كما هو واضح في (4.10). وبالتالي يصبح متوسط العينة ($\overline{X}$) تقدير كفء في حالة السحب بدون إرجاع، كذلك يعتبر ($\overline{X}$) غير كفء إذا كان السحب مع الإرجاع.

رابعاً: الكفاية   **Sufficiency***

* لمزيد من التفصيل أنظر المرجع:
Mood. A.. Graybill. F.. and Boes. D. (1963): "Introduction to The Theory of Statistics" , McGraw-Hill.

إذا كان التقدير $\hat{\theta}$ يستخدم جميع المعلومات الموجودة في العينة والملائمة لتقدير المعلمة θ، فإنه يقال أن التقدير $\hat{\theta}$ تقدير كافي **Sufficient**. فإذا تم حساب التقدير $\hat{\theta}$ من بيانات العينة $x_1, x_2, x_3,.....,x_n$ فإنه يقال أن التقدير $\hat{\theta}$ تقدير كافي للمعلمة θ إذا كانت دالة الاحتمال الشرطية $P_r(x_1, x_2, x_3,.....,x_n \mid \hat{\theta})$ لمفردات العينة لا تعتمد على المعلمة θ.

ومما سبق يتضح أنه إذا توافرت الخصائص الأربعة السابقة (التحيز – الاتساق – الكفاءة – الكفاية) في التقدير المحسوب من بيانات العينة فإنه يقال أن التقدير تقدير جيد.

وفي الفصول التالية سوف نقدم بعض الطرق التي يمكن باستخدامها إيجاد تقديرات لبعض معلمات المجتمع.

وتنقسم طرق التقدير **Methods of Estimation** لمعلمة أو أكثر من معلمات المجتمع إلى قسمين هما:

١- التقدير بنقطة **Point Estimation**.

٢- التقدير بفترة **Interval Estimation**.

أولاً: التقدير بنقطة

هو أن نقوم بحساب قيمة واحدة من بيانات العينة لتكون تقدير لمعلمة المجتمع المجهولة المناظرة لها، بحيث إذا توافرت في هذه القيمة المحسوبة من بيانات العينة خصائص التقدير الجيد فإن هذه القيمة تعتبر تقدير جيد لمعلمة المجتمع المجهولة.

ولكن من عيوب هذه الطريقة أنه في أكثر الحالات تكون قيمة التقدير غير مساوية لقيمة المعلمة في المجتمع. وبالتالي لا نستطيع تحديد مدي بعدها من القيمة الحقيقية للمعلمة المجهولة أو تحديد مدي دقة هذا التقدير.

ثانياً: التقدير بفترة

وللتغلب على عيوب طرق التقدير بنقطة تستخدم طرق التقدير بفترة، وطرق التقدير بفترة تعتمد كل منها على تحديد الفترة التي تقع فيها المعلمة وتسمي هذه الفترة بفترة الثقة Confidence Interval. فإذا كان a , b هما الحد الأدنى Lower Limit والحد الأعلى Upper Limit على الترتيب لفترة الثقة، فإن a , b يسميان بحدي الثقة Confidence Limits.

ويسمي احتمال وقوع المعلمة المطلوب تقديرها داخل هذه الفترة بدرجة الثقة Degree of Confidence. فإذا كانت المعلمة θ وكل من a , b هما حدا الثقة فإن:

$$P_r(a \leq \theta \leq b) = 1 - \alpha \qquad \textbf{(4.12)}$$

حيث $1 - \alpha$ تشير إلى درجة الثقة.

وبالتالي فإن احتمال عدم وقوع المعلمة θ داخل فترة الثقة يساوي α حيث:

$$P_r\left[(\theta < a) \cup (\theta > b)\right] = \alpha \qquad \textbf{(4.13)}$$

وتسمي α بمستوي المعنوية Significance Level.

مثال (٤-٣)

إذا كان متوسط الإنفاق الشهري للفرد الذي تم حسابه من بيانات عينة عشوائية بسيطة يساوي $\overline{X} = 150$ جنية. فإن ($\overline{X}$) ممكن أن تستخدم كتقدير لمتوسط

الإنفاق الشهري للفرد في المجتمع المسحوب منه العينة. ويعتبر ($\overline{X}$) في هذه الحالة تقدير لنقطة.

اما إذا استخدمت الفترة [120-170] كفترة ثقة بدرجة ثقة 95% أي ($1-\alpha = 0.95$)، فإنه يمكن القول أن متوسط الإنفاق الشهري للفرد في المجتمع يقع داخل الفترة [120-170] باحتمال 0.95 أي أن:

$$P_r[120 \le \mu \le 170] = 0.95$$

حيث μ هو متوسط الإنفاق الشهري للفرد بالجنية في المجتمع محل الدراسة المحسوب منه العينة.

(٤-٣) تقدير توقع المجتمع

The Estimation of Population′s Mean

إذا كان الوسط الحسابي (المتوسط) للمتغير محل الدراسة في العينة المسحوبة يساوي $\overline{X}$ من مجتمع معتاد بمتوسط μ وتباينه σ^2، فإن $\overline{X}$ تستخدم كتقدير غير متحيز لمتوسط (الوسط الحسابي) المجتمع μ ويعتبر $\overline{X}$ تقدير غير متحيز ومتسق وكفء للمعلمة μ.

لكن إذا كان المطلوب تقدير الفترة التي يقع داخلها المتوسط في المجتمع μ (أي تحديد الحد الأدنى والحد الأعلى b , a) أي التقدير بفترة بدرجة ثقة ($1-\alpha$) فإننا يمكن تحديد فترة الثقة للمعلمة μ على النحو التالي:

أولاً: إذا كان تباين المجتمع σ^2 معلوم

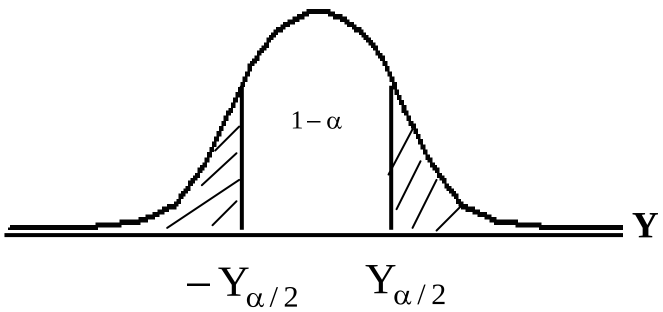
إذا كان $\overline{X}$ هو الوسط الحسابي (المتوسط) للمتغير محل الدراسة في عينة عشوائية حجمها n مسحوبة من مجتمع يتبع فيه المتغير محل الدراسة التوزيع المعتاد بمتوسط μ وتباين σ^2. حيث σ^2 قيمة معلومة فإن المتغير Y حيث:

$$Y = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \textbf{(4.14)}$$

فإن المتغير Y يتبع التوزيع المعتاد القياسي (أنظر نظرية (٣-٤)).
فإذا كانت درجة الثقة $(1-\alpha)$ فإن:

$$P_r = \left(-Y_{\alpha/2} < Y < Y_{\alpha/2}\right) = 1 - \alpha \qquad \textbf{(4.15)}$$

كما هو موضح بالشكل التالي:



شكل (٤-٢)

حيث $\pm Y_{\alpha/2}$ هما حدا الثقة ويتم حساب $Y_{\alpha/2}$ من جدول التوزيع المعتاد القياسي (أنظر ملحق رقم ٤).

ومن المعادلتين (4.14) ، (4.15) نجد أنه عند درجة ثقة $1-\alpha$ يوجد حدا الثقة $+Y_{\alpha/2}$ ، $-Y_{\alpha/2}$ وبالتالي فإن:

$$-Y_{\alpha/2} \le \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \le Y_{\alpha/2}$$

$$\overline{X}-Y_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X}+Y_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \textbf{(4.16)}$$

ومن العلاقة (4.16) نجد أن الحد الأدنى لفترة الثقة يساوي $\overline{X}-Y_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

والحد الأعلى لفترة الثقة يساوي $\overline{X}+Y_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

مثال (٤-٤)

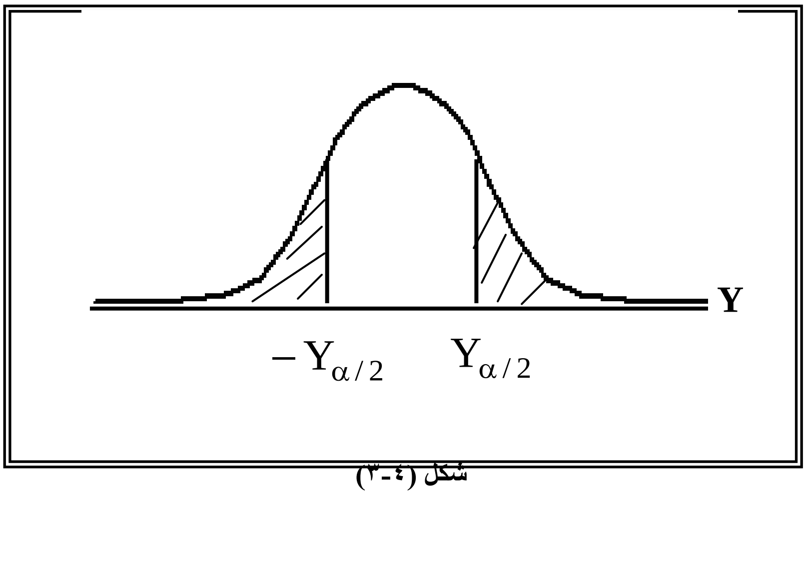
إذا كانت القيمة المتوسطة للمتغير محل الدراسة المحسوبة من عينة عشوائية بسيطة حجمها n بحيث أن $n=100$ **مفردة تساوي** $\overline{X}=25$**. فإذا كانت العينة مسحوبة من مجتمع معتاد تباينه** $\sigma^2=25$**.**

قدر فترة الثقة لمتوسط المجتمع μ **المسحوب منه العينة عند درجة ثقة 95%.**

الحل:

بما أن درجة الثقة $(1-\alpha)=0.95$**.**

إذن $Y_{\alpha/2}=1.96$ **من جدول التوزيع القياسي (ملحق رقم ٤) كما هو موضح بالشكل التالي:**



شكل (٤-٣)

حيث نجد أن الحد الأدنى =

$$\overline{X} - Y_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 25 - (1.96)\frac{5}{\sqrt{100}} = 24.02$$

والحد الأعلى =

$$\overline{X} + Y_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 25 + (1.96)\frac{5}{\sqrt{100}} = 25.98$$

وبالتالي فإن:

$$24.02 \le \mu \le 25.98$$

وذلك باحتمال 0.95 (أو بدرجة ثقة 95%).

ثانياً: إذا كان تباين المجتمع σ^2 غير معلوم وحجم العينة صغير

أما إذا كان تباين المتغير محل الدراسة في المجتمع المسحوب منه العينة غير معلوم (وهي الحالة الأكثر استخداماً حيث في معظم الحالات يكون التباين في المجتمع غير معلوم) فإن تباين المتغير محل الدراسة المسحوب من بيانات العينة S^2 يستخدم كتقدير لتباين المجتمع σ^2.

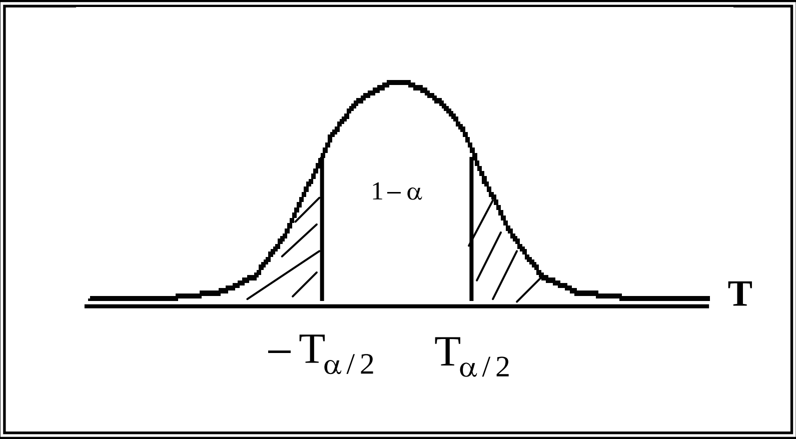
فإذا كان $\overline{X}$ هو القيمة المتوسطة للمتغير محل الدراسة من بيانات عينة حجمها **n** فإن المتغير **T** حيث:

$$T = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \qquad \textbf{(4.17)}$$

يتبع توزيع ستيودنت بدرجات حرية **(n-1)** (أنظر ملحق ٢) وبالتالي إذا كانت درجة الثقة $(1-\alpha)$ فإن:

$$P_r(-T_{\alpha/2} \le \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} \le T_{\alpha/2}) = 1-\alpha \qquad \textbf{(4.18)}$$

حيث تحسب القيمة $T_{\alpha/2}$ من جداول توزيع ستيودنت (ملحق رقم ٦). والشكل التـالي يوضح ذلك:



شكل (٤-٤)

وبالتالي فعند درجة الثقة $(1-\alpha)$ نجد أن:

$$\overline{X} - T_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + T_{\alpha/2}\frac{S}{\sqrt{n}} \qquad \textbf{(4.19)}$$

حيـث نجـد أن الحـد الأدنـى لفتـرة الثقـة $\overline{X} - T_{\alpha/2}\frac{S}{\sqrt{n}}$ والحـد الأعلـى يسـاوي $\overline{X} + T_{\alpha/2}\frac{S}{\sqrt{n}}$.

مثال (٤-٥)

إذا سحبت عينـة عشـوائية مكونـة من 4 مفـردات مـن مجتمـع معتـاد، وسـجلت قـيم المشاهدات فكانت على النحو التالي:

6 , 12 , 8 , 10

١- أوجد الوسط الحسابي والتباين للمتغير محل الدراسة المحسوبين من بيانـات العينـة.

٢- أوجد فترة الثقة للوسط الحسابي للمتغير محل الدراسة فـي المجتمـع المسـحوب منـه العينة وذلك عند درجة ثقة 95%.

الحل:

١- نفرض أن $\overline{X}, S^2$ **هما توقع وتباين المتغير المحسوبين من العينة فإن:**

$$\overline{X} = \frac{10+8+12+6}{4} = \frac{36}{4} = 9$$

$$S^2 = \frac{\sum (x-\overline{X})^2}{n-1}$$

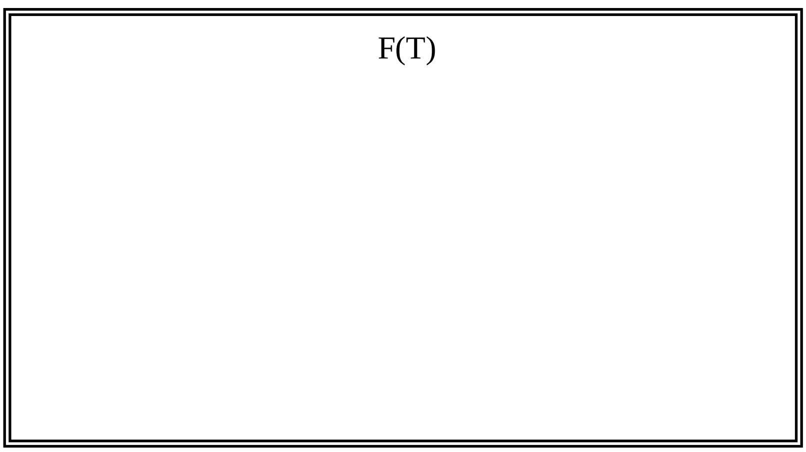
$$= \frac{(10-9)^2 + (8-9)^2 + (12-9)^2 + (6-9)^2}{4-1}$$

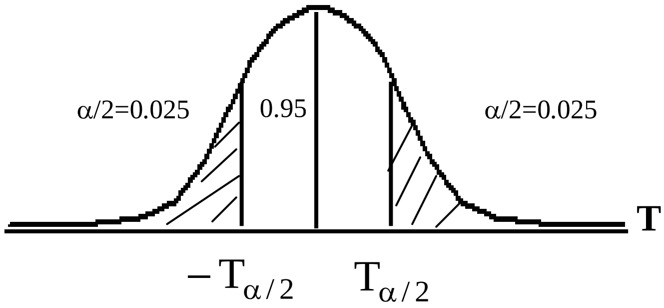
$$= \frac{1+1+9+9}{3} = \frac{20}{3} = 6.67$$

$$\therefore S = \sqrt{6.67} = 2.58$$

٢- بما أن درجة الثقة $(1-\alpha) = 0.95$

إذن من جدول توزيـع اسـتيودنت بملحق (٦) نجد ان $T_{\alpha/2} = 3.18$ **كمـا هـو موضح في الشكل التالي:**





شكل (٤-٥)

وبالتالي فإن الحد الأدني لفترة الثقة يساوي

$$\overline{X}-T_{\alpha/2}\frac{S}{\sqrt{n}}=9-(3.18)\frac{2.58}{\sqrt{4}}=4.9$$

والحد الأعلى يساوي

$$\overline{X}+T_{\alpha/2}\frac{S}{\sqrt{n}}=9+(3.18)\frac{2.58}{\sqrt{4}}=13.1$$

إذن:

$$4.9\le\mu\le13.1$$

وذلك بدرجة ثقة 95% (أي مستوي معنوية 5%)

(٤-٤) تقدير تباين المجتمع

The Estimation of Population′s Variance

إذا كـان S^2 هـو تبـاين المتغيـر محـل الدراسـة المسـحوب مـن بيانـات عينـة عشـوائية حجمهـا n مـن مجتمـع معتـاد توقعـه μ وتباينـه σ^2. فـإن تبـاين العينـة S^2 يسـتخدم كتقدير غير متحيز (أنظر الفصل ٤-٢) لتباين المتغير محل الدراسة في المجتمع (σ^2) حيث أن:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})}{n-1} \qquad \textbf{(4.20)}$$

ويمكننا تحديد فترة الثقة لتباين المجتمع σ^2 على النحو التالي:

بما أن المتغير Y حيث:

$$Y = \frac{\sum\left(x_i - \bar{X}\right)^2}{\sigma^2} = \frac{S^2(m-1)}{\sigma^2} \qquad \textbf{(4.21)}$$

يتبع توزيع χ^2_{m-1} (أنظر الفصل ٣-٣) وبالتالي فإذا كانت درجة الثقة $(1-\alpha)$ فإن:

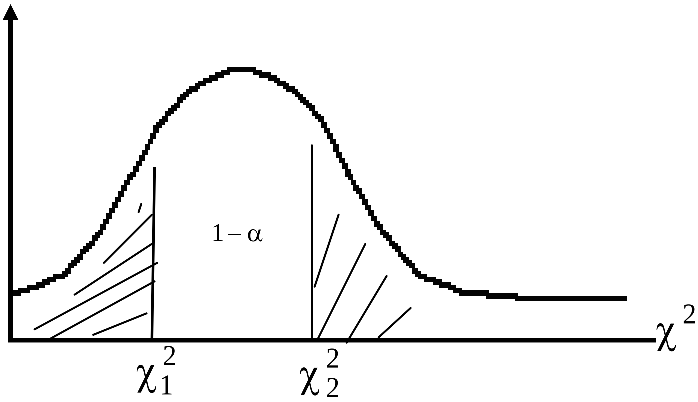
$$P_r(\chi_1^2 \le \frac{S^2(m-1)}{\sigma^2} \le \chi_2^2) = 1-\alpha \qquad \textbf{(4.22)}$$

وبالتالي فإن عند درجة الثقة $(1-\alpha)$ نجد أن:

$$\frac{S^2(n-1)}{\chi_2^2} \le \sigma^2 \le \frac{S^2(n-1)}{\chi_1^2} \qquad \textbf{(4.23)}$$

حيث يتم حساب القيمتين χ_1^2, χ_2^2 من جدول توزيع χ^2 عند درجات حرية (n-1) بملحق رقم (٥). والشكل التالي يوضح قيم χ_1^2, χ_2^2 عند درجة الثقة $(1-\alpha)$.

$f(\chi^2)$



شكل (٤-١٠)

وبالتالي فإن الحد الأدنى لفترة الثقة لتباين المجتمع σ^2 يساوي $\frac{S^2(n-1)}{\chi_2^2}$ والحد الأعلى لفترة الثقة يساوي $\frac{S^2(n-1)}{\chi_1^2}$.

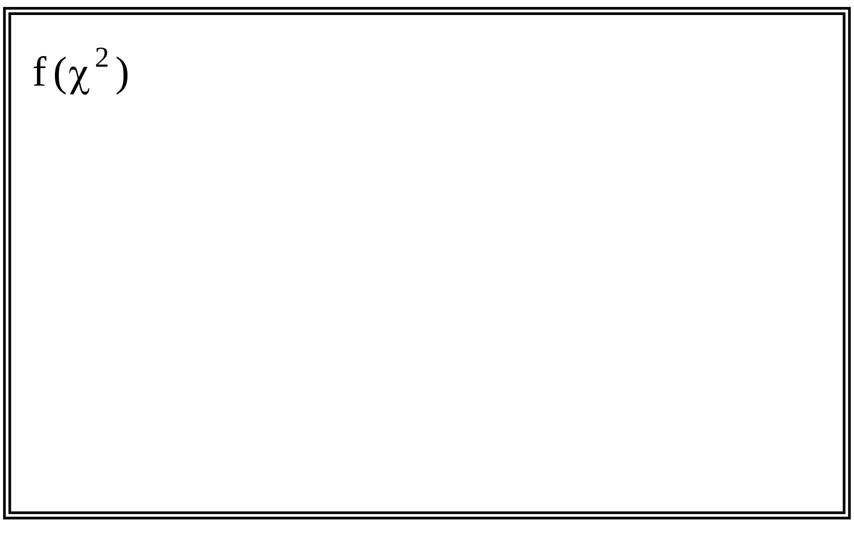
مثال (٤-٦)

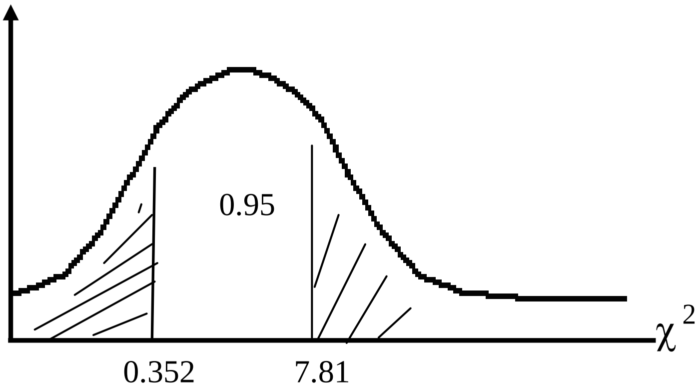
في المثال السابق (٤-٥) أوجد فترة الثقة لتباين المجتمع σ^2 بدرجة ثقة 90%.

الحل:

بما أن درجة الثقة تساوي 90% ، n-1=3.

مـــن جـــدول توزيـــع χ^2 بملحـــق (٥) وعنـــد درجـــة الثقـــة 90% نجـــد أن $\chi_1^2 = 0.352$ ، $\chi_2^2 = 7.81$ كما هو موضح بالشكل التالي:





شـ ( )

ومن العلاقة (4.23) نجد أن الحد الأدنى يساوي

$$\frac{S^2(n-1)}{\chi_2^2} = \frac{6.67(4-1)}{7.81} = 2.56$$

والحد الأعلى لفترة الثقة يساوي

$$\frac{S^2(n-1)}{\chi_1^2} = \frac{6.67(4-1)}{0.352} = 56.82$$

وبالتالي فإن:

$$2.56 \le \sigma^2 \le 56.82$$

$$1.6 \le \sigma^2 \le 7.54$$

(٤-٥) تقدير النسبة في المجتمع

The Estimation of Population Proportion

إذا كانت θ هي النسبة في المجتمع (نسبة المفردات في المجتمع التي تملك خاصية ما)، $\hat{\theta}$ هي النسبة في العينة (نسبة المفردات في العينة التي تملك هذه الخاصية) المسحوبة من المجتمع. فإن $\hat{\theta}$ تستخدم كتقدير نقطة للنسبة θ.

ولإيجاد الفترة التي تقع فيها المعلمة θ (أي تقدير فترة) فإنه يتضح من الفصل (٤-٣) أن $\hat{\theta}$ متغير عشوائي يقترب من التوزيع المعتاد بتوقع θ وتباين $\frac{\theta(1-\theta)}{n}$

حيث:

$$\mu(\hat{\theta}) = \theta \quad , \quad \sigma^2(\hat{\theta}) = \frac{\theta(1-\theta)}{n}$$

وبالتالي فإن المتغير

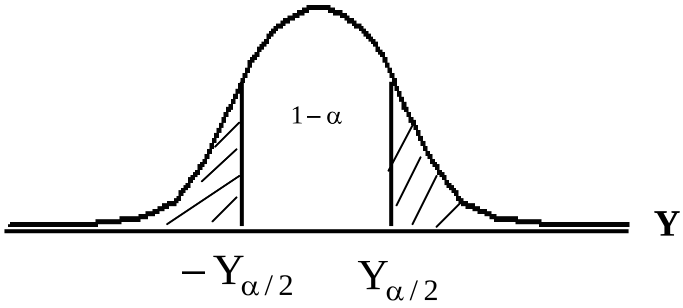
$$Y = \frac{\hat{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \qquad \textbf{(4.24)}$$

متغير يتبع التوزيع المعتاد القياسي (أنظر ملحق ٢)

وبالتالي عند درجة الثقة ($1-\alpha$) نجد أن:

$$P_r(-Y_{\alpha/2} \le \frac{\hat{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \le Y_{\alpha/2}) = 1-\alpha \qquad \textbf{(4.25)}$$

حيث يتم حساب $\pm Y_{\alpha/2}$ من جدول التوزيع المعتاد القياسي بملحق رقم (٤). كما هو موضح بالشكل التالي:



شكل (٤-٨)

وبما أن θ معلمة مجهولة فإن $\hat{\theta}$ تستخدم كتقدير لها في حساب الانحراف المعياري للمتغير $\hat{\theta}$ أي يستخدم $\sqrt{\dfrac{\hat{\theta}(1-\hat{\theta})}{n}}$ كتقدير للانحراف المعياري $\sqrt{\dfrac{\theta(1-\theta)}{n}}$ وبالتالي عند درجة الثقة $(1-\alpha)$ نجد أن:

$$-Y_{\alpha/2} \le \frac{\hat{\theta}-\theta}{\sqrt{\dfrac{\hat{\theta}(1-\hat{\theta})}{n}}} \le Y_{\alpha/2}$$

$$\hat{\theta}-Y_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \le \theta \le \hat{\theta}+Y_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \qquad \textbf{(4.26)}$$

ومن العلاقة (4.26) نجد أن الحد الأدنى لفترة الثقة للنسبة في المجتمع يساوي $\hat{\theta}-Y_{\alpha/2}\sqrt{\dfrac{\hat{\theta}(1-\hat{\theta})}{n}}$ والحد الأعلى يساوي $\hat{\theta}+Y_{\alpha/2}\sqrt{\dfrac{\hat{\theta}(1-\hat{\theta})}{n}}$

مثال (٤-٧)

إذا أخذت عينة عشوائية مكونة من 200 طالب وطالبة فوجد ان 144 طالب وطالبة منهم اجتازوا امتحان الفصل الدراسي الأول. المطلوب:

١- أوجد نسبة الناجحين في العينة.

٢- أوجد فترة الثقة لنسبة الناجحين في المجتمع المسحوب منه العينة بدرجة ثقة 95%.

الحل:

إذا كانت θ ، $\hat{\theta}$ همل نسبة النجاحين في المجتمع والعينة على الترتيب حيث:

$$1-\hat{\theta}=\frac{144}{200}=0.72$$

٢- بما أن درجة الثقة $(1-\alpha)=0.95$ فإن:

$Y_{\alpha/2}=1.96$ **(من جدول التوزيع المعتاد القياسي)**

إذن الحد الأدنى لفترة الثقة يساوي

$$\hat{\theta}-Y_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}=0.72-1.96\sqrt{\frac{0.72(0.28)}{200}}=0.688=68.8\%$$

والحد العلى لفترة الثقة يساوي

$$\hat{\theta}+Y_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}=0.72+1.96\sqrt{\frac{0.72(0.28)}{200}}=0.752=75.2\%$$

وبالتالي فإن فترة الثقة للنسبة في المجتمع θ تصبح على النحو التالي:

$$68.8\%\leq\theta\leq75.2\%$$

وذلك بدرجة ثقة 95%.

(٤-٦)خطأ المعاينة والدقة

Sampling Error and Precision

في الفصول السابقة تناولنا بالدراسة كيفية الحصول على تقديرات لبعض معالم المجتمع محل الدراسة التي يتم حسابها من بيانات العينة وعادة تكون قيمة التقدير تخلف عن قيمة المعلمة الفعلية. وفي هذا الفصل سوف نتناول دقة هذا التقدير المسحوب من بيانات العينة.

فإذا كانت $\hat{\theta}$ تمثل تقدير للمعلمة θ فإن الفرق بين المعلمة الفعلية θ والقيمة التقديرية لها $\hat{\theta}$ يمسي بخطأ المعاينة وعادة يرمز له بالرمز ε حيث:

$$\varepsilon = \theta - \hat{\theta} \qquad (4.27)$$

أو بعبارة أخري فإن القيمة الفعلية θ تساوي القيمة التقديرية $\hat{\theta}$ مضافاً إليها خطأ المعاينة، أي:

$$\theta = \hat{\theta} + \varepsilon \qquad (4.28)$$

ومن المعادلة (4.27) نجد أن خطأ المعاينة ممكن أن يساوي قيماً موجبة، وممكن أن يساوي قيماً سالبة، وممكن أن يساوي الصفر. وبالتالي فإن خطأ المعاينة ε يمثل متغير عشوائي حيث أن قيمته تختلف من عينة لأخرى (حيث أن قيمة ε تختلف من عينة لأخرى). ونظراً لأن قيمة المعلمة الفعلية θ غير معلومة في أغلب الأحيان فإن خطأ المعاينة ε يكون قيمة غير معلومة **Unknown Value** أيضاً.

تعريف (٤-١): إذا كان خطأ المعاينة ε بحيث:

$$\left|\theta - \hat{\theta}\right| = \varepsilon'$$

فإذا كان ε' هو أقصى قيمة للفرق المطلق بين المعلمة الفعلية θ والقيمة التقديرية لها $\hat{\theta}$ أي أن ε' هو أقصى قيمة مطلقة للمتغير ε أي:

$$\varepsilon' = \text{Max}\left|\theta - \hat{\theta}\right| \qquad \textbf{(4.29)}$$

فإن المقدار ε' يسمي بدقة التقدير Precision

مثال (٤-٨)

إذا سحبت عينة عشوائية حجمها 100 مفردة من مجتمع معتاد توقعه μ وتباينه $\sigma^2 = 225$ ، فإذا كان متوسط العينة $\overline{X} = 20$. والمطلوب:

أوجد بدرجة ثقة 95% دقة التقدير $\overline{X}$ لتوقع المجتمع μ.

الحل:

بما أن المعلمة μ غير معلومة، $\overline{X}$ هو تقدير لها. فمن الفصل (٤-٣) نجد ان:

$$\overline{X} - Y_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Y_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

وبما أن درجة الثقة 95% فإن $Y_{\alpha/2} = 1.96$ (من جدول التوزيع المعتاد القياسي). ومن العلاقة (4.16) وبطرح قيمة $\overline{X}$ من أطراف المتباينات نجد أن:

$$-1.96\left(\frac{15}{10}\right) \le (\mu - \overline{X}) \le 1.96\left(\frac{15}{10}\right)$$

$$\left|\mu - \overline{X}\right| \le 1.96\left(\frac{15}{10}\right) = 2.94$$

$$\therefore \varepsilon' = 2.94$$

وهذا يعني أن أقصى قيمة مطلقة للفرق بين μ ، $\overline{X}$ لا يتجاوز 2.94.

(٤-٧) حجم العينة الأمثل لتقدير متوسط المجتمع

Optimum Sample Size for Estimating Population's Mean

يعتبر تحديد حجم العينة لإجراء أي دراسة إحصائية من اهم العوامل التي تؤخذ في الاعتبار عند إجراء التحليل الإحصائي، نظراً لأنه يترتب على هذا الحجم المؤشرات

التي يتم حسابها من العينة والتي تستخدم كتقديرات لمعلمات المجتمع محل الدراسة، وبالتالي يتوقف على هذه التقديرات القرارات المبنية على النتائج المستخلصة من الدراسة.

وفيما يلي سوف نقدم كيفية تقدير حجم العينة العشوائية لتقدير متوسط المجتمع.

إذا كان $\overline{X}$ هو توقع المتغير المسحوب من عينة حجمها **n** مسحوبة من مجتمع معتاد توقعه μ وتباينه σ^2، فمن الفصل (٤-٣) نجد ان:

$$\overline{X} - Y_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Y_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

أو بعبارة أخرى:

$$P_r\left[\overline{X} - Y_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Y_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha \quad \textbf{(4.30)}$$

حيث $(1-\alpha)$ هي درجة الثقة، و $Y_{\alpha/2}$ يتم حسابها من جدول التوزيع المعتاد القياسي بملحق رقم (٤).

ومن المعادلة (4.30) نجد أن:

$$P_r\left[\left|\mu - \overline{X}\right| \le Y_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha \qquad \textbf{(4.31)}$$

وبما أن $\left|\mu - \overline{X}\right|$ يمثل خطأ المعاينة، فمن المعادلة (4.31) يتضح أن الحد الأعلى لخطأ المعاينة (أو الدقة ε') في هذه الحالة يساوي $Y_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ حيث:

$$\varepsilon \le Y_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \rightarrow \quad \varepsilon' = Y_{\alpha/2}\frac{\sigma}{\sqrt{n}} \qquad \textbf{(4.32)}$$

أو بعبارة أخرى خطأ المعاينة ε لا يزيد عن $Y_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ بدرج ثقة $(1-\alpha)$.

من العلاقة (4.32) نجد أن:

$$n = \left[\frac{Y_{\alpha/2}\sigma}{\varepsilon'}\right]^2 \qquad (4.33)$$

ومن العادلة (4.33) يتضح أن تحديد حجم العينة العشوائية في هذه الحالة يتطلب:

١- افتراض درجة ثقة معينة ولتكن ($1-\alpha$)، ووفقاً لهذه القيمة ($1-\alpha$) يتم تحديد قيمة $Y_{\alpha/2}$ من جدول التوزيع المعتاد القياسي.

٢- معرفة الانحراف المعياري للمجتمع (σ).

٣- افتراض الدقة (ε') أي الحد الأعلى لخطأ المعاينة المسموح به.

ملحوظة: في حالة إذا كان الانحراف المعياري للمجتمع σ غير معروف ففي هذه الحالة يستخدم الانحراف المعياري في العينة S كتقدير للانحراف المعياري في المجتمع σ إذا كان حجم العينة كبير (أكثر من 30 مفردة) حيث يؤول توزيع استيودنت إلى التوزيع المعتاد في هذه الحالة انظر ملحق (٦)

مثال (٤-٩)

قدر حجم عينة عشوائية مسحوبة من مجتمع معتاد تباينه 81 لتقدير متوسط المجتمع μ بدرجة ثقة 95% وخطأ معاينة لا يزيد عن 1.5 (أو دقة $\varepsilon' = 1.5$)

الحل:

بما ان:

$$1-\alpha = 0.95 \quad \rightarrow \quad Y_{\alpha/2} = 1.96$$

$$\sigma = \sqrt{81} = 9 \quad \rightarrow \quad \varepsilon' = 1.5$$

فإذا فرضنا أن حجم العينة المقدر يساوي n فإن:

$$n = \left[\frac{Y_{\alpha/2}\sigma}{\varepsilon'}\right]^2 = \left[\frac{1.96(9)}{1.5}\right]^2 = 138.3 \approx 139 \text{ مفردة}$$

(٤-٨) حجم العينة الأمثل لتقدير النسبة في المجتمع

Optimum Sample Size for Estimating A Population Proportion

إذا فرضنا أن θ ، $\hat{\theta}$ هما النسبة في المجتمع والنسبة في العينة على الترتيب. فمن الفصل (٤-٥) نجد انه عند درجة الثقة $(1-\alpha)$ وباستخدام جدول التوزيع المعتاد القياسي (لحساب $Y_{\alpha/2}$) أن:

$$\hat{\theta} - Y_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \le \theta \le \hat{\theta} + Y_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}} \qquad \textbf{(4.34)}$$

فمن العلاقة (4.34) نجد أن الحد الأعلى لخطأ المعاينة (الدقة)

$$\varepsilon' = Y_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$$

حيث أن:

$$P_r\left[\left|\theta - \hat{\theta}\right| \le Y_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}\right] = 1-\alpha \qquad \textbf{(4.35)}$$

وبالتالي فإن:

$$n = \hat{\theta}(1-\hat{\theta})\left[\frac{Y_{\alpha/2}}{\varepsilon'}\right]^2 \qquad \textbf{(4.36)}$$

مثال (٤-١٠)

قدر حجم العينـة العشـوائية التـي يمكن سـحبها لتقـدير نسـبة الطلاب الراسـبين فـي إحدى المـواد الدراسـية. وإذا كانـت نسـبة الراسـبين فـي العينـة 0.25 وذلـك بدرجـة ثقـة 95% إذا كان:

١- الحد الأعلى لخطأ المعاينة (الدقة) = 0.01.

٢- الحد الأعلى لخطأ المعاينة (الدقة) = 0.1.

الحل:

١- بما أن

$$1-\alpha = 0.95 \quad \rightarrow \quad Y_{\alpha/2} = 1.96$$

$$\varepsilon' = 0.01 \qquad \hat{\theta} = 0.25$$

فإن:

$$n = \hat{\theta}(1-\hat{\theta})\left[\frac{Y_{\alpha/2}}{\varepsilon'}\right]^2 = 0.25(0.75)\left[\frac{1.96}{0.01}\right]^2 = 7203$$

٢- بما أن

$$1-\alpha = 0.95 \quad \rightarrow \quad Y_{\alpha/2} = 1.96$$

$$\varepsilon' = 0.1 \qquad \hat{\theta} = 0.25$$

فإن:

$$n = \hat{\theta}(1-\hat{\theta})\left[\frac{Y_{\alpha/2}}{\varepsilon'}\right]^2 = 0.25(0.75)\left[\frac{1.96}{0.1}\right]^2 = 72.03 \approx 72$$

من المعادلتين (4.37) ، (4.38) نجد أنه كلما زادت القيمة للحد الأعلى لخطـأ المعاينـة (الدقة ε') أدى ذلك إلى نقص حجم العينـة. أو بعبارة أخـرى العلاقـة بـين حجـم العينـة n وقيمـة ε' علاقة عكسية.

(٤-٩) تطبيقات Applications

تطبيق (٤-١): في دراسة عن متوسط الدخل اليومي المتوقع للأسرة في إحدى محافظات الجمهورية أخذت عينة مكونه من 200 أسرة فوجد أن متوسط الدخل اليومي للأسرة في العينة يساوي 12.5 جنية بانحراف معياري 3 جنيهات.

أوجد فترة الثقة التي يقع فيها الدخل اليومي المتوقع للأسرة في هذه المحافظة وذلك بدرجة ثقة 95%.

الحل:-

بما أن:

$$1-\alpha = 0.95 \quad \rightarrow \quad Y_{\alpha/2} = 1.96$$

$$n = 2000 \quad , \quad \bar{X} = 12.5$$

$$S = 3$$

ومن العلاقة (4.19) بالفصل (٤-٣) نجد أن:

$$\bar{X} - Y_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + Y_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$12.5 - 1.96\frac{3}{\sqrt{200}} \le \mu \le 12.5 + 1.96\frac{3}{\sqrt{200}}$$

$$12.084 \le \mu \le 12.916$$

أي أن متوسط الدخل اليومي المتوقع للأسرة في هذه المحافظة يتراوح بين (12.084 ، 12.916) جنية وذلك بدرجة ثقة 95%.

تطبيق (٤-٢): **في أحدي المراكز الصحيحة أجريت دراسة على الكمية المتوقعة للأكسجين (باللتر في الدقيقة) الذي يستهلكه الفرد الذي عمره يتراوح بين [17-21] سنة. وكان المطلوب تحديد حجم العينة (عينة البحث) بافتراض أن تباين المجتمع المسحوب منه العينة 0.09 لتري في الدقيقة. ودرجة الثقة 95% بحيث لا يزيد خطأ المعاينة عن 0.10 لتري في الدقيقة.**

الحل:-

بما أن المطلوب تحديد حجم العينة n لدراسة توقع المجتمع μ حيث:

$$1-\alpha = 0.95 \quad \rightarrow \quad Y_{\alpha/2} = 1.96$$

$$\varepsilon' = 0.10$$

$$\sigma^2 = 0.09$$

$$n = \left[\frac{Y_{\alpha/2}\sigma}{\varepsilon'}\right]^2 = \left[\frac{1.96 \times 0.3}{0.10}\right]^2 = 34.6 \approx 35$$ **مفردة**

تطبيق (٤-٣): **يقوم أحد مراكز قياس السمع للأطفال بتحديد الزمن المتوقع لاستجابة الطفل للرد على سؤال معين. فإذا أخذت عينة مكونه من 25 طفل وبسؤال كل طفل وتسجيل الفترة التي استغرقها الطفل بين سماع السؤال والرد عليه وجد أن متوسط الاستجابة للطفل في العينة 160 ثانية بانحراف معياري 5 ثوان وبافتراض أن زمن الاستجابة للأطفال في المركز يتبع التوزيع المعتاد. والمطلوب:**

قدر فترة الثقة لمتوسط زمن الاستجابة للطفل في المركز بدرجة ثقة 99%.

الحل:-

بما ان:

$$n = 25 \quad , \quad S = 5 \quad , \quad \overline{X} = 160$$

$$(1-\alpha) = 0.95 \quad \rightarrow \quad \alpha/2 = 0.005$$

وبما أن العينة أقل من 30 مفردة و σ مجهولة

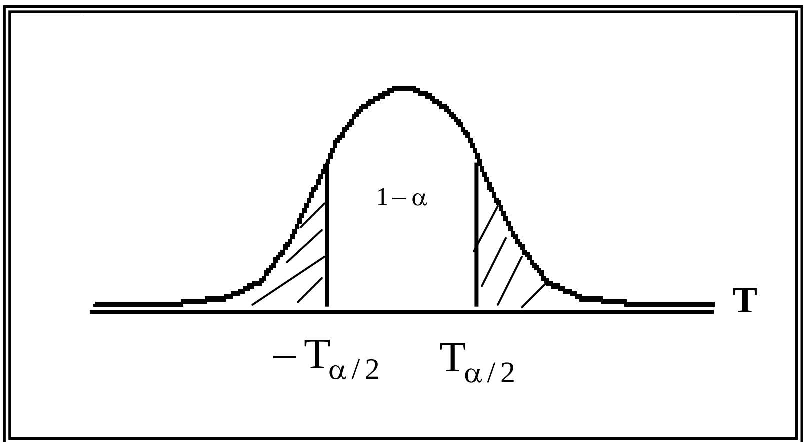
$\therefore T_{\alpha/2} = 2.7969$ **(من جدول توزيع ستيودنت بملحق رقم ٦)**

ومن العلاقة (4.19) بالفصل (٤-٣) نجد أن:

$$\overline{X} - T_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + T_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$160 - 2.797\frac{5}{\sqrt{25}} \le \mu \le 160 + 2.797\frac{5}{\sqrt{25}}$$

$$157.203 \le \mu \le 162.797$$



شكل (٤-٩)

تطبيق (٤-٤): في إحدى الدراسات عن متوسط درجة الطالب في مادة الإحصاء في كليات التجارة بجمهورية مصر العربية في إحدى السنوات الدراسية. فإذا وجد ان 20000 يدرسون مادة الإحصاء في السنة فإذا أخذت عينة عشوائية بسيطة حجمها 20 طالب فكان متوسط درجة الطالب في العينة 12.5 درجة بانحراف معياري 2 درجة.

أوجد فترة الثقة لمتوسط درجة الطالب في كليات التجارة في هذه السنة بدرجة ثقة .99%, 98%, 95%

الحل:-

بما أن:

$$n = 20 \quad , \quad S = 2 \quad , \quad \overline{X} = 12.5$$

١- عند درجة الثقة 95% أي $(1-\alpha) = 0.95$، وعند درجات حرية m=10-1=a نجد أن $T_{\alpha/2} = 2.09$.

$$\overline{X} - T_{\alpha/2}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \le \mu \le \overline{X} + T_{\alpha/2}\frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$12.5 - 2.09\frac{2}{\sqrt{20}}\sqrt{\frac{20000-20}{20000-1}} \le \mu \le 12.5 + 2.09\frac{2}{\sqrt{20}}\sqrt{\frac{20000-20}{20000-1}}$$

$$12.5 - 0.94 \le \mu \le 12.5 + 0.94$$

$$11.56 \le \mu \le 13.44 \qquad (1)$$

٢- بالمثل عندما تساوي درجة الحرية 98% فإن $T_{\alpha/2} = 2.54$

$$\overline{X} - T_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + T_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$12.5 - 2.54\frac{2}{\sqrt{20}} \le \mu \le 12.5 + 2.54\frac{2}{\sqrt{20}}$$

$$12.5 - 1.14 \le \mu \le 12.5 + 1.14$$

$$11.36 \le \mu \le 13.64 \qquad (2)$$

حيث أن المقدار $\sqrt{\frac{N-n}{N-1}}$ **يؤول إلى 1.**

٣- وعند درجة حرية 99% نجد ان $T_{\alpha/2} = 2.86$

$$\overline{X} - T_{\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + T_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$12.5 - 2.86\frac{2}{\sqrt{20}} \le \mu \le 12.5 + 2.86\frac{2}{\sqrt{20}}$$

$$12.5 - 1.28 \le \mu \le 12.5 + 1.28$$

$$11.22 \le \mu \le 13.78 \qquad (3)$$

ومن العلاقات الثلاثة (1)، (2)، (3) يتضح أنه كلما زادت درجة الثقة كلما قل الحد الأدنى وزاد الحد الأعلى لفترة الثقة للمعلمة μ (أي كلما اتسعت فترة الثقة).

تطبيق (٤-٥): في التطبيق السابق (٤-٤) قدر حجم العينة المطلوب سحبها لتقدير متوسط الدرجة في الكليات بحيث لا يزيد الحد الأعلى لخطأ المعاينة عن 1.5 درجة.

الحل:-

بما أن الحد الأعلى لخطأ المعاينة المسموح به يساوي ε' حيث $\varepsilon' = 1.5$ كذلك الانحراف المعياري في العينة S = 2. فإذا فرضنا أن حجم العينة n فإن:

١- عند درجة الثقة 95% نجد أن $T_{\alpha/2} = 2.09$.

$$n = \left[\frac{T_{\alpha/2}S}{\varepsilon'}\right]^2 = \left[\frac{2.09\times 2}{1.5}\right]^2 = 7.77 \approx 8 \text{ مفردات} \qquad (4)$$

٢- عند درجة الحرية 98% فإن $T_{\alpha/2} = 2.54$.

$$n = \left[\frac{T_{\alpha/2}S}{\varepsilon'}\right]^2 = \left[\frac{2.54\times 2}{1.5}\right]^2 = 11.5 \approx 12 \text{ مفردة} \qquad (5)$$

٣- عند درجة حرية 99% نجد ان $T_{\alpha/2} = 2.86$

$$n = \left[\frac{T_{\alpha/2}S}{\varepsilon'}\right]^2 = \left[\frac{2.86\times 2}{1.5}\right]^2 = 12.5 \approx 13 \text{ مفردة} \qquad (6)$$

من (4)، (5)، (6) يتضح أنه كلما زادت درجة الثقة كلما زاد حجم العينة.

تطبيق (٤-٦): تقوم شركة بإنتاج 4 أنواع من المنظفات الصناعية **A, B, C, D** فإذا أخذت عينة عشوائية من العملاء مستخدمي منتجات هذه الشركة حجمها 400 عميل من 1000 عميل توجه أن 310 منهم يفضلون النوع **B**. والمطلوب:

١- قدر بدرجة ثقة 95% نسبة العملاء الذين يفضلون النوع **B**.

٢- قدر بدرجة ثقة 99% نسبة العملاء الذين يفضلون الأنواع **A, C, D**.

٣- قدر عدد العملاء الذين يفضلون النوع **B** عند درجة ثقة 98%.

الحل:-

إذا فرضنا أن $\bar{\theta}, \theta$ هي نسبة العملاء الذين يفضلون استخدام المنتج **B** في المجتمع والعينة على الترتيب.

١- $$\bar{\theta} = \frac{310}{400} = 0.775$$

وعند درجة الثقة 95% نجد أن $Y_{\alpha/2} = 1.96$

$$\bar{\theta} - Y_{\alpha/2}\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}} \le \theta \le \bar{\theta} + Y_{\alpha/2}\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}$$

$$0.775 - 1.96\sqrt{\frac{0.775\times 0.225}{400}} \le \theta \le 0.775 + 1.96\sqrt{\frac{0.775\times 0.225}{400}}$$

$$0.775 - 0.041 \le \theta \le 0.775 + 0.041$$

$$\therefore 0.734 \le \theta \le 0.816$$

بدرجة ثقة 95%.

٢- بما ان نسبة الذين يفضلون النوع **B** في العينة يساوي **0.775** وبالتالي فإن نسبة الذين يفضلون الأنواع **A, C, D** يساوي P' حيث:

$$P_r(P') = 1 - 0.775 = 0.225$$

وعند درجة حرية **99%** نجد ان $Y_{\alpha/2} = 2.58$

فإذا كانت p نسبة الذين يفضلون الأنواع A, C, D في المجتمع (من العملاء) فإن:

$$0.225 - 2.58\sqrt{\frac{0.225 \times 0.775}{400}} \le P \le 0.225 + 2.58\sqrt{\frac{0.225 \times 0.775}{400}}$$

$$0.225 - 0.054 \le P \le 0.225 + 0.054$$

$$0.171 \le P \le 0.269$$

$$\therefore 17.1\% \le P \le 26.9\%$$

٣- عند درجة حرية 98% نجد ان $Y_{\alpha/2} = 2.34$

$$0.775 - 2.34\sqrt{\frac{0.775 \times 0.225}{400}} \le \theta \le 0.775 + 2.34\sqrt{\frac{0.775 \times 0.225}{400}}$$

$$0.726 \le \theta \le 0.824$$

$$\therefore 27.6\% \le \theta \le 82.4\%$$

تطبيق (٤-٧): في أحد الأعوام وجد أن عدد النزلاء في مؤسسات رعاية الأحداث بجمهورية مصر العربية يوجد 5000 نزيل، فإذا أخذت عينة عشوائية مكونه من 225 نزيل وأجريت دراسة عن تحديد سبب وطول المؤسسة فوجد أن 75% من العينة يرجع سبب دخول الحدث المؤسسة إلى فقدان الرعاية الأسرية. والمطلوب:

قدر نسبة الأحداث في المؤسسات الذين يرجع سبب دخولهم المؤسسات إلى عدم توافر الرعاية الأسرية بدرجة ثقة 95%.

الحل:-

بما أن

$$n = 225 \quad , \quad N = 5000 \quad , \quad \bar{\theta} = 0.75$$

$$(1-\alpha) = 0.95 \quad \rightarrow \quad Y_{\alpha/2} = 1.96$$

فإذا كانت θ هي نسبة الأحداث في المؤسسات الذين يرجع سبب دخولهم المؤسسات إلى عدم توافر الرعاية الأسرية

$$\bar{\theta} - Y_{\alpha/2}\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}} \le \theta \le \bar{\theta} + Y_{\alpha/2}\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}}$$

$$0.75 - 1.96\sqrt{\frac{0.75 \times 0.25}{225}} \le \theta \le 0.75 + 1.96\sqrt{\frac{0.75 \times 0.25}{225}}$$

$$0.693 \le \theta \le 0.807$$

$$\therefore 69.3\% \le \theta \le 80.7$$

بدرجة ثقة 95%

وبالتالي فإن عدد الأحداث في المؤسسات لفقدان الرعاية الأسرية يتراوح بين:

$$0.693 \times 5000 \le \theta \le 0.807 \times 5000$$

$$3465 \le \theta \le 4035$$

تطبيق (٤-٨): أخذت عينـة مكونـه مـن 20 طالب وسجلت درجـاتهم (X) فـي مـادة الإحصاء، فوجد أن متوسط الدرجة في العينة $\overline{X} = 65$ درجة. ووجد أن:

$$\sum(x-\overline{X})^2 = 2230$$

بـافتراض أن يتبـع الـدرجات المسـحوبة مـن العينـة التوزيـع المعتـاد أوجـد بدرجـة ثقـة 95% فترة لتباين درجات الطلاب في المجتمع المسحوب منه العينة.

الحل:-

إذا فرضنا أن S^2 هي تباين الدرجات في العينة فإن:

$$S^2 = \frac{\sum(x-\overline{X})^2}{n-1} = \frac{2230}{20-1} = \frac{2230}{19} = 117.37$$

$\Theta\ \chi^2$ متغير يتبع توزيع χ^2 بدرجات حرية n-1 حيث:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

وبالتالي نجد أن:

$$\frac{(n-1)S^2}{\chi_2^2} \le \sigma^2 \le \frac{(n-1)S^2}{\chi_1^2}$$

ومن جدول توزيع χ^2 عند درجات حرية 19 نجد أن:

$$\chi_1^2 = 8.91 \quad , \quad \chi_2^2 = 32.9$$

وبالتالي فإن:

$$\frac{19\times117.37}{32.9} \le \sigma^2 \le \frac{19\times117.37}{8.91}$$

$$67.78 \le \sigma^2 \le 250.28$$

$$\therefore 8.23 \le \sigma \le 15.82$$

(٤-١٠) تمرينـــــــــــــــــــــــــــــــــات Exercises

(٤-١): في إحدى الدراسات عن متوسط درجة الطالب في مـادة الإحصـاء فـي كليـات التجارة بجمهوريـة مصر العربيـة في إحدى السنوات الدراسية. فإذا وجد أن 20000 يدرسون مادة الإحصاء في هذه السنة. فإذا أخذت عينـة عشوائية بسيطة حجمهـا 20 طالــب فكــان متوسـط درجــة الطالــب فــي العينــة 12.5 بــانحراف معيــاري 2 درجــة. والمطلوب:

أوجد فترة الثقة لمتوسط درجة الطالب في كليات التجارة في هذه السنة بدرجة ثقة .99%, 98%, 95%

(٤-٢): في التمرين السابق قدر حجم العينة المطلوب سحبها لتقدير متوسط الدرجة في الكليات بحيث لا يزيد الحد الأعلى لخطأ المعاينة عن 1.5 درجة.

(٤-٣): فـي دراسـة عن سبب فشل بعض الطـلاب فـي اسـتكمال دراسـتهم بالكليـات العسكرية. أخذ عينـة مكونـه من 200 طالب فشـلوا فـي اجتيـاز مراحل الدراسـة بالكليـات العسكرية من 1500 طالب لم يجتازوا مراحل الدراسة. فوجد أن 140 طالـب فـي العينـة يرجع سبب عدم اجتيازهم مراحل الدراسة إلى أسباب صحية. والمطلوب:

قدر نسبة الطلاب في الكليات العسكرية الذين لم يجتازوا المراحل الدراسية بالكليات العسكرية لأسباب صحية وذلك بدرجة ثقة 95%.

(٤-٤): تقوم شركة بإنتاج 4 أنواع من المنظفات الصناعية A, B, C, D فإذا أخذت عينة عشوائية من العملاء مستخدمي منتجات هذه الشركة حجمها 400 عميل من 1000 عميل فوجد أن 310 منهم يفضلون النوع B. والمطلوب:

١- قدر بدرجة ثقة 95% نسبة العملاء الذين يفضلون النوع B.

٢- قدر بدرجة ثقة 99% نسبة العملاء الذين يفضلون النوع A, C, D.

٣- قدر عدد العملاء الذين يفضلون النوع B عند درجة ثقة 98%.

(٤-٥): إذا كان سعر الوحدة الواحدة من سلعة ما يتغير من محافظة إلى أخرى داخل الجمهورية (لاختلاف تكلفة النقل)، وبافتراض أن السعر متغير يتبع التوزيع المعتاد، فإذا أخذت عينة من الأسعار لهذه السلعة في 15 محافظة فكانت على النحو التالي:

53, 90, 53, 71, 78, 55, 66, 55, 94, 98, 74, 85, 74, 74, 58

قدر التباين والانحراف المعياري لسعر الوحدة من هذه السلعة داخل الجمهورية بدرجة ثقة 95%.

الباب الخامس
اختبارات الفروض المعلمية
Parametric Tests of Hypotheses

(٥-١) الفروض الإحصائية
Statistical Hypotheses
(٥-٢) اختبارات الفروض عن متوسط المجتمع
Tests of Hypotheses of A Population's Mean
(٥-٣) اختبارات الفروض عن تباين المجتمع
Tests of Hypotheses of A Population's Variance
(٥-٤) اختبارات الفروض عن النسبة في المجتمع
Tests of Hypotheses of A Population's Propration
(٥-٥) اختبارات الفروض عن الفرق بين متوسطي مجتمعين باستخدام عينتين مستقلتين
Tests of Hypotheses of The Difference Between Two Population Mean: Independent Samples
(٥-٦) اختبارات الفروض عن الفرق بين متوسطي مجتمعين باستخدام عينتين غير مستقلتين
Tests of Hypotheses of The Difference Between Two Population Mean: Dependent Samples
(٥-٧) تطبيقات
Applications
(٥-٨) تمرينات
Exercises

(٥-١) الفـــــروض الإحصــــــائية Statistical Hypotheses

في معظم المشاكل الفعلية **Real Problems** يتم صناعة القرارات بشأن المجتمع محل الدراسة بناء على معلومات ومعرفة مستمدة من عينة مسحوبة من هذا المجتمع . فـي هـذه الحالـة تسـمى هـذه القـرارات المبنيـة علـى المعلومـات المسـتمدة مـن العينـة بالقرارات الإحصائية.

فمثلاً إذا كنا نريد أن نقرر ما إذا حدثت زيادة فعلية في دخل الأسرة الشهري أم لم تحدث زيادة فعلية ، كذلك تقرير ما إذا أرتفع مستوى جودة المنتج المحلي من الأدويـة أم لا من خلال بيانات عينة عن الأدوية المنتجة محلياً خلال فترة معينة.

وللوصول إلى قرار بشأن المؤشرات (أو المعلمات **Parameters**) أو العلاقات **Relationships** في المجتمع محل الدراسة فإنه يتم وضع فروض (تخمينات) عن هذه المؤشرات أو العلاقات في المجتمع محل الدراسة وصياغتها في شكل رياضي وتسمى هذه الفروض بالفروض الإحصائية **Statistical Hypothesis** وقد تكون هذه الفروض صحيحة وقد تكون خاطئة.

وعملية اتخاذ القرار بقبول أو رفض فرض بناء على المعلومات المستخلصة من عينة (أو عينات) يسمى باختبارات الفروض الإحصائية **Test of Statistic Hypothesis** .

ونظراً لأن المعلمة أو العلاقات موضع الاختبار تكون غير معلومة **Unknown** فإن القرار الذي نتوصل إليه بناء على المعلومات من العينة (أو العينات) قد يكون صحيحاً وقد يكون خاطئاً.

وعند اتخاذ قرار بشأن فرض معين يكون لدينا الحالات الممكنة التالية:-

أولاً: أن يكون الفرض صحيح واتخاذ قرار بقبوله (يكون القرار في هذه الحالة قرار صحيح).

ثانياً: أن يكون الفرض صحيح واتخاذ قرار برفضه (يكون القرار في هذه الحالة غير صحيح "خاطئ").

ثالثاً: أن يكون الفرض خاطئ واتخاذ قرار برفضه (يكون القرار في هذه الحالة قرار صحيح).

رابعاً: أن يكون الفرض خاطئ واتخاذ قرار بقبوله (يكون القرار في هذه الحالة غير صحيح "خاطئ").

ومما سبق يتضح أنه توجد حالتين يكون فيها القرار غير صحيح (خاطئ)، وبالتالي يوجد لدينا نوعين من الأخطاء (أو نوعين من المخاطرة Risk) هما:-

- **النوع الأول من الخطأ**

وهو الحالة الثانية أي يكون الفرض صحيح واتخاذ قرار برفضه فيكون القرار خاطئ . فإذا كان احتمال أن يكون الفرض صحيح واتخاذ قرار بقبوله هو $(1-\alpha)$ حيث تسمى $(1-\alpha)$ بدرجة الثقة (الحالة الأولى) فإن احتمال أن يكون الفرض صحيح واتخاذ قرار برفضه يساوي α حيث تسمى α بمستوى المعنوية Significance's Level ويطلق على الخطأ في هذه الحالة بالخطأ من النوع الأول(أو المخاطرة من النوع الأول)٠ وهذا النوع من الخطأ قد لا يترتب عليه خسائر أو مشاكل٠

- **النوع الثانى من الخطأ**

وهو الحالة الرابعة أي يكون الفرض خاطئ واتخاذ قرار بقبوله فيكون القرار خاطئ . فإذا كان احتمال أن يكون الفرض خاطئ واتخاذ قرار برفضه هو $(1-\beta)$ (الحالة الثالثة) وبالتالي فإن احتمال أن يكون الفرض خاطئ واتخاذ قرار بقبوله يساوي β حيث تسمى β بمميز الفاعلية **Operating Characteristic** ويطلق على الخطأ في هذه الحالة أي قبول فرض خاطئ باحتمال β بالخطأ من النوع الثاني. وهذا النوع من الخطأ يمثل أهمية كبيرة نظراً لما يترتب عليه من خسائر.

والجدول التالي يوضح الحالات الممكنة عن اتخاذ القرار واحتمالات حدوث كل منها:-

جدول (٥-١)

الفرض / القرار	صحيح	خاطئ
قبول	قرار سليم باحتمال $(1-\alpha)$	قرار غير سليم باحتمال (β)
رفض	قرار غير سليم باحتمال (α)	قرار سليم باحتمال $(1-\beta)$

ووقوع متخذ القرار في النوع الأول أو الثاني من الأخطاء يمثل مخاطرة **Risk**. ومن هنا يتضح أهمية الأساليب الإحصائية في تحديد وقياس المخاطرة المترتبة على القرار الذي يجعل المخاطرة أقل ما يمكن **Minimum Risk** وهي ما تسمى بأساليب البرمجة الاحتمالية* **Probabilistic Programming Techniques**.

يوجد نوعين من الاختبارات الإحصائية:-

- **النوع الأول**

* Vojda (1972) : "Probabilistic Programming" ,Academic Press, New York

وهو ما يسمى بالاختبارات المعلمية ويتطلب إجراء هذا النوع من الاختبارات عن بعض المؤشرات للمتغير في المجتمع محل الدراسة وجود معرفة رقمية **Numerical Information** عن التوزيع الاحتمالي للمتغير محل الدراسة في المجتمع . وسوف نتناول بعض الاختبارات من هذا النوع في هذا الباب.

- **النوع الثاني**

وهو ما يسمى بالاختبارات اللامعلمية ولا يتطلب إجراء هذا النوع من الاختبارات عن بعض المؤشرات للمتغير في المجتمع وجود معرفة رقمية عن التوزيع الاحتمالي للمتغير محل الدراسة في المجتمع وسوف نتناول بعض الاختبارات من هذا النوع في الباب التالي.

وفي هذا الباب سوف تقتصر دراستنا على بعض اختبارات الفروض الإحصائية التي تجعل احتمال الوقوع في النوع الأول من الخطأ مساوياً بالقيمة المعلومة α (أو بعبارة أخرى عند درجة ثقة $(1-\alpha)$).

وعند إجراء الاختبار الإحصائي يكون لدينا فرضين أولهما الفرض محل الاختبار وهو ما يسمى بالفرض العدمي **Null Hypothesis** ويرمز له بالرمز H_0 والفرض الأخر يسمى بالفرض البديل **Alternative Hypothesis** ويرمز له بالرمز H_1.

ورفض الفرض العدمي يؤدي إلى قبول الفرض البديل ، كذلك قبول الفرض العدمي يؤدي إلى رفض الفرض البديل.

ويلاحظ أن رفض الفرض العدمي **Rejection of Null Hypothesis** يعني أن ذلك الفرض خاطئ في حين أن قبول الفرض العدمي يعني أنه ليس لدينا دليل كافي لرفضه لذلك غالباً ما يوضع الفرض العدمي في صيغة ينتظر رفضها وسوف نوضح ذلك من خلال الأمثلة في الفصول التالية.

وفيما يلي سوف نلخص عملية الاختبار الإحصائي واتخاذ القرار بناء على الاختبار على النحو التالي:-

١- صياغة الفرض العدمي والفرض البديل بحيث يؤدي رفض الفرض العدمي إلى قبول الفرض البديل أو العكس.
٢- تحديد المقياس الذي يتم حسابه من العينة لاستخدامه في الاختبار ومعرفة الخصائص الإحصائية لهذا المقياس.
٣- تحديد درجة الثقة $(1-\alpha)$.
٤- تحديد منطقة القبول ومنطقة الرفض للفرض العدمي.
٥- اتخاذ القرار بالقبول أو الرفض وفقاً للخطوات السابقة.

(٥-٢) اختبارات الفروض عن قيمة متوسط المتغير في المجتمع

Test of Hypotheses of a Population's Mean

في هذا الفصل سوف نتناول اختبارات الفروض بالنسبة لمتوسط (الوسط الحسابي) للمتغير محل الدراسة في المجتمع μ عندما يكون تباين المتغير في المجتمع σ^2 معلوم وغير معلوم.

أولاً: إذا كان تباين المجتمع σ^2 معلوم والمراد اتخاذ قرار بشأن افتراض قيمة معينة لمتوسط المجتمع ولتكن μ ولإجراء هذا الاختبار نتبع الخطوات المذكورة أعلاه على النحو التالي:-

١- الفرض العدمي

$$H_0 : \mu = \mu_0 \qquad (5.1)$$

ويمكن أن يأخذ الفرض البديل إحدى الصور التالية:-

$$H_1 : \mu > \mu_0 \qquad (5.2)$$

$$H_1 : \mu < \mu_0 \qquad (5.3)$$

$$H_1 : \mu \neq \mu_0 \qquad (5.4)$$

فإذا أخذ الفرض البديل الصورة (5.2) أو (5.3) فإن الاختبار يكون في هاتان الحالتين يسمى باختبار الذيل الواحد (أو الطرف الواحد) One Tailed Test أما إذا أخذ الفرض البديل الصورة (5.4) فإن الاختبار يسمى باختبار الذيلين (أو الطرفين) Two Tailed Test.

٢- حساب المقياس Y حيث

$$Y = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad (5.5)$$

حيث:

هو الوسط الحسابي (المتوسط) في العينة : $\overline{X}$

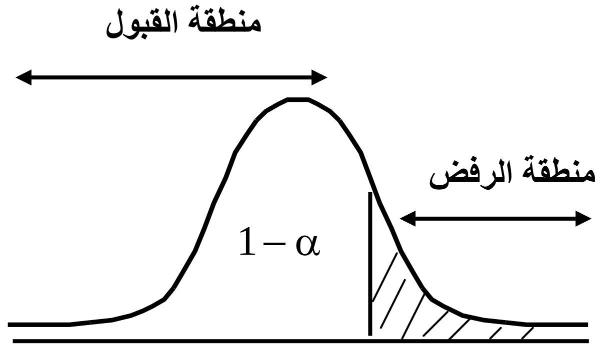
حجم العينة : n

فنجد أن المتغير Y يتبع التوزيع المعتاد القياسي (أنظر التوزيع الاحتمالي لمتوسط العينة بالفصل (٢-٢)).

٣- إذا فرضنا أن درجة الثقة $(1-\alpha)$

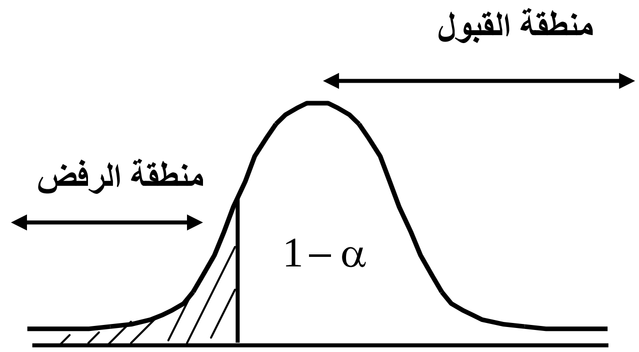
٤- عند درجة الثقة $(1-\alpha)$ يوضح كل شكل من الأشكال (٥-١) – (٥-٣) منطقة القبول ومنطقة (أو مناطق) الرفض عند الفروض البديلة المختلفة

$$H_1 : \mu > \mu_0$$



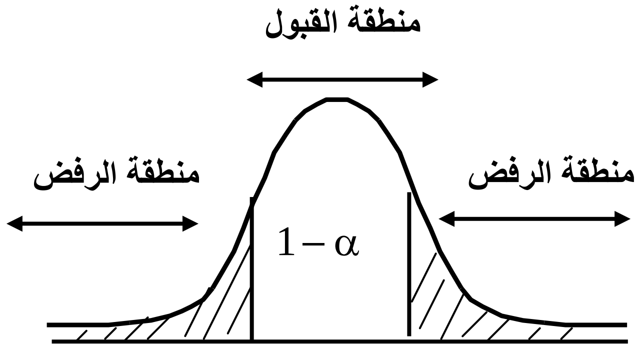
شكل (٥-١)

$$H_1 : \mu < \mu_0$$



شكل (٥-٢)

$H_1 : \mu \neq \mu_0$



شكل (٥-٣)

٥- إذا كانت قيمة **Y** المحسوبة في الخطوة (٢) تقع في منطقة القبول فإننا نقبل الفرض العدمي ونرفض الفرض البديل بدرجة ثقة $1-\alpha$ أما إذا كانت قيمة **Y** تقع في منطقة الرفض فإننا نرفض الفرض العدمي ونقبل الفرض البديل بدرجة ثقة $1-\alpha$.

مثال(٥-١)

إذا سحبت عينة مكونة من 100 عامل في إحدى الشركات وحسب متوسط الإنتاجية اليومية للعامل في العينة فوجد $\overline{X} = 28$ وحدة منتجة .فإذا كان التباين لإنتاجية العامل في هذه الشركة $\sigma^2 = 25$.

المطلوب

اختبر الفرض القائل بأن متوسط إنتاجية العامل اليومية في هذه الشركة تزيد عن 30 وحدة يومياً بدرجة ثقة 95% (أي متوسط معنوية 5%)

الحل

$$\because \overline{X} = 28 \quad , n = 100 \quad , \sigma^2 = 25$$

١- الفرض العدمي

$$H_0 : \mu = 30$$

الفرض البديل

$$H_1 : \mu > 30$$

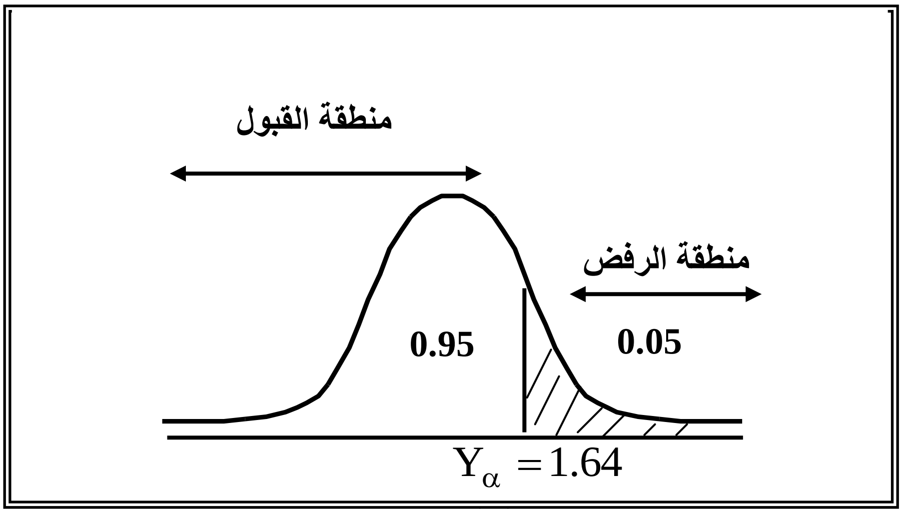
٢- وحيث أن:

$$Y = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{28 - 30}{\frac{5}{10}} = -4$$

٣- من جدول التوزيع المعتاد القياسي "بملحق رقم ٤"

$$\Theta\ (1 - \alpha) = 0.95 \rightarrow \therefore Y_\alpha = 1.64$$

٤- والشكل التالي يوضح منطقة الرفض ومنطقة القبول للفرض العدمي:-



شكل (٥-٤)

وبما أن Y تقع في منطقة الرفض، إذن نرفض الفرض القائل بأن متوسط إنتاجية العامل في الشركة أكبر من 30 وحدة وذلك بدرجة ثقة 95%

ثانياً: إذا كان تباين المجتمع σ^2 غير معلوم فإننا نستخدم تباين المتغير في الدراسة العينة S^2 كتقدير للتباين في المجتمع وفي هذه الحالة نحسب المقياس t حيث

$$t = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} \qquad (5.6)$$

فنجد أن المتغير t يتبع توزيع استيودنت بدرجات حرية (n-1) وباستخدام جدول توزيع t (استيودنت) بملحق رقم (٦) ، ويتم حساب قيمة t الجدولية عند درجة ثقة معينة.

ويلاحظ هنا أنه في حالة معلومية التباين في المجتمع تستخدم جداول التوزيع المعتاد والقياس لتحديد قيمة Y الجدولية ، أما في حالة عندما يكون تباين المجتمع مجهول نستخدم جداول توزيع t.

مثال (٥-٢)

إذا سحبت عينة مكونة من 9 طلاب وحسب الوسط الحسابي والانحراف المعياري لدرجة الطالب في مادة الإحصاء في العينة فوجد $\overline{X} = 8 , S = 3$. اختبر الفرض القائل أن متوسط درجة الطالب في الإحصاء في المجتمع المسحوب منه العينة أقل من 6.5 بدرجة ثقة 95%.

الحل

$$\Theta\ n = 9 , \overline{X} = 8 , S = 3$$

١- الفرض العدمي H_0 بحيث:-

$$H_0 : \mu = 6.5$$

الفرض البديل H_1 بحيث:-

$$H_1 : \mu < 6.5$$

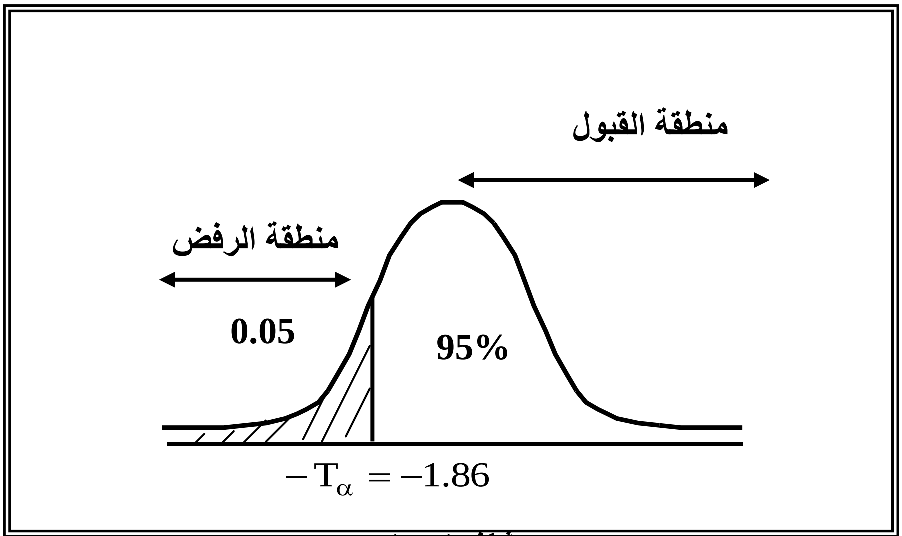
٢- وحيث أن:

$$t = \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{8 - 6.5}{\frac{3}{3}} = 1.5$$

٣- من جدول توزيع استيودنت "بملحق ٦" عند درجة الثقة 95% نجد أن

$$t_\alpha = -1.86$$

كما هو موضح بالشكل (٥-٥):-



شكل (٥-٥)

٤- بما أن t المحسوبة في (٢) تقع في منطقة القبول حيث $1.5 > -1.86$

إذن نقبل الفرض العدمي ونرفض الفرض البديل القائل بأن متوسط درجة الطالب في المجتمع أقل من 6.5 درجة وذلك بدرجة ثقة 95%.

(٥-٣) اختبارات الفروض عن قيمة تباين المجتمع

Tests of Hypotheses of Population's Variance

إذا كان σ^2 هو تباين المتغير محل الدراسة في المجتمع المطلوب اتخاذ قرار بشأنه، S^2 تباين المتغير محل الدراسة في العينة العشوائية المسحوبة من المجتمع بحجم n .

لإجراء الاختبار الإحصائي بشأن قيمة معينة لتباين المتغير في المجتمع بناء على معلومات من العينة نتبع الخطوات المذكورة في الفصل (٥-١) على النحو التالي:-

١- الفرض العدمي

$$H_0 : \sigma^2 = \sigma_0^2$$

الفرض البديل

$$H_1 : \sigma^2 > \sigma_0^2$$

أو

$$H_1 : \sigma^2 < \sigma_0^2$$

أو

$$H_1 : \sigma^2 \neq \sigma_0^2$$

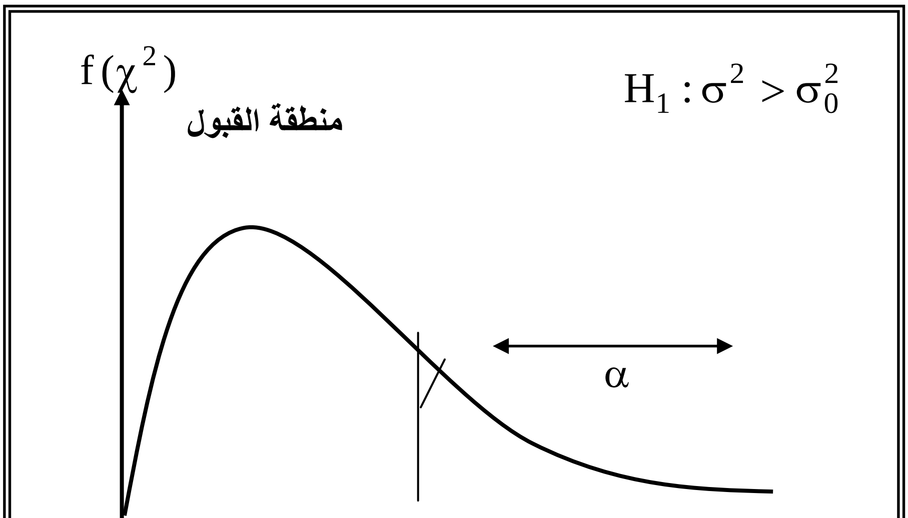
٢- من الفصل (٣-٣) نجد أن المتغير χ^2 حيث:-

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \qquad (5.7)$$

متغير يتبع توزيع χ^2_{n-1} لذلك نحسب المقياس χ^2 حيث:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

٣- عند درجة الثقة $(1-\alpha)$ نحدد منطقة القبول ومنطقة (أو مناطق الرفض للفرض العدمي كما هو موضح بأشكال (٥-٦)-(٥-٨)



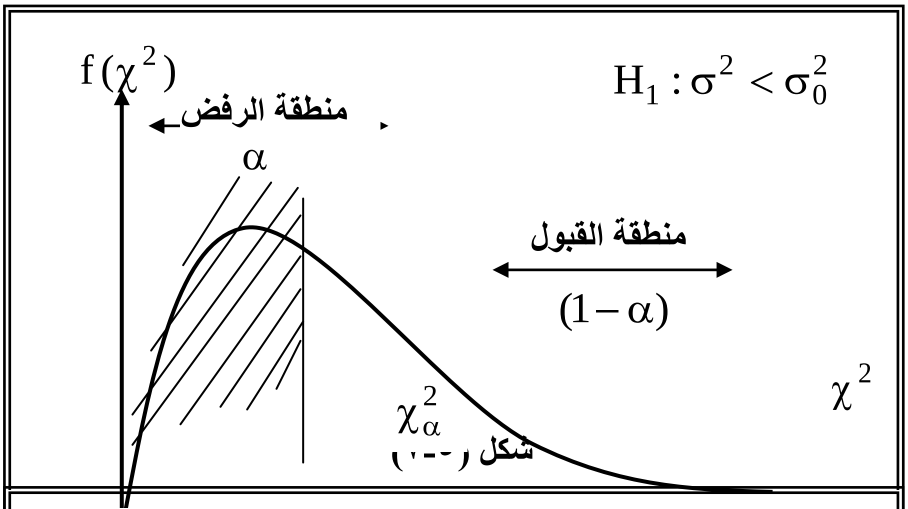
$(1-\alpha)$

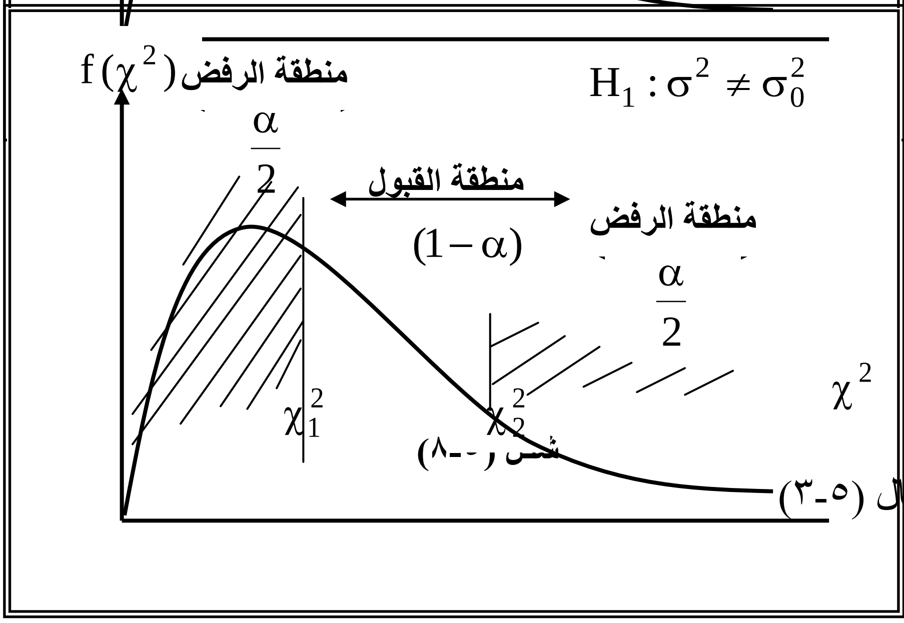
منطقة الرفض

χ^2

χ^2_α

شكل





مثال (٥-٣)

إذا سحبت عينة عشوائية من مجتمع معتاد حجمها 5 مفردات وتباين المتغير محل الدراسة في العينة $S^2 = 25$.

المطلوب

1- **اختبر الفرض القائل بأن تباين المجتمع المسحوب منه العينة يزيد عن 20 وذلك بدرجة ثقة 95%.**
2- **اختبر الفرض القائل بأن تباين المجتمع يختلف عن 20 بدرجة ثقة 95% أيضاً.**

الحل

أولاً:-

$$\because S^2 = 25 , n = 5 , (1- \alpha) = 0.95$$

1- **الفرض العدمي**

$$H_0 : \sigma^2 = 20$$

الفرض البديل

$$H_1 : \sigma^2 > 20$$

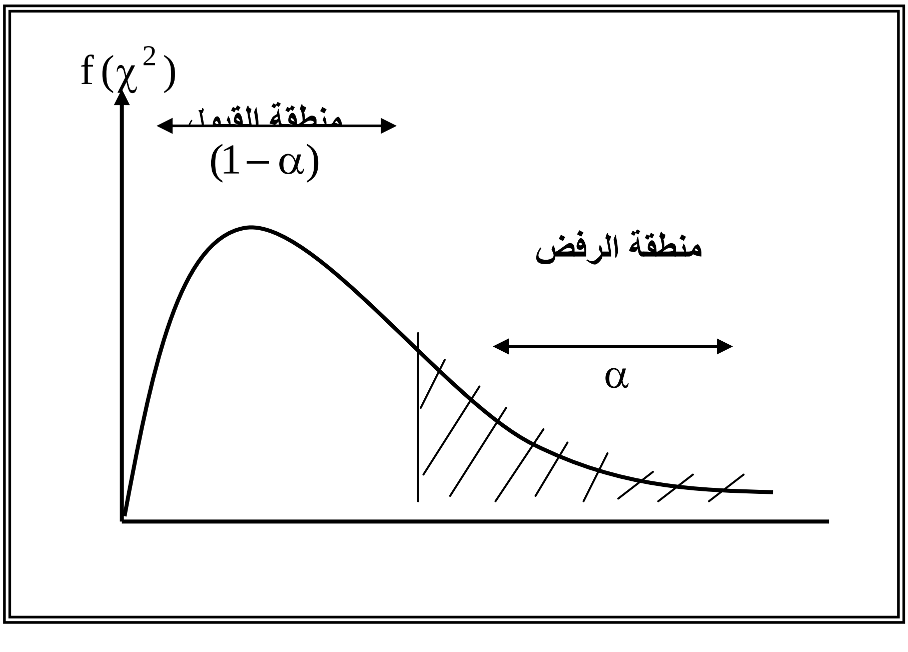
2- **نحسب قيمة χ^2 حيث**

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

$$= \frac{(5-1)(25)}{20} = 5$$

3- **عند درجة الثقة 95% نجد أن $\chi^2 = 9.49$**

4- **فتصبح منطقة الرفض والقبول كما هو موضح بالشكل التالي:-**



χ^2

$$\chi_{\alpha}^{2} = 9.49$$

شكل (٥-٦)

٥- وبما أن قيمة χ^2 المحسوبة في الخطوة (٢) تساوي 5 أي تقع في منطقة القبول إذن نقبل الفرض العدمي ونرفض الفرض البديل القائل بأن التباين في المجتمع يزيد عن 20 وذلك بدرجة ثقة 95%.

ثانياً:-

١- في هذه الحالة نجد أن

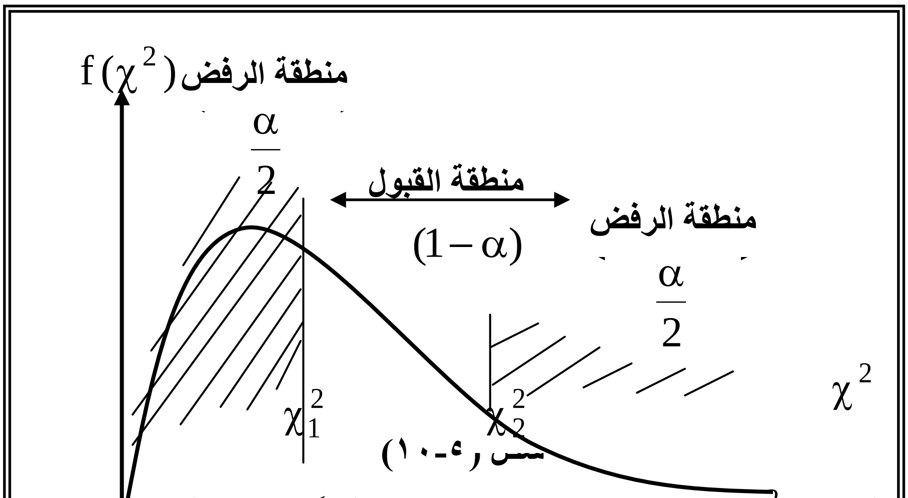
الفرض العدمي

$$H_0 : \sigma^2 = 20$$

والفرض البديل

$$H_1 : \sigma^2 \neq 20$$

٢- في هذه الحالة توجد منطقتين للرفض كما هو موضح بالشكل التالي:-



شكل (٥-١٠)

وبما أن قيمة χ^2 المحسوبة في الخطوة رقم (٢) بـ أولاً تساوي 5 أي تقع في منطقة القبول حيث:-

$$\chi_1^2 < \chi^2 < \chi_2^2$$

إذن نقبل الفرض العدمي ونرفض الفرض البديل القائل بأن التباين في المجتمع يختلف عن 20.

(٥-٤) اختبارات الفروض عن قيمة النسبة في المجتمع

Tests of Hypotheses of a Population's Preparation

إذا كانت θ هي نسبة خاصية ما في المجتمع ، $\hat{\theta}$ هي النسبة لهذه الخاصية في العينة العشوائية من هذا المجتمع فنجد أن:-

١- الفرض العدمي

$$H_0 : \theta = \theta_0 \qquad (5.9)$$

والفرض البديل

$$H_1 : \theta > \theta_0 \qquad (5.9)$$

أو

$$H_1 : \theta < \theta_0 \qquad (5.10)$$

أو

$$H_1 : \theta \neq \theta_0 \qquad (5.11)$$

والأشكال التالية (٥-١١) – (٥-١٣) توضح مناطق القبول والرفض في كل حالة.

٢- ومن الباب الثاني نجد أن النسبة في العينة $\hat{\theta}$ متغير عشوائي توقعه يساوي النسبة في المجتمع أي:-

$$\mu(\hat{\theta}) = \theta$$

وتباينه

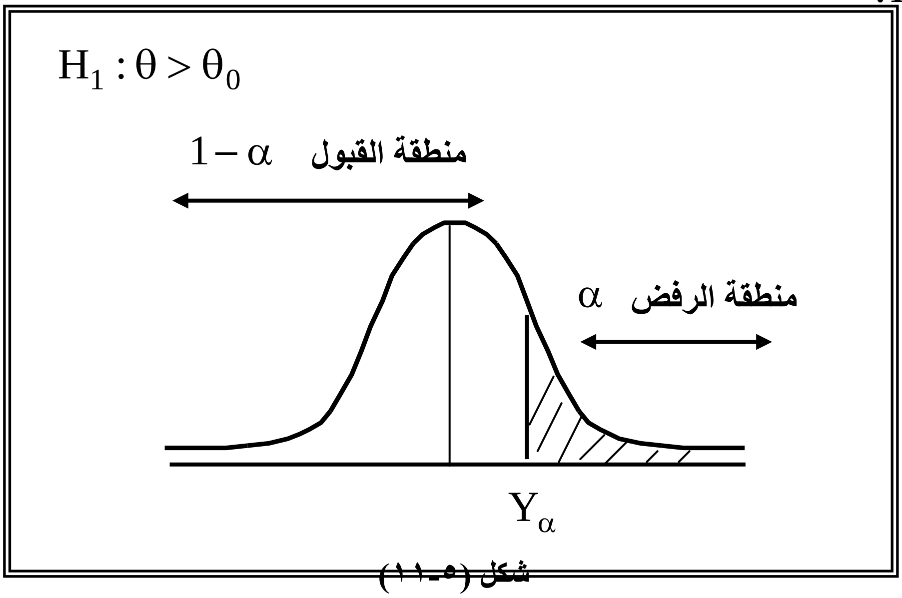
$$\sigma^2(\hat{\theta}) = \frac{\theta(1-\theta)}{n}$$

وبالتالي إذا فرضنا أن المتغير Y بحيث:-

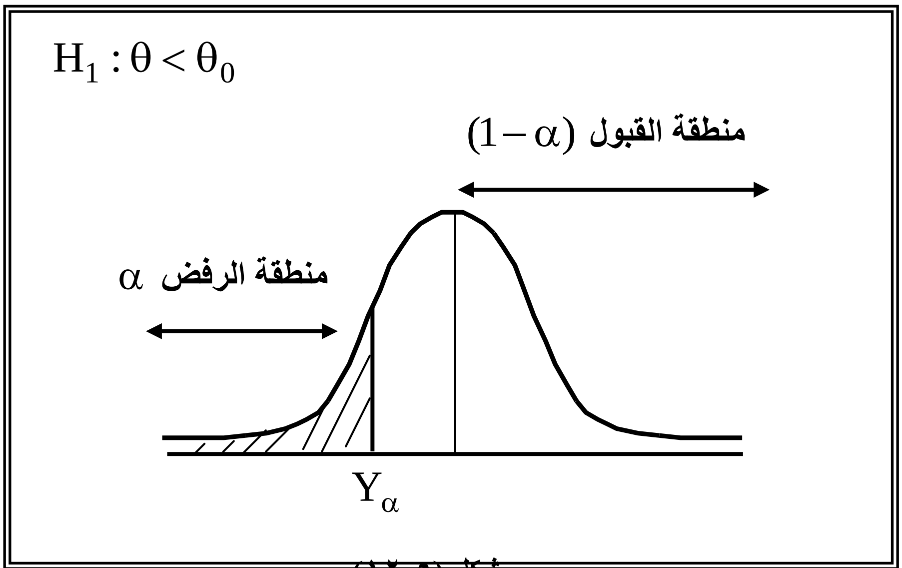
$$Y = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}}$$

فإن Y متغير يتبع التوزيع المعتاد القياسي.

ونظراً لأن θ غير معلومة فإن $\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$ يستخدم كتقدير للانحراف المعياري لـ**Y**.

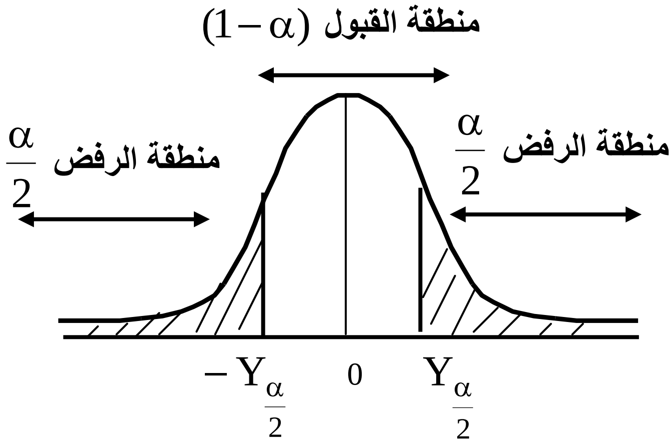


شكل (٥-١١)



شكل (٥-١٢)

$$H_1 : \theta \neq \theta_0$$



شكل (٥-١٣)

فإذا كانت قيمة Y المحسوبة تقع في منطقة القبول فإننا نقبل الفرض العدمي، أما إذا وقعت في منطقة الرفض فإننا نرفض الفرض العدمي ونقبل الفرض البديل.

مثال (٥-٤)

سحبت عينة عشوائية مكونة من 100 شخص فكانت نسبة المدخنين في العينة $\hat{\theta} = 56\%$ اختبر الفرض القائل بأن نسبة المدخنين في المجتمع المسحوب منه العينة أقل من 50% وذلك بمستوى معنوية 5% .

<u>الحل</u>

$$\because \hat{\theta} = 0.56 \,, n = 100 \,, 1- \ \alpha = 0.95$$

١- الفرض العدمي

$$H_0 : \theta = 0.50$$

الفرض البديل

$$H_1 : \theta < 0.50$$

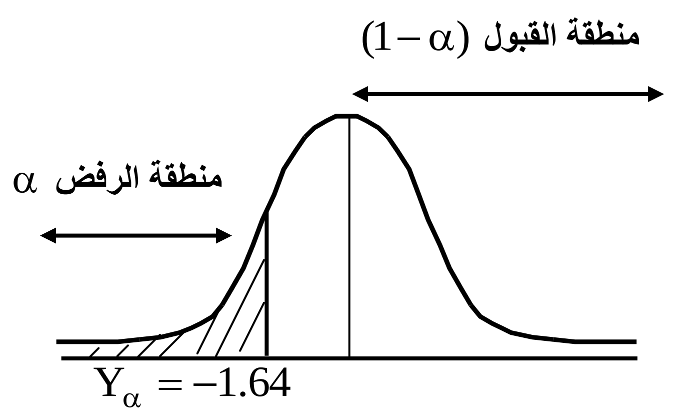
٢- المقياس Y حيث:-

$$Y=\frac{\hat{\theta}-\theta_0}{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}}=\frac{0.56-0.50}{\sqrt{\frac{0.56(44)}{100}}}$$

$$=\frac{0.06}{0.4964}=1.209$$

٣- مـن جـدول التوزيـع المعتـاد القياسـي وعنـد درجـة الثقـة **0.95** نجـد أن $Y_\alpha = 1.67$ كما هو موضح بالشكل (٥-١٤).

وبمـا أن $Y > Y\alpha$ أي تقـع فـي منطقـة القبـول وبالتـالي نقبـل الفـرض العـدمي ونرفض الفرض البديل القائل بأن نسبة المدخنين في المجتمع المسحوب منه العينة أقل من **50%** وذلك بدرجة ثقة **95%**.



شكل (٥-١٤)

مثال (٥-٥)

في الاختبارات اليومية لقياس جودة الوحدات المنتجة في أحد المصانع أخذت عينة مكونة من 400 وحدة وبفحص كل وحدة في العينة وجد أن نسبة الوحدات غير المطابقة للمواصفات 15% . اختبر الفرض القائل بأن نسبة الوحدات غير المطابقة للمواصفات في الإنتاج اليومي للمصنع أكبر من 10% وذلك بدرجة ثقة 99%.

الحل

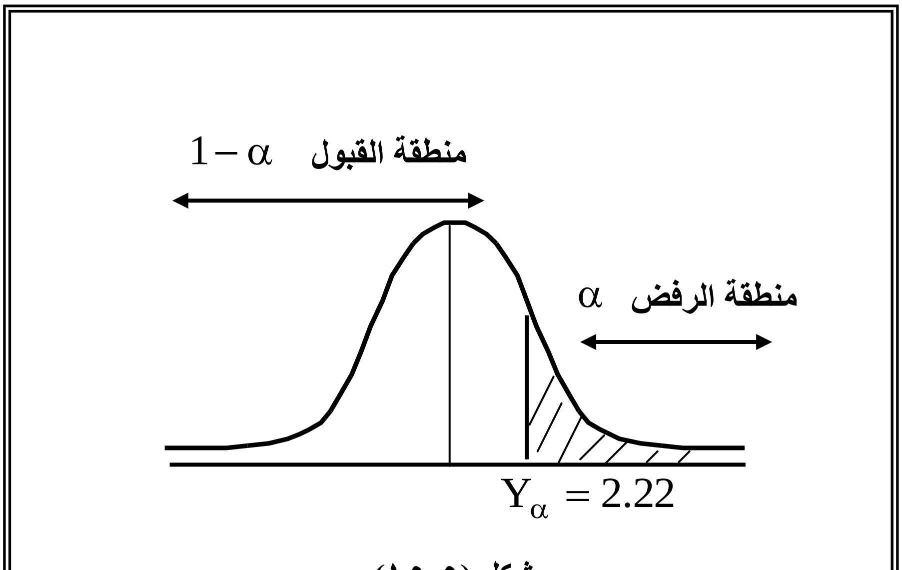
$$\because n = 400 , \hat{\theta} = 0.15 , (1-\alpha) = 0.99$$

الفرض العدمي

$$H_0 : \theta = 0.10$$

الفرض البديل

$$H_1 : \theta > 0.10$$



شكل (٥-١٥)

وحيث أن:

$$Y = \frac{\hat{\theta} - \theta_0}{\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}} = \frac{0.15 - 0.10}{\sqrt{\frac{0.15(0.85)}{400}}}$$

$$= \frac{0.05}{0.36} = 1.39$$

وبما أن $Y < Y_\alpha$ أي أن Y تقع في منطقة القبول إذن نقبل الفرض العدمي ونرفض البديل القائل بأن نسبة الوحدات غير المطابقة للمواصفات في الإنتاج اليومي للمصنع تزيد عن 10% وذلك بدرجة ثقة 99%.

(٥-٥) اختبارات الفروض عن الفرق بين متوسطي مجتمعين باستخدام عينتين مستقلتين

Tests of Hypotheses of the Difference between Two Population Means: Independent Samples

إذا فرضنا أن μ_1, μ_2 هما متوسطي المتغير محل الدراسة في مجتمعين 1,2 على الترتيب والمراد إجراء اختبارات للفروض عن الفرق بينهما $(\mu_2 - \mu_1)$ باستخدام عينتين مستقلتين مسحوبتين منت هذين المجتمعين حجمهما n_1, n_2 ، ومتوسط المتغير في العينة $\overline{X}_1, \overline{X}_2$ على الترتيب

في هذا الفصل سوف تقتصر دراستنا لاختبارات الفروض عن الفرق بين متوسطي المجتمعين μ_1, μ_2 باستخدام عينتين مستقلتين مسحوبتين من مجتمعين معتادين (أي المتغير محل الدراسة في كل منهما يتبع التوزيع المعتاد الطبيعي Normal Probability Distribution) في الحالات التالية:-

<u>الحالة الأولى</u>

إذا كانا تباينا المجتمعين σ_1^2, σ_2^2 معلومتين.

<u>الحالة الثانية</u>

إذا كانا تباينا المجتمعين σ_1^2, σ_2^2 مجهولتين.

الحالة الأولى:- **إذا كانا $\overline{X}_1, \overline{X}_2$ هما متوسطي المتغير محل الدراسة في العينتين العشوائيتين المستقلتين المسحوبتين من مجتمعي الدراسة فإن المتغير Y حيث:-**

$$Y = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad (5.12)$$

متغير يتبع التوزيع المعتاد القياسي* ويأخذ Y كمقياس للاختبار ، ويتم إجراء الاختبار على النحو التالي:-

١- صياغة الفرض العدمي والفرض البديل على النحو:-
الفرض العدمي:-

$$H_0 : (\mu_1 - \mu_2) = h \tag{5.13}$$

الفرض البديل:-

$$H_1 : (\mu_1 - \mu_2) > h \tag{5.14}$$

أو

$$H_1 : (\mu_1 - \mu_2) < h \tag{5.15}$$

أو

$$H_1 : (\mu_1 - \mu_2) \neq h \tag{5.16}$$

حيث h هي الفرق بين متوسطي المتغير محل الدراسة في المجتمعين وفي حالة افتراض تساوي المتوسطين أي $\mu_1 = \mu_2$ فهذا يعني أن h تساوي صفر.

٢- عند مستوى المعنوية α "أي درجة الثقة $(1-\alpha)$" يتم تحديد حد (أو حدي الثقة) Y_α من جدول التوزيع المعتاد القياسي وتحديد منطقة (أو مناطق) الرفض ومنطقة القبول كما هو موضح في الأشكال من (٥-١٦) - (٥-١٨).

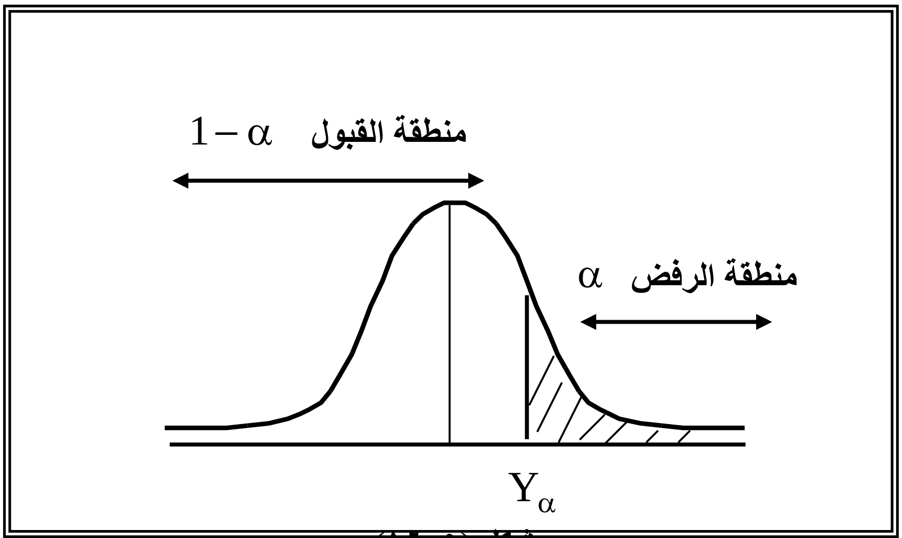
٣- نحسب المقياس Y في المعادلة (5.12) بوضع $(\mu_1 - \mu_2) = h$.

---

* Mood, A.M., Grayhill. F.A.; and Boes D.C.(1963): "Introduction to the Theory of
Statistics" McGraw – Hill International Book Company , London

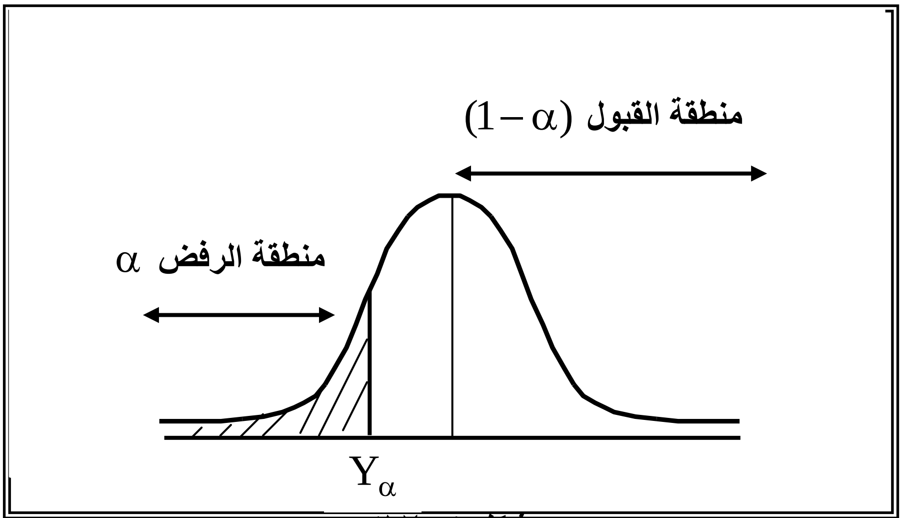
٤- مقارنة المقياس Y بـ Y_α (أو $Y_{\alpha/2}$) وتحديد هل Y تقع في منطقة القبول أو في منطقة (أو مناطق) الرفض ووفقاً لذلك يتم قبول أو رفض الفرض العدمي.

إذا كان الفرض البديل $H_1 : (\mu_1 - \mu_2) > h$



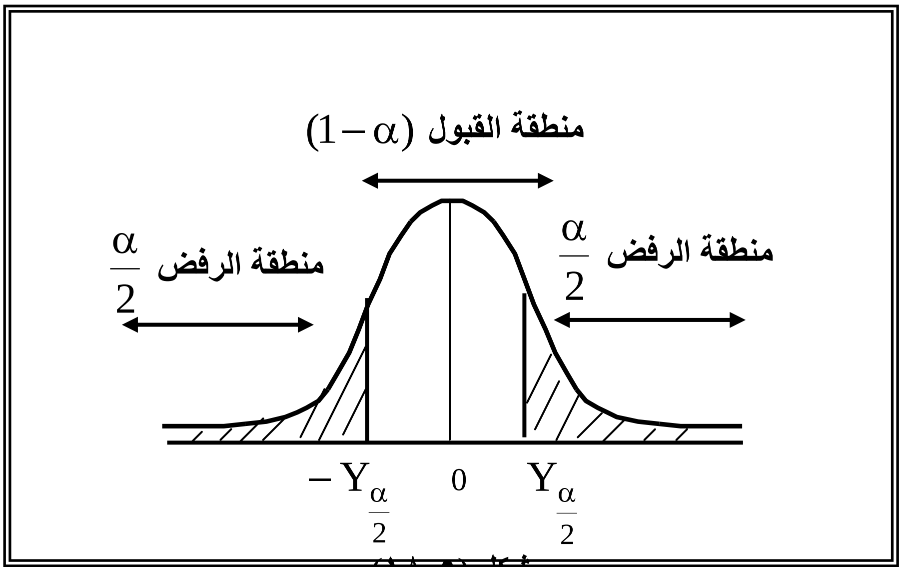
شكل (٥-١٦)

إذا كان الفرض البديل $H_1 : (\mu_1 - \mu_2) < h$



شكل (٥-١٧)

إذا كان الفرض البديل $H_1 : (\mu_1 - \mu_2) \neq h$



شكل (٥-١٨)

مثال (٥-٦)

أخذت عينة عشوائية حجمها n_1 حيث $n_1 = 10$ من العاملين بأحد البنوك الحكومية فكان متوسط الأجر الشهري في إحدى السنوات في العينة يساوي $\overline{X}_1 = 800$ جنيه . وأخذت عينة عشوائية أخرى حجمها $n_2 = 10$ من العاملين بأحد البنوك الاستثمارية فكان متوسط الأجر الشهري للعامل في نفس السنة يساوي $\overline{X}_2 = 870$ جنيه . فإذا كان الانحراف المعياري للأجر في البنك الحكومي يساوي $\sigma_1 = 25$ والانحراف المعياري في البنك الاستثماري يساوي $\sigma_2 = 40$ جنيه.

المطلوب:-

عند درجة ثقة 95% اختبر الفرض القائل:-

<u>أولاً</u>: أن متوسط الأجر الشهري في البنك الاستثماري يزيد عن متوسط الأجر الشهري في البنك الحكومي بـ50 جنيه على الأكثر.

<u>ثانياً</u>: أن متوسط الأجر الشهري في البنك الاستثماري يزيد عن متوسط الأجر الشهري في البنك الحكومي بـ30 جنيه على الأقل.

<u>ثالثاً</u>: أن متوسط الأجر الشهري في البنك الحكومي يختلف عن البنك الاستثماري.

الحل

إذا فرضنا ان μ_1, μ_2 تشير إلى متوسط الأجر الشهري في البنك الحكومي والبنك الاستثماري على الترتيب

$$\Theta\ \overline{X}_1 = 800\ ,\ n_1 = 10\ ,\ \sigma_1 = 25$$
$$\overline{X}_2 = 870\ ,\ n_2 = 10\ ,\ \sigma_2 = 40$$

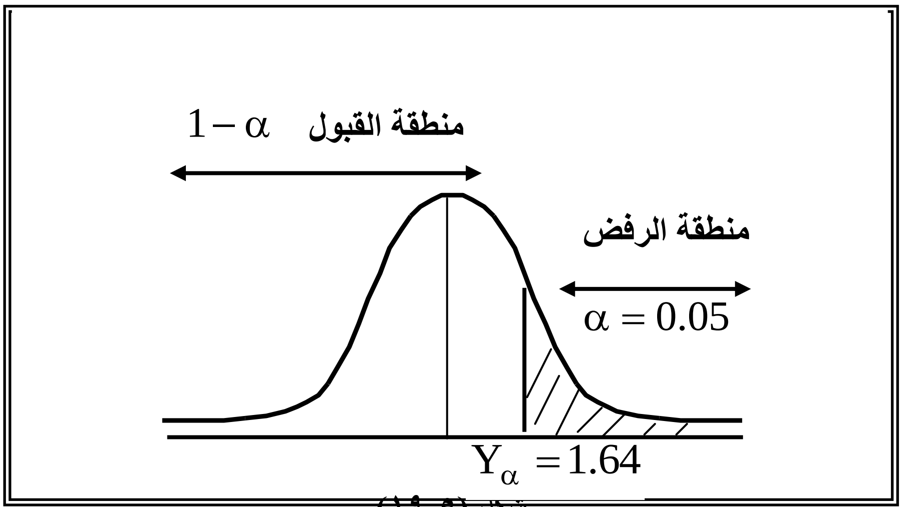
<u>أولاً</u> ١- الفرض العدمي

$$H_0 : (\mu_1 - \mu_2) = -50$$

٢- الفرض البديل

$$H_1 : (\mu_1 - \mu_2) > -50$$

٣- بما أن درجة الثقة تساوي 95% وباستخدام جدول التوزيع المعتاد القياسي (ملحق ٤) نجد أن $Y_\alpha = 1.64$ والشكل التالي يوضح منطقة الرفض والقبول (للفرض العدمي)



شكل (٥-١٩)

٤- نحسب المقياس Y حيث:-

$$Y = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

$$= \frac{(800 - 870)^2 - (-50)}{\sqrt{\dfrac{625}{10} + \dfrac{1600}{10}}} = \frac{-20}{14.92} = -1.34$$

وبما أن $Y>1.64$ إذن نقبل الفرض العدمي ونرفض الفرض البديل القائل بأن متوسط الأجر الشهري في البنك الاستثماري يزيد عن متوسط الأجر الشهري وفي البنك الحكومي بـ50 جنيه فأكثر.

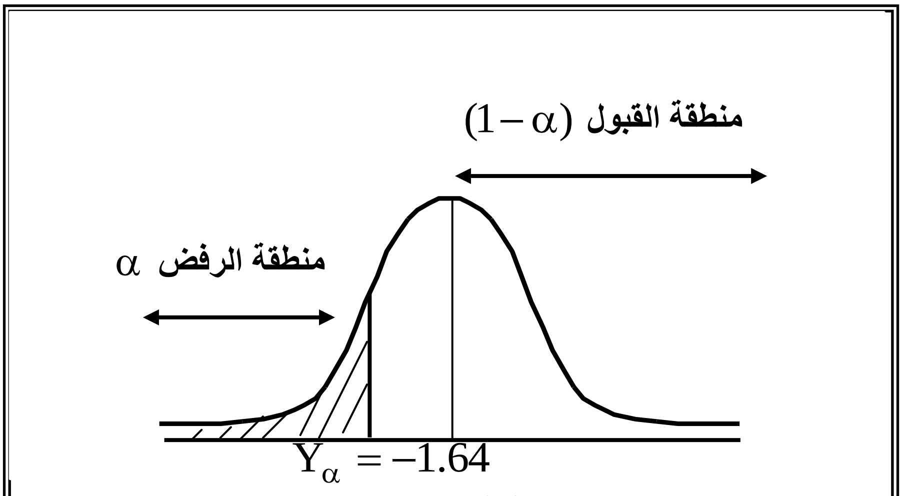
ثانياً: ١- الفرض العدمي

$$H_0 : (\mu_1 - \mu_2) = -30$$

الفرض البديل

$$H_1 : (\mu_1 - \mu_2) < -30$$

٢- عند درجة الثقة 95% نجد أن $Y_\alpha = 1.64$ كما هو موضح بالشكل التالي:-



شكل (٥-٢٠)

$$Y = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{(800 - 870) - (-30)}{\sqrt{\frac{625}{10} + \frac{1600}{10}}} = -2.68$$

وبما أن Y تقع في منطقة الرفض إذن نرفض الفرض العدمي ونقبل الفرض البديل القائل بأن متوسط الأجر الشهري بالبنك الاستثماري يزيد عن البنك الحكومي بـ30 جنيه شهرياً على الأقل.

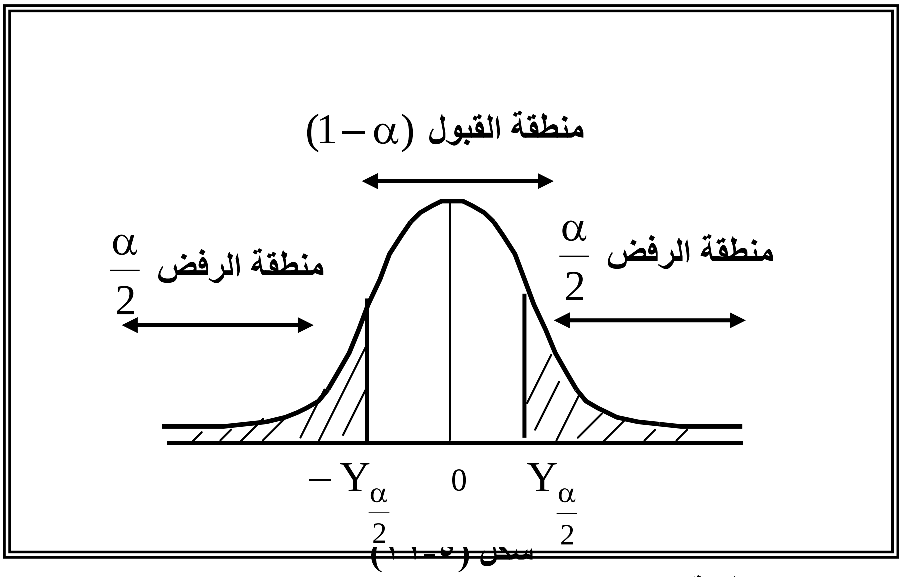
<u>ثالثاً</u>: ١- الفرض العدمي

$$H_0 : (\mu_1 - \mu_2) = 0$$

الفرض البديل

$$H_1 : (\mu_1 - \mu_2) \neq 0$$

٢- عند درجة الثقة 95% نجد أن $Y_{\alpha/2} = \pm 1.96$ **كما هو بالشكل التالي:-**



٣- نحسب قيمة Y حيث:-

$$Y = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{(800 - 870) - 0}{\sqrt{\frac{625}{10} + \frac{1600}{10}}} = -4.7$$

٤- وبما أن Y=-4.7 أي تقع في منطقة الرفض إذن نرفض الفرض العدمي ونقبل الفرض البديل القائل بأنه يوجد اختلاف بين متوسط الأجر الشهري في البنك الاستثماري والبنك الحكومي وذلك بدرجة ثقة 95%.

مثال (٥-٧)

تدعى إحدى الشركات لإنتاج السخانات الكهربائية في إعلاناتها أن مدة صلاحية السخانات من إنتاجها أكبر من مدة صلاحية السخان من إنتاج الشركات الأخرى المنافسة لها ، وللتأكد من صحة إعلان الشركة قامت إحدى الجهات المعنية بأخذ عينة من إنتاج هذه الشركة حجمها $n_1 = 20$ سخان وبفحصها وجد أن متوسط فترة الصلاحية للسخان $\overline{X}_1 = 11$ بالسنوات كذلك أخذت عينة عشوائية أخرى حجمها $n_2 = 15$ سخان من إنتاج الشركات الأخرى المنافسة وبفحصها وجد أن مدة الصلاحية للسخان تساوي $\overline{X}_2 = 10.5$ بالسنوات فإذا كانت مدة الصلاحية للسخان في الشركة متغير معتاد تباينه $\sigma_1^2 = 4$ كذلك كانت مدة الصلاحية للسخان في الشركات المنافسة متغير معتاد تباينه $\sigma_2^2 = 6.25$.

المطلوب:

اختبر ادعاء الشركة بدرجة ثقة 95%.

الحل

إذا فرضنا أن μ_1, μ_2 تشير إلى متوسط فترة الصلاحية للسخان في الشركة وفي الشركات المنافسة على الترتيب

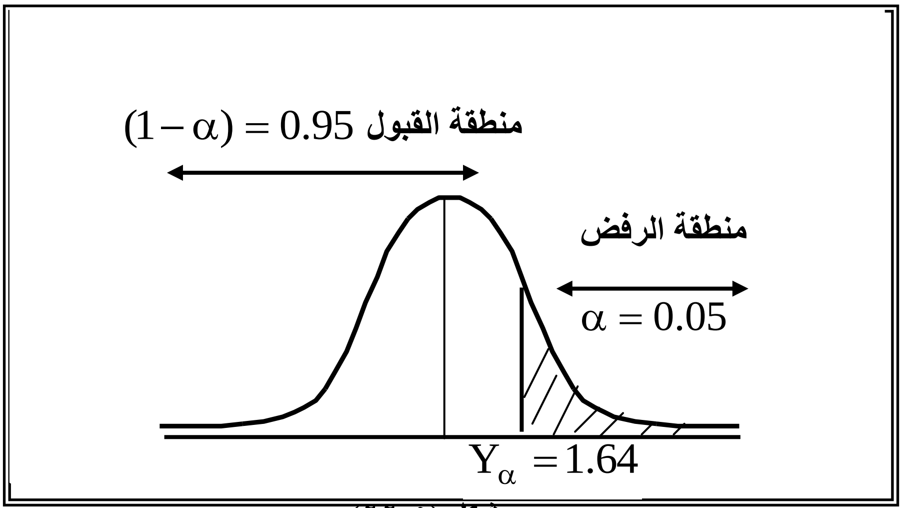
١- الفرض العدمي

$$H_0 : (\mu_1 - \mu_2) = 0$$

الفرض البديل

$$H_1 : (\mu_1 - \mu_2) > 0$$

٢- عند درجة الثقة $(1-\alpha) = 0.95$ فإن $Y_\alpha = 1.64$ كما هو موضح بالشكل التالي:-



شكل (٥-٢٢)

$$Y = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{(11 - 10.5) - 0}{\sqrt{\frac{4}{10} + \frac{6.25}{15}}} = \frac{0.5}{0.904} = 0.553$$

٣- بمقارنة قيمة Y بقيمة Y_α نجد أن Y=0.553 أقل من $Y_\alpha = 1.64$ أي أن Yتقع في منطقة القبول.

٤- وبالتالي نقبل الفرض العدمي ونرفض الفرض البديل القائل بأن متوسط فترة الصلاحية للسخان من إنتاج الشركة أكبر من متوسط فترة الصلاحية للسخان من الشركات المنافسة وذلك بدرجة ثقة 95% أي أن ادعاء الشركة غير صحيح.

الحالة الثانية: **إذا كان $\bar{X}_1, \bar{X}_2$ هما متوسطي المتغير محل الدراسة في العينتين العشوائيتين المسحوبتين من مجتمعي الدراسة عندما يكون تباين المتغير محل**

الدراسة في المجتمعين σ_1^2, σ_2^2 على الترتيب مجهولتين (أي غير معلومين) وتحت افتراض أن قيمة التباين في المجتمعين متساوية أي أن:-

$$\sigma^2 = \sigma_1^2 = \sigma_2^2 \qquad (5.17)$$

في هذه الحالة يستخدم S^2 كتقدير للتباين σ^2 (أي كتقدير لتباين المتغير في المجتمع الأول (أي تقدير σ_1^2)) كذلك كتقدير لتباين المتغير في المجتمع الثاني (أي تقدير لـ σ_2^2) حيث:-

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \qquad (5.18)$$

S_1^2, S_2^2 هما تباينا المتغير من العينة المسحوبة من المجتمع الأول والمجتمع الثاني على الترتيب

ويعتبر S^2 متغير عشوائي (حيث كل من S_1^2, S_2^2 متغيران عشوائيين) يتبع توزيع χ^2 بدرجات حرية $(n_1 + n_2 - 2)$ وبالتالي فإن المتغير **t** حيث:-

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (5.19)$$

متغير يتبع توزيع استيودنت (توزيع **t**) بدرجات حرية $(n_1 + n_2 - 2)$ وملحق رقم ٦ يعطي التوزيع الاحتمالي للمتغير **t**.

ويأخذ المقياس **t** في المعادلة **(5.19)** كمقياس للاختبار عندما عندما يكون حجم العينتين n_1, n_2 صغير (عادة يعتبر حجم العينة صغير عندما يكون عدد مفرداتها أقل من **30** مفردة).

وعندما يكون حجم العينتين n_1, n_2 كبير فإن المتغير **t** المعرف في المعادلة (5.19) يؤول إلى التوزيع المعتاد القياسي وبالتالي يستخدم جدول التوزيع المعتاد القياسي في ملحق رقم (٤) بدلاً من جدول توزيع **t** بملحق (٦).

ويتم إجراء الاختبار بنفس الخطوات المتبعة في الحالة الأولى – مع مراعاة:-

١- حساب المقياس **t** بدلاً من **Y**.
٢- استخدام جداول توزيع استيودنت ف ملحق (٦) عندما يكون حجم العينتين صغير وتستخدم جداول التوزيع المعتاد القياسي في ملحق (٤) عندما يكون حجم العينتين كبير . وسوف نوضح ذلك في الأمثلة التالية.

مثال (٥-٨)

أجريت دراسة عن متوسط استهلاك الكهرباء للأسر محدودة الدخل بالمدينة ومتوسط استهلاك الكهرباء للأسرة محدودة الدخل بالقرية . فأخذت عينة عشوائية مكونة من 5 أسر بالقرى فكان متوسط الاستهلاك الشهري للكهرباء بالكيلووات على النحو التالي:-

54 , 51 , 50 , 49 , 46

وأخذت عينة عشوائية مكونة من 6 أسر محدودة الدخل بالمدن فكان متوسط الاستهلاك الشهري للكهرباء بالكيلووات على النحو التالي:-

59, 52, 68, 61, 55, 65

بافتراض أن متوسط استهلاك الأسرة من محدودي الدخل في القرية أو في المدينة متغير يتبع التوزيع المعتاد.

المطلوب

١- اخبر الفرض القائل أن متوسط استهلاك الكهرباء الشهري للأسرة في المدن أكبر منه في القرى بدرجة ثقة 95%.
٢- اختبر الفرض القائل أن متوسط استهلاك الكهرباء الشهري للأسرة في القرية لا يختلف عن متوسط استهلاك الكهرباء الشهري للأسرة في المدينة بدرجة ثقة 95%.

الحل

١- إذا فرضنا أن μ_1 , μ_2 هما متوسطي استهلاك الكهرباء الشهري للأسرة في القرية والمدينة على الترتيب حيث $n_1 = 5 , n_2 = 6$

<u>أولاً</u>

نحسب المؤشرات $\overline{X}_1 , \overline{X}_2 , S_1^2 , S_2^2$ على النحو التالي:-

$$\overline{X}_1 = \frac{\sum_{i=1}^{5} x_i}{n_1} = \frac{54 + 51 + 50 + 49 + 46}{5} = 50 \text{ كيلووات}$$

$$\overline{X}_2 = \frac{\sum_{i=1}^{6} x_i}{n_2} = \frac{59 + 52 + 68 + 61 + 55 + 65}{6} = 60 \text{ كيلووات}$$

وبما أن

$$S_1^2 = \frac{\sum_{i=1}^{5}(x_i - \overline{X}_1)^2}{n_1 - 1}$$

$$= \frac{(46-50)^2 + (49-50)^2 + (50-50)^2 + (51-50)^2 + (54-50)^2}{5-1}$$

$$= \frac{16+1+0+1+16}{4} = \frac{34}{4} = 8.5$$

وكذلك

$$S_2^2 = \frac{\sum_{i=1}^{6}(x_i - \overline{X}_2)}{n_2 - 1}$$

$$= \frac{(65-60)^2 + (55-60)^2 + (61-60)^2 + (52-60)^2 + (59-60)^2}{6-1}$$

$$= \frac{25+25+1+64+64+1}{5} = \frac{180}{5} = 36$$

$$\therefore S^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

$$= \frac{(5-1)8.5 + (6-1)36}{5+6-2} = \frac{214}{9} = 23.78$$

$$S = \sqrt{23.78} = 4.88$$

ثانياً

١- صياغة الفرض العدمي والفرض البديل على النحو التالي:-

(١)

الفرض العدمي

$$H_0 : (\mu_1 - \mu_2) = 0$$

الفرض البديل

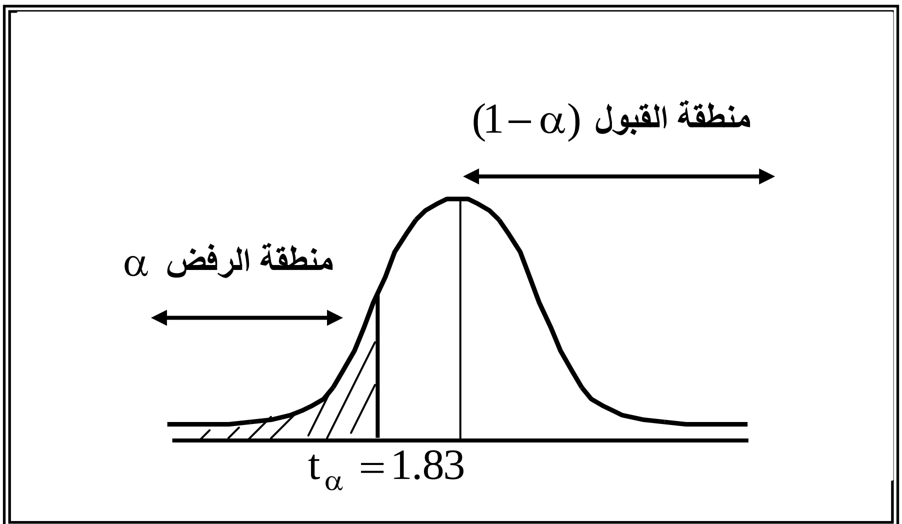
$$H_1 : (\mu_1 - \mu_2) < 0$$

٢- يحسب المقياس t حيث:-

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{(50-60)-0}{4.88\sqrt{\frac{1}{5}+\frac{1}{6}}} = \frac{-10}{2.8012} = -4.7$$

٣- عند درجة الثقة 95% وباستخدام جدول توزيع t بملحق رقم (٦) عند درجات الحرية $(n_1 + n_2 - 2)$ نجد أن $t_\alpha = -1.83$.



شكل (٥-٢٣)

٤- بمقارنة t=-3.75 بقيمة $t_\alpha = -1.83$ نجد أن t تقع في منطقة الرفض وبالتالي نرفض الفرض العدمي ونقبل الفرض البديل القائل بان متوسط

استهلاك الأسرة الشهري من الكهرباء في المدينة أكبر من متوسط الاستهلاك الشهري للكهرباء للأسرة بالقرية.

(٢)

الفرض العدمي

$$H_0 : (\mu_1 - \mu_2) = 0$$

الفرض البديل

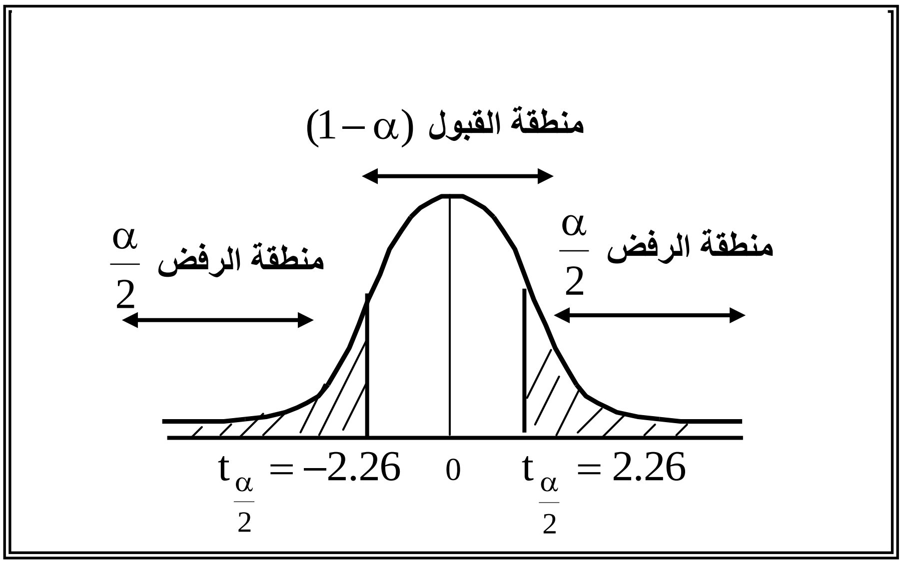
$$H_1 : (\mu_1 - \mu_2) \neq 0$$

٢- من أولاً نجد أن t=-3.57

٣- عند درجة الثقة 95% وباستخدام جدول توزيع t بملحق رقم (٦) عند درجات الحرية9 $n_1 + n_2 - 2 = 9$ نجد ن $t_{\alpha/2} = \pm 2.26$.

٤- بمقارنة t=-3.57 بقيمة $t_{\alpha/2} = \pm 2.26$ نجد أن t تقع في منطقة الرفض

بالتالي نرفض الفرض العدمي ونقبل الفرض البديل القائل بأن متوسط استهلاك الأسر الشهري من الكهرباء في القرية يختلف عن متوسط استهلاك الأسرة للكهرباء في المدينة وذلك بدرجة ثقة 95%.



شكل (٥-٢٤)

مثال (٥-٩)

في المثال السابق إذا أخذت عينة من القرى حجمها 200 أسرة فكان متوسط استهلاك الأسرة من الكهرباء والتباين هما $\overline{X}_1 = 55$ كيلووات , $\sigma_1^2 = 40$ على التوالي كذلك أخذت عينة أخرى من المدن حجمها 100 أسرة فكان متوسط استهلاك الأسرة والتباين هما $\overline{X}_2 = 68$ كيلووات , $\sigma_2^2 = 38$ على التوالي.

<u>المطلوب</u>

اختبر الفرض القائل بان متوسط استهلاك الأسرة الشهري للكهرباء في القرية مختلف عن متوسط استهلاك الأسرة الشهري للكهرباء في المدينة بدرجة ثقة 95%.

<u>الحل</u>

١- الفرض العدمي

$$H_0 : (\mu_1 - \mu_2) = 0$$

الفرض البديل

$$H_1 : (\mu_1 - \mu_2) \neq 0$$

٢- نحسب التباين المشترك S^2 ثم نحسب المقياس t على النحو التالي:-

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

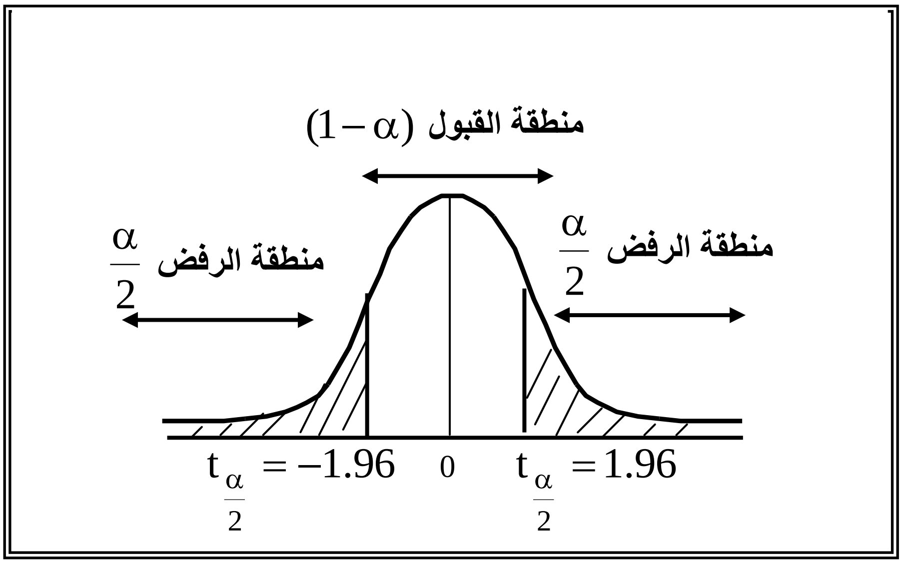
$$= \frac{(200-1)40 + (100-1)38}{(200+100-2)} = \frac{11722}{298} = 39.336$$

$$S = \sqrt{39.336} = 6.272 \text{ كيلووات}$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{(55-68) - 0}{6.272\sqrt{\frac{1}{100} + \frac{1}{200}}} = \frac{-13}{0.768} = -16.9$$

٣- بما أن درجة الثقة 95% وحجم العينتين $n_1 = 200, n_2 = 100$ **حجم كبير لذلك نستخدم جدول التوزيع المعتاد القياسي بملحق رقم (٤) لتحديد قيمة** $t_{\alpha/2} = \pm 1.96$ **كما هو موضح بالشكل التالي:-**



شكل (٥-٢٤)

وبما أن $t=16.9$ أي تقع في منطقة الرفض بالتالي نرفض الفرض العدمي ونقبل الفرض البديل القائل بأن متوسط استهلاك الأسرة الشهري من الكهرباء في القرية مختلف عن متوسط استهلاك الأسرة الشهري للكهرباء في المدينة وذلك بدرجة 95%.

(٥-٦) اختبارات الفروض عن الفرق بين متوسطي مجتمعين باستخدام عينتين غير مستقلتين

Tests of hypotheses of the Difference Between Two Populations Mean : Dependent Samples

في الفصل السابق تناولنا بالدراسة اختبارات الفروض الإحصائية عن الفرق بين متوسطي مجتمعين باستخدام عينتين مستقلين مسحوبتين من المجتمعين ، وفي هذا الفصل سوف نتناول اختبارات الفروض الإحصائية عن الفرق بين متوسطي مجتمعين باستخدام عينتين غير مستقلين حجم كل منهما يساوي n:-

فإذا فرضنا أن مشاهدات العينة الأولى هي:-

$$x_1, x_2, \ldots, x_n$$

ومشاهدات العينة الثانية هي:-

$$y_1, y_2, \ldots, y_n$$

وعدم استقلال العينتين هنا يعني أن المفردة رقم (j) المأخوذة منها المشاهدة x_j في العينة الأولى هي نفس المفردة المأخوذ منها المشاهدة y_j في العينة الثانية ومن هنا نجد أن الأزواج المرتبة Matched Pairs من المشاهدات:-

$$(x_1, y_1), (x_2, y_2), \ldots, (x_j, y_j), \ldots (x_n, y_n)$$

فإذا أشرنا للفرق المشاهدتين x_j, y_j لنفس المفردة رقم (j) بالرمز d_j فإن:-

١-

$$d_j = x_j - y_j \qquad , j = 1,2,...,n \qquad (5.20)$$

يمثل متغير عشوائي يتبع التوزيع المعتاد (الطبيعي) في حالة إذا كان كل من المتغيرين x,y في المجتمعين المسحوب منهما العينتين الأولى والثانية يتبع كل منهما التوزيع المعتاد (الطبيعي).

٣- إذا كان μ_1, μ_2 يشيران إلى متوسطي المجتمعين المسحوب منهما العينتين فإن المتوسط للفرق d سوف نشير له بالرمز μ_d حيث:-

$$\boxed{\mu_d = \mu_1 - \mu_2 \qquad (5.21)}$$

٤- إذا أشرنا إلى المتوسط والتباين للفرق d في العينة بالرمز $\bar{d}, S_d^2$ على الترتيب فإن:-

$$\bar{d} = \frac{\sum_{j=1}^{n} d_j}{n} \qquad (5.22)$$

$$S_d^2 = \frac{\sum_{j=1}^{n} (d_j - \bar{d})^2}{n-1} \qquad (5.23)$$

٥- المتغير t حيث:-

$$t = \frac{\bar{d} - (\mu_1 - \mu_2)}{\frac{S_d}{\sqrt{n}}}$$

$$= \frac{\sqrt{n}[\bar{d} - (\mu_1 - \mu_2)]}{S_d} \qquad (5.24)$$

متغير يتبع توزيع t (استيودنت) بدرجات حرية (n-1) أنظر نظرية (٣-٤) بالباب الثالث.

وفيما يلي سوف نلخص خطوات إجراء اختبار الفرق بين متوسطي مجتمعين باستخدام عينتين غير مستقلتين مسحوبتين من مجتمعين لكل منهما يتبع التوزيع المعتاد.

١- الفرض العدمي

$$\boxed{H_0 : (\mu_1 - \mu_2) = h \qquad (2.24)}$$

الفرض البديل

$$H_1 : (\mu_1 - \mu_2) > h \qquad (5.25)$$

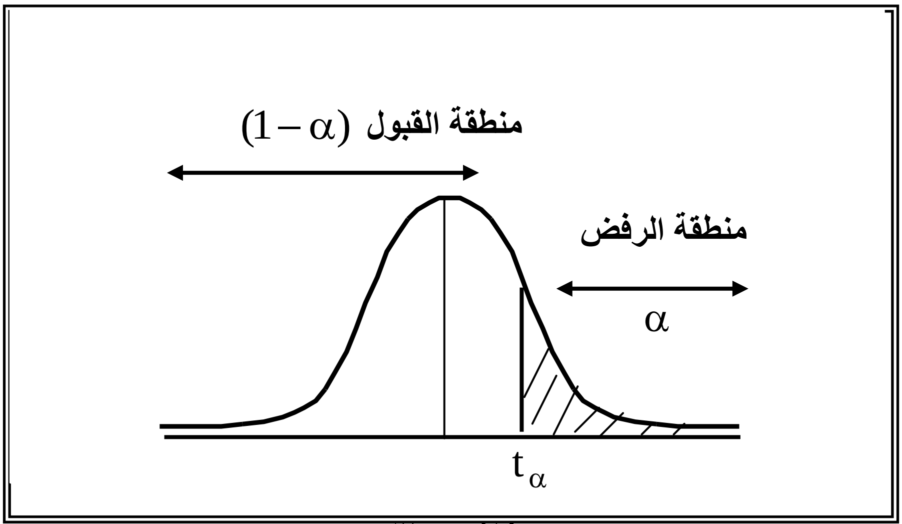
$$H_1 : (\mu_1 - \mu_2) < h \qquad (5.26)$$

$$H_1 : (\mu_1 - \mu_2) \neq h \qquad (5.27)$$

٢- نحسب المقياس t حيث:-

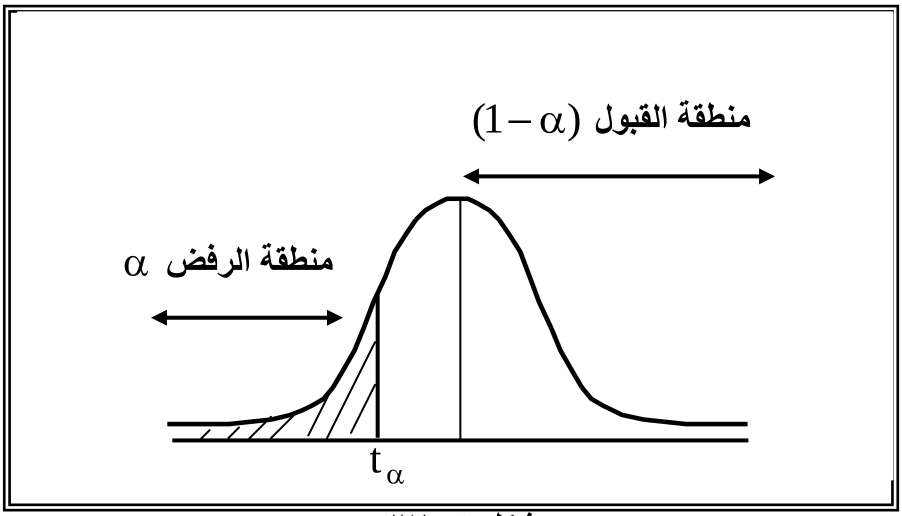
$$\boxed{t = \frac{\sqrt{n}(\bar{d} - (\mu_1 - \mu_2))}{S_d} = \frac{\sqrt{n}(\bar{d} - h)}{S_d} \qquad (5.29)}$$

٣-عند درجة الثقة $(1-\alpha)$ باستخدام جدول توزيع t بملحق رقم (٨) عند درجات الحرية (n-1) نحسب t_{α} إذا كان الفرض البديل كما في العلاقة (2.26) كما هو موضح بالشكل (٥-٢٦).



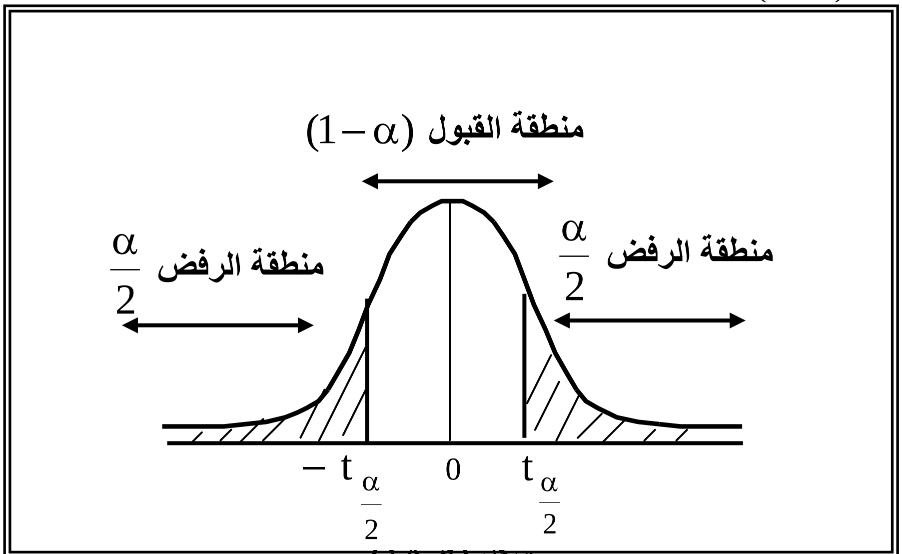
شكل(٥-٢٦)

أو إذا كان الفرض البديل كما هو في العلاقة (5.27) نحسب t_{α} كما هو موضح بشكل (٥-٢٧)



شكل (٥-٢٧)

وإذا كـان الفرض البـديل كمـا هو فـي العلاقـة (5.28) نحسـب $\pm t_{\frac{\alpha}{2}}$ كمـا هو موضـح بشكل (٥-٢٨).



شكل (٥-٢٨)

٤- نقارن قيمة t المحسوبة في الخطوة رقم (٢) بقيمة t_{α} أو $t_{\alpha/2}$ فإذا كانت قيمـة t المحسوبة تقع في منطقة القبـول نقبل الفرض العدمي أمـا إذا كانـت قيمـة t المحسوبة تقع في منطقة الرفض فإننا نرفض الفرض العدمي ونقبل الفرض البديل.

ملحوظات

- إذا كان عدد درجات الحرية (n-1) عدد كبير (وعادة يكون عدد درجات الحرية كبيـر أي حجـم العينــة كبيـر عنــدما 30<(n-1)) فــإن المتغيـرات بالمعادلــة (5.24) يؤول إلى التوزيع المعتاد القياسي وهذا يعني أننـا في هذه الحالـة (أي عندما 30<(n-1)) نستخدم جداول التوزيـع المعتـاد القياسـي بملحق رقم (٤) بـــدلاً مـــن جـــداول توزيـــع t بملحـــق رقـــم (٦) لحســـاب حـــدود الثقـــة $\pm t_{\alpha/2}$ أو $\pm t_{\alpha}$.

- عندما يكون المتغير محل الدراسة في مجتمع أو في المجتمعين معاً:-

أ- غير معروف التوزيع.

ب- يتبع توزيع غير التوزيع المعتاد.

ففي هذه الحالات لا يمكن استخدام اختبار الفرق بين متوسطي مجتمعين المذكور أعلاه وفي هذه الحالات يستخدم اختبار لا معلمي كما سوف نستعرض ذلك بالتفصيل في الفصل (٦-٤) بالباب القادم.

مثال (٥-١٠)

قامت إدارة التدريب بأحد البنوك بدراسة فاعلية أحدى الدورات التدريبية للعاملين ، والجدول التالي يوضح مستوى 10 من العاملين (حجم العينة n=10) قبل الدورة التدريبية وبعد الدورة.

رقم العامل	1	2	3	4	5	6	7	8	9	10
المستوى قبل الدورة	50	80	67	73	95	65	77	53	79	61
المستوى بعد الدورة	55	85	70	75	98	72	75	58	72	70

المطلوب

١- اختبر الفرض القائل بأن الدورة التدريبية أدت إلى رفع متوسط مستوى أداء العامل بالبنك وذلك بدرجة ثقة 95%.

٢- اختبر الفرض القائل بأنه لا يوجد تأثير فعلى للدورة على متوسط مستوى اداء العامل بالبنك وذلك عند درجة ثقة 95%.

الحل

- نكون الجدول التالي

x_j	y_j	d_j	$(d_j - d)$	$(d_j - d)^2$
50	55	-5	-5-(-3)=-2	4

80	**85**	**-5**	**-5-(-3)=-2**	**4**
67	**70**	**-3**	**-3-(-3)=0**	**0**
73	**75**	**-2**	**-2-(-3)=1**	**1**
95	**98**	**-3**	**-3-(-3)=0**	**0**
65	**73**	**-7**	**-7-(-3)=-4**	**16**
77	**75**	**2**	**2-(-3) =5**	**25**
53	**58**	**-5**	**-5-(-3)=-2**	**4**
79	**72**	**7**	**7-(-3) =-1**	**100**
61	**70**	**-9**	**-9-(-3)=-6**	**36**
		$\sum d_j = -30$		$\sum (d_j - d)^2 = 190$

$$\bar{d} = \frac{\sum_{j=1}^{10} d_j}{n} = \frac{-30}{10} = -3$$

$$S_d^2 = \frac{\sum_{j=1}^{10} (d_j - d)^2}{n-1} = \frac{190}{10-1} = \frac{190}{9} = 21.11$$

$$S_d = \sqrt{21.11} = 4.5$$

(١) ١- الفرض العدمي

$$H_0 : (\mu_1 - \mu_2) = 0$$

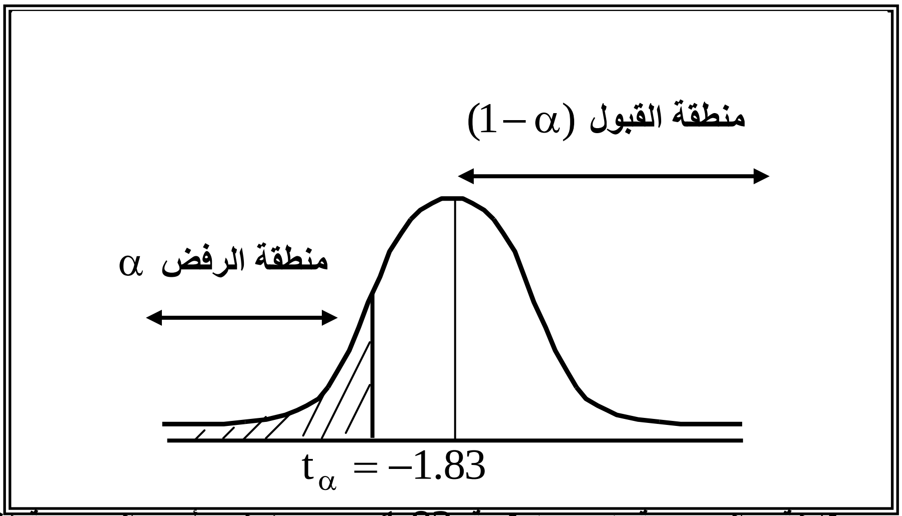
الفرض البديل

$$H_1 : (\mu_1 - \mu_2) < 0$$

٢- نحسب المقياس t حيث

$$t = \frac{\sqrt{n}(\bar{d} - h)}{S_d} = \frac{\sqrt{10}(-3-0)}{4.5} = \frac{-9.49}{4.5} = -2.11$$

عند درجة الثقة 95% وباستخدام جدول توزيع t عند درجات الحرية 9=(n-1)
نجد أن حد الثقة يساوي1.83-



شكل (٥-٢٩)

وبمقارنة t المحسوبة (11 [illegible] t_{α} = [illegible] نجد أن t المحسوبة تقع في منطقة الرفض وبالتالي نرفض الفرض العدمي ونقبل الفرض البديل القائل بأن متوسط الأداء بعد التدريب أكبر من قبل التدريب وهذا يعني أن للدورة تأثير فعال على مستوى أداء العاملين وذلك بدرجة ثقة 95%.

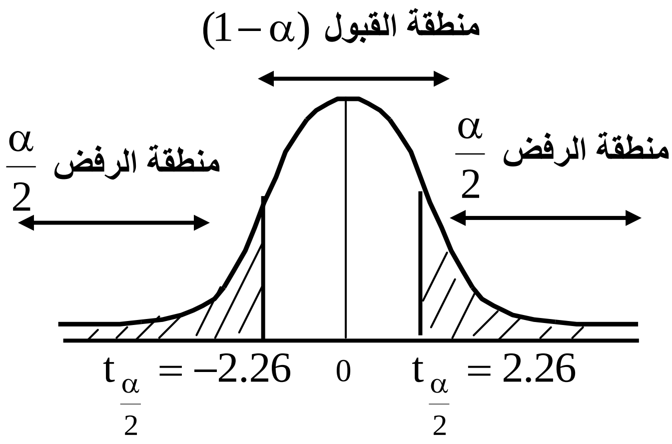
(٢) ١- الفرض العدمي

$$H_0 : (\mu_1 - \mu_0) = 0$$

الفرض البديل

$$H_1 : (\mu_1 - \mu_0) \neq 0$$

عند درجة ثقة 95% نجد أن $t_{\alpha/2} = \pm 2.26$



شكل (٥-٣٠)

وبما أن t المحسوبة حيث

t=-2.11

أي تقع في منطقة القبول وبالتالي نقبل الفرض العدمي القائل بأنه لا يوجد تأثير فعلي للدورة على متوسط أداء العامل بالبنك

مثال(٥-١١)

إذا كانت إنتاجية الفدان من أحد المحاصيل متغير يتبع التوزيع المعتاد القياسي فإذا أجريت دراسة على احد الأسمدة على إنتاجية الفدان من أحد المحاصيل الزراعية فأخذت عينة مكونة من 100 فدان وحسب إنتاجية الفدان في العينة ثم حسب إنتاجية الفدان في العينة بعد إضافة السماد فوجد أن متوسط الفرق بين إنتاجية الفدان بعد إضافة السماد وإنتاجية الفدان قبل إضافة السماد يساوي d حيث d=6 طن كذلك تباين $S_d^2 = 20.25$.

المطلوب

اختبر الفرض القائل بأن إضافة السماد أدى إلى زيادة متوسط إنتاجية الفدان بـ5 طن على الأقل وذلك بدرجة ثقة 95%.

الحل

إذا أشرنا إلى متوسط إنتاجية الفدان بعد إضافة السماد وقبل الإضافة بالرمز μ_1, μ_2 **على الترتيب. فإن**

١- **الفرض العدمي**

$$H_0 : (\mu_1 - \mu_2) = 5$$

الفرض البديل

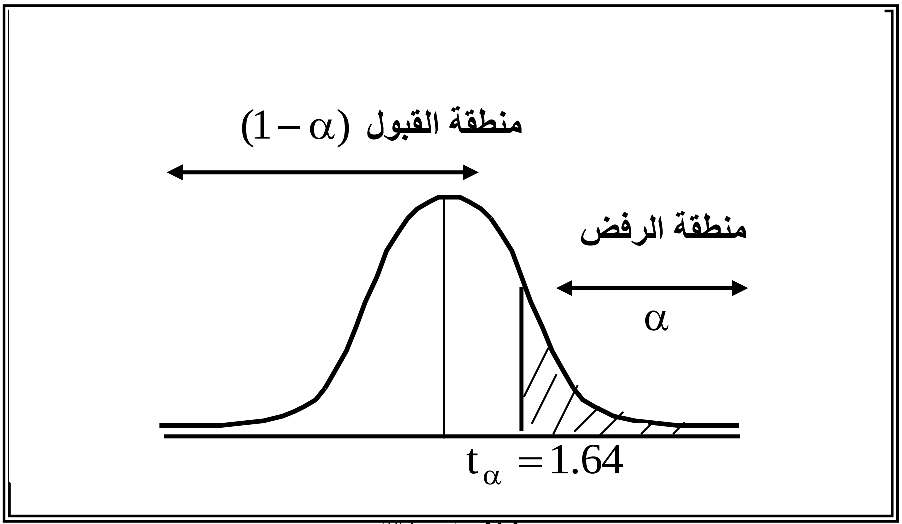
$$H_1 : (\mu_1 - \mu_2) > 5$$

٢- **نحسب المقياس t حيث:-**

$$t = \frac{\sqrt{n}(\bar{d} - h)}{S_d} = \frac{\sqrt{10}(6-5)}{4.5} = 2.22$$

٣- **وبما أن حجم العينة n=100 أي تعتبر عينة كبير بالتالي نستخدم جدول التوزيع المعتاد القياسي بملحق (٤) لتحديد حد الثقة عند درجة الثقة 95% ومن جدول التوزيع المعتاد القياسي نجد أن حد الثقة يساوي** $t_\alpha = 1.64$

كما هو موضح بالشكل التالي:-



شكل (٥-٣١)

٤- بمقارنة قيمة t=2.22 بقيمة $t_{\alpha} = 1.64$ نجد أن t المحسوبة تقع في منطقة الرفض إذن نرفض الفرض العدمي ونقبل الفرض البديل القائل بأن إضافة السماد يؤدي إلى زيادة إنتاجية الفدان الواحد بأكثر من 5 طن وذلك بدرجة ثقة 95%.

(٥-٧) تطبيقات
Applications

تطبيق (٥-١): **يقوم أحد مصانع الحديد والصلب بجمهورية مصر العربية بإنتاج شرائح الحديد التي ينتج منها هياكل السيارات ، فإذا كان متوسط السمك للشريحة يجب ألا يزيد أو يقل عن 0.10 بوصة (حيث تؤدي الزيادة عن 0.10 إلى زيادة التكاليف ويؤدي النقص عن 0.10 إلى خلل في التصميم) ، ولاختبار جودة المنتج تم فحص عينة مكونة من 100 شريحة وبفحصها وجد أن متوسط السمك في الشريحة الواحدة في العينة يساوي 0.102 بوصة وانحراف معياري 0.10**

المطلوب

اختبر الفرض القائل بأن إنتاج الشرائح في هذا المصنع غير مطابق للمواصفات وذلك بدرجة ثقة 95%.

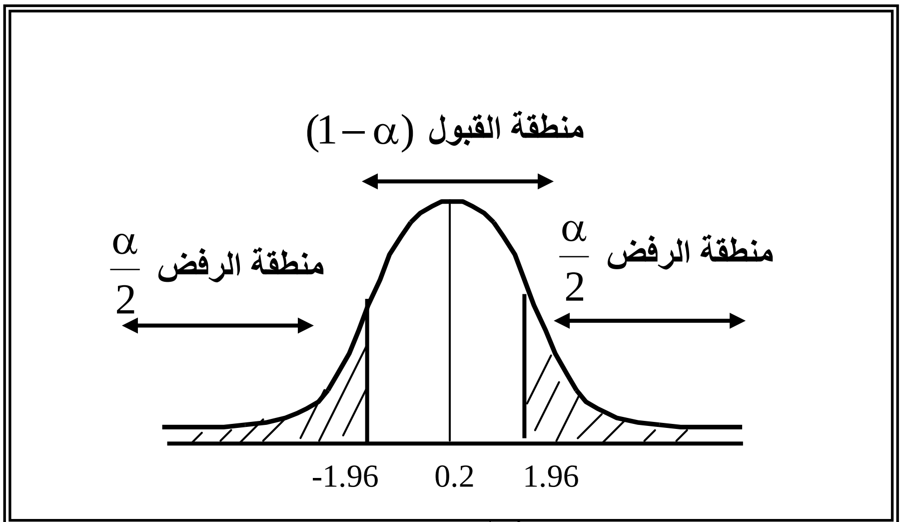
الحل

$$\Theta \; n = 100 \; , \; \overline{X} = 0.102 \; , \; S = 0.10$$

$$H_0 : \mu = 0.10$$
$$H_1 : \mu \neq 0.10$$

$$Y = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{0.102 - 0.10}{\frac{0.10}{\sqrt{100}}}$$
$$= \frac{0.002}{0.01} = 0.2$$

Θ درجة الثقة $95\% \leftarrow Y_{\alpha/2} = \pm 1.96$



شكل (٥-٣٢)

Θ Y تقع في منطقة القبول

$\therefore$ نقبل الفرض العدمي القائل بأن متوسط سمك الشريحة 0.10 بوصة

أي نرفض الفرض القائل بان سمك الشرائح من إنتاج المصنع يختلف عن 10 بوصة وذلك بدرجة ثقة95%.

تطبيق (٥-٢): **في إحدى المحافظات الزراعية كانت مساحة الأرض المزروعة قطن 50000 فدان ، فقامت وزارة الزراعة بعمل مسح لـ500 فدان من هذه المحافظة لمعرفة نسبة الإصابة بدودة القطن فوجد أن 15% من الـ 200 فدان مصابة.**

<u>المطلوب</u>

اختبر الفرض القائل بان نسبة الإصابة في هذه المحافظة تزيد عن 20% من المساحة المزروعة قطن وذلك بدرجة ثقة 95% . ثم قدر المساحة المتوقع إصابتها.

<u>الحل</u>

$$(1-\alpha) = 0.95, n = 200, \theta = 0.15$$

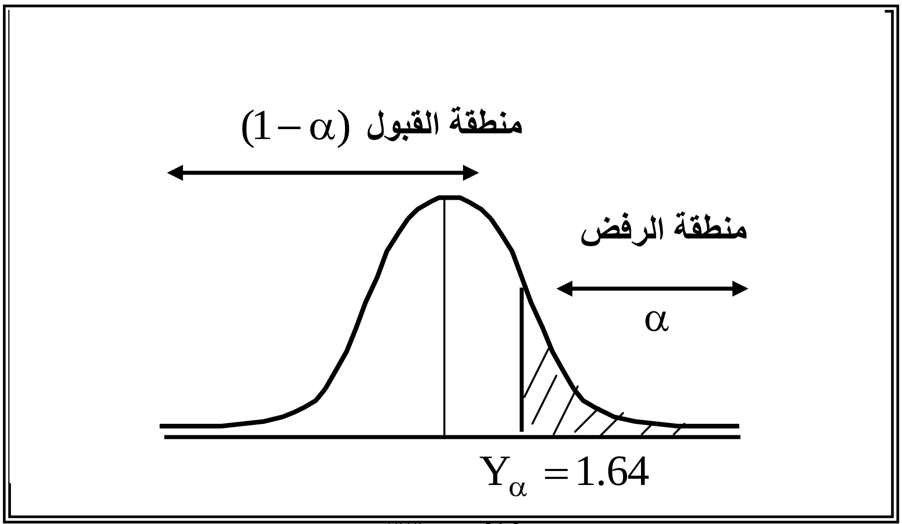
الفرض العدمي

$$H_0 : \theta = 0.20$$

الفرض البديل

$$H_1 : \theta > 0.20$$

$$Y = \frac{\bar{\theta} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} = \frac{0.15 - 0.20}{\sqrt{\frac{0.2(0.8)}{200}}} = -1.77$$



شكل (٥-٣٣)

$\Theta\ Y < Y_{\alpha}$

أي أن Y تقع في منطقة القبول

∴ نقبل الفرض العدمي القائل بان نسبة الإصابة تساوي 20% ،

أي نرفض الفرض البديل القائل بأن نسبة الإصابة في هذه المحافظة تزيد عن 20%

وذلك بدرجة ثقة 95%.

ولإيجاد تقدير المساحة المتوقع إصابتها فإن النسبة في المجتمع θ وعند درجة ثقة 95% فإن $Y_{\alpha/2} = \pm 1.96$

$$(\bar{\theta} - Y_{\alpha/2}\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}} \le \theta \le \bar{\theta} + Y_{\alpha/2}\sqrt{\frac{\bar{\theta}(1-\bar{\theta})}{n}})$$

$$(0.15 - 1.96\sqrt{\frac{0.15(0.85)}{200}} \le \theta \le 0.15 + 1.96\sqrt{\frac{0.15(0.85)}{200}})$$

$$0.10 \le \theta \le 0.20$$

$\because$ المساحة المزروعة قطن في هذه المحافظة تساوي 50000 فدان

$\therefore$ أقل مساحة ممكن أن تكون مصابة

$= 0.10 \times 50000 = 5000$ فدان

وأقصى مساحة

$= 0.20 \times 50000 = 10000$ فدان

أي أن المساحة المتوقع إصابتها تتراوح بين 10000,5000 فدان.

تطبيق (٥-٣): **في إحدى الدراسات عن متوسط درجة الطالب في مادة الإحصاء في كليات التجارة بجمهورية مصر العربية في سنة 1995 وجد أن 20000 طالب يدرسون الإحصاء فإذا أخذت عينة عشوائية حجمها 20 طالب فكان متوسط درجة الطالب في العينة 12.5 درجة بانحراف معياري 3 درجات.**

المطلوب

١- **اختبر الفرض القائل بان متوسط درجة الطالب في كليات التجارة في مادة الإحصاء اقل من 11 درجة وذلك بدرجة ثقة 95%.**

٢- **اختبر الفرض القائل بأن متوسط درجة الطالب اكبر من 12 درجة بدرجة ثقة 95%.**

الحل

$$\bar{X} = 12.5, S = 3, n = 20, (1-\alpha) = 0.95$$

١- الفرض العدمي

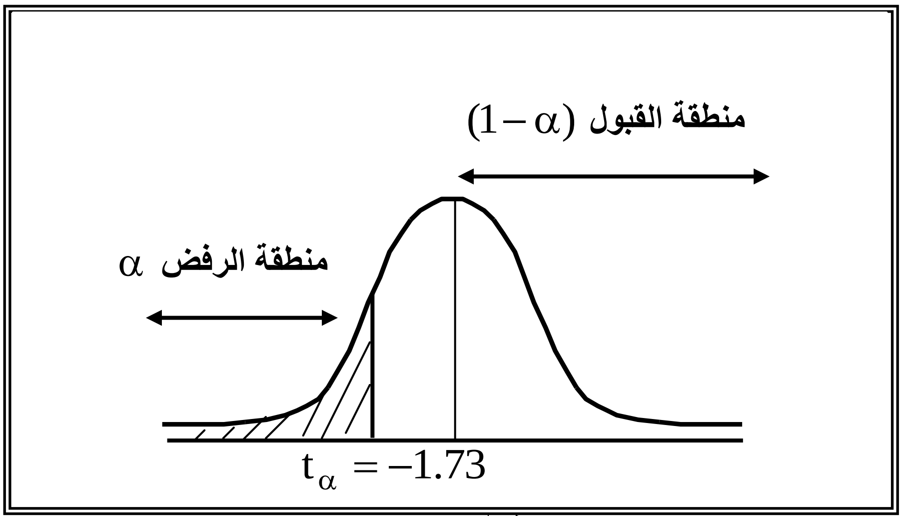
$$H : \mu = 11$$

الفرض البديل

$$H_1 : \mu < 11$$

$$t_{\alpha} = -1.73.$$

$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{12.5 - 11}{\frac{3}{\sqrt{20}}} = 2.23$$



شكل (٥-٣٤)

$\Theta\ t > t_{\alpha}$

∴ نقبل الفرض العدمي القائل بأن متوسط الدرجة تساوي 11 درجة

أي نرفض الفرض القائل بأن متوسط درجة الطالب في مادة الإحصاء في كليات التجارة

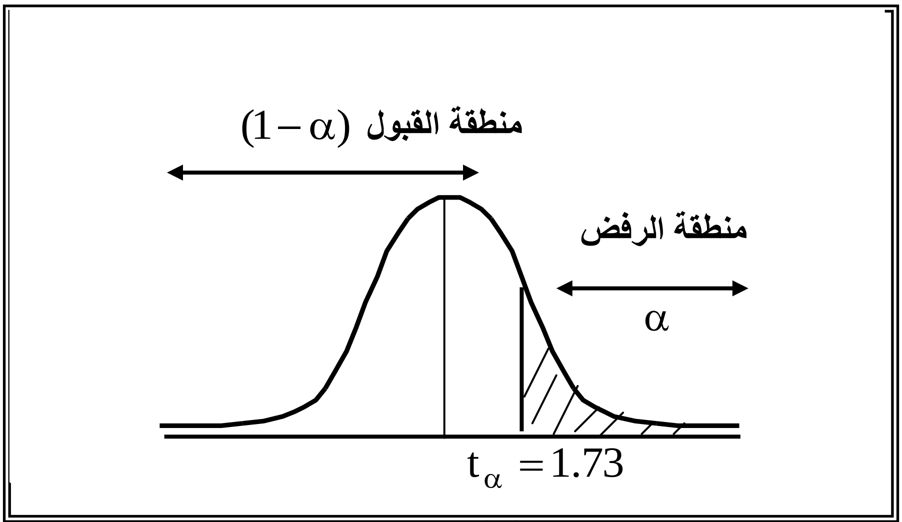
أقل من 11 درجة وذلك بدرجة ثقة 95%.

٢- الفرض العدمي

$$H_0 : \mu = 12$$

الفرض البديل

$$H_1 : \mu > 12$$



شكل (٥-٣٥)

$$t_\alpha = 1.73$$

$$t = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} = \frac{12.5 - 12}{\frac{3}{\sqrt{20}}} = 0.75$$

∴ نقبل الفرض العدمي القائل بأن متوسط الدرجة هو 12 درجة ونرفض الفرض البديل القائل بأن متوسط الدرجة أكبر من 12 وذلك بدرجة ثقة 95%.

تطبيق (٥-٤): **تقوم إحدى شركات إنتاج المنظفات الصناعية بإنتاج 4 أنواع من مختلفة من المنظفات A,B,C.D ، فإذا أخذت عينة من 400 مستهلك من 10000 مستهلك من مستهلكي منتجات الشركة فوجد أن 310 منهم يفضلون النوع B.**

المطلوب

١- **اختبر الفرض القائل بأن 80% من المستهلكين يفضلون النوع B بدرجة ثقة 80%.**

٢- **اختبر الفرض القائل بأن 50% من المستهلكين يفضلون استخدام الأنواع A,C,D بدرجة ثقة 90%.**

٣- اختبر الفرض القائل بأن 70% على الأقل يفضلون النوع B بدرجة ثقة 85%.

الحل

إذا فرضنا أن θ_1 هي نسبة المستهلكين الذين يفضلون النوع B ، θ_2 هي نسبة المستهلكين الذين يفضلون باقي الأنواع (A,C,D) كذلك θ_1, θ_2 هما النسبتين السابقتين في العينة على الترتيب

$$\theta_1 = \frac{310}{400} = 0.775$$

$$\theta_2 = 1 - 0.775 = 0.225$$

١- الفرض العدمي

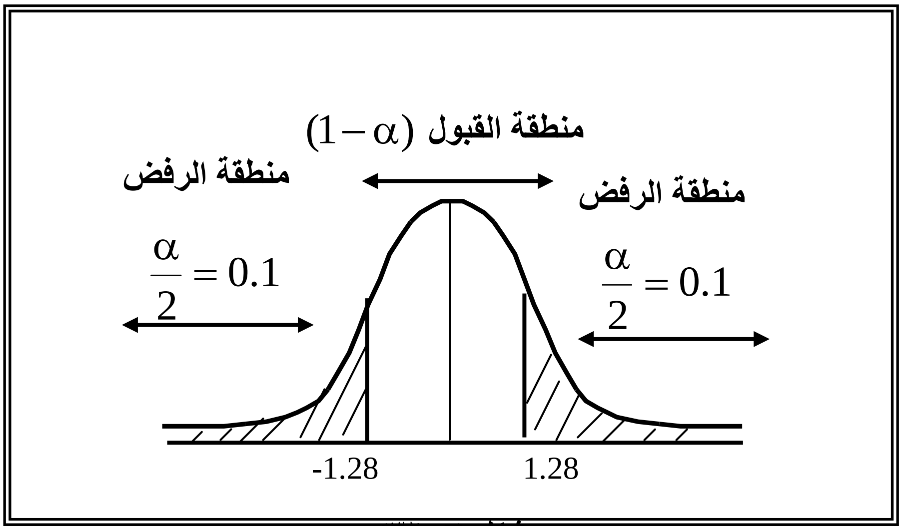
$$H_0 : \theta_1 = 0.80$$

الفرض البديل

$$H_1 : \theta_1 \neq 0.80$$

$$\Theta\ (1 - \alpha) = 0.80$$

$$Y_{\alpha/2} = \pm 1.28$$



شكل (٥-٣٦)

$$Y=\frac{0.775-0.80}{\sqrt{\frac{0.80(0.20)}{400}}}=-1.25$$

∴ نقبل الفرض العدمي القائل بأن نسبة الذين يفضلون النوع B هي 80% وذلك بدرجة ثقة 80%.

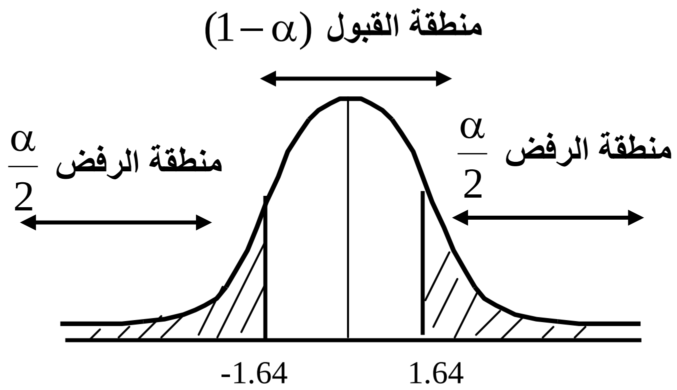
٢- الفرض العدمي

$$H_0:\theta_2=0.50$$

الفرض البديل

$$H_1:\theta_2\neq 0.50$$

$$Y=\frac{0.775-0.500}{\sqrt{\frac{0.50(0.50)}{0.025}}}=11$$



شكل (٥-٣٧)

$Y\Theta$ تقع في منطقة الرفض

$\therefore$ نقبل أن نسبة المستهلكين الذين يفضلون باقي الأنواع (A,C,D) لا تساوي 50%

وذلك بدرجة ثقة 90%.

٣- الفرض العدمي

$$H_0 : \theta_1 = 0.70$$

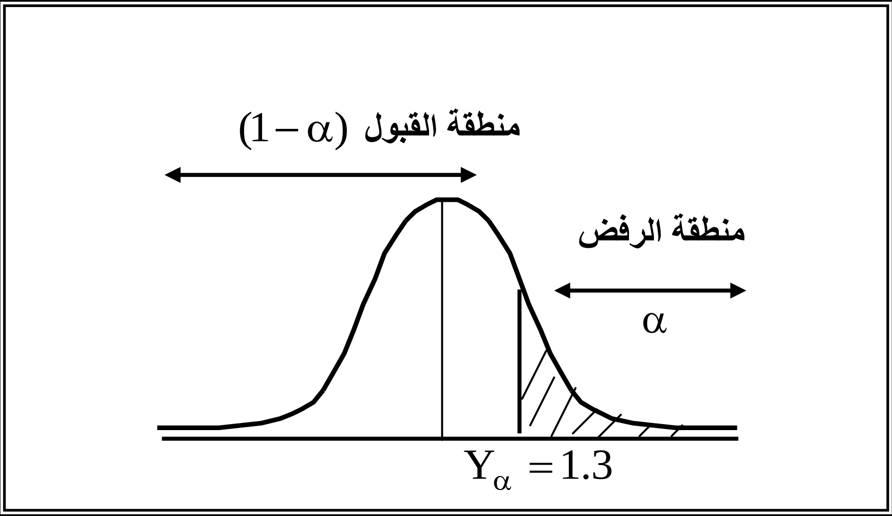
الفرض البديل

$$H_1 : \theta_1 > 0.70$$

$$\Theta\ (1-\alpha) = 0.85$$

$$Y_\alpha = 1.03$$

$$Y = \frac{0.775 - 0.700}{\sqrt{\frac{0.7(0.3)}{400}}} = 3.27$$



شكل (٥-٣٨)

$\therefore$ نرفض الفرض القائل بأن نسبة 70% من المستهلكين يفضلون النوع **B** أي نقبل الفرض القائل بأن نسبة 70% على الأقل يفضلون النوع **B** وذلك بدرجة ثقة 85%.

تطبيق (٥-٥): في أحد مراكز تعليم غير المبصرين (المكفوفين) يوجد نظامين **A,B** لتعليم القراءة فإذا كان التباين لمستوى القراءة بالنسبة للدارسين عن طريق النظام **A** يساوي **1.44** فإذا أخذت عينة مكونة من **21** دارس ممن يدرسون عن طريق النظام **B** فوجد تباين مستوى القراءة في العينة **1.05**

المطلوب: اختبر الفرض القائل بأن التدريس باستخدام النظام **B** أفضل من التدريس باستخدام النظام **A** حيث أن التشتت باستخدام النظام **B** أقل منه باستخدام النظام **A** وذلك بدرجة ثقة 95%.

الحل

الهدف من الاختبار معرفة هل التباين في المجتمع بالنسبة للدراسة عن طريق النظام **B** أقل منه عن طريق النظام **A** وبالتالي:-

١- الفرض العدمي

$$H_0 : \sigma^2 = 1.44$$

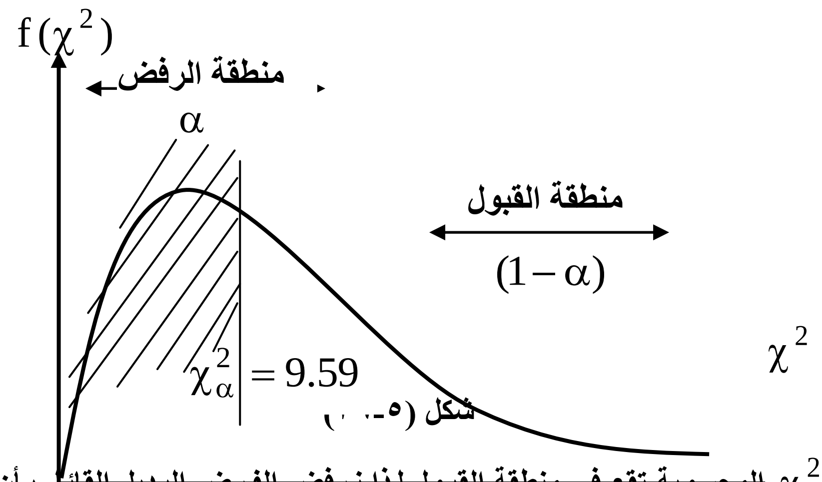
الفرض البديل

$$H_1 : \sigma^2 < 1.44$$

٢- المقياس χ^2 حيث:-

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(21-1)1.05}{1.44} = 14.58$$

عند درجة الثقة 95% فإن $\chi^2 = 9.59$



شكل (٥-

Θ χ^2 **المحسوبة تقع في منطقة القبول لذا نرفض الفرض البديل القائل بأن تباين مستوى القراءة باستخدام النظام B أقل من تباين مستوى القراءة باستخدام النظام A وذلك بدرجة ثقة 95%.**

(٥-٨) تمرينات Exercises

(٥-١): في إحدى الدراسات عن متوسط درجة الطالب في مادة الإحصاء في كليات التجارة بجمهورية مصر العربية في سنة 1995 وجد أن 20000 طالب يدرسون مادة الإحصاء فإذا أخذت عينة عشوائية حجمها 20 طالب فكان متوسط درجة الطالب في العينة 12.5 درجة بانحراف معياري 3 درجات.

<u>المطلوب</u>

١. **اختبر فترة الثقة لمتوسط درجة الطالب في كليات التجارة في هذه السنة بدرجة ثقة 99%.**

٢. **اختبر الفرض القائل بأن متوسط درجة الطالب في كليات التجارة أقل من ١١ درجة.**

(٥-٢): في التمرين السابق قدر حجم العينة المطلوب سحبها لتقدير متوسط الدرجة في الكليات بحيث لا يزيد الحد الأعلى لخطأ المعاينة عن 1.5 درجة.

(٥-٣): تقوم إحدى شركات إنتاج المنظفات الصناعية بإنتاج 4 أنواع من مختلفة من المنظفات A,B,C.D ، فإذا أخذت عينة من 400 مستهلك من 10000 مستهلك من مستهلكي منتجات الشركة فوجد أن 310 منهم يفضلون النوع B.

<u>المطلوب</u>

١- **قدر بدرجة ثقة 95% نسبة المستهلكين الذين يفضلون النوع B.**

٢- **قدر بدرجة ثقة 99% نسبة المستهلكين الذين يفضلوا الأنواع A,C,D.**

٣- **قدر عدد العملاء الذين يفضلون النوع B عند درجة ثقة 95%.**

(٥-٤): إذا كان سعر الوحدة الواحدة من سلعة ما يتغير من محافظة إلى أخرى داخل الجمهورية في فترة زمنية معينة ، وبافتراض أن السعر متغير يتبع التوزيع المعتاد

بتوقع μ وتباين σ^2 فإذا أخذت عينة من الأسعار لهذه السلعة من 16 محافظة فكانت على النحو التالي:-

74, 98 , 94 , 55 , 66 , 55 , 78 , 71 , 53 , 90 , 53 , 85 , 74 , 74 , 58

المطلوب

١- قدر التباين لسعر الوحدة من هذه السلعة داخل الجمهورية بدرجة ثقة 95%.

٢- بافتراض أن سعر الوحدة المحدد من وزارة التموين هو 70، ارسم خريطة مراقبة الأسعار بدرجة ثقة 95%.

(٥-٥): في دراسة عن سبب فشل بعض الطلاب في المرحلة الجامعية أخذت عينة من 20 طالب من الذين فشلوا في اجتياز هذه المرحلة من 1500 طالب لم يجتازوا هذه المرحلة، فإذا وجد أن 140 طالب من طلاب العينة يرجع سبب فشلهم إلى أسباب أسرية.

المطلوب

قدر نسبة الطلاب الذين لم يجتازوا المرحلة الجامعية لأسباب أسرية بدرجة ثقة 95%.

(٥-٦): سحبت عينتين حجمهما $n_1 = 15, n_2 = 20$ من مجتمعين معتادين بتباينان $\sigma_1^2 = 9, \sigma_2^2 = 15$ على الترتيب ، فإذا كان متوسط العينة الأولى يساوي $\overline{X}_1 = 18$ ومتوسط العينة الثانية $\overline{X}_2 = 20$.

المطلوب

١- اختبر الفرض القائل بأن الفرق بين متوسطي المجتمعين المسحوب منهما العينتين $\mu_1 - \mu_2$ يساوي 4 بدرجة ثقة 90%,

٢- اختبر الفرض القائل بأنه لا يوجد فرق بين متوسطي المجتمعين المسحوب منهما العينتين بدرجة ثقة 95%.

٣- اختبر الفرض القائل بأن متوسط المجتمع الأول يقل على الأقل بمقدار 4 عن متوسط المجتمع الثاني بدرجة ثقة 95%.

(٥-٧)

في أحد القطاعات الإنتاجية يرغب متخذ القرار في معرفة تأثير تغيير نظام الإجازة الأسبوعية من يوم واحد إلى يومين على عدد أيام العينات خلال العام للعاملين بهذا القطاع فأخذت عينة مكونة من 10 عاملين والجدول التالي يوضح عدد أيام العينات السنوي للعامل في عامين متتاليين في العام الأول وفقاً للنظام القديم(الإجازة الأسبوعية يوم واحد) وفي العام الثاني وفقاً للنظام الجديد (الإجازة الأسبوعية يومين):-

رقم العامل	1	2	3	4	5	6	7	8	9	10
متوسط عدد أيام الغياب للعامل وفقاً للنظام القديم	10	7	11	13	9	15	20	5	4	16
متوسط عدد أيام الغياب للعامل وفقاً للنظام الجديد	3	2	5	5	3	5	8	0	0	4

بافتراض أن عدد أيام الغياب السنوية للعامل يمثل متغير معتاد عند درجة ثقة 95%

١- اختبر الفرض القائل بأن متوسط عدد أيام الغياب للعامل وفقاً للنظام الجديد أقل من متوسط عدد أيام الغياب وفقاً للنظام الجديد.

٢- اختبر الفرض القائل بأن متوسط عدد أيام الغياب للعامل وفقاً للنظام القديم أقل من متوسط عدد أيام الغياب للعامل وفقاً للنظام الجديد.

٣- اختبر الفرض القائل بأنه لا يوجد تأثير لتغير نظام الإجازة الأسبوعية على متوسط عدد أيام الغياب السنوية للعامل بهذا القطاع.

(٥-٨): أخذت عينة من درجات طلاب الفرقة الثالثة بشعبة المحاسبة في إحدى السنوات الدراسية بإحدى كليات التجارة في مادة الإحصاء فكانت على النحو الموضح بالجدول التالي:-

رقم الطالب	1	2	3	4	5	6	7	8
درجة الطالب في مادة الإحصاء	85	98	57	78	82	75	94	81

كذلك أخذت عينة من درجات طلاب الفرقة الثالثة بشعبة الإدارة في نفس السنة ونفس الكلية في مادة الإحصاء فكانت على النحو التالي:-

رقم الطالب	1	2	3	4	5	6	7	8	9	10
درجة الطالب في مادة الإحصاء	68	52	91	59	64	56	78	62	51	54

بافتراض أن درجة الطالب في مادة الإحصاء تتبع التوزيع المعتاد (الطبيعي) ، عند درجة ثقة 90%

١- اختبر الفرض القائل بأن متوسط درجة الطالب في الإحصاء في شعبة المحاسبة لا يختلف عن متوسط درجة الطالب في الإحصاء في شعبة الإدارة.

٢- اختبر الفرض القائل بأن متوسط درجة الطالب في الإحصاء في شعبة المحاسبة أكبر من متوسط درجة الطالب في الإحصاء بشعبة الإدارة.

٣- اختبر الفرض القائل بأن متوسط درجة الطالب في الإحصاء في شعبة المحاسبة يزيد عن متوسط درجة الطالب في الإحصاء بشعبة الإدارة بـ 5 درجات على الأقل.

٤- اختبر الفرض القائل بأن متوسط درجة الطالب في الإحصاء في شعبة المحاسبة يزيد عن متوسط درجة الطالب في الإحصاء بشعبة الإدارة بـ 8 درجات على الأكثر.

(٥-٩): إذا كانت مدة انتظار العميل بأحد البنوك التجارية حتى يتم إنجاز الخدمة له يمثل متغير عشوائي يتبع التوزيع المعتاد (الطبيعي) وأجريت بعض التعديلات على نظام العمل بالبنك بالنسبة للعملاء ونظراً لأهمية تصغير مدة انتظار العميل قامت إدارة

البنك بأخذ عينـة عشوائية مـن العمـلاء المستديمين بالبنـك لمعرفـة مـدة انتظـار العميـل حتى يتم إنجاز خدمته (بالدقائق) قبل إجراء التعديل وبعد إجراء التعديل كما هو موضح بالجدول التالي:-

رقم العميل	1	2	3	4	5	6	7	8
زمن الانتظار بالدقائق قبل التعديل	7	20	23	10	25	24	11	30
زمن الانتظار بالدقائق بعد التعديل	9	15	10	5	10	10	3	17

المطلوب

١- اختبر الفرض القائل بأنـه لا يوجد تـأثير للتعديل علـى متوسط زمـن انتظـار العميل بدرجة ثقة 95%.

٢- اختبر الفرض القائل بأن التعديل أدى إلى تصغير متوسط زمـن انتظـار العميل (أي يوجد أثر إيجابي للتعديل) بدرجة ثقة 90%.

٣- اختبر الفرض القائل بأنـه يوجد أثر سلبي للتعديل علـى متوسط مـدة انتظـار العميل بدرجة ثقة 95%.

الباب السادس
الاختبارات اللامعلمية
Nonparametric Tests

(٦-١) استخدام الاختبارات اللامعلمية
Using Nonparametric Test

(٦-٢) اختبارات الإشارة للوسيط
Sign Test of A Median

(٦-٣) اختبار ويل كاكسون لمجموع الرتب
Wilcoxon Rank-Sum Test

(٦-٤) اختبار ويل كاكسون للرتب ذات الإشارة
Wilcoxon Signed-Rank Test

(٦-٥) تمرينات
Exercises

(٦-١) استخدام الاختبارات اللامعلمية
Using Nonparametric Tests

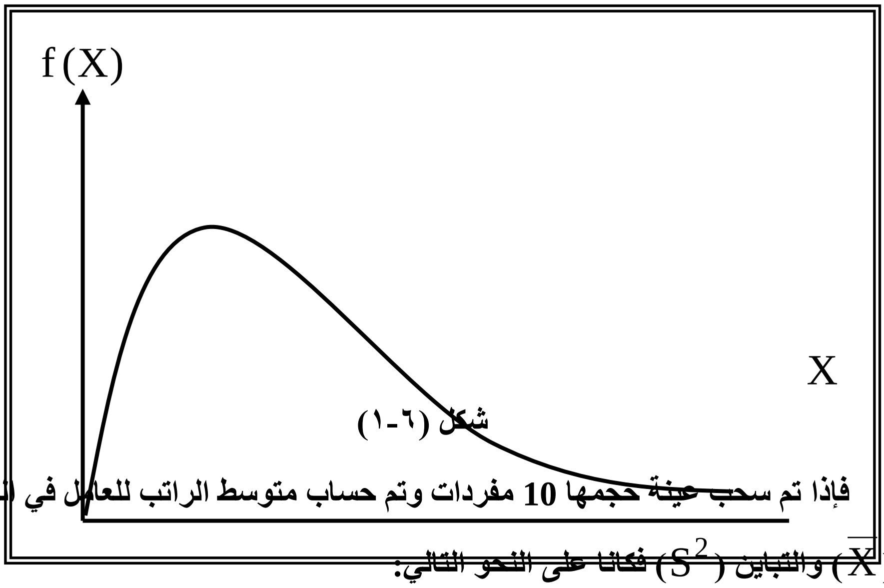
في الباب الرابع تناولنا بالدراسة التفصيلية بعض الاختبارات الإحصائية، ووجدنا أن جميع هذه الاختبارات تخضع لبعض الشروط التي يجب توافرها حتي يمكن إجراء الاختبار.

فبالنسبة للاختبارات عن قيم بعض معلمات المجتمع مثل المتوسط μ أو النسبة θ أو التباين σ^2 باستخدام عينة عشوائية مسحوبة من مجتمع يتبع التوزيع المعتاد أو يؤول إلى التوزيع المعتاد (وللاختصار يقال مجتمع معتاد). ولكن في معظم الدراسات التطبيقية لا يتوافر هذا الشرط – بمعني أن المتغير محل الدراسة في المجتمع لا يتبع التوزيع المعتاد أو يؤول إليه. بل قد يكون غير معروف التوزيع الاحتمالي.

وفي هذه الحالة أي عندما لا يتوافر شرط أن يكون المتغير محل الدراسة يتبع التوزيع المعتاد فإنه لا يمكن إجراء الاختبارات التي سبق تناولها في الفصول (٥-٢) – (٥-٦).

مثال (٦-١)

إذا كان التوزيع الاحتمالي للراتب الشهري (X) لعاملين بإحدى المؤسسات كما هو موضح بالشكل التالي.



شكل (٦-١)

فإذا تم سحب عينة حجمها **10** مفردات وتم حساب متوسط الراتب للعامل في العينـة ($\overline{X}$) والتباين (S^2) فكانا على النحو التالي:

$$\overline{X} = 500 \text{ جنية} \qquad S^2 = 25$$

والمطلوب: إجراء اختبار عن متوسط المجتمع (μ) وتباين المجتمع (σ^2) للراتـب الشهري في المؤسسة.

في هذه الحالة لا يمكن استخدام اختبار المتوسط في المجتمع المقدم في الفصل (٥-٢)أو اختبار χ^2 للتبـاين فـي المجتمـع المقدم فـي الفصـل (٥-٣). وذلـك يرجـع إلـى أن الراتـب الشـهري (X) للعـاملين فـي المؤسسـة لا يتبـع التوزيـع المعتـاد كمـا هـو موضـح بشـكل (٦-١). ومـن الشـكل يتضـح أيضـاً أن التوزيـع لا يـؤول إلـى المعتـاد فهـو ملتـوي ناحية اليمين.

مثال (٦-٢)

إذا أخذت عينة من المرتبات الشهرية للعاملين بقطاع الأعمال فكان المتوسط للراتب الشهري للعامل يساوي 900 جنية، وأخذت عينة من المرتبات الشهرية للعاملين بالقطاع الحكومي فكان متوسط الراتب الشهري للعامل بالحكومة 400 جنية. فإذا كان المطلوب إجراء بعض الاختبارات عن الفرق بين متوسط الراتب الشهري للعامل بقطاع الأعمال ومتوسط الرتب الشهري للعامل بالحكومة – حيث أن التوزيع الاحتمالي للراتب الشهري للعاملين بقطاع الأعمال وكذلك للقطاع الحكومي غير معروفين.

في هذه الحالة لا يمكن استخدام اختبار للفرق بين متوسطي مجتمعين $(\mu_1 - \mu_2)$ الذي سبق تناوله في الفصل (٥-٥) وذلك يرجع إلى أن استخدام الاختبار يتطلب أن يكون التوزيع الاحتمالي لكل من الراتب الشهري للعاملين بقطاع الأعمال يتبع التوزيع المعتاد، كذلك الراتب الشهري للعاملين بالقطاع الحكومي يتبع التوزيع المعتاد أيضاً. وفي هذا المثال لا يتوافر هذا الشرط.

مما سبق يتضح أنه في حالة عدم توافر شروط إجراء الاختبارات المعلمية التي سبق تناولها في الباب الخامس فإننا نستخدم نوع أخر من الاختبارات التي تسمي بالاختبارات اللامعلمية Nonparametric Tests. والاختبارات اللامعلمية سواء لمعلمات مجتمع واحد أو عدة مجتمعات لا تتطلب:

١- وجود أي فروض على التوزيع الاحتمالي للمتغير محل الدراسة في المجتمع (أو المجتمعات) المسحوب منه (أو منها) العينة (أو العينات).

٢- معرفة التوزيع الاحتمالي للمتغير محل الدراسة في المجتمع (أو المجتمعات) المسحوب منه العينة (أو العينات).

٣- أن تكون البيانات بيانات كمية كما في الاختبارات المعلمية ولكن ممكن أن تكون البيانات كمية أو بيانات نوعية ترتيبية **Ordinal** كما سوف يتضح في الاختبارات التي سوف نقوم بعرضها في الفصول التالية.

ورغم المزايا المذكورة أعلاه للاختبارات اللامعلمية **Nonparametric Tests** إلا إننا لا نستخدمها إلا في حالة تعذر استخدام الاختبارات المعلمية **Parametric Tests** وذلك للأسباب التالية:

١- بما أن قوة الاختبار (أنظر الفصل (٥-٢)) هي احتمال رفض الفرض العدمي عندما يكون خاطئ وقبول الفرض البديل عندما يكون صحيح $(1-\beta)$ أي احتمال اتخاذ قرار سليم – ونظراً لأن الاختبار اللامعلمي يفترض عدم وجود معلومات عن التوزيع الاحتمالي للمتغير محل الدراسة في المجتمع – وبالتالي فإنه عادة يكون الاختبار اللامعلمي اقل قوة من الاختبار المعلمي.

٢- نظراً لأن الاختبار اللامعلمي يعتمد على التوزيع الحر **Free Distribution** لتوزيع المتغير محل الدراسة في المجتمع (حيث لا يوجد أي فروض على توزيع المتغير في المجتمع) ويترتب على ذلك أن الاختبار اللامعلمي أقل حساسية من الاختبار المعلمي لمعلمات توزيع المتغير محل الدراسة في المجتمع ويترتب على ذلك أن الاختبار اللامعلمي أقل حساسية للخطأ في الفروض من الاختبار المعلمي.

ومما سبق يتضح أنه إذا توافر لنا إمكانية استخدام اختبار معلمي واختبار لا معلمي فإننا يجب استخدام الاختبار المعلمي وذلك للأسباب المذكورة أعلاه.

(٦-٢) اختبار الإشارة للوسيط
Sign Test of A Mediam

في كثير من المشاكل الفعلية يرغب متخذ القرار في اتخاذ قرار بشأن القيمة المتوسطة (μ) لمتغير معين وبالتالي اختبار القيمة المتوسطة. ولكن لا يتوافر لديه معلومات وبيانات دقيقة عن التوزيع الاحتمالي للمتغير في المجتمع محل الدراسة أو قد يكون التوزيع الاحتمالي للمتغير توزيع ملتوي ولا يؤول إلى التوزيع المعتاد أو قد تكون البيانات المتاحة بيانات وصفية وليست كمية في كل هذه الحالات لا يمكن إجراء الاختبارات المعلمية بشأن الوسط الحسابي (المتوسط) التي سبق تناولها في الباب الخامس. ولكن يمكن استخدام مقياس أخر من مقاييس النزعة المركزية وليكون الوسيط* كبديل للوسط الحسابي في المجتمع. حيث يمكن إجراء اختبار لا معلمي **Nonparametric Tests** عن الوسيط كما سوف نوضح فيما يلي.

ويتصف الوسيط بعدة خصائص ومن أهمها أنه مهما اختلف التوزيع الاحتمالي للمتغير محل الدراسة في المجتمع فإن احتمال أن تزيد قيم المفردات عن الوسيط يساوي 0.5 وبالتالي احتمال أن تقل قيم المفردات الوسيط يساوي 0.5 أيضاً. وبعبارة أخرى إذا أشرنا للوسيط بالرمز (M) والرمز (X) للمتغير محل الدراسة فإن:

$$\sum_{X\geq M}^{\infty} P_r(X) = \sum_{X=-\infty}^{M} P_r(X) = 0.5 \qquad \textbf{(6.1)}$$

حيث $P_r(X)$ هي دالة الاحتمال للمتغير المتقطع X.

أما إذا كان المتغير X متغير متصل فإن:

---

* الوسيط هو القيمة التي 50% من المفردات تأخذ قيم أقل منها (إذا كان عدد المفردات زوجي) أو تساويها (إذا كان عدد المفردات فردي). وبالتالي 50% من المفردات تأخذ قيم أكبر منها (إذا كان عدد المفردات زوجي) أو تساويها (إذا كان عدد المفردات فردي). لمزيد من المعرفة عن خصائص وطرق حساب الوسيط أنظر مرجع رقم [21].

$$\int_{M}^{\infty} f(X)d(X) = \int_{-\infty}^{M} f(X)d(X) = 0.5 \qquad \textbf{(6.2)}$$

حيث $f(X)$ هي دالة كثافة الاحتمال للمتغير X في هذه الحالة.

أولاً: اختبار الإشارة المبني على توزيع ذي الحدين

إذا كانت $X_1, X_2, X_3, ..., X_n$ تمثل مفردات عينة عشوائية بسيطة تم سحبها من المجتمع محل الدراسة (بغض النظر عن التوزيع الاحتمالي للمتغير في المجتمع) فإن الفرق بين كل قيمة من قيم مفردات العينة وقيمة الوسيط (M) يمثل متغيرات عشوائية يتبع كل منها توزيع برنولي بمعلمة (0.5) بعبارة أخرى فإن المتغير Y_j حيث:

$$\boxed{Y_j = X_j - M \quad , \quad j = 1,2,3,...N \qquad \textbf{(6.3)}}$$

يتبع توزيع برنولي حيث أن القيمة X_j إما أن تكون أكبر من (أو تساوي) M وبالتالي تكون القيمة Y_j قيمة موجبة (+) وذلك باحتمال يساوي 0.5، وإما أن تكون أقل من (أو تساوي) M وبالتالي تكون القيمة Y_j قيمة سالبة (-) وذلك باحتمال يساوي (1-P) أي 0.5 أيضاً.

فإذا أشرنا إلى عدد الفروق الموجبة في العينة بالرمز n(+)، وعدد الفروق السالبة في العينة بالرمز n(-). فإننا نجد ان المتغير n(+)* يتبع توزيع ذات الحدين بمعلمتين P, n (حيث n تساوي حجم العينة، و0.5 = P يعني أن الوسيط يساوي M). وبالتالي فإن دالة الاحتمال للمتغير n(+) على النحو التالي:

* ممكن استخدام المتغير n(-) بدلاً من n(+) وتتبع نفس الطريقة.

$$P_r\left(n(+)\right)=C^{n}_{n(+)}(0.5)^{n(+)}(0.5)^{n-n(+)}$$
$$,n(+)=0,1,2,...,n \quad (6.4)$$

ملحوظة:

١- بما ان $P = 0.5$ فتجد أن التوزيع في العلاقة (6.4) متماثل عند القيمة المتوقعة والتي تساوي ($nP = 0.5n$)، وتباين يساوي:

$P(1-P)n = 0.5(1-0.5)n = 0.5(0.5)n = 0.25n$

٢- وبما أن التوزيع متماثل فإن التوقع = الوسيط = المنوال.

ومما سبق يتضح أن اختبار أن الوسيط يساوي قيمة معينة ولتكن M يكافئ اختبار أن المتغير $n(+)$ يتبع توزيع ذات الحدين بمعلمتين (n, 0.5). ومن ثم يمكن اعتبار أن $n(+)$ هو المقياس الذي يتم حسابه من العينة لمعرفة معلومات عن الوسيط في المجتمع (M). ويتم إجراء الاختبار على النحو التالي:

١- صياغة الفرض العدمي والفرض البديل على النحو التالي:

الفرض العدمي

$$H_0 : M = M_0 \quad (6.5)$$

ويمكن أن يأخذ الفرض البديل إحدى الصياغات التالية:

$$H_1 : M > M_0 \quad (6.6)$$
$$H_1 : M < M_0 \quad (6.7)$$
$$H_1 : M \neq M_0 \quad (6.8)$$

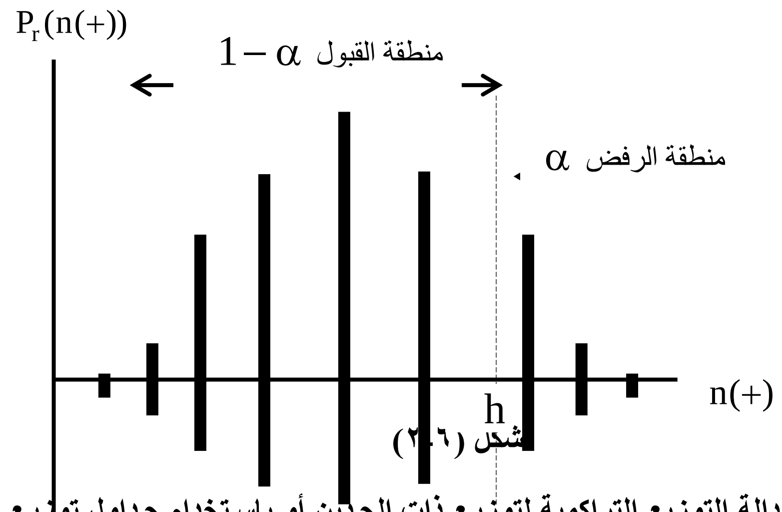
٢- حساب المقياس $n(+)$ وذلك عن طريق تحديد الفرق

$$Y_j = X_j - M_0 \quad (6.9)$$

ثم حساب عدد الفروق الموجبة أي حساب $n(+)$.

٣- تحديد درجة الثقة ($1-\alpha$). ووفقاً للفرض البديل ودرجة الثقة تحدد منطقة القبول ومنطقة (أو مناطق) الرفض على النحو التالي:

٤- إذا كان الفرض البديل كما هو في العلاقة (6.6) أي $H_1 : M > M_0$. فنجد أن منطقة الرفض في الطرف الأيمن كما هو موضح في شكل (٦-٢).



شكل (٦-٢)

وباستخدام دالة التوزيع التراكمية لتوزيع ذات الحدين أو باستخدام جداول توزيع ذات الحدين بملحق رقم (٧) نحسب قيمة h حيث:

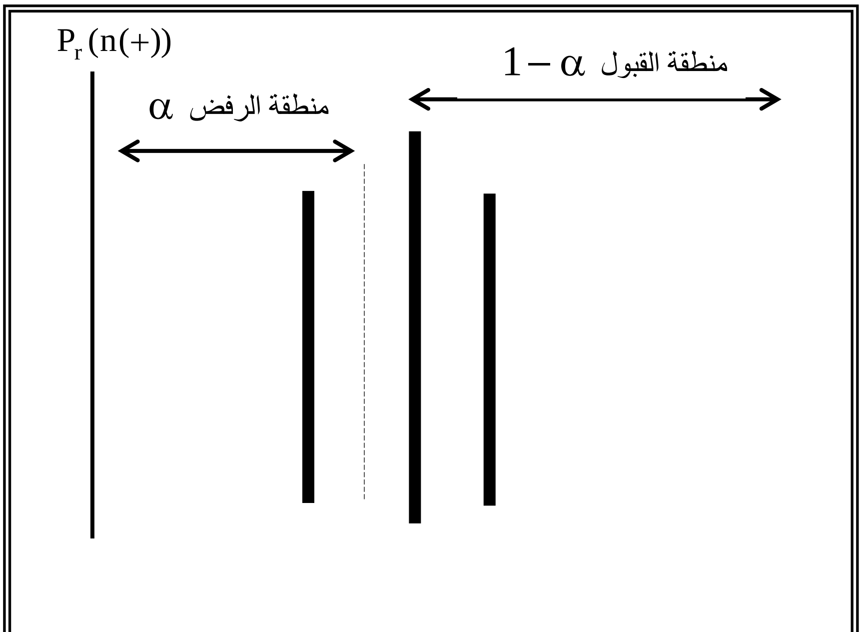
$$P_r\left(n(+) \geq h\right) < \alpha \qquad (6.10)$$

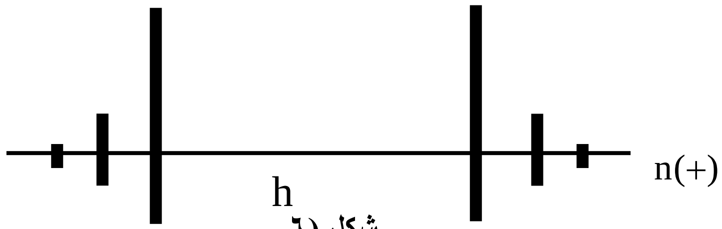
فإذا كانت قيمة $n(+)$ المحسوبة من العينة أقل من h أي أن $n(+)$ تقع في منطقة القبول وبالتالي نقبل الفرض العدمي ونرفض الفرض البديل.

٥- أما إذا كان الفرض البديل كما في العلاقة (6.7) أي أن

$$H_1 : M < M_0$$

فنجد أن منطقة الرفض في الطرف الأيسر كما هو موضح في شكل (٦-٣)





شكل (٦-

وباستخدام دالة التوزيع التراكمية لتوزيع ذات الحدين أو باستخدام جداول توزيع ذات الحدين نحسب قيمة **h** حيث:

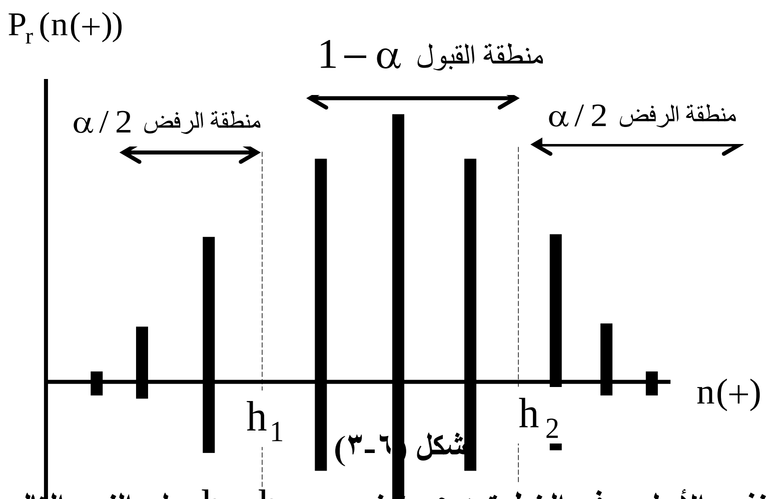
$$P_r\left(n(+) \le h\right) < \alpha \qquad (6.11)$$

فإذا كانت قيمة **n(+)** المحسوبة أكبر من **h** أي أن **n(+)** تقع في منطقة القبول وبالتالي نقبل الفرض العدمي ونرفض الفرض البديل.

٦- أما إذا كان الفرض البديل كما في العلاقة **(6.8)** أي أن

$$H_1 : M \ne M_0$$

فإننا نجد في هذه الحالة وجود منطقتين للرفض في الطرف الأيمن والطرف الأيسر كما هو موضح في شكل (٦-٤)



شكل (٦-٣)

٧- وبنفس الأسلوب في الخطوتين ٥، ٦ نحسب h_1, h_2 على النحو التالي:

$$P_r\left(n(+) \geq h_2\right) = \frac{\alpha}{2} \qquad (6.12)$$

$$\boxed{P_r\left(n(+) \leq h_1\right) = \frac{\alpha}{2} \qquad (6.13)}$$

فإذا كانت قيمة $n(+)$ المحسوبة من بيانـات العينـة تقـع بين h_1, h_2. أي تقـع في منطقة القبول أي أن:

$$\boxed{h_1 < n(+) < h_2 \qquad (6.14)}$$

فإننــا نقبـل الفـرض العـدمي ونـرفض الفـرض البـديل وسـوف نوضـح الخطـوات المذكورة أعلاه من خلال الأمثلة التالية:

ومما هو جدير بالذكر أن هذا الاختبار يسمي باختبار الإشـارة Sign Test لأن المقياس الذي يتم حسابه من العينـة هو عدد الإشـارات الموجبـة $n(+)$ أو عدد الإشارات السالبة $n(-)$.

مثال (٦-٣)

أخذ عينه من الدخل الشهري لـ 8 أفراد بإحدى المؤسسات فكانت على النحو التالي:
780, 650, 540, 620, 750, 500, 820, 550

عند درجة الثقة 96% اجري الاختبارات التالية:

أولاً: أن تكون القيمة الوسيطية للدخل أكبر من 680 جنية.

ثانياً: أن تكون القيمة الوسيطية للدخل أقل من 800 جنية.

ثالثاً: أن تكون القيمة الوسيطية لا تساوي 700 جنية.

الحل:

أولاً:

١- الفرض العدمي $H_0 : M = 680$

الفرض البديل $H_1 : M > 680$

٢- نكون الجدول التالي:

جدول (٦-١)

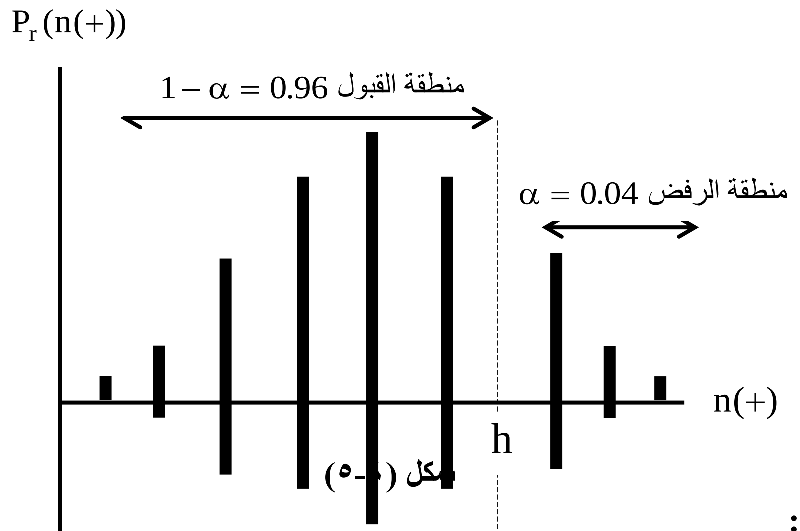
X_j	$X_j - M = X_j - 680$
550	-
820	+
500	-
750	+
620	-
540	-
650	-
780	+

من الجدول نجد أن:

$$n(+) = 3 \quad , \quad n(-) = 5$$

٣- بما أن درجة الثقة $\alpha = 0.04 \leftarrow (1-\alpha) = 0.96$

وبما أن الفرض البديل $H_1 : M > 680$. بالتالي فإن منطقة الرفض تكون في الطرف الأيمن كما هو موضح بالشكل التالي:



شكل (٦-٥)

وبما أن:

$$P_r\left(n(+) \geq h\right) < 0.04 \rightarrow h > 6$$

وبما أن 3 = n(+) أي أقل من h. أي تقع في منطقة القبول بالتالي نقبل الفرض العدمي ونرفض الفرض البديل القائل أن الوسيط للدخل أكبر من 680 جنية.

ثانياً:

١- الفرض العدمي $H_0 : M = 680$

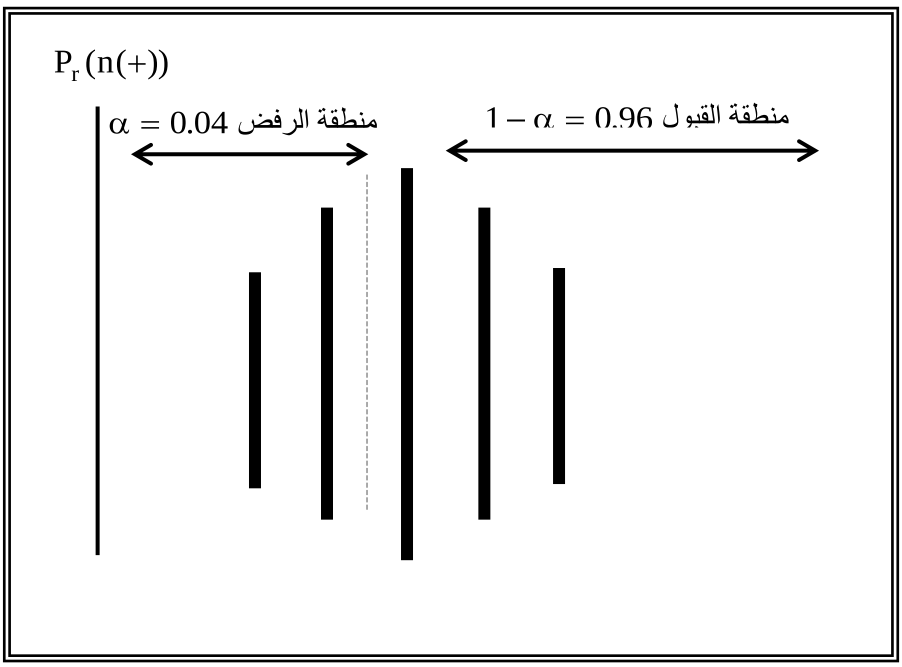
الفرض البديل $H_1 : M < 680$

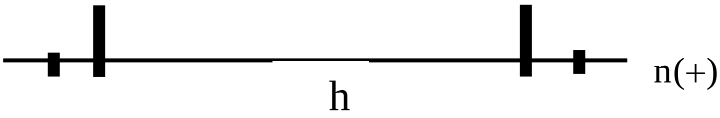
٢- من جدول (٦-١) نجد أن:

N(+) = 3 , n(-) = 5

٣- بما أن درجة الثقة $\alpha = 0.04 \leftarrow (1-\alpha) = 0.96$

وبما أن الفرض البديل $H_1 : M < 680$ **. وبالتالي فإن منطقة الرفض تكون في الطرف الأيسر كما هو موضح بالشكل التالي:**





شحں (٦-١)

وبما أن:

$$P_r\left(n(+) \le h\right) < 0.04 \rightarrow h < 2$$

وبما أن n(+) = 3 أي أكبر من h. أي أن n(+) = 3 تقع في منطقة القبول كما هو موضح بشكل (٦-٦). وبالتالي نقبل الفرض العدمي ونرفض الفرض البديل القائل بأن الوسيط للدخل أقل من 680 جنية.

<u>ثالثاً:</u>

١- **الفرض العدمي** $H_0 : M = 680$

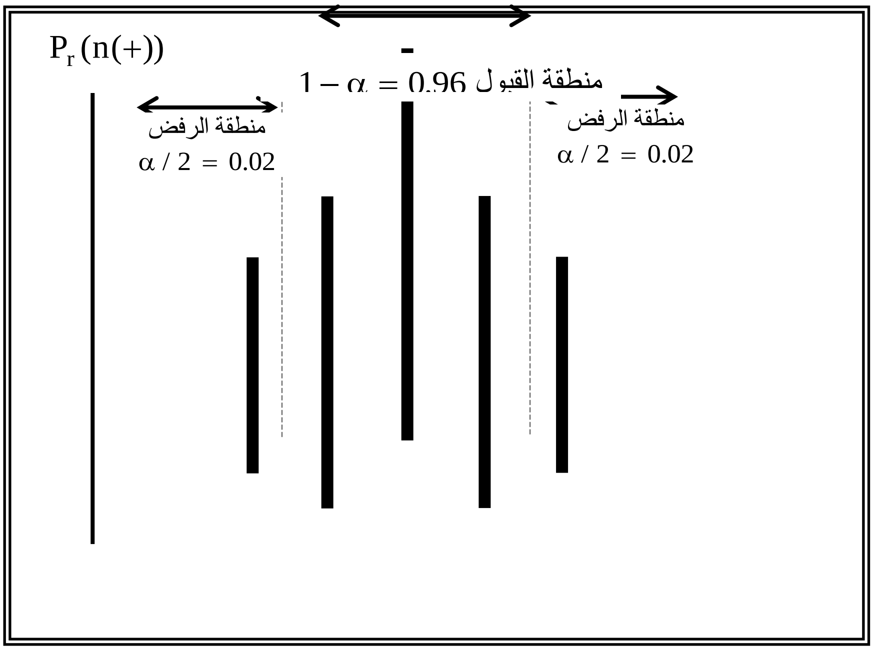
الفرض البديل $H_1 : M \neq 680$

٢- **من جدول (٦-١) نجد أن:**

N(+) = 3 , n(-) = 5

٣- **بما أن درجة الثقة** $\alpha = 0.04 \leftarrow (1-\alpha) = 0.96$

وبما أن الفرض البديل $H_1 : M \neq 680$**. وبالتالي فإنه يوجد منطقتين للرفض واحدة في الطرف الأيمن والأخرى في الطرف الأيسر كما هو موضح بالشكل (٦-٧).**



شكل (٦-٧)

وبمـا أن 3 = n(+) أي أن $h_1 < n(+) < h_2$ أي تقـع فـي منطقـة القبـول.

وبالتالي نقبل الفرض العدمي ونـرفض الفرض البـديل القائـل بـأن قيمـة الوسيط للـدخل يختلف عن 680 جنية وذلك بدرجة ثقة 96%.

مثال (٦-٤)

ترغب إدارة إحدى البنـوك التجاريـة فـي استطلاع رأى العمـلاء المستديمين بالبنـك بالنسبة لمستوي أداء العاملين على الشبابيك مع العملاء وذلك من خلال عينـة عشوائية بسيطة مكونـه من 10 عملاء للبنك فكانـت أرائـهم بالنسـبة لمسـتوي أداء العـاملين علـى الشبابيك بالبنك على النحو التالي:

متوسط، ممتاز، جيد، جيد جداً، متوسط، ممتاز، جيد جداً، جيد، متوسط، ممتاز.

أختبـر الفـرض القائـل بـأن آراء 50% مـن عمـلاء البنـك أن مسـتوي الأداء للعـاملين بالشبابيك أفضل من جيد وذلك بدرجة ثقة 96%.

الحل:

١- **الفرض العدمي** $H_0 : M =$ جيد

الفرض البديل $H_1 : M >$ جيد

٢- **نكون الجدول التالي:**

جدول (٦-٢)

الفرق	الرأي	
-	متوسط	
+	ممتاز	
صفر	جيد	تستبعد هذه المفردة
+	جيد جداً	
-	متوسط	
+	ممتاز	
+	جيد جداً	
صفر	جيد	تستبعد هذه المفردة
-	متوسط	
+	ممتاز	

وبالتالي يصبح حجم العينة:

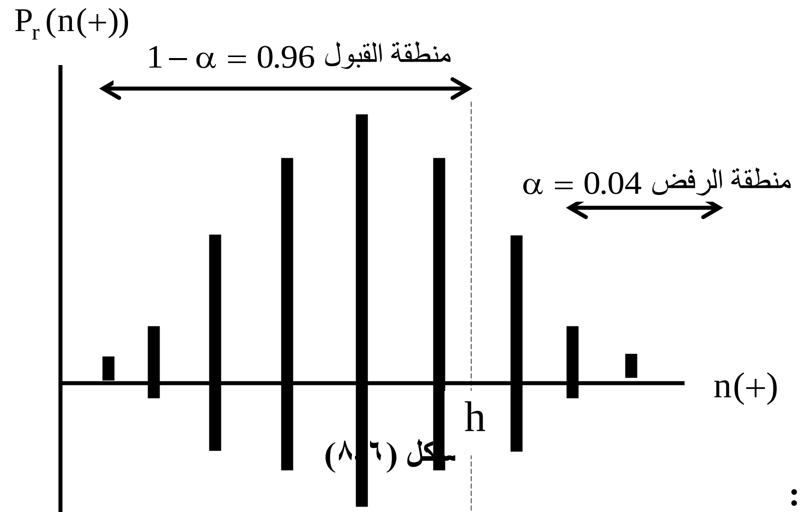
مفردات $10 - 2 = 8$

ومن الجدول نجد أن:

$$n(+) = 5 \qquad , \qquad n(-) = 3$$

٣- بما أن درجة الثقة $\alpha = 0.04 \leftarrow (1-\alpha) = 0.96$

وبما أن الفرض البديل $H_1 : M >$ **جيد ، وبالتالي فإن منطقة الرفض تقع في الطرف الأيمن كما هو موضح بشكل (٦-٨).**



شكل (٦-٨)

وبما ان:

$$P_r\left(n(+) \geq h\right) < 0.04 \rightarrow h > 6$$

وبما أن $n(+) = 5$ أي تقع في منطقة القبول. بالتالي نقبل الفرض العدمي ونرفض الفرض البديل القائل بان 50% من العملاء يرون أن مستوي الأداء للعاملين بالبنك على الشبابيك أفضل من الجيد.

ثانياً: اختبار الإشارة المبني على التوزيع المعتاد القياسي

وكما ذكرنا سابقاً أن المتغير $n(+)$ يتبع توزيع ذات بمعلمتين (n, P) بتوقع $\mu = 0.5n$ ، وتباين $\sigma^2 = 0.25n$. وعندما يزيد حجم العينة n عن 10 مفردات فإن التوزيع الاحتمالي للمتغير $n(+)$ يمكن أن يقترب من (أي يؤول إلى) التوزيع المعتاد بتوقع $\mu = 0.5n$ وانحراف معياري σ حيث:

$$\sigma = \sqrt{P(1-P)n} = \sqrt{0.5(0.5)n} = 0.5\sqrt{n}$$

وبالتالي فإن المتغير Y حيث:

$$Y = \frac{\left(n(+) \pm 0.5\right) - 0.5n}{0.5\sqrt{n}} \qquad (6.15)$$

يؤول إلى المعتاد القياسي*

ملحوظة*:

١- عندما n(+) يؤول إلى التوزيع المعتاد فإن:

$$a) P_r\left(n(+) = h\right) \rightarrow P_r\left(h - 0.5 \le n(+) \le h + 0.5\right) \qquad (6.16)$$

$$b) P_r\left(n(+) \ge h\right) \rightarrow P_r\left(n(+) \ge h - 0.5\right) \qquad (6.17)$$

$$c) P_r\left(n(+) \le h\right) \rightarrow P_r\left(n(+) \le h + 0.5\right) \qquad (6.18)$$

٢- إذا كان $n(+) > 0.5n$ فإن المقياس Y يحسب على النحو التالي:

$$Y = \frac{\left(n(+) - 0.5\right) - 0.5n}{0.5\sqrt{n}} \qquad (6.19)$$

٣- أما إذا كان $n(+) < 0.5n$ فإن المقياس Y يحسب على النحو التالي:

$$Y = \frac{\left(n(+) + 0.5\right) - 0.5n}{0.5\sqrt{n}} \qquad (6.20)$$

وفي هذه الحالة يتم إجراء الاختبار على النحو التالي:

١- صياغة الفرض العدمي والبديل على النحو التالي:-

الفرض العدمي: $H_0 : M = M_0$

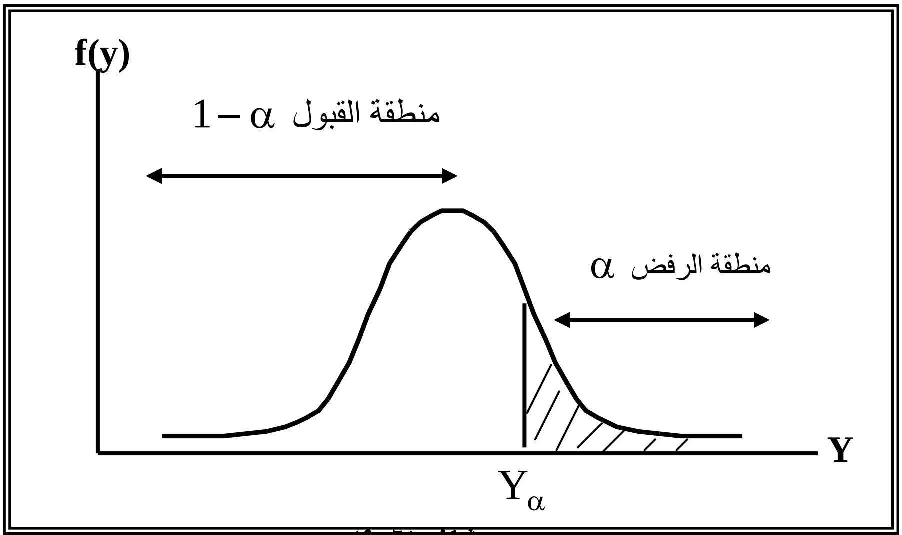
الفرض البديل: $H_1 : M > M_0 \quad \text{or} \quad M < M_0 \quad \text{or} \quad M \neq M_0$

٢- حساب n(+) من العينة ثم حساب المقياس Y من المعادلة (6.19) أو (6.20).

* Lawrence B. Morse (1993): Statistics for "Business and Economics" Harper Collins College Publishers. New York.

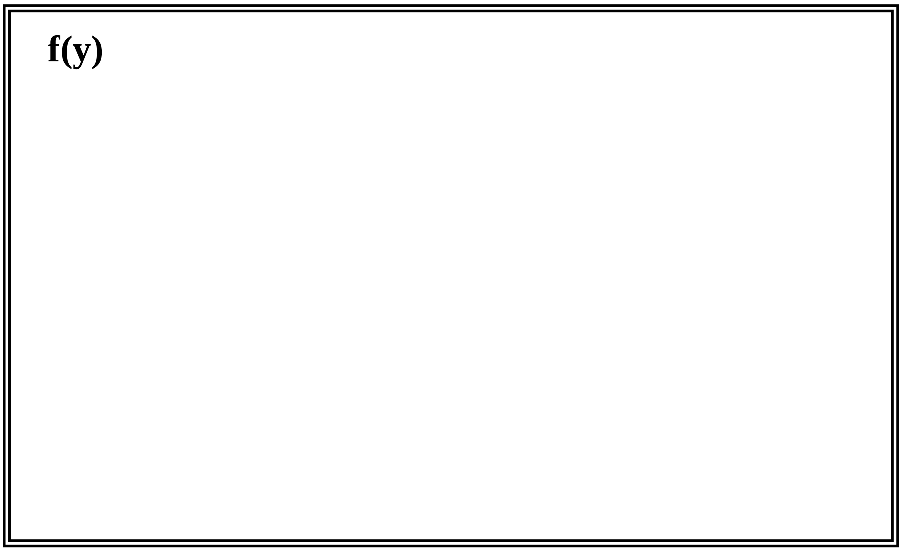
٣- عند درجة الثقة $(1-\alpha)$ أي مستوي المعنوية α من جدول المعتاد القياسي نحسب Y_α.

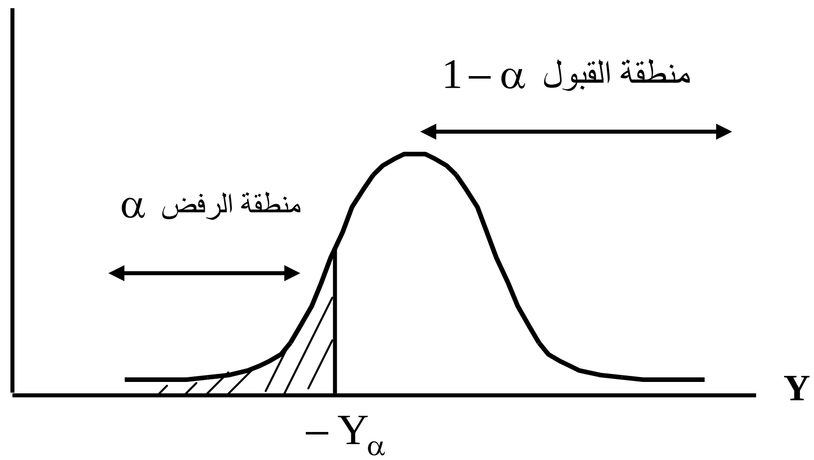
٤- إذا كان الفرض البديل $H_1 : M > M_0$ فإن منطقة الرفض تكون في الطرف الأيمن كما هو موضح بالشكل التالي:



شكل (٦-٩)

ويرفض الفرض العدمي إذا كانت Y المحسوبة أكبر من Y_α. وإذا كان الفرض البديل $H_1 : M < M_0$. فإن منطقة الرفض تقع في الطرف الأيسر كما هو موضح بشكل (٦-١٠).



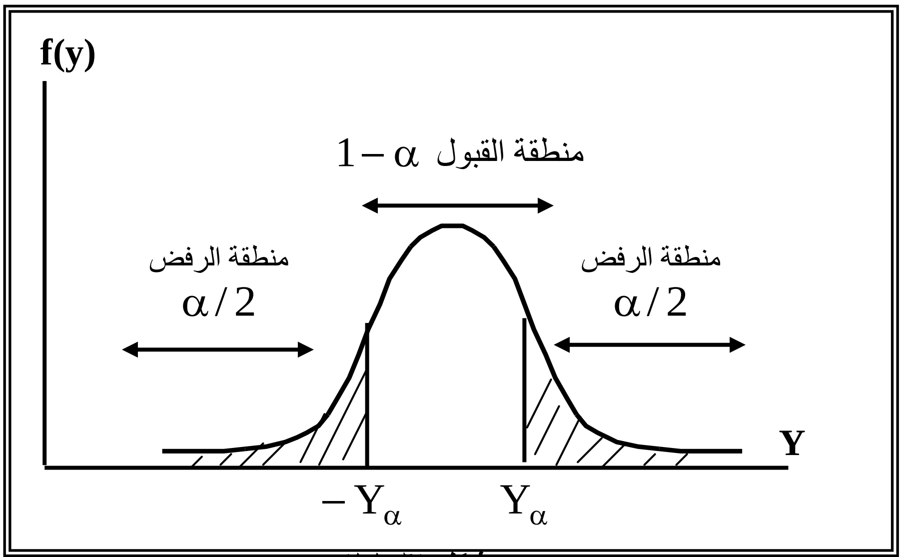


شكل (٦-١٠)

ويرفض الفرض العدمي في هذه الحالة إذا كانت قيمة Y المحسوبة أقل من ($-Y_{\alpha}$).

وإذا كان الفرض البديل $H_1 : M \neq M_0$ فإنه يوجد منطقتين للرفض كما هو موضح في شكل (٦-١١).

ويرفض الفرض العدمي في هذه الحالة إذا كانت قيمة Y المحسوبة أكبر من $Y_{\alpha/2}$ أو أقل من $-Y_{\alpha/2}$.



شكل (٦-١١)

وسوف نوضح خطوات إجراء الاختبار في هذه الحالة من خلال المثال التالي:

مثال (٦-٥)

سحبت عينة عشوائية بسيطة 25 مشاهدة من الأجور الشهرية للعاملين في إحدى القطاعات الإنتاجية بالجنية – فكانت على النحو التالي:

300, 320, 125, 270, 310, 250, 270, 245, 190, 197, 200, 210, 180, 150, 180, 310, 270, 280, 295, 175, 193, 240, 305,410,255

اختبر الفرض القائل بأن الأجر الشهري لـ 50% من العاملين بهذا القطاع تزيد عن 275 جنية شهرياً. وذلك بدرجة ثقة 95%.

<u>الحل:</u>

١- **الفرض العدمي** $H_0 : M = 275$

الفرض البديل $H_1 : M > 275$

٢- **نكون الجدول التالي:**

جدول (٦-٣)

X_j	الفرق $= X_j - 200$

X_j	الفرق $= X_j - 200$
300	+
320	+
125	-
270	-
310	+
250	-
270	-
245	-
190	-
197	-
200	-
210	-
180	-
150	-
180	-
310	+
270	-
180	+
295	+
175	-
193	-
240	-
305	+
410	+
255	-

ومن الجدول يتضح أن:

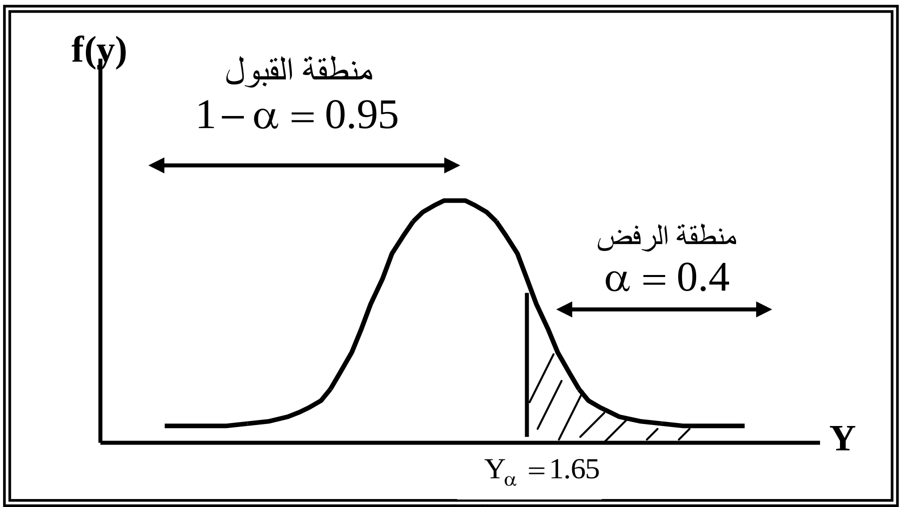
$$n(+) = 8 \quad , \quad n(-) = 17$$

وبما أن 8 = n(+)، وبما أن $0.5n = 0.5 \times 25 \approx 13$ **أي أن** $n(+) < 0.5n$ **وبالتالي فمن المعادلة (6.20) نجد أن:**

$$Y = \frac{\left(n(+) + 0.5\right) - 0.5n}{0.5\sqrt{n}}$$

$$= \frac{\left(8 + 0.5\right) - 0.5 \times 25}{0.5\sqrt{25}} = \frac{-4}{2.5} = -1.6$$

٣- عند درجة الثقة 95% أي $\alpha = 5\%$ نجد أن منطقة الرفض تقع في الطرف الأيمن ومن جدول التوزيع المعتاد القياسي بملحق (٤) نجد أن $Y_\alpha = 1.65$.



شكل (٦-١٢)

وبما أن قيمة Y المحسوبة تساوي (-1.6) أي أقل من قيمة Y_α بالتالي نقبل الفرض العدمي ونرفض الفرض البديل القائل بأن الأجر الشهري للعامل في هذا القطاع أكبر من 275 جنية.

(٦-٣) اختبار ويل كاكسون لمجموع الرتب

The Wilcoxon-Rank Sum Test

في الفصل (٥-٥) تناولنا اختبار الفروض المعلمية للفرق بين متوسطي مجتمعين $(\mu_1 - \mu_2)$ عن طريق عينتين مستقلتين من مجتمعي الدراسة ويتطلب إجراء الاختبار توافر الشروط التالية:

١- أن يكون التوزيع الاحتمالي للمتغير محل الدراسة في كل من المجتمعين يتبع التوزيع المعتاد.

٢- أن يكون التبـاين للمتغيـر فـي المجتمعين σ_1^2, σ_2^2 **معلـومين، أو أن التبـاين متساوي في المجتمعين أي** $\sigma_1^2 = \sigma_2^2$**.**

وفي حالة عدم توافر الشرطين السابقين فإنه لا يمكن استخدام الاختبارات المعلمية عن الفرق بين متوسط المجتمعين $(\mu_1 - \mu_2)$**.**

في سنة ١٩٤٥ قدم العالم فرانك ويل كاكسون اختبـار لا معلمـي لاختبـار الفرض القائـل بـأن التوزيـع التكـراري النسـبي Relative Frequency Distribution للمتغير فـي المجتمـع الأول يطـابق Identical To التوزيـع التكـراري النسـبي للمتغيـر فـي المجتمـع الثـاني. وفـي حالـة صـحة الفـرض أي أن التوزيـع النسـبي للمتغيـر فـي المجتمـع الأول مطـابق للتوزيـع النسـبي للمتغيـر فـي المجتمـع الثـاني. فهـذا يكـافئ أن متوسط المتغير في المجتمع الأول يساوي متوسط المتغير في المجتمـع الثـاني، أي ان $\mu_1 = \mu_2$ **أو** $\mu_1 - \mu_2 = 0$**.**

وصمم ويل كاكسون الاختبار على النحو التالي:

١- سـحب عينتـين عشـوائيتين مسـتقلتين مـن مشـاهدات المجتمعـين حجـم كـل منـهمـا n_2, n_1 **على الترتيب ولتكونـا:**

$$X_1, X_2, X_3, \ldots\ldots, X_{n_1} \quad , \quad n_1 \ge 10$$

$$Y_1, Y_2, Y_3, \ldots\ldots\ldots, Y_{n_2} \quad , \quad n_2 \ge 10$$

٢- خلـط مفـردات العينتـين ثـم ترتيـب قيمهمـا ترتيـب تصـاعدي بحيـث يـتم أعطـاء المفـردات ذات أصـغر قيمـة الترتيـب (1) والمفـردات ذات أكبـر قيمـة الترتيـب $(n_1 + n_2)$**.**

٣- تسجيل رتب كل عينة ثم إيجاد مجموع رتب المشاهدات للعينة المسحوبة من المجتمع الأول وليكن U_1 وإيجاد مجموع رتب المشاهدات للعينة المسحوبة من المجتمع الثاني ولتكن U_2.

٤- أثبت ويل كاكسون أن المتغير U_1 (أو U_2) يتبع التوزيع المعتاد بتوقع μ_{U_1} بحيث:

$$\mu_{U_1} = \frac{n_1(n_1+n_2+1)}{2} \tag{6.21}$$

$$\sigma_{U_1} = \sqrt{\frac{n_1(n_2)(n_1+n_2+1)}{12}} \tag{6.22}$$

وبالتالي فإن المتغير U بحيث:

$$U = \frac{U_1 - \mu_{U_1}}{\sigma_{U_1}}$$

أي

$$U = \frac{U_1 - \left[\dfrac{n_1(n_1+n_2+1)}{2}\right]}{\sqrt{\dfrac{n_1(n_2)(n_1+n_2+1)}{12}}} \tag{6.23}$$

متغير يتبع التوزيع المعتاد القياسي.

ملحوظة:

في حالة استخدام المتغير U_2 فإن توقعه μ_{U2} وانحرافة المعياري σ_{U2} بحيث:

$$\mu_{U_2} = \frac{n_2(n_1+n_2+1)}{2} \tag{6.24}$$

$$\sigma_{U_2} = \sqrt{\frac{n_2(n_1)(n_1 + n_2 + 1)}{12}} \qquad (6.25)$$

وبالمثل فإن المتغير U بحيث

$$U = \frac{U_2 - \mu_{U_2}}{\sigma_{U_2}}$$

أي

$$\boxed{U = \frac{U_2 - \left[\dfrac{n_2(n_1 + n_2 + 1)}{2}\right]}{\sqrt{\dfrac{n_2(n_1)(n_1 + n_2 + 1)}{12}}}} \qquad (6.26)$$

واعتبر ويل كاكسون المتغير U المحسوب من بيانات العينتين هو المقياس المستخدم في الاختبار. ولإجراء الاختبار تتبع الخطوات التالية:

١- صياغة الفرض العدمي والفرض البديل على النحو التالي:

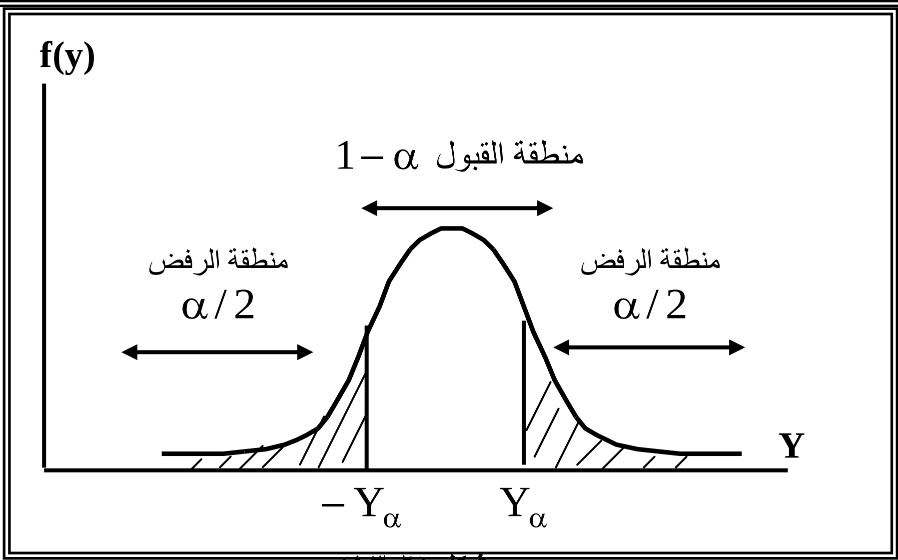
الفرض العدمي $H_0 : \mu_1 = \mu_2$

الفرض البديل $H_1 : \mu_1 \neq \mu_2$

٢- حساب المقياس U.

٣- تحديد درجة الثقة $(1-\alpha)$ وباستخدام جدول المعتاد القياسي يحسب حدي الثقة $-U_{\alpha/2}$ ، $U_{\alpha/2}$.

وتحدد منطقتي الرفض ومنطقة القبول كما هو موضح في الشكل التالي.



شكل (٦-١٣)

٤- إذا كانت قيمة U المحسوبة تقع في منطقة القبول أي أن:

$$-U_{\alpha/2} < U < U_{\alpha/2} \qquad \textbf{(6.27)}$$

فإننا نقبل الفرض العدمي القائل بأن متوسطي التوزيعين للمجتمعين متساويين أي $\mu_1 = \mu_2$. أما إذا كان:

$$U > U_{\alpha/2}$$

فإننا نرفض الفرض العدمي أي أن $\mu_1 \neq \mu_2$. وسوف نوضح خطوات الاختبار من خلال المثال التالي:

مثال (٦-٦)

المطلوب اختبار الفرض القائل بأنه لا يوجد فرق بين متوسط الأجر الشهري للعامل في القطاع الحكومي ومتوسط الأجر الشهري للعامل في قطاع الأعمال. وذلك بدرجة ثقة 95%. فأخذ عينة عشوائية بسيطة حجمها 12 عامل بالقطاع الحكومي، وسجل

الأجر الشهري لكل منهم عن شهر يناير سنة ٢٠٠١ فكانت أجورهم على النحو التالي بالجنية:

650, 320, 210, 200, 470, 590, 325, 600, 280, 310, 450, 215

كذلك أخذت عينة عشوائية بسيطة من أجور 13 عامل بقطاع الأعمال عن يناير سنة ٢٠٠١ بالجنية فكانت على النحو التالي:

800, 500, 250, 611, 750, 350, 410, 211, 359, 520, 379, 630, 700

الحل:

لإجراء اختبار ويل كاكسون لاختبار الفرض القائل بأنه لا يوجد فرق بين متوسط الأجر الشهري للعامل في القطاع الحكومي ومتوسط الأجر الشهري للعامل في قطاع الأعمال نتبع الخطوات التالية:

١- إذا فرضنا أن μ_2, μ_1 هما الوسط الحسابي للأجر الشهري للعامل في القطاع الحكومي وفي قطاع الأعمال على الترتيب فإن:

الفرض العدمي $H_0 : \mu_1 = \mu_2$

الفرض البديل $H_1 : \mu_1 \neq \mu_2$

٢- لحساب المقياس U نكون جدول (٦-٤).

وبما أن

$$a)\mu_{U_1} = \frac{n_1(n_1 + n_2 + 1)}{2} = \frac{12(12+13+1)}{2}$$

$$= \frac{12(26)}{2} = 12 \times 13 = 156$$

$$b)\sigma_{U_1} = \sqrt{\frac{n_1(n_2)(n_1+n_2+1)}{12}}$$

$$= \sqrt{\frac{12(13)(12+13+1)}{12}} = \sqrt{338} = 18.39$$

$$U = \frac{U_1 - \mu_{U_1}}{\sigma_{U_1}} = \frac{107-156}{18.39} = -2.67$$

٣- وبما أن درجة الثقة 95% $\leftarrow \alpha = 0.05 \leftarrow \frac{\alpha}{2} = 0.025$

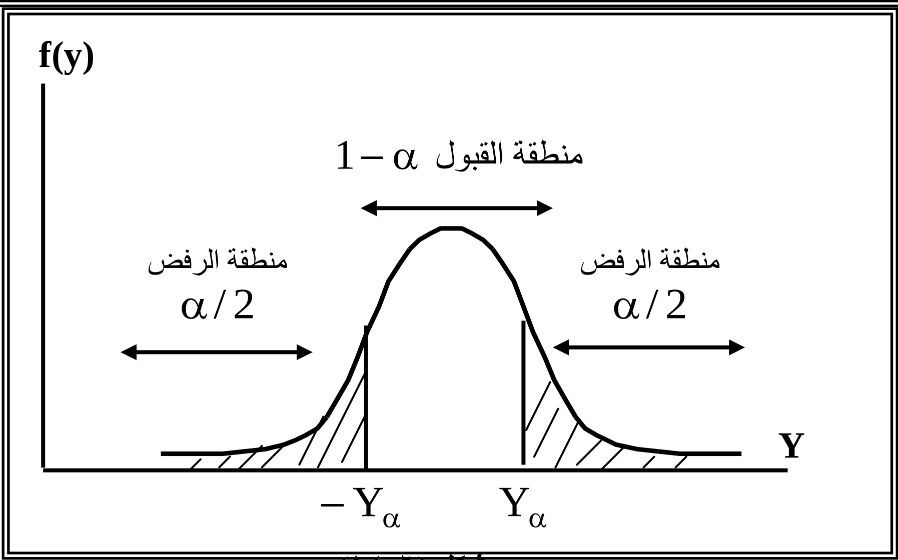
وباستخدام جدول التوزيع المعتاد نجد أن:

$$U_{\alpha/2} = \pm 1.96$$

كما هو موضح بالشكل (٦-٤)

جدول (٦-٤)

العينة من قطاع الحكومة (A)	العينة من قطاع الأعمال (B)	الترتيب التصاعدي في العينتين	رتب القيم	رتب العينة الأولي (A)	رتب العينة الثانية (B)
650	800	200	$A \rightarrow 1$	1	
320	500	210	$A \rightarrow 2$	2	
210	250	211	$B \rightarrow 3$		3
200	611	215	$A \rightarrow 4$	4	
470	750	250	$B \rightarrow 5$		5
590	350	280	$A \rightarrow 6$	6	
325	420	310	$A \rightarrow 7$	7	
600	211	320	$A \rightarrow 8$	8	
280	359	325	$A \rightarrow 9$	9	
310	510	350	$B \rightarrow 10$		10
450	379	359	$B \rightarrow 11$		11
215	600	379	$B \rightarrow 12$		12
	700	420	$B \rightarrow 13$		13
		450	$A \rightarrow 14$	14	
		470	$A \rightarrow 15$	15	
		500	$B \rightarrow 16$		16
		510	$B \rightarrow 17$		17
		590	$B \rightarrow 18$		18
		600	$A \rightarrow 19$	19	
		611	$B \rightarrow 20$		20
		630	$B \rightarrow 21$		21
		650	$A \rightarrow 22$	22	
		700	$B \rightarrow 23$		23
		750	$B \rightarrow 24$		24
		800	$B \rightarrow 25$		25
			المجموع	$U_1 = 107$	$U_2 = 218$



شكل (٦-١٤)

وبما أن:

$$U < -U_{\alpha/2}$$

أي أن U تقع في منطقة الرفض – بالتالي نرفض الفرض العدمي ونقبل الفرض البديل القائل بأن متوسط الأجر الشهري للعامل في القطاع الحكومي يختلف عن متوسط الأجر الشهري للعامل في قطاع الأعمال.

ملحوظة:

إذا كررت إحدى القيم أكثر من مرة فتأخذ كل منها نفس الرتبة وهو المتوسط للرتب للمواقع التي تحتلها هذه القيم ـ كما سوف نوضح ذلك من خلال المثال التالي:

مثال (٦-٧)

اختبر الفرض القائل بان متوسط الاستهلاك الشهري للكهرباء بالكيلو وات للأسرة الواحدة في الريف لا يختلف عن متوسط الاستهلاك الشهري للأسرة في الحضر في سنة معينة بدرجة ثقة 90% فإذا أخذت عينة من 12 أسرة من سكان الريف وتم حساب الاستهلاك الشهري لكل منها في هذه السنة فكانت على النحو التالي:

170, 178, 200, 250, 178, 370, 300, 178, 150, 190, 180, 292

كذلك أخذت عينة مكونه من 10 أسر من سكان الحضر فكان الاستهلاك الشهري لكل منها في هذه السنة على النحو التالي:

210, 200, 320, 190, 260, 280, 300, 290, 285, 288

الحل:

١- الفرض العدمي العدمي $H_0 : \mu_1 = \mu_2$

الفرض البديل $H_1 : \mu_1 \neq \mu_2$

٢- لحساب المقياس U نتبع الخطوات التالية:

أ) نكون جدول (٦-٥).

جدول (٦-٥)

(1) العينة الأولي (A)	(2) العينة الثانية (B)	(3) الترتيب التصاعدي في العينتين	(4) رتب القيم	(5) رتب العينة الأولي (A)	(6) رتب العينة الثانية (B)
170	210	150	$A \to 1$	1	
178	200	170	$A \to 2$	2	
200	320	178	$A \to 4$	3	
250	190	178	$A \to 4$	4	
178	260	178	$A \to 4$	5	
370	280	180	$A \to 6$	6	
300	300	190	$A \to 7.5$	7.5	
178	290	190	$B \to 7.5$		7.5
150	285	200	$A \to 9.5$	9.5	
190	288	200	$B \to 9.5$		9.5
180		210	$B \to 11$		11
		250	$A \to 12$	12	
		260	$B \to 13$		13
		280	$B \to 14$		14
		285	$B \to 15$		15
		288	$B \to 16$		16
		290	$B \to 17$		17
		300	$A \to 18.5$	18.5	
		300	$B \to 19.5$		18.5
		320	$B \to 20$		20
		370	$A \to 21$	21	
			المجموع	$U_1 = 89.5$	$U_2 = 141.5$

ملحوظة:

في العمود رقم (3) نجد أن:

١) القيمة 178 تكررت ثلاث مرات والرتب المناظرة لها هي 5 ,4 ,3 فتأخذ كل قيمة من القيم الثلاثة التي تساوي كل منها 178 الرتبة المتوسطة وهي

$$\frac{3+4+5}{3}=\frac{12}{3}=4$$

٢) بالمثل نجد أن القيمة 190 تكررت مرتين فتأخذ لكل قيمة منها الرتبة

$$\frac{7+8}{2}=\frac{15}{2}=7.5$$

٣) بالمثل بالنسبة للقي 300 ,200.

ب) من الجدول نجد أن:

$$U_1=89.5 \quad ، \quad U_2=141.5$$

$$a)\mu_{U_1}=\frac{n_1(n_1+n_2+1)}{2}=\frac{11(11+10+1)}{2}$$

$$=\frac{11(22)}{2}=121$$

$$b)\sigma_{U_1}=\sqrt{\frac{n_1(n_2)(n_1+n_2+1)}{12}}$$

$$=\sqrt{\frac{11(10)(11+10+1)}{12}}=\sqrt{\frac{11(10)(22)}{12}}=\sqrt{201.67}=14.2$$

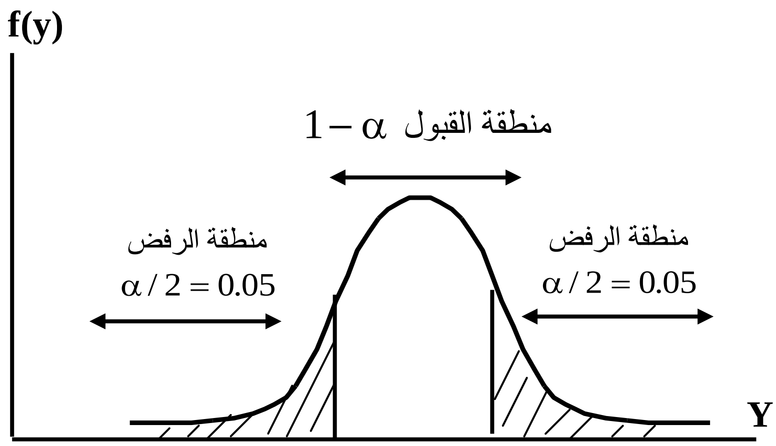
$$c)U=\frac{U_1-\mu_{U_1}}{\sigma_{U_1}}=\frac{89.5-121}{14.2}=-2.22$$

٣- وبما أن درجة الثقة 90% $\leftarrow \alpha = 0.10 \leftarrow 0.05=\frac{\alpha}{2}$

وباستخدام جدول التوزيع المعتاد نجد أن:

$$U_{\alpha/2}=1.64$$

كما هو موضح بالشكل (٦-١٥)



شكل (٦-١٥)

وبما أن $U < -1.64$ أي تقع في منطقة الرفض وبالتالي نرفض الفرض العدمي ونقبل الفرض البديل القائل بأن متوسط الاستهلاك الشهري للأسرة للكهرباء في الريف يختلف عن متوسط الاستهلاك الشهري للأسرة للكهرباء في الحضر وذلك بدرجة ثقة 90%.

(٦-٤) اختبار ويل كاكسون للرتب ذات الإشارة

Wilcoxon Signed-Rank Test

إذا كان لـدينا عينتيـن عشوائيتيـن غيـر مستقليـن مـن المشـاهـدات علـى النحـو التالـي:

$$X_1, X_2, X_3, \ldots\ldots, X_n$$

$$Y_1, Y_2, Y_3, \ldots\ldots, Y_n$$

قـدم العـالم ويـل كاكسـون اختبـار لا معلمـي لاختبـار الفـرض القائـل بـأن التوزيـع التكـراري النسـبي للمتغيـر محـل الدراسـة فـي المجتمـع المسـحوب منـه العينـة $X_j \to j = 1,2,\ldots,n$ يطابق Identical To التوزيع التكراري النسبي للمتغير في المجتمع الثاني المسحوب منه العينة $Y_j \to j = 1,2,\ldots,n$. وفي حالـة صحة الفـرض أي أن التوزيـع النسـبي للمتغيـر فـي المجتمـع الأول مطـابق للتوزيـع النسـبي للمتغيـر فـي المجتمـع الثـاني فهـذا يكـافئ أيضـاً أن متوسـط المتغيـر فـي المجتمـع الأول يساوي متوسط المتغير في المجتمع الثاني أي أن $\mu_1 = \mu_2$.

وعادة ما يستخدم هذا الاختبار لدراسة تأثير معالجة معينة على مجتمع الدراسة – أي يستخدم في الدراسات التجريبية.

وصمم ويل كاكسون الاختبار على النحو التالي:

١- **دمج العينتين في عينة واحدة من المشاهدات المزدوجة عل النحو التالي:**

$$(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \ldots\ldots, (X_n, Y_n)$$

٢- **حساب الفرق بين المشـاهدات المتنـاظرة فـي العينتين وسوف نشير إليـه بـالرمز d_j على النحو التالي:**

$$d_j = X_j - Y_j \qquad , \qquad j = 1,2,3,\ldots, n$$

ويتم استبعاد $d_j = 0$ ويعدل حجم العينة n وفقاً لذلك.

٣- **ثم حساب $|d_j|$، ثم ترتب $|d_j|$ ترتيب تصاعدي**

٤- ضرب الرتبة المناظرة للقيمة الموجودة للفرق في إشارة الفرق ثم حساب مجموع الرتب ذات الإشارة ولتكن تساوي (**d**).

وأثبت ويل كاكسون أن (**d**) متغير توقعه μ_d وانحرافة المعياري σ_d. وعندما $n \geq 10$ فإن:

$$\mu_d = 0 \qquad \textbf{(6.28)}$$

$$\sigma_d = \sqrt{\frac{n(n+1)(2n+1)}{6}} \qquad \textbf{(6.29)}$$

ولأن المتغير **d** يؤول إلى التوزيع المعتاد وبالتالي فإن المتغير **U** بحيث:

$$U = \frac{d-\mu_d}{\sigma_d} = \frac{d}{\sqrt{\frac{n(n+1)(2n+1)}{6}}} \qquad \textbf{(6.30)}$$

ولإجراء الاختبار تتبع الخطوات التالية:

١- صياغة الفرض العدمي صياغة الفرض العدمي والفرض البديل على النحو التالي:

الفرض العدمي $H_0 : \mu_1 = \mu_2$

الفرض البديل $H_1 : \mu_1 \neq \mu_2$

٢- حساب المقياس **d** ثم المقياس **U**.

٣- تحديد درجة الثقة $(1-\alpha)$ وباستخدام جدول التوزيع المعتاد القياسي يحسب حدي الثقة $-U_{\alpha/2}$ ، $U_{\alpha/2}$.

٤- نقبل الفرض العدمي إذا كانت قيمة **U** المحسوبة بحيث:

$$-U_{\alpha/2} < U < U_{\alpha/2}$$

فيما عدا ذلك يرفض الفرض العدمي ويقبل الفرض البديل.

وسوف نوضح الخطوات من خلال المثال التالي:

مثال (٦-٨)

ترغب إحدى المؤسسات في اختبار الفرض القائل أنه يوجد تأثير للدورات التدريبية على متوسط مستوي كفاءة العاملين بالمؤسسة. فأخذت عينة مكونه من 15 عامل وقيس مستوي كفاءة كل منهم قبل تلقيهم للدورات التدريبية فكانت على النحو التالي:

$X_j = 6.3, 6.7, 5.5, 6.4, 4.6, 4.5, 5.6, 7.8, 6.5, 5.5, 7.2, 8.1, 10, 9.1, 6.2$

وبعد إجراء التدريب تم قياس كفاءة نفس العاملين – فكانت على النحو التالي:

$Y_j = 6.5, 7, 6, 5.9, 5.3, 4.6, 7, 8.4, 6.9, 6.3, 8.1, 9.11, 10, 9.6, 8$

اختبر الفرض القائل بأنه لا يوجد تأثير للدورات على متوسط كفاءة العامل بالمؤسسة وذلك بدرجة ثقة 95%.

الحل:

يعتبر المتغير محل الدراسة هو مستوي كفاءة العامل بالمؤسسة. فإذا فرضنا أن μ_1 تشير إلى متوسط كفاءة العامل بالمؤسسة فبل تلقي التدريب، و μ_2 تشير إلى متوسط كفاءة العامل بالمؤسسة بعد تلقي التدريب.

١- الفرض العدمي $H_0 : \mu_1 = \mu_2$

الفرض البديل $H_1 : \mu_1 \neq \mu_2$

٢- لحساب المقياس U نكون الجدول (٦-٦).

جدول (٦-٦)

(1) X_j	(2) Y_j	(3) d_j	(4) $\lvert d_j \rvert$	(5) ترتيب $\lvert d_j \rvert$	(6) ترتيب $\lvert d_j \rvert$ ذو الإشارة
6.3	**6.5**	**-0.2**	**0.2**	**2**	**-2**
6.7	**7**	**-0.3**	**0.3**	**3**	**-3**

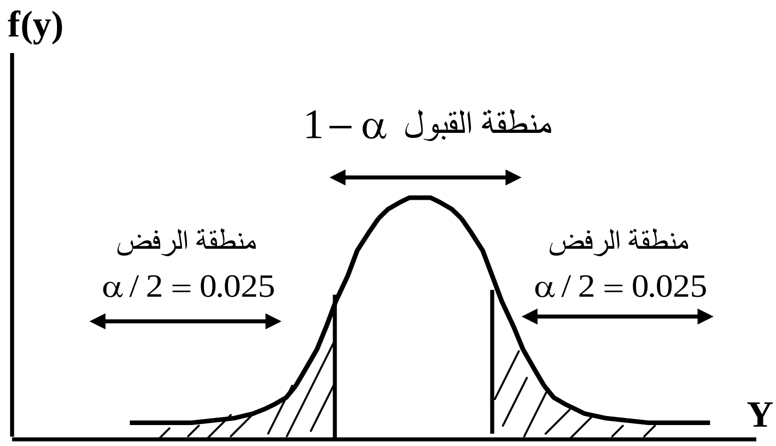
5.5	**6**	**-0.5**	**0.5**	**6**	**-6**
6.4	**5.9**	**+0.5**	**0.5**	**6**	**+6**
4.6	**5.3**	**-0.7**	**0.7**	**9**	**-9**
4.5	**4.6**	**-0.1**	**0.1**	**1**	**-1**
5.6	**7**	**-1.4**	**1.4**	**13**	**-13**
7.8	**8.4**	**-0.6**	**0.6**	**8**	**-8**
6.5	**6.9**	**-0.4**	**0.4**	**4**	**-4**
5.5	**6.3**	**-0.8**	**0.8**	**10**	**-10**
7.2	**8.1**	**-0.9**	**0.9**	**11**	**-11**
8.1	**9.11**	**-1.0**	**1.0**	**12**	**-12**
10	**10**	**0**	**–**	**–**	**–**
9.1	**9.6**	**-0.5**	**0.5**	**6**	**-6**
6.2	**8**	**-1.8**	**1.8**	**14**	**-14**
					d=-93

وبما أن $n = 15 - 1 = 14 > 10$ **فإن:**

$$\sigma_d = \sqrt{\frac{n(n+1)(2n+1)}{6}} = \sqrt{\frac{14(15)(29)}{6}} = \sqrt{1015} = 31.86$$

$$U = \frac{d - \mu_d}{\sigma_d} = \frac{d}{\sqrt{\frac{n(n+1)(2n+1)}{6}}} = \frac{-93}{31.86} = -2.92$$

٣- وعنـد درجـة الثقـة 95% نجـد أن حـدي الثقـة $\pm U_{\alpha/2}$ **تسـاوي** ± 1.96.
والشكل التالي يوضح مناطق الرفض والقبول.



شكل (٦-١٦)

وبما أن قيمة U المحسوبة تساوي 2.92- أي تقع في منطقة الرفض – بالتالي نرفض الفرض العدمي القائل أنه لا يوجد تأثير للدورات التدريبية ونقبل الفرض البديل القائل أنه يوجد تأثير للدورات التدريبية على مستوي كفاءة العاملين.

ملاحظات:

١- **تستبعد المشاهدات التي لكل منها الفرق** $d_j=0 \;,\; j=1,2,3,...,n$ **ويعاد حجم العينة وفقاً لعدد المفردات بعد استبعاد المفردات التي لكل منها** $d_j=0$.

ففي المثال السابق كان عدد المفردات يساوي 15 وبعد استبعاد المفردات التي لها الفرق $d_{13}=0$ **أصبح حجم العينة** $n=15-1=14$.

٢- **عندما يكون الفرض العدمي** $\mu_1=\mu_2$ **فإن** $\mu_d=0$.

(٦-٥) تمرينات
Exercises

(٦-١): أخذت عينة من 10 طلاب بالمرحلة الثانوية وسجل متوسط عدد الساعات اليومية لاستذكار الطالب فكانت على النحو التالي:

2 , 5 , 7 , 8 , 3 , 1 , 2 , 4 , 5 , 6

المطلوب:

١- **اختبر الفرض القائل بأن 50% من الطلاب بالمرحلة الثانوية يكون متوسط ساعات استذكار الطالب أكبر من 5 ساعات يوميا. وذلك بدرجة ثقة 90%.**

٢- **اختبر الفرض القائل بأن الوسيط لمتوسط ساعات استذكار الطالب يساوي 4 ساعات يومياً. وذلك بدرجة ثقة 95%.**

٣- **اختبر الفرض القائل بأن 50% من الطلاب بالمرحلة الثانوية يكون متوسط ساعات استذكار الطالب أقل من 6 ساعات يومياً. وذلك بدرجة ثقة 90%.**

(٦-٢): في دراسة عن مستوي أداء العاملين بأحد القطاعات أخذت عينة مكونه من 10 عاملين وحدد مستوي كل عامل وسجل فكانت على النحو التالي:

جيد، ضعيف، ضعيف، متوسط، جيد، جيد، جيد، جيدجداً، جيد، متوسط، ممتاز، ممتاز، متوسط، جيد، ممتاز، جيد.

المطلوب:

١- **اختبر الفرض القائل بان 50% من العاملين بهذا القطاع مستوي أداء كل منهم أعلي من المتوسط وذلك بدرجة ثقة 95%.**

٢- **اختبر الفرض القائل بأن مستوي أداء العامل بهذا القطاع أقل من الجيد وذلك بدرجة ثقة 90%.**

(٦-٣): في أحد الأحياء بالمدن الجديدة – أجريت دراسة عن أسعار الشقق التي مساحتها 100 متر مربع تقريباً. فأخذت عينة عشوائية حجمها 8 شقق وسجلت أسعارها بالجنية فكانت على النحو التالي:

112000, 125000, 116000, 118000, 12200, 125000, 130000, 129000

اختبر الفرض القائل بأن 50% من شقق هذا الحي التي مساحة كل منها 100 متر تقريباً تزيد عن 120000 جنية. وذلك بدرجة ثقة 90%.

(٦-٤): أخذت عينة عشوائية من 20 طفل حديثي الولادة في إحدى المستشفيات وسجلت أوزانهم فكانت على النحو التالي:

2.70	3.95	2.75	3.50	3.10
2.50	4.10	3.25	3.10	2.50
3.50	2.90	2.50	2.70	2.10
4.20	2.80	3.90	5.00	4.15

اختبر الفرض القائل بأن 50% من الأطفال حديث الولادة في هذه المستشفي يقل وزن كل منهم عن 2.5 كيلوا جرام.

(٦-٥): أجريت دراسة عن متوسط عدد الساعات اليومية لاستخدام شبكة الإنترنت للمشتركين في إحدى الشركات فإذا كان المشتركين من الباحثين ومن غير الباحثين – فأخذت عينة عشوائية من الباحثين فكان متوسط عدد الساعات اليومية للمشترك على النحو التالي:

6 , 5 , 76 , 10 , 8 , 4 , 3 , 2 , 4

كذلك كان متوسط عدد الساعات اليومية للمشترك من غير الباحثين في عينة عشوائية مكونه من 10 عملاء على النحو التالي:

2 , 3 , 4 , 7. 5 , 7 , 3 , 2 , 1 , 1

المطلوب:

١- بافتراض أن متوسط عدد الساعات اليومية للعميل لاستخدام شبكة الإنترنت يتبع التوزيع المعتاد – اختبر الفرض القائل بأن متوسط عدد الساعات اليومية لاستخدام شبكة الإنترنت للعميل من الباحثين أكبر من متوسط عدد الساعات اليومية لاستخدام الشبكة من العملاء غير الباحثين وذلك بدرجة ثقة 90%.

٢- إذا كان التوزيع الاحتمالي لمتوسط عدد ساعات استخدام شبكة الإنترنت للعملاء غير معروف.

أ) اختبر الفرض القائل بأن متوسط عدد ساعات استخدام العميل من الباحثين لشبكة الإنترنت أكبر من متوسط عدد ساعات استخدام العميل من غير الباحثين للشبكة. وذلك بدرجة ثقة 90%.

ب) أختبر الفرض القائل بأن الفرق بين متوسطي عدد ساعات استخدام العملاء الباحثين وغير الباحثين لشبكة الإنترنت يساوي صفر. وذلك بدرجة ثقة 95%.

(٦-٦): في إحدى الدراسات عن تأثير زيادة الجرعة من أحد الأدوية على سرعة الشفاء للمرضي فأخذت عينة من 10 من المرضي المتماثلين فكانت حالة كل منهم قبل زيادة الجرعة على النحو التالي:

سيئة جدا، سيئة، متوسطة، عادية، سيئة جدا، سيئة، متوسطة، عادية، متوسطة.

وبعد زيادة الجرعة سجلت حالة نفس المرضي فكانت على النحو التالي:

سيئة، متوسطة، متوسطة، عادية، متوسطة، متوسطة، متوسطة، متوسطة، عادية، عادية.

اختبر الفرض القائل بان زيادة الجرعة يؤدي إلى تحسين حالة المرضي (أي لزيادة الجرعة تأثير إيجابي على المرضي). وذلك بدرجة ثقة 90%.

الجزء الثاني

الاستدلال الإحصائي للعلاقات الإحصائية

Statistical Inference of Statistical Relationships

الباب السايع
توفيق المنحنيات
Curve Fitting

(٧-١) العلاقة بين المتغيرات
Relationship Between Variables

(٧-٢) شكل الإنتشار
S-Catter Diagram

(٧-٣) طرق توفيق المنحنيات
Methods of Curve Fitting

(٧-٤) طريقة العزوم
Moment´s Method

(٧-٥) طريقة المربعات الصغري
Least Square Method

(٧-٦) توفيق منحنيات تؤول إلى الخط المستقيم
Curve Fitting Tending to Linear Curve

(٧-٧) تمرينات
Exercises

(٧-١) العلاقة بين المتغيرات

Relation Between Variables

في الجزء الأول من هذا الكتاب تناولنا بالدراسة كيفية تحويل المعلومات المستخلصة من العينة عن المتغيرات محل الدراسة إلى معرفة رقمية باستخدام أساليب الاستدلال الإحصائي، ورغم أن المعرفة التي تم استخلاصها عن المتغيرات محل الدراسة تعتبر ذو أهمية لمتخذ القرار، فإنه يعتبر أيضاً معرفة شكل العلاقة بين المتغيرات محل الدراسة ذو أهمية بالغة في صناعة القرار حيث يمكن متخذ القرار من التنبؤ بقيم أحد المتغيرات عن طريق قيم المتغيرات الأخرى المرتبط بها.

فمثلاً

عادة تكون القيم المطلوبة من منتج معين مرتبط بسعر بيع الوحدة الواحدة من هذا المنتج فتعتبر الكميات المطلوبة متغير تابع وسعر بيع الوحدة متغير أساسي (مستقل) وبمعرفة شكل العلاقة بين الكميات المطلوبة والسعر فإنه يمكن تقدير الكميات المطلوبة عند سعر معين. فإذا أشرنا إلى المتغير الذي يمثل الكمية المطلوبة بالرمز (y) كذلك إذا أشرنا إلى المتغير الذي يمثل سعر الوحدة بالرمز (x) فإن y تعتبر دالة في المتغير x وتكتب على النحو التالي:-

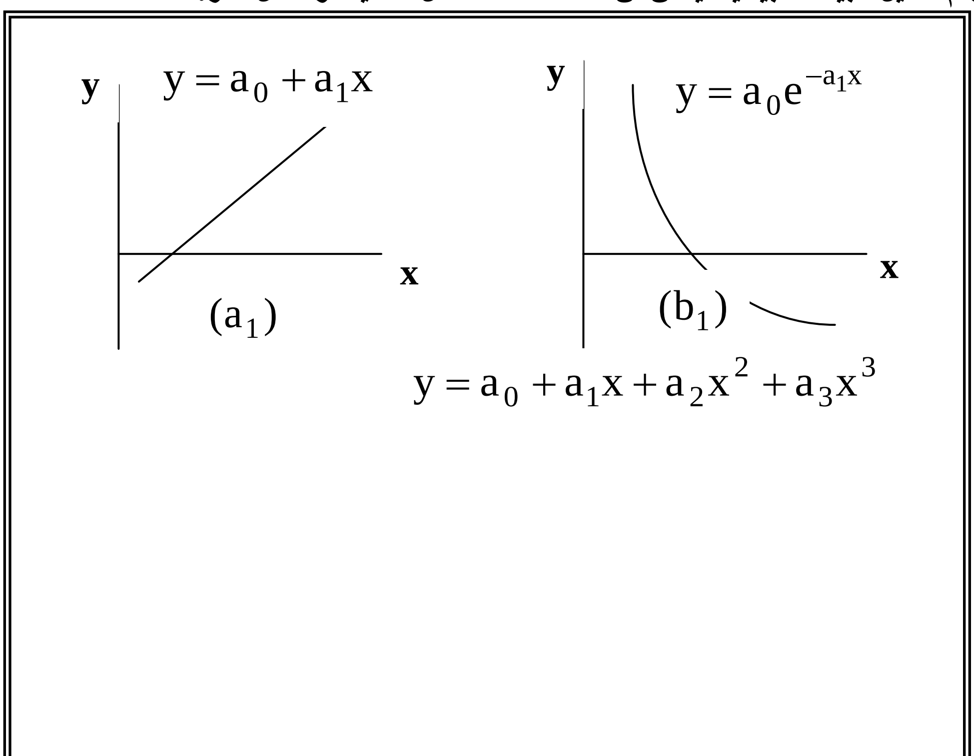
$$y = f(x) \qquad (7.1)$$

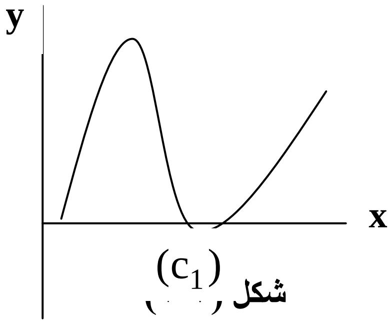
وبمعرفة صياغة الدالة f(x) يمكن تحديد الكميات المطلوبة (y) عند قيمة معينة للسعر (x)، لذا سوف نتناول في هذا الجزء الأساليب والطرق التي تمكننا من صياغة العلاقة بين متغيرين أو أكثر صياغة إحصائية أو تقدير هذه العلاقة.

إذا اعتبرنا أن x المتغير الأساسي (المستقل) Independent Variable، y المتغير التابع Dependent Variable وتم جمع بيانات عن قيم المتغير x والقيم المناظرة لها للمتغير y فكانت على النحو التالي:-

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$$

وتم تمثيل البيانات بيانياً فيمكن أن تأخذ أحد الأشكال التالية أو شكل أخر:-





شكل (c_1)

ومن الشكل (a_1) نجد أن العلاقة بين **x , y** علاقة خطية على النحو التالي:-

$$y = f(x) = a_0 + a_1 x \qquad (7.2)$$

بالمثل نجد أن جميع النقط تقع على المنحنى في الشكل (b_1) وبالتالي فإن العلاقة بين **x , y** علاقة أسية على النحو التالي:-

$$y = f(x) = a_0 e^{-a_1 x} \qquad (7.3)$$

وكذلك الشكل (c_1) العلاقة بين **x , y** علاقة من الدرجة الثالثة على النحو التالي:-

$$y = f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \qquad (7.4)$$

وتسمى بالنموذج غير الخطي من الدرجة الثالثة

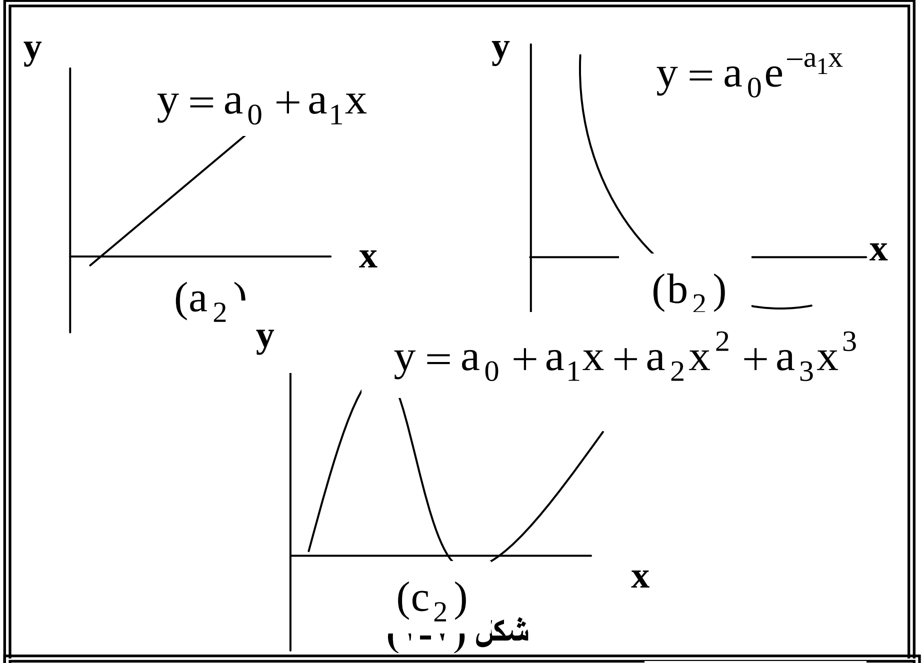
ونلاحظ بالنسبة لكل دالة من الدوال الموضحة في (7.4) : (7.2) أنه عند قيمة معينة من قيم المتغير ولتكن x_i توجد قيمة واحدة مناظرة لها للمتغير y_i لذا تسمى النماذج في (8.4) : (8.2) بالنماذج اليقينية **Deterministic Models.**

(٧-٢) شكل الانتشار **Scattor Diagram**

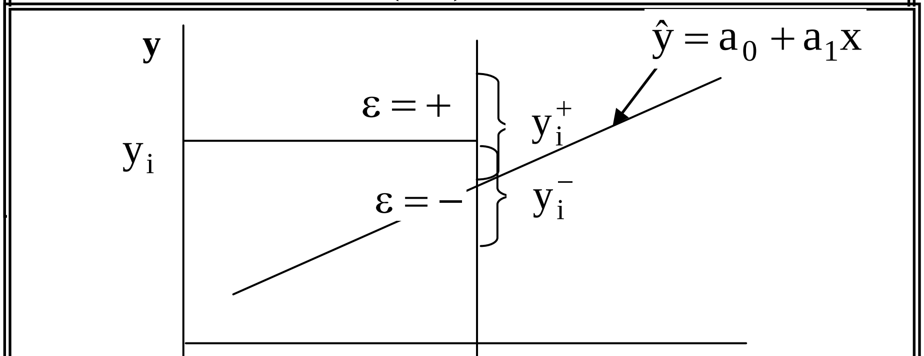
إذا كان المتغير التابع **y** مرتبط بالمتغير المستقل **x** وتم جمع بيانات عن قيم **x** والقيم المناظرة له للمتغير **y** فكانت على النحو التالي:-

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$$

فإذا تم تمثيل البيانات بيانياً وأخذت أحد الأشكال التالية الموضحة في الشكل (٧-٢) أو شكل أخر. ويسمى كل من الأشكال $(a_2),(b_2),(c_2)$ بشكل الانتشار **Scatter Diagram**. فنلاحظ من الأشكال التالية أنه توجد علاقة بين المتغير **y** والمتغير **x** ولكنها تختلف عن العلاقات في (٧-١) ، حيث نجد أن اتجاه العلاقة في الشكل (a_2) يأخذ الشكل الخطي ولكن بعض نقط الانتشار تقع على الخط والبعض الأخر يقع فوق أو تحت الخط فنجد عند نقطة معينة يأخذ المتغير **x** القيمة x_i نجد أنه يوجد عدة قيم مناظرة لها للمتغير التابع **y** ولتكن y_i التي تقع على الخط ، y_i^+ تقع فوق الخط ومناظرة لـ x_i ، y_i^- تقع تحت الخط ومناظرة لـ x_i أيضاً كما هو موضح بشكل (٧-٣).



شكل (٧-٢)



شكل (٧-٣)

ولكي نأخذ جميع النقط في شكل الانتشار في الاعتبار فإننا نضيف متغير حقيقي ونشير له بالرمز ε إلى الطرف الأيمن للعلاقة (8.2) فتصبح على النحو التالي

$$y = a_0 + a_1x + \varepsilon \qquad (7.5)$$

حيث ε متغير عشوائي Random Variable (وأحياناً يسمى بالخطأ Error) ويأخذ قيم موجبة للنقط التي تقع فوق الخط ، وقيم سالبة للنقط التي تقع تحت الخط ، ويأخذ القيمة صفر للنقط التي تقع على الخط ، وبمقارنة (7.2) بـ (7.5) نجد أن ε_i

هــو الفــرق بــين قيمــة y_i الفعليــة وقيمــة $\hat{y}_i$ الواقعــة علــى الخــط أي $\varepsilon_i = (y_i - \hat{y}_i)$.

بالمثل بالنسبة للعلاقات في شكل $(b_2),(c_2)$ حيث:-

$$y = a_0 e^{a_1 x} + \varepsilon \qquad (7.6)$$

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \varepsilon \qquad (7.7)$$

على الترتيب

وتسمى العلاقات في (7.7) - (7.5) بالنمـاذج الاحتماليـة Probabilistic Models ، حيث المتغير التابع y دالة في المتغير المستقل x والمتغير العشوائي ε أي أن

$$y = f(x, \varepsilon) \qquad (7.8)$$

ونجد أن الخط المستقيم (أي النموذج اليقيني) في شكل (a_2) :-

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x \qquad (7.9)$$

يعتبر تقدير Estimate للعلاقة في (8.5) وبالمثل النماذج التالية:

$$\hat{y} = \hat{a}_0 e^{-\hat{a}_1 x} \qquad (7.10)$$

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2 + \hat{a}_3 x^3 \qquad (7.11)$$

تعتبر تقديرات Estimates للعلاقات في (7.7) , (7.6) على الترتيب.

ونظراً لأنه في كثير من الأحيان يتم استخدام أسلوب العينـة فإنـه فـي حالـة سحب عينة عشوائية حجمها n من المجتمع محل الدراسة وتم جمع البيانات

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)$$

وأخذ شكل الانتشـار الاتجـاه الخطـي (مـثلاً) وباسـتخدام بيانـات العينـة تـم حسـاب قـيم $\hat{a}_0, \hat{a}_1$ لتقدير الخط المستقيم الذي يمثل العلاقة في العينة فكانت على النحو

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x$$

فإن العلاقة (7.9) تعتبر تقدير للعلاقة بين x , y في المجتمع المسحوب منه العينة.

ومن ثم تتلخص المشكلة في تقدير العلاقة بين المتغيرين x , y باستخدام بيانات العينة ثم إجراء الاستدلال الإحصائي لهذه العلاقة المقدرة ، وهو ما يتم دراسته بالتفصيل في هذا الباب والأبواب التالية.

ومما سبق يتضح التالي:-

١- الصياغة الرياضية للعلاقة بين متغيرين أو أكثر تسمي بالنموذج **Model**.

٢- يتم افتراض الشكل الرياضي للعلاقة (النموذج) عن طريق شكل الانتشار.

٣- تنقسم النماذج إلى مجموعتين هما:-

أ- نماذج يقينية.

ب- نماذج احتمالية.

النموذج اليقيني يمثل العلاقة بين متغيرات يقينية**Deterministic** أما النموذج الاحتمالي فهو يمثل العلاقة بين المتغيرات ويكون أحد هذه المتغيرات متغير عشوائي.

٤- النموذج اليقيني الذي يتم اشتقاقه باستخدام بيانات العينة يعتبر تقدير للعلاقة بين المتغيرات محل الدراسة في المجتمع المسحوب منه العينة (أي تقدير للنموذج الاحتمالي في المجتمع).

(٧-٣) طرق توفيق المنحنيات

Methods of Curve Fitting

وتسمى الطرق التي يمكن باستخدامها تقدير العلاقة الرياضية (النموذج) بين متغيرين سواء باستخدام بيانات المجتمع أو بيانات العينة بطرق توفيق المنحنيات**Curve Fitting**.

وتوجد طرق متعددة لتوفيق المنحنيات منها على سبيل المثال:-

١- طريقة العزوم **Moment Method**

٢- طريقة المربعات الصغرى **Least Square Method**

٣- طريقة أقل مجموع انحرافات **Lest Sum Absolute Method**

٤- طريقة الإمكان الأكبر Maximum Likelihood Method

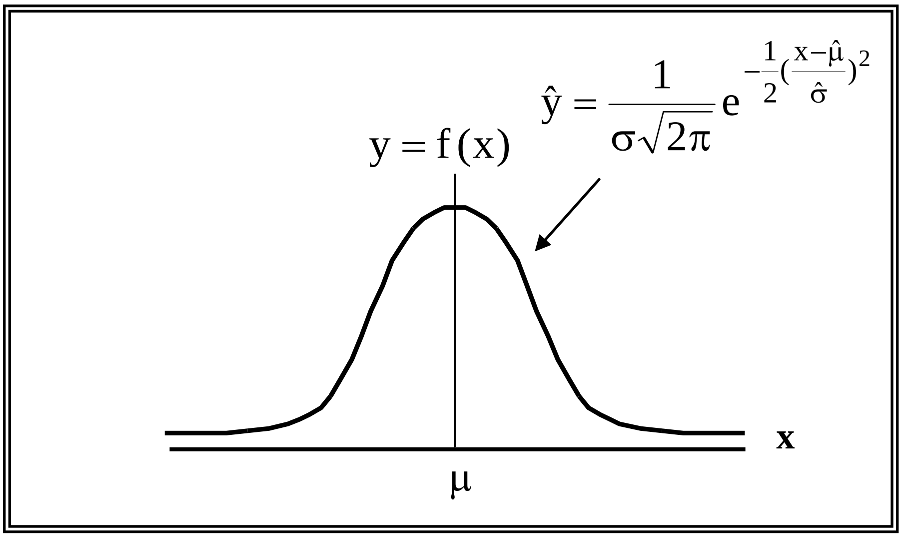
ويتم تحديد الطريقة المستخدمة في تقدير العلاقة إلى طبيعة المتغيرات محل الدراسة (متصلة، متقطعة، ...، الخ) إلى طبيعة العلاقة المفترضة (خطية، غير خطية، ...، الخ) ، مدى توافر البيانات عن متغيرات العلاقة ، خصائص التقديرات التي يتم الحصول عليها باستخدام العلاقة المقدرة ، وسوف نوضح ذلك بالتفصيل في الفصول التالية.

ومما هو جدير بالذكر انه يمكن في بعض الحالات استخدام أكثر من طريقة لتقدير العلاقة ونحصل على نفس التقدير فمثلاً إذا استخدمت طريقة المربعات الصغرى لتقدير العلاقة الخطية سوف نصل إلى نفس العلاقة المقدرة التي نحصل عليها باستخدام طريقة الإمكان الأكبر.وسوف تقتصر دراستنا في هذا الباب على طريقة العزوم وطريقة المربعات الصغرى.

(٧-٤) طريقة العزوم Moment Method

في الكثير من الحالات قد يكون المتغير التابع y هو عبارة عن السلوك الاحتمالي للمتغير الأساسي x أي $y = f(x)$ فقد تكون $f(x)$ دالة الاحتمال للمتغير x إذا كان x متغير متقطع أو قد تكون $f(x)$ دالة كثافة الاحتمال للمتغير x إذا كان x متغير متصل.

ومثال ذلك إذا أخذت بيانات عن x ، f(x) وتم تمثيل البيانات في شكل الانتشار التالي:-



شكل (٧-٤)

فمن الشكل نجد أن أنسب صياغة رياضية تمثل البيانات هي المنحنى الطبيعي

$$\hat{y} = \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\hat{\mu}}{\hat{\sigma}})^2} \qquad -\infty < x < \infty$$

فهنا نجد أن $\hat{y}$ دالة صريحة في $\hat{\mu}$ ، $\hat{\sigma}^2$ وهما تقديرين للتوقع والتباين على الترتيب للمتغير x حيث يعتبر كل من $\hat{\mu}$ ، $\hat{\sigma}^2$ تقديرات للعزم الأول غير المركزي μ والعزم الثاني المركزي σ^2 للمتغير x.

ويمكن أن يتم توفيق المنحنيات عن طريق مساواة العزوم المحسوبة من بيانات العينة بالعزوم المناظرة لها في المجتمع وتسمى هذه الطريقة بطريقة العزوم.

لذلك سوف نوضح باختصار تعريف العزوم ثم نوضح كيفية استخدام طريقة العزوم في توفيق المنحنيات من خلال الأمثلة التالية.

العزوم غير المركزية

إذا كان x متغير عشوائي فإن العزم من الترتيب r غير المركزي للمتغير x ويرمز له بالرمز μ_r يعرف على النحو التالي:-

$$\mu_r = E(x^r) \qquad (7.12)$$

فإذا كان x متغير متقطع فإن:-

$$\mu_r = \sum_x x^r P_r(x) \qquad (7.13)$$

حيث $P_r(x)$ دالة الاحتمال للمتغير x كذلك إذا كان x متغير متصل فإن:-

$$\mu_r = \int_x x^r f(x)dx \qquad (7.14)$$

حيث $f(x)$ هي دالة كثافة الاحتمال للمتغير x ويلاحظ أن العزم الأول غير المركزي للمتغير x هو توقع المتغير x حيث

$$\mu_1 = E(x) = \mu \qquad (7.15)$$

العزوم المركزية

إذا كان μ هي القيمة المتوقعة للمتغير x فإن العزم المركزي من الترتيب r يرمز له بالرمز μ'_r ويعرف على النحو التالي:-

$$\mu'_r = E(x-\mu)^r \qquad (7.16)$$

فإذا كان x متغير متقطع فإن:-

$$\mu'_r = E(x-\mu)^r = \sum_x (x-\mu)^r P_r(x) \qquad (7.17)$$

كذلك إذا كان x متغير متصل فإن:-

$$\mu'_r = E(x-\mu)^r = \int_x (x-\mu)^r f(x)dx \qquad (7.18)$$

ويلاحظ أن العزم الثاني المركزي μ'_2 يساوي التباين σ^2 حيث:-

$$\mu'_2 = E(x-\mu)^2 = \sigma^2 \tag{7.19}$$

مثال (٧-١)

إذا ألقيت 4 قطع نقود عشوائيا 60 مرة وفى كل مرة يتم تسجيل النتائج التي تظهر على سطح القطع ، والجدول التالي يوضح نتائج الرمي , فإذا كانت x تشير إلى الصور التي تظهر على القطع الأربعة في الرمية الواحدة حيث x=0,1,2,3,4 , كذلك f_i تمثل عدد مرات الرمي التي يأخذ فيها المتغير x القيمة x_i (أي التكرارات المناظرة للقيمة x_i).

جدول (٧-١)

عدد الصور x_i	0	1	2	3	4	$\sum$
عدد مرات الرمي (التكرارات f_i)	3	5	6	16	30	60

<u>المطلوب</u>

١- أوجد التوزيع التكراري النسبي للمتغيرx ووضح ذلك بيانيا.

٢- من الرسم في (١) وفق دالة الاحتمال المناسبة للمتغيرx باستخدام طريقة العزوم.

٣- كون جدول التوزيع الاحتمالى للمتغير في (٢) ووضح ذلك بيانيا ثم قارن هذا التوزيع بالتوزيع في(١).

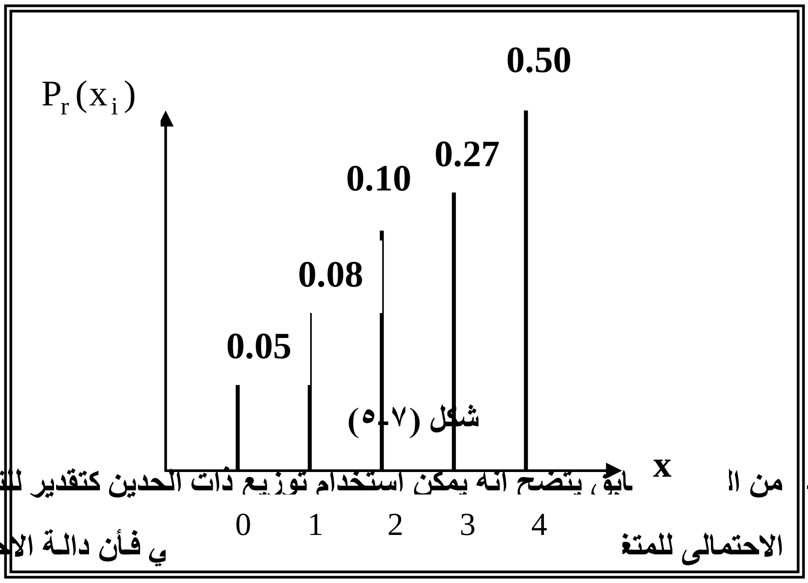
<u>الحل</u>

١- الجدول التالي يعطى التكرار النسبي للمتغير x والذي يمثل احتمال المتغير x عند القيمة المناظرة , وسوف نشير إليه بالرمز $P_r(x_i)$:-

جدول (٧-٢)

x_i	0	1	2	3	4	$\sum$
$P_r(x_i)$	**0.05**	**0.08**	**0.10**	**0.27**	**0.50**	**1**

والشكل التالي يوضح هذا التوزيع:-



شكل (٧-٥)

٢- من ال ابقة، يتضح انه يمكن استخدام توزيع ذات الحدين كتقدير للتوزيع الاحتمالى للمتغ ي فأن دالة الاحتمال المقدرة $P_r = (x)$ تأخذ الصيغة التالية:-

$$P_r(x) = C_x^n P^x (1-P)^{n-x} \quad , x = 1,2,3,4 \qquad (7.20)$$

ونجد أن المعلمة المجهولة في الدالة (8.20) هي (P) أي معلمة واحدة ، لذا باستخدام العزوم سوف نساوي العزم الأول غير المركزي في العينة $\hat{\mu}$ بالعزم الأول غير المركزي في المجتمع μ للحصول على تقدير للمعلمة (P) وسوف يتم ذلك على النحو التالي:-

١- من جدول (٧-٢) نحسب قيمة $\hat{\mu}$ كما يلي:-

$$\hat{\mu} = E(x_i)$$
$$= 0(\frac{3}{60}) + 1(\frac{5}{60}) + 2(\frac{6}{60}) + 3(\frac{16}{60}) + 4(\frac{30}{60}) = 3.083 \qquad (7.21)$$

٢- وبما أن القيمة المتوقعة لمتغير ذات الحدين في التوزيع (8.20) يساوي μ حيث:-

$$\boxed{\mu = nP \qquad (7.22)}$$

٣- بمساواة التوقع في العينة $\hat{\mu}$ في (7.21) بالتوقع في المجتمع μ في (7.22) نجد أن:-

$$\mu = \hat{\mu} \rightarrow nP = 3.083 \rightarrow P = 0.771 \approx 0.8$$

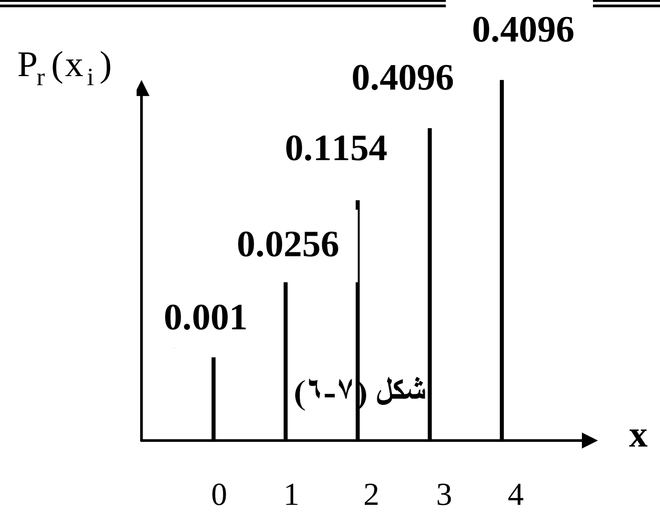
وبالتالي يصبح تقدير دالة الاحتمال $P_r(x)$ والتي نرمز لها بالرمز $\hat{P}_r(x)$ على النحو التالي:-

$$\hat{P}_r(x) = C_x^4 (0.8)(0.2)^{4-x} \quad , x = 1,2,3,4 \qquad (7.23)$$

٤- وباستخدام توزيع ذات الحدين في (7.23) نكون الجدول التالي:-

جدول (٧-٣)

x_i	0	1	2	3	4	$\sum$
$P_r(x_i)$	**0.0016**	**0.0256**	**0.11536**	**0.4096**	**0.4096**	**1**



شكل (٧-٦)

وشكل (٧-٦) يعطي تقدير لدالة الاحتمال للمتغير x (النموذج أو التوفيق) المبنية على بيانات العينة ، وبمقارنة الشكل البياني للتوزيع الاحتمالي للعينة في شكل (٧-٥) مع الشكل البياني للتوزيع في شكل (٧-٦) نجد أن كل توزيع من التوزيعين ملتوي ناحية اليسار* حيث:-

$$P = 0.80 > 0.05$$

مثال (٧-٢)

الجدول التالي يوضح سعر الوحدة الواحدة من سلعة معينة بالجنيه والكميات المعروضة من هذه السلعة (بالألف وحدة)

جدول (٧-٤)

فئات السعر بالجنيه (x)	20-	25-	30-	35-	40-45	$\sum$
الكميات المعروضة (بالألف وحدة)	7	25	33	27	8	100

<u>المطلوب</u>

١- كون المنحنى التكراري النسبي ثم وضح ذلك بيانياً.

٢- وفق دالة كثافة الاحتمال للسعر (x) بافتراض أن السعر متغير معتاد (طبيعي) باستخدام طريقة العزوم ثم وضح ذلك بيانياً.

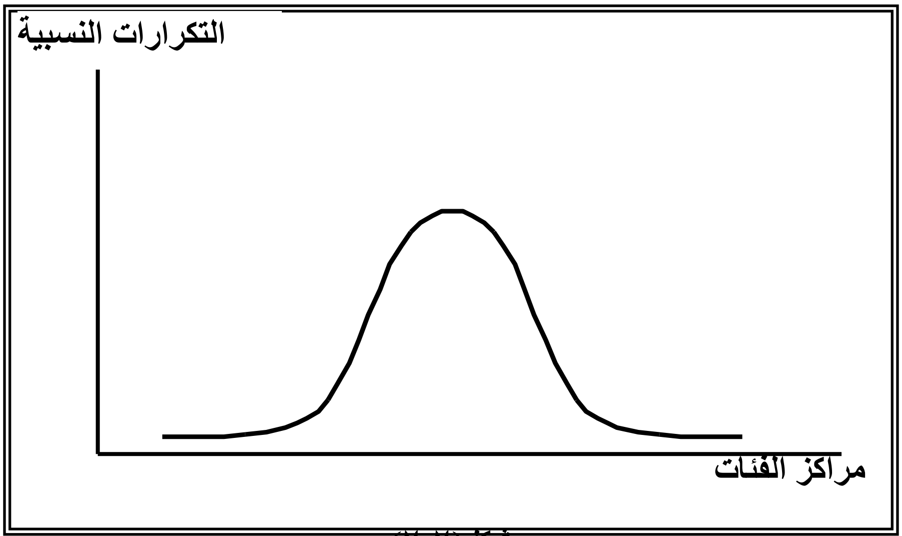
<u>الحل</u>

١- الجدول التالي يوضح التكرارات النسبية كذلك مراكز الفئات المناظرة لها.

* أ.د. عفاف الدش (٢٠٠٥): " الإحصاء التطبيقي" – الجزء الأول – الطبعة السادسة – جهاز نشر وتوزيع الكتاب الجامعي - جامعة حلوان.

جدول (٧-٥)

الفئات	التكرارات f_i	مراكز الفئات x_i	التكرارات النسبية
20-	**7**	**(20+25)/2=22.5**	**0.07**
25-	**25**	**(25+30)/2=27.5**	**0.25**
30-	**33**	**(30+35)/2=32.5**	**0.33**
35-	**27**	**(35+40)/2=37.5**	**0.27**
40-45	**8**	**(40+45)/2=42.5**	**0.08**
$\sum$	**100**		**1**



شكل(٧-٧)

٢- من شكل (٧-٧) نجد أن المنحنى التكراري النسبي لتكرارات عينة الأسعار يقترب من التوزيع المعتاد . ولتوفيق (تقدير) دالة كثافة الاحتمال للمتغير x (السعر) بدالة كثافة الاحتمال للمتغير المعتاد f(x) حيث:-

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \qquad -\infty < x < \infty \qquad (7.24)$$

نجد أن دالة كثافة الاحتمال f(x) تتضمن معلمتان هما التوقع μ والانحراف المعياري σ للمتغير x ويمكن تقدير قيمة μ و σ من بيانات العينة (بيانات جدول (٧-٤)) على النحو التالي:-

جدول (٧-٦)

x_i مراكز الفئات	f_i التكرارات	$x_i f_i$	$x_i^2 f_i$
22.5	**7**	**157.5**	**3543.75**
27.5	**25**	**687.5**	**18906.25**
32.5	**33**	**1072.5**	**34856.25**
37.5	**27**	**1012.5**	**37968.75**
42.5	**8**	**340**	**14450**
$\sum$	**100**	**3270**	**109725**

من الجدول نحسب كل من التوقع والتباين للعينة فنجد أن:-

$$\hat{\mu} = \sum x_i P_r(x_i) = \frac{3270}{100} = 32.7 \qquad (7.25)$$

$$\hat{\sigma}^2 = \sum (x_i - \mu)^2 P_r(x_i) = \frac{\sum x_i^2 f_i}{\sum f_i} - (\hat{\mu})^2$$

$$= \frac{109725}{100} - (32.7)^2 = 1097.254 - 1069.29 = 27.96 \qquad (7.26)$$

$$\hat{\sigma} = \sqrt{27.96} = 5.29$$

وبمساواة عزوم العينة بعزوم المجتمع نجد أن:-

$$\mu = \hat{\mu} = 32.7$$

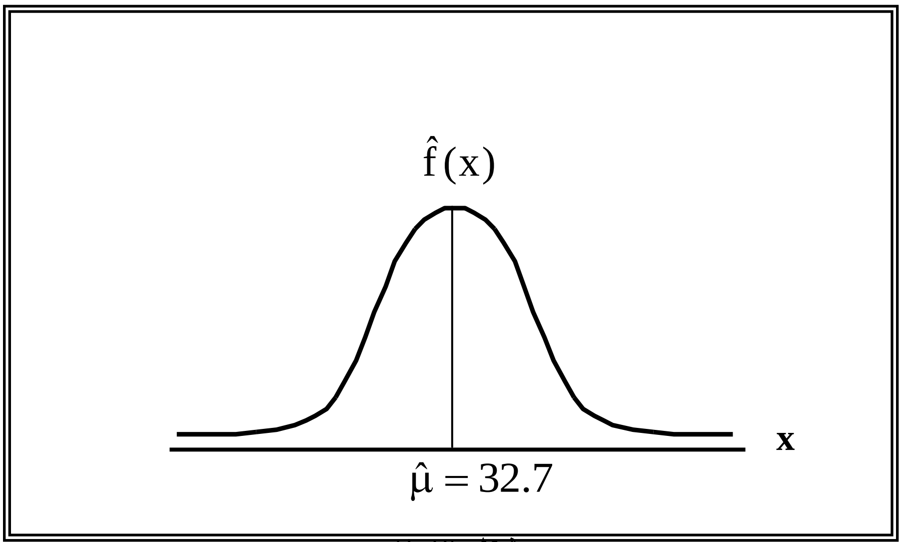
$$\sigma = \hat{\sigma} = 5.29$$

وبالتالي تصبح دالة كثافة الاحتمال المقدرة التي تم توفيقها على النحو التالي:-

$$\hat{f}(x) = \frac{1}{5.29\sqrt{2\pi}} e^{-\frac{1}{55.92}(x-32.7)^2} \qquad (7.27)$$

$$\pi = \frac{22}{7}$$

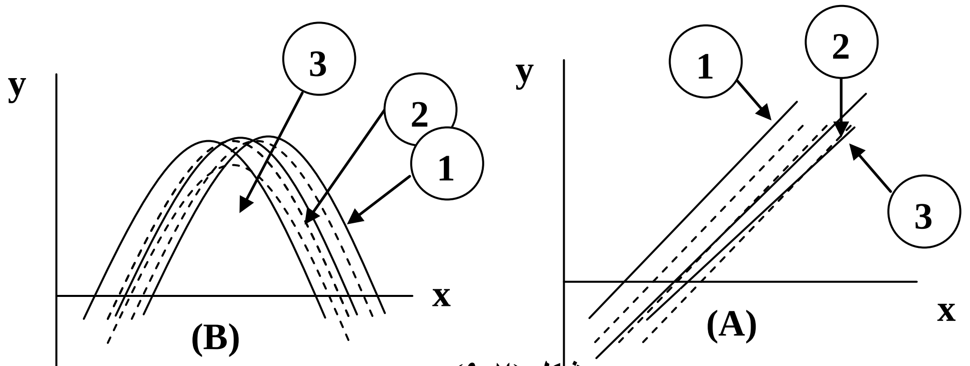
والشكل التالي يوضح التوزيع المقدر في العلاقة (7.27)



شكل (٧-٨)

(٧-٥) طريقة المربعات الصغرى
Least Square Method

مما سبق يتضح أن شكل الانتشار يوضح فقط اتجاه العلاقة بين المتغير التابع والمتغير المستقل فمثلاً إذا كان اتجاه بين المتغير التابع y والمتغير المستقل x اتجاه خطي أو منحنى كما في الشك (A) ، (B) بشكل (٧-٩) . ولكن شكل الانتشار لا يعطي أفضل خط مستقيم في (A) أو أفضل منحنى في (B) بشكل (٧-٩).



شكل (٧-٩)

ويصبح السؤال ما هو أفضل خط مستقيم في شكل (A) هل المستقيم 1 أو 2 أو 3 أو خط أخر - بالمثل في شكل (B) ما هو أفضل منحنى يمثل العلاقة هل المنحنى 1 أو 2 أو 3 أو منحنى أخر.

فإذا اعتبرنا أن العلاقة بين المتغير التابع y والمتغير المستقل x على النحو التالي:-

$$y = f(x \mid a) + \varepsilon \qquad (7.28)$$

حيث تشير f(x|a) إلى دالة في المتغير المستقل (أو المتغيرات المستقلة) والتي تعتمد على المعلمات (المقادير الثابتة) a ، وتشير ε إلي مجموعة المتغيرات العشوائية المتجانسة (أي التي لها نفس التباين) التي تؤثر على المتغير التابع Y فقط ولا تؤثر على المتغير (أو المتغيرات) المستقل X.

إذا فرضنا أن أفضل منحنى يمثل العلاقة (7.28) هو $\hat{y}$ حيث:-

$$\hat{y} = f(x \mid a) \qquad (7.29)$$

والأفضلية هنا تعني أن المنحنى $\hat{y}$ هو المنحنى الذي يتوسط نقط الانتشار بحيث يكون مجموع مربعات انحرافات القيم الفعلية للمتغير التابع y عن القيم المقدرة لها $\hat{y}$ (أي المناظرة لها على المنحنى) أقل ما يمكن أو بعبارة أخرى فإن المنحنى $\hat{y}$ في (7.29) هو المنحنى الذي يجعل المقدار E حيث

$$\begin{aligned} E &= \sum (y_i - \hat{y}_i)^2 = \sum \varepsilon_i^2 \\ &= \sum [y - f(x \mid a)]^2 \qquad (7.30) \end{aligned}$$

وتبنى طريقة المربعات الصغرى على إيجاد التقديرات $\hat{a}$ لمعلمات الدالة f(x|a) التي تجعل المقدار E $(\sum \varepsilon^2)$ اقل ما يمكن ولذا سميت بطريقة المربعات الصغرى حيث يتم الحصول على هذه التقديرات.

وسوف نوضح كيفية الحصول على أفضل منحنى باستخدام طريقة المربعات الصغرى بالنسبة لبعض العلاقات على النحو التالي:-

أولاً : نموذج الانحدار الخطي

نظرية (٧-١): إذا كانت العلاقة بين x , y على النحو التالي:-

$$y = a_0 + a_1 x + \varepsilon$$

أو (7.30)

$$y_i = a_0 + a_1 x_i + \varepsilon_i \qquad , i = 1,2,\dots,n$$

بحيث توقع المتغير ε يساوي صفر أي أن :-

$$E(\varepsilon)=0 \qquad (7.31)$$

وتباين المتغير ε يساوي σ^2 أي أن:-

$$Var(\varepsilon)=\sigma^2 \qquad (7.32)$$

كذلك تغاير $\varepsilon_i,\varepsilon_j$ يساوي صفر أي أن :-

$$Cov(\varepsilon_i,\varepsilon_j)=0 \qquad ,i\neq j \qquad (7.33)$$

فإن أفضل مستقيم يمثل العلاقة (7.30) باستخدام بيانات عينة عشوائية حجمها n هو:-

$$\hat{y}=\hat{a}_0+\hat{a}_1 x \qquad (7.34)$$

حيث يتم حساب $\hat{a}_0,\hat{a}_1$ كتقدير للمعلمة a_0,a_1 على الترتيب من حل المعادلتين التاليين:-

$$\sum_{i=1}^{n} y_i = n\hat{a}_0+\hat{a}_1\sum_{i=1}^{n} x_i \qquad (7.35)$$

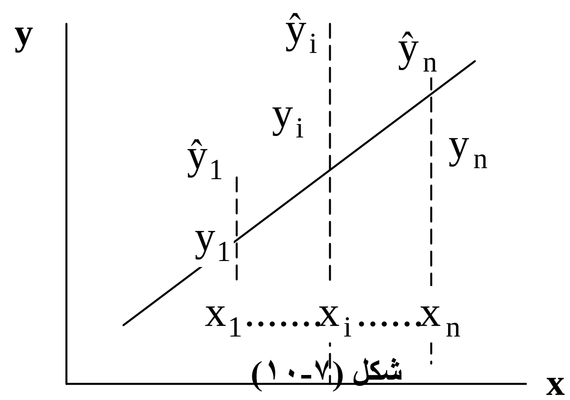
$$\sum_{i=1}^{n} x_i y_i = \hat{a}_0\sum_{i=1}^{n} x_i+\hat{a}_1\sum_{i=1}^{n} x_i^2 \qquad (7.36)$$

وتسمى المعادلتين (7.36) , (7.35) بالمعادلات الطبيعية **normality Equation**.

<u>**الإثبات:**</u> وكما ذكرنا سابقاً أن أفضل مستقيم أي المستقيم الذي يكون مجموع مربعات انحرافات القيم الفعلية y عن القيم المقدرة $\hat{y}_i$ باستخدامه اقل ما يمكن أي المستقيم الذي يجعل المقدار ε اقل ما يمكن حيث:-

$$E=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)=\sum_{i=1}^{n}(y_i-\hat{a}_0-\hat{a}_1x_i)^2 \quad (7.37)$$

انظر شكل (٧-١٠)



شكل (٧-١٠)

وبالتالي تصبح المشكلة تحديد قيم $\hat{a}_0,\hat{a}_1$ التي تجعل الدالة E نهاية صغيرة* Minimum Value ويمكن إيجاد قيمة النهاية الصغرى للدالة E في (7.37) بإيجاد المعاملات التفاضلية الجزئية* للدالة E بالنسبة لكل من $\hat{a}_0,\hat{a}_1$ ومساواة هذه المعاملات بالصفر (حيث تساوي هذه المعاملات الصفر عندما تكون E نهاية صغري) فنحصل على التقديرات المطلوبة على النحو التالي:-

$$\frac{\partial E}{\partial \hat{a}_0}=-2\sum_{i=1}^{n}(y_i-\hat{a}_0-\hat{a}_1x_i)=0 \qquad (7.38)$$

$$\frac{\partial E}{\partial \hat{a}_1}=-2\sum_{i=1}^{n}x_i(y_i-\hat{a}_0-\hat{a}_1x_i)=0 \qquad (7.39)$$

* أ.د. عفاف الدش (١٩٩٩) "رياضيات العمال" جهاز نشر وتوزيع الكتاب الجامعي – جامعة حلوان - القاهرة

وبقسمة طرفي كل من المعادلتين (7.39) , (7.38) على (-2) نحصل عل المعادلتين الطبيعيتين على النحو التالي:-

$$\sum_{i=1}^{n} y_i = n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^{n} x_i \qquad (7.40)$$

$$\sum_{i=1}^{n} x_i y_i = \hat{a}_0 \sum_{i=1}^{n} x_i + \hat{a}_1 \sum_{i=1}^{n} x_i^2 \qquad (7.41)$$

وبحل المعادلتين الطبيعيتين نحصل على قيم $\hat{a}_0, \hat{a}_1$ **كأفضل تقديرات لكل من** a_0, a_1 **في الدالة (7.34).**

نتيجة (١)

من المعادلتين الطبيعيتين نجد أن:-

$$\hat{a}_0 = \frac{\sum_{i=1}^{n} y_i}{n} - \hat{a}_1 \frac{\sum_{i=1}^{n} x_i}{n}$$

أو

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} \qquad (7.42)$$

حيث $\bar{y}, \bar{x}$ **هما الوسط الحسابي للمتغير x والوسط الحسابي للمتغير y على الترتيب كذلك:-**

$$\hat{a}_1 = \frac{n\sum_{i=1}^{n} xy - (\sum_{i=1}^{n} x)(\sum_{i=1}^{n} y)}{n\sum_{i=1}^{n} x^2 - ((\sum_{i=1}^{n} x))^2} \qquad (7.43)$$

مثال(٧-٣)

الجدول التالي يعطي قيم المتغير المستقل x والقيم المناظرة له للمتغير التابع y.

جدول (٧-٩)

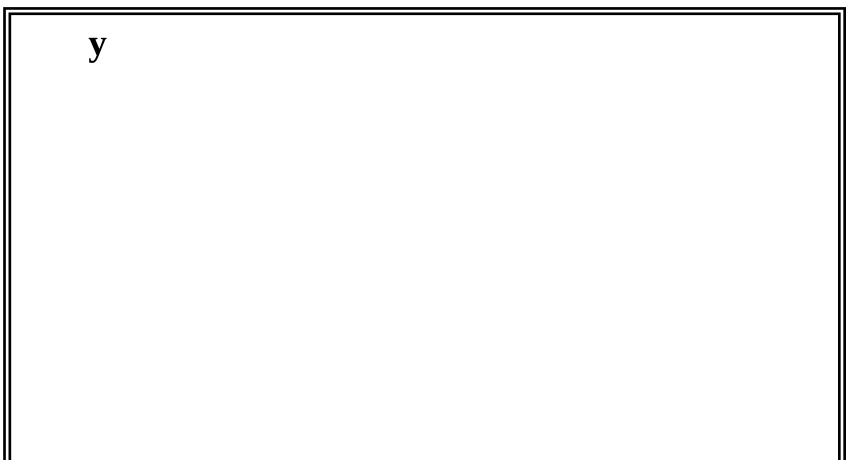
X	0	2	3	4	5	10
y	2	2.4	2.6	3	3	4

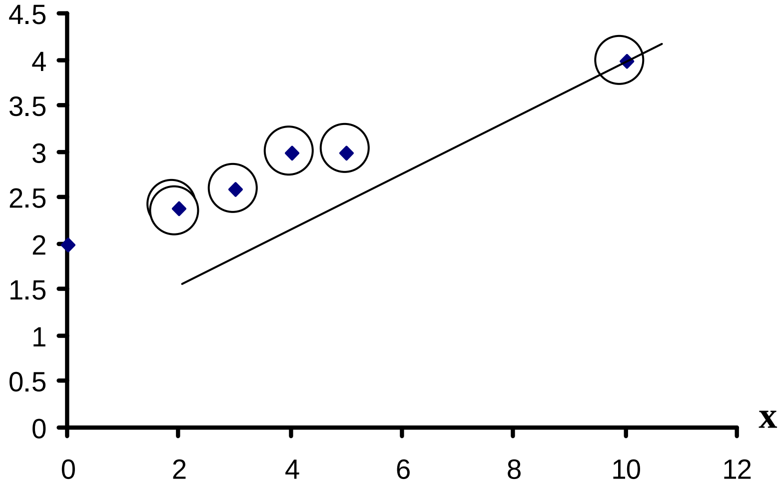
المطلوب

1. **ارسم شكل الانتشار الذي يوضح العلاقة بين x ,y .**
2. **أوجد باستخدام طريقة المربعات الصغرى نموذج انحدار y على x ووضح ذلك بيانياً.**
3. **باستخدام النموذج في (٢) قدر قيمة y عندما تكون x=15.**

الحل

1. **الشكل التالي يوضح العلاقة بين x , y**





شكل(٧-١١)

٢- من الشكل يتضح أنه يمكن افتراض العلاقة الخطية بين y , x على النحو التالي:-

$$y = a_0 + a_1 x + \varepsilon$$

وبالتالي فإن نموذج الانحدار y على x يصبح على النحو التالي:-

$$y = \hat{a}_0 + \hat{a}_1 x$$

ولحساب قيم $\hat{a}_0, \hat{a}_1$ كتقديرات للمعلمتين a_0, a_1 باستخدام طريقة المربعات الصغرى نكون الجدول التالي:-

جدول (٧-١٠)

X	Y	Xy	x^2
0	2	0	0
2	2.4	4.8	4
3	2.6	7.8	9
4	3	12	16
5	3	15	25
10	4	40	100
$\sum x = 24$	$\sum y = 14$	$xy = 79.6$	$\sum x^2 = 154$

وبما أن المعادلات الطبيعية هي:-

$$\sum_{i=1}^{n} y_i = n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i y_i = \hat{a}_0 \sum_{i=1}^{n} x_i + \hat{a}_1 \sum_{i=1}^{n} x_i^2$$

بالتعويض من الجدول السابق في المعادلات الطبيعية نجد أن :-

$$17 = 6\hat{a}_0 + 24\hat{a}_1 \qquad (7.44)$$

$$79.6 = 24\hat{a}_0 + 154\hat{a}_1 \qquad (7.45)$$

وبحل المعادلتين (7.45) , (7.44) نجد أن:-

$$\hat{a}_0 = 2.03 \quad , \hat{a}_1 = 0.2$$

وبالتعويض في النموذج الخطي نجد أن:-

$$\hat{y} = 2.03 + 0.2x \qquad (7.46)$$

وشكل (٧-١١)يوضح الخط $\hat{y} = 2.03 + 0.2x$

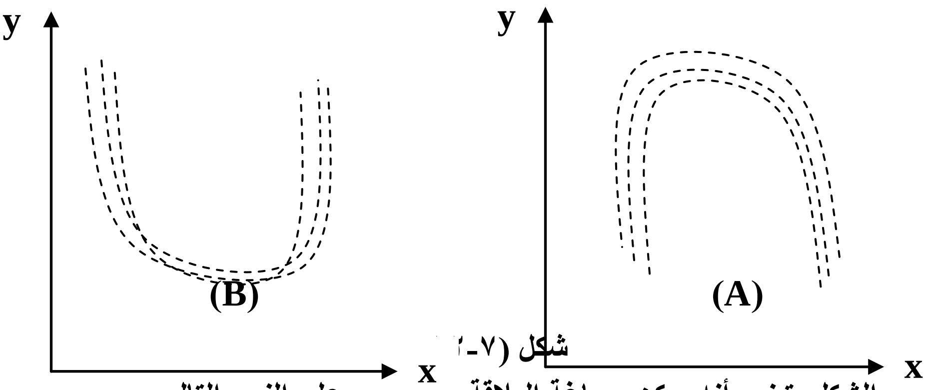
٣- عندما x=15 فإن:-

$$\hat{y} = 2.03 + 0.2(15) = 5.03$$

ثانياً : نموذج انحدار الدرجة الثانية

Quadratic Regression Model

إذا تم جمع البيانات من عينة حجمها n عن قيم المتغير الأساسي x والقيم المناظرة لها للمتغير التابع y ورسمنا شكل الانتشار فاخذ أحد الشكلين التاليين:-



شكل (٧-٢)

ومن الشكل يتضح أنه يمكن صياغة العلاقة بين x , y على النحو التالي:-

$$y_i = a_0 + a_1 x + a_2 x^2 + \varepsilon \qquad (7.47)$$

أو

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 + \varepsilon_i \qquad (7.48)$$

حيث a_0, a_1, a_2 **معلمات (مقادير ثابتة) ، ε مجموعة من المتغيرات العشوائية المستقلة تؤثر فقط على المتغير التابع y توقعها يساوي صفر وتباينها مقدار ثابت وليكن** γ^2.

ومن الشكل السابق نجد أن أفضل تقدير للعلاقة y=f(x|a) في (7.47) هي:-

$$\hat{y}_i = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2 \qquad (7.49)$$

نظرية(٧-٢): إذا كانت العلاقة بين x , y على النحو التالي:-

$$y = a_0 + a_1 x + a_2 x^2 + \varepsilon$$

أو

$$y_i = a_0 + a_1 x_i + a_2 x_i^2 + \varepsilon_i \qquad , i = 1,2,...,n$$

بحيث:-

$$E(\varepsilon) = 0 \quad , Var(\varepsilon) = \gamma^2 \quad , Cov(\varepsilon_i , \varepsilon_j) = 0 \qquad (7.50)$$

فإن أفضل منحنى يمثل العلاقة (7.49) باستخدام بيانات عينة حجمها n هو:-

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2$$

حيث يتم حساب $\hat{a}_0, \hat{a}_1, \hat{a}_2$ كتقديرات للمعلمات a_0, a_1, a_2 على الترتيب من

المعادلات الطبيعية التالية :-

$$\sum y = n\hat{a}_0 + \hat{a}_1 \sum x + \hat{a}_2 \sum x^2 \qquad (7.51)$$

$$\sum xy = \hat{a}_0 \sum x + \hat{a}_1 \sum x^2 + \hat{a}_2 \sum x^3 \qquad (7.5\ 2)$$

$$\sum x^2 y = \hat{a}_0 \sum x^2 + \hat{a}_1 \sum x^3 + \hat{a}_2 \sum x^4 \qquad (7.53)$$

الإثبات

بنفس طريقة إثبات نظرية (٧-١)

مثال (٧-٤)

الجدول التالي يوضح صافي الربح y (بالالف جنيه) لاحدى المنشآت و حجم المنشأة بالالف جنيه (x) .

جدول (٧-١١)

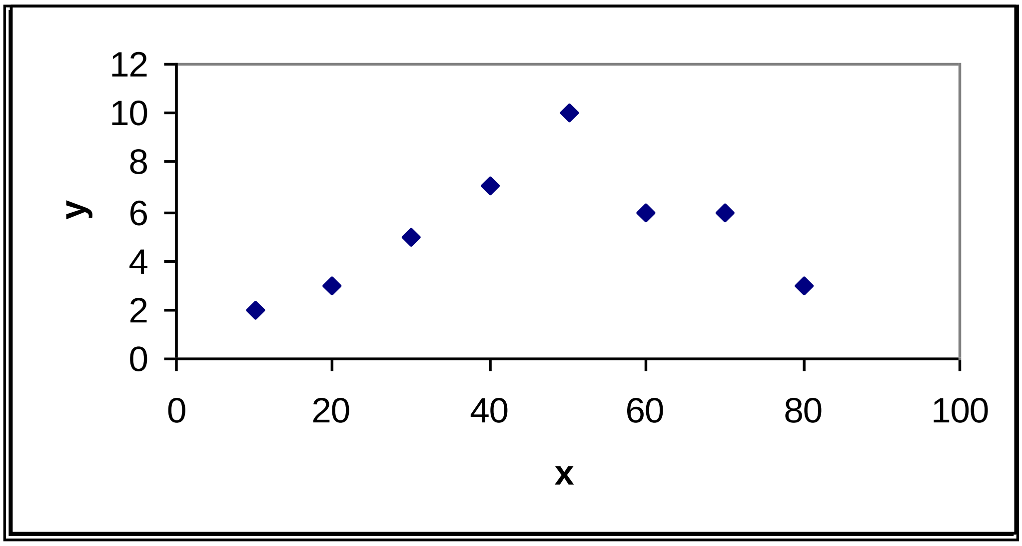
حجم المنشأة (x) بالألف جنيه	10	20	30	40	50	60	70	80
صافي الربح (y) بالألف جنيه	2	3	5	7	10	6	6	3

المطلوب

١- ارسم شكل الانتشار الذى يوضح العلاقة بين x,y .

٢- باستخدام طريقة المربعات الصغرى ابني نموذج انحدار مناسب (أو وفق منحنى انحدار)يمثل العلاقة بين x,y .

٣- قدر قيمة y عندما x=90 , x=80.



شكل (٧-١٣)

٢-من الشكل يمكن افتراض أن العلاقة بين x , y تأخذ الصياغة التالية:-

$$y = a_0 + a_1 x + a_2 x^2$$

وبالتالي فإن نموذج انحدار y على x يصبح على النحو التالي:-

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x + \hat{a}_2 x^2$$

ولحساب قيم $\hat{a}_0, \hat{a}_1, \hat{a}_2$ كتقدير للمعلمات a_0, a_1, a_2 نكون الجدول التالي:-

جدول (٧-١٢)

x	y	x^2	x^3	x^4	xy	x^2y
10	2	100	1000	10000	20	200
20	3	400	8000	160000	60	1200
30	5	900	27000	810000	150	4500
40	7	1600	64000	2560000	280	11200
50	10	2500	125000	6250000	500	25000
60	6	3600	216000	12960000	360	21600
70	6	4900	343000	24010000	420	29400
80	3	6400	512000	40960000	240	19200
360	42	20400	1296000	87720000	2030	112300

وبالتعويض في المعادلات الطبيعية (7.53) : (7.51) من الجدول السابق نحصل على المعادلات التالية:-

$$42 = 8\hat{a}_0 + 360\hat{a}_1 + 20400\hat{a}_3 \qquad (7.54)$$

$$2030 = 360\hat{a}_0 + 20400\hat{a}_1 + 1296000\hat{a}_2 \qquad (7.55)$$

$$1123000 = 20400\hat{a}_0 + 1296000\hat{a}_1 + 87720000\hat{a}_3 \qquad (7.56)$$

وبحل المعادلات (7.56) : (7.54) نجد أن:-

$$\hat{a}_0 = -2.19 \quad , \ \hat{a}_1 = 0.392 \quad , \hat{a}_2 = -0.004$$

وبالتالي فإن نموذج انحدار y على x يأخذ الصورة التالية:-

$$\hat{y} = -2.19 + 0.392x - 0.004x^2$$

٣- عندما X=80 فإن:-

$$\hat{y} = -2.19 + 0.392(80) - 0.004(80)^2$$
$$= -2.19 + 31.36 - 25.6 = 3.57$$

ونلاحظ أن القيمة الفعلية لـ Y=3 عندما X=80 أي أن الفرق بين القيمة الفعلية والقيمة التقديرية $\hat{y}$ يصبح:-

$$\varepsilon = y - \hat{y} = 3 - 3.57 = 0.57$$

كذلك عند X=90 فإن:-

$$\hat{y} = -2.19 + 0.392(90) - 0.004(90)^2$$
$$= -2.19 + 35.28 - 32.4 = 0.69$$

ثالثاً : نموذج الانحدار المتعدد الخطى
Linear Multi-regression Model

نظريـة (٧-٣): إذا كانـت العلاقـة بـين المتغيـرين التـابع y والمتغيـرات المسـتقلة $x_1, x_2, \ldots, x_n$ على النحو:-

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n + \varepsilon \qquad (7.54)$$

أو

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \ldots + a_n x_{ni} + \varepsilon_i \qquad (7.55)$$

بحيث:-

$$E(\varepsilon)=0 \quad , \ \text{Var}(\varepsilon)=\gamma^2$$
$$\text{Cov}(\varepsilon_i,\varepsilon_j)=0 \quad , i\neq j \qquad (7.56)$$

فإن أفضل منحنى (نموذج) يمثل العلاقة (7.54) باستخدام عينة حجمها n هو:-

$$\hat{y}=\hat{a}_0+\hat{a}_1x_1+\hat{a}_2x_2+\ldots+\hat{a}_nx_n \qquad (7.57)$$

حيث يتم حساب التقديرات $\hat{a}_j, j=1,2,3,\ldots,n$ من المعادلات الطبيعية التالية:-

$$\sum y=n\hat{a}_0+\hat{a}_1\sum x_1+\hat{a}_2\sum x_2+\ldots.+\hat{a}_n\sum x_n$$
$$\sum x_1y=\hat{a}_0\sum x_1+\hat{a}_1\sum x_1^2+\hat{a}_2\sum x_1x_2+\ldots.+\hat{a}_n\sum x_1x_n$$

.

.

$$.\sum x_ny=\hat{a}_0\sum x_n+\hat{a}_1\sum x_nx_1+\hat{a}_2\sum x_nx_2+\ldots.+\hat{a}_n\sum x_n^2$$

الإثبات

بنفس طريقة إثبات نظرية (٧-١).

مثال (٧-٥)

ترغب إحدى الشركات الإنتاجية في تقدير العلاقة بين الكمية المطلوبة (y) على إحدى منتجاتها وسعر بيع الوحدة الواحدة من هذا المنتج (x_1) وسعر بيع الوحدة الواحدة من السلعة البديلة لهذا المنتج (x_2) . والجدول التالي يعطي بيانات (عينة) للكميات المطلوبة (y) بالألف وحدة سعر بيع الوحدة والأسعار المناظرة لهذه الكميات x_1, x_2 :-

جدول (٧-١٣)

y	**10**	**15**	**20**	**25**	**27**	**33**	**35**	**40**	**50**	**55**

x_1	15	12	11	10	9	8	7	5	3	2
x_2	2	3	5	7	8	10	11	12	17	20

بافتراض أن العلاقة بين y وكل من x_1, x_2 علاقة خطية أي:-

$$y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$$

حيث ε متغير عشوائي.

المطلوب

١- **باستخدام طريقة المربعات الصغرى كون نموذج انحدار y على كل من x_1, x_2.**

٢- **باستخدام النموذج قدر قيمة y عندما $x_1 = 17, x_2 = 1$.**

٣- **احسب الخطأ المعياري للتقدير $\hat{y}$.**

الحل

١- **بما أن العلاقة بين y, x_1, x_2 على النحو:-**

$$y = a_0 + a_1x_1 + a_2x_2 + \varepsilon$$

فإن العلاقة التقديرية لها على النحو:-

$$\hat{y} = \hat{a}_0 + \hat{a}_1x_1 + \hat{a}_2x_2$$

ولحساب التقديرات $\hat{a}_0, \hat{a}_1, \hat{a}_2$ نكون الجدول التالي:-

جدول (٧-١٤)

y	x_1	x_2	yx_1	yx_2	x_1^2	x_2^2	x_1x_2	$\hat{y}$	$(y-\hat{y})$	$(y-\hat{y})^2$
1.7	1	1	1.7	1.7	1	1	1	1.7	0	0
0.9	1	0	0.9	0	1	0	0	0.9	0	0
2.20	2	3	4.4	6.60	4	9	6	2.20	0	0
4.1	1	4	4.1	16.4	1	16	4	4.1	0	0
2.5	3	5	7.5	12.5	9	25	15	2.7	-0.2	0.04
7.1	10	20	71	142	100	400	200	7	0.1	0.01
0.8	5	5	4	4	25	25	25	0.5	0.3	0.09
3.2	6	10	19.2	32	36	100	60	3.4	-0.2	0.04
1.8	11	15	19.8	27	121	225	165	1.9	-0.1	0.01
0.3	3	2	0.9	0.6	9	4	6	0.3	0	0
$\sum$ 0.19			$\sum$ 482	$\sum$ 805	$\sum$ 307	$\sum$ 242.8	$\sum$ 133.5	$\sum$ 15	$\sum$ 43	$\sum$ 24.6

وبما أنه يوجد متغير تابع y ومتغيرين مستقلين x_1, x_2 فإنه يوجد ثلاثة معادلات طبيعية على النحو التالي:-

$$\sum y = n\hat{a}_0 + \hat{a}_1 \sum x_1 + \hat{a}_2 \sum x_2$$

$$\sum x_1 y = \hat{a}_0 \sum x_1 + \hat{a}_1 \sum x_1^2 + \hat{a}_2 \sum x_1 x_2$$

$$\sum x_2 y = \hat{a}_0 \sum x_2 + \hat{a}_1 \sum x_1 x_2 + \hat{a}_2 \sum x_2^2$$

وبالتعويض من الجدول السابق في المعادلات الطبيعية نحصل على:-

$$24.6 = 10\hat{a}_0 + 43\hat{a}_1 + 65\hat{a}_2$$

$$133.5 = 43\hat{a}_0 + 307\hat{a}_1 + 482\hat{a}_2$$

$$242.8 = 65\hat{a}_0 + 48\hat{a}_1 + 805\hat{a}_2$$

وبحل المعادلات أعلاه نحصل على قيم $\hat{a}_0, \hat{a}_1, \hat{a}_2$ على النحو التالي:-

$$\hat{a}_0 = 2 \;,\; \hat{a}_1 = -1.10 \;,\; \hat{a}_2 = 0.8$$

$$\hat{y} = 2 - 1.1x_1 + 0.8x_2$$

٢- عندما $x_1 = 3, x_2 = 5$ فإن:-

$$\hat{y} = 2 - 1.1(3) + 0.8(5) = 2.7$$ ألف وحدة

٣- لحساب الخطأ المعياري للتقدير $\hat{\gamma}$ باستخدام نموذج الانحدار:-

$$\hat{y} = 2 - 1.1x_1 + 0.8x_2$$

نحسب $\hat{y}$ عند قيم x_1, x_2 المعطاة ثم نحسب $(y - \hat{y})^2$ كما هو موضح بالجدول

$$\hat{\gamma} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 2}}$$

$$= \sqrt{\frac{0.19}{10 - 2}} = \sqrt{0.02375} = 0.154$$

<u>ملحوظة:</u> من الجدول نلاحظ أن الفرق بين القيم الفعلية للمتغير y والقيم المقدرة $\hat{y}$ المناظرة لها قيم صغيرة جداً وهذا يعني أن القيم التقديرية $\hat{y}$ أقرب ما يمكن إلى القيم الفعلية y مما أدى إلى أن قيمة الخطأ المعياري للتقدير $\hat{\gamma}$ صغيرة ($\hat{\gamma} = 0.154$) وهذا يدل على أن نموذج الانحدار

$$\hat{y} = 2 - 1.1x_1 + 0.8x_2$$

$$y = 2 - 1.1x_1 + 0.8x_2 + \varepsilon$$

<u>خصائص تقديرات المربعات الصغرى</u>

بما أن التقديرات $\hat{a}_0, \hat{a}_1, \hat{a}_3, ...$ والتي يتم حسابها من بيانات عينة بالتالي فهي تختلف من عينة لأخرى أي تمثل متغيرات ، وفيما يلي سوف نلخص أهم خصائص التقديرات التي يتم الحصول عليها باستخدام المربعات الصغرى من خلال النظريات التالية:-

<u>نظرية (٧-٤)</u>: إذا كانت العلاقة بين المتغير x والمتغير y على النحو التالي:

$$y_i = a_0 + a_1 x_i + \varepsilon_i$$

والعلاقة التقديرية لهذه العلاقة هي:-

$$\boxed{\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i \qquad (7.58)}$$

حيث ε متغير عشوائي بحيث $E(\varepsilon) = 0, Var(\varepsilon) = \gamma^2$ حيث γ^2 مقدار ثابت ، فإن التقديرات $\hat{y}, \hat{a}_0, \hat{a}_1$ في العلاقة (7.58) تمثل تقديرات غير متحيزة Unbiased Estimator لكل من y, a_0, a_1 على الترتيب.

أي أن:-

$$\boxed{E(\hat{y}) = y, E(\hat{a}_0) = a_0, E(\hat{a}_1) = a_1 \qquad (7.59)}$$

<u>نظرية (٧-٥)</u>: تباينات التقديرات $\hat{a}_1, \hat{a}_0$ هي:-

$$\mathrm{Var}(\hat{a}_0) = \frac{\gamma^2 (\sum x_i)^2}{n \sum (x_i - \bar{x})^2} \qquad (7.60)$$

$$\mathrm{Var}(\hat{a}_1) = \frac{\gamma^2}{\sum (x_i - \bar{x})^2} \qquad (7.61)$$

<u>نظرية (٧-٦)</u>: إذا كان γ^2 يمثل تباين المتغير العشوائي ε فإن $\hat{\gamma}$ يمثل تقدير لـ γ^2 حيث:-

$$\hat{\gamma}^2 = \frac{\sum \varepsilon_i^2}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

حيث $\hat{\gamma}^2$ تقدير غير متحيز للمعلمة γ^2

<u>نظرية (٧-٧)</u>: تسمى هذه النظرية بنظرية جاوس ماركوف Gauss – Markov Theorem نسبة إلى العالم جاوس.

إذا كانت $\hat{a}_0, \hat{a}_1$ تقديرات تم الحصول عليها باستخدام طريقة المربعات الصغرى فإن $\hat{a}_0, \hat{a}_1$ تعتبر تقديرات مثلى وغير متحيزة Best and Unbiased Estimators (تقديرات مثلى هنا يعني أن لكل منها أصغر تباين).

ملحوظة

ويمكن تعميم النظريات المذكورة أعلاه بالنسبة لنموذج الانحدار غير الخطي ، ونموذج الانحدار المتعدد.

(٧-٦) توفيق منحنيات تؤول إلى الخط المستقيم

Curve Fitting Tending to Linear Curve

كثير من الظواهر الاقتصادية والاجتماعية ...الخ تكون العلاقة بين المتغير الأساسي x والمتغير التابع y علاقة غير خطية ولكن في بعض الحالات يمكن تحويل العلاقة غير الخطية إلى علاقة خطية فعلى سبيل المثال إذا كانت العلاقة الفعلية تأخذ الشكل التالي:-

$$y = a_0 x^{a_1} \varepsilon \qquad (7.62)$$

حيث a_0, a_1 **هما معلمات العلاقة ،** ε **المتغير العشوائي.**

فإن العلاقة التقديرية (النموذج) للعلاقة في (7.62) يصبح:

$$\hat{y} = \hat{a}_0 x^{\hat{a}_1} \varepsilon \qquad (7.63)$$

حيث $\hat{a}_0, \hat{a}_1$ **هما تقديرين لـ** a_0, a_1 **ويمكن حساب** $\hat{a}_0, \hat{a}_1$ **باستخدام طريقة المربعات الصغرى على النحو التالي:-**

إذا فرضنا أن :-

$$y' = \ln y \rightarrow y = e^{y'} \qquad (7.64)$$

$$a'_0 = \ln a_0 \rightarrow a_0 = e^{a'_0} \qquad (7.65)$$

$$x' = \ln x \rightarrow x = e^{x'} \qquad (7.66)$$

$$\varepsilon' = \ln x \rightarrow \varepsilon = e^{\varepsilon'} \qquad (7.67)$$

ملحوظة: e هي الأساس الطبيعي حيث $e \approx 2.718$

وبأخذ لوغاريتمات الطرفين للعلاقة (7.62) نجد أن:-

$$\ln y = \ln a_0 + a_1 \ln x + \ln \varepsilon$$

أو

$$y' = a_0' + a_1 x' + \varepsilon' \qquad (7.68)$$

ويتضح أن العلاقة (7.68) علاقة خطية وباستخدام طريقة المربعات الصغرى نوجد تقدير للمعلمات a_1, a_0 وسوف نشير إليها بالرمز $\hat{a}_1, \hat{a}_0$ من المعادلتين الآتيتين.

$$\sum y' = n\hat{a}_0' + \hat{a}_1 \sum x' \qquad (7.69)$$

$$\sum x'y' = \hat{a}_0' \sum x' + \hat{a}_0 \sum x'^2 \qquad (7.70)$$

وبحل المعادلتين (7.70) , (7.69) يمكن الحصول على $\hat{a}_1, \hat{a}_0$ ، وسوف نوضح ذلك من خلال المثال التالي:-

مثال(٧-٦)

الجدول التالي يعطي قيم المتغير الأساسي x والقيم المناظرة لها للمتغير التابع y:

جدول (٧-١٥)

X	**0**	**1**	**2**	**3**	**4**	**5**	**6**
y	**0**	**0.5**	**2**	**4.5**	**8**	**12.5**	**18**

المطلوب

١- **ارسم شكل الانتشار.**

٢- **أوجد باستخدام طريقة المربعات نموذج الانحدار إذا كانت العلاقة بين x,y على النحو التالي:-**

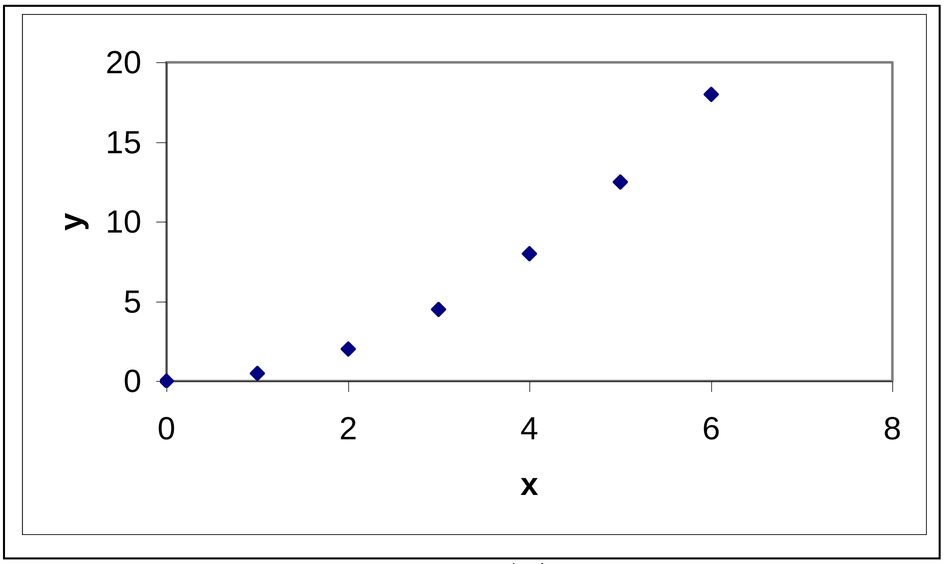
$$y = a_0 x^{a_1} \varepsilon$$

حيث ε متغير عشوائي.

٣- قدر قيمة y عندما x=4.

الحل

١- الشكل التالي يعطي شكل الانتشار الذي يوضح العلاقة بين x,y.



شكل (٧-١٤)

٢- بأخذ لوغاريتمات طرفي العلاقة بين y', x فإن:-

$$\ln y = \ln a_0 + a_1 \ln x + \ln \varepsilon \qquad (7.71)$$

وبافتراض أن:-

$$\varepsilon' = \ln \varepsilon$$

$$y' = \ln y \rightarrow y = e^{y'}$$

$$x' = \ln x \rightarrow x = e^{x'}$$

وبالتالي تصبح العلاقة:-

$$y' = a_0' + a_1 x' + \varepsilon' \qquad (7.72)$$

ويكون النموذج المناظر للعلاقة (8.72) على النحو التالي:-

$$\hat{y}' = \hat{a}_0' + \hat{a}_1 x' \qquad (7.73)$$

ولتكوين المعادلات الطبيعية التالية نكون الجدول التالي:-

$$\sum y' = n\hat{a}'_0 + \hat{a}_1 x' \tag{7.74}$$

$$\sum x'y' = \hat{a}'_0 \sum x' + \hat{a}_1 \sum x'^2 \tag{7.75}$$

جدول (٧-١٦)

x	y	$x' = \ln x$	$y' = \ln y$	$x'y'$	x'^2
1	**0.5**	**0**	**-0.69**	**0**	**0**
2	**2**	**0.69**	**0.69**	**0.48**	**.48**
3	**4.5**	**1.10**	**1.50**	**1.65**	**1.21**
4	**8**	**1.39**	**2.08**	**2.89**	**1.93**
5	**12.5**	**1.61**	**2.53**	**4.07**	**2.59**
6	**18**	**1.79**	**2.89**	**5.17**	**3.20**
$\sum$		**6.58**	**9**	**17.26**	**9.41**

$$9 = 6\hat{a}'_0 + 6.58\hat{a}_1$$

$$17.26 = 6.58\hat{a}'_0 + 9.41\hat{a}_1$$

وبحل المعادلتين نجد أن:-

$$\hat{a}'_0 = -2.23 \rightarrow \hat{a}_0 = e^{-2.23} = 0.11$$

$$\hat{a}_1 = 3.39$$

$$\hat{y} = (0.11)x^{3.39}$$

عندما X = 4 فإن:-

$$y' = (0.11)(4)^{3.39} \approx 12.09$$

Standard Error of Estimate الخطأ المعياري

مما سبق يتضح أنه عند قيمة معينة للمتغير المستقل x ولتكن x_i فإنه توجد قيم مختلفة للمتغير التابع y تكون فيما بينها توزيعاً تكرارياً وسطه الحسابي هو القيمة $\hat{y}_i$ التي تقع على خط الانحدار * أي أن:-

$$\bar{y} = \hat{y}_i \qquad (7.76)$$

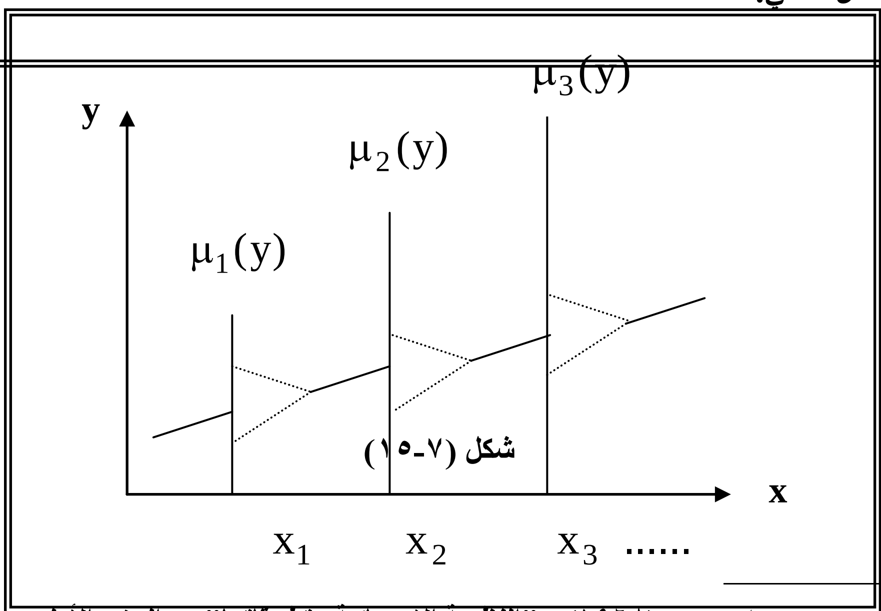
وبالمثل إذا فرضنا أن $x = x_i$ فإن قيم المتغير التابع y المناظرة لقيمة x_i تكون أيضاً توزيعاً تكرارياً وسطه الحسابي يقع أيضاً على خط الانحدار أي أن:-

$$\bar{y} = \hat{y}_i \qquad (7.77)$$

وبالتالي فإنه يمكن القول أن خط الانحدار $\hat{y}$ على x (نموذج الانحدار الخطي) حيث:-

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x \qquad (7.78)$$

يمر بجميع الأوساط الحسابية $(\mu_1(y), \mu_2(y), \ldots.)$ للتوزيعات التكرارية للمتغير العشوائي y التي تناظر قيم $(x_1, x_2, \ldots.)$ للمتغير المستقل X . كما هو موضح بالشكل التالي:-



شكل (٧-١٥)

* أ.د محرم وهبه محمود (١٩٦٨) : "النظرية الإحصائية وتطبيقاتها" – الجزء الأول – معهد التخطيط القومي - القاهرة

ويسمى المقدار الذي يمثل الجذر التربيعي لمتوسط مربعات انحرافات القيم المقدرة $\hat{y}$ عن القيم الفعلية للمتغير التابع **y** في عينة حجمها **n** بخطأ التقدير ويرمز له $\hat{\gamma}$ حيث يمثل $\hat{\gamma}^2$ تقدير لتباين المتغير العشوائي ε أي تقدير للمعلمة γ^2 حيث:-

$$\hat{\gamma}^2 = \frac{\sum \varepsilon^2}{n} = \frac{\sum_i (y_i - \hat{y})}{n} \qquad (7.79)$$

نظرية (٧-٨): يعتبر الخطأ في التقدير $\hat{\gamma}$ تقدير غير متحيز للانحراف المعياري للمتغير العشوائي ε (أي تقدير غير متحيز للمعلمة γ) إذا كان:-

$$\hat{\gamma} = \sqrt{\frac{\sum_i (y_i - \hat{y})}{n-2}} \qquad (7.80)$$

مثال(٧-٧)

الجدول التالي يوضح توزيع 100 عامل حسب كفاءة الداء للعامل بإحدى المؤسسات الصناعية:-

جدول (٧-١٧)

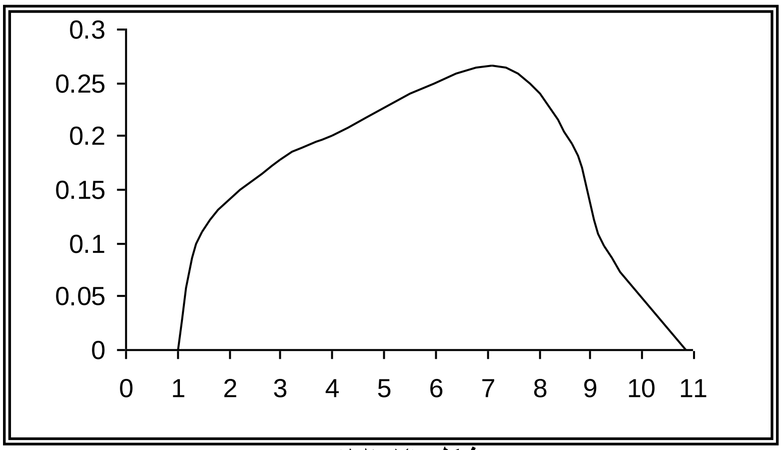
مستوى الأداء (x)	0-	2-	4-	6-	8-10	$\sum$
عدد العمال (التكرارات)f	15	18	22	25	20	100

المطلوب

١- ارسم المنحنى التكراري النسبي للتوزيع السابق.
٢- وفق دالة كثافة احتمال التوزيع المنتظم لتمثل التوزيع الاحتمالي لمستوى الداء في المؤسسة (المسحوب منه العينة).
٣- احسب التكرارات المتوقعة باستخدام التوزيع في (٢) ثم قارنها بالتكرارات الفعلية **f**.

الحل

١- الشكل التالي يوضح التكرار النسبي للتوزيع



شكل (٧-١٦)

بما أن أقل مستوى أداء يساوي صفر وأعلى مستوى أداء يساوي 10 بالتالي فإن دالة كثافة الاحتمال للتوزيع المنتظم f(x) تأخذ الشكل التالي:-

$$f(x) = \frac{1}{10-0} = \frac{1}{10} \qquad 0 \le x \le 10$$

وبالتالي فإن:-

$$P_r(x \le 2) = \int_0^2 \frac{1}{10} dx = \frac{1}{10}[x]_0^2$$

$$= \frac{1}{10}[2-0] = 0.2$$

بالمثل

$$P_r(2 \le x \le 4) = \int_2^4 \frac{1}{10} dx = 0.2$$

وحيث أن أطوال الفئات متساوية فإن احتمال وقوع x في أي فئة يساوي 0.2 وبما أن التكرارات المتوقعة يساوي مجموع التكرارات × الاحتمال المناظر.

وبما أن أطوال الفئـات متسـاوية والمتغيـر x يتبـع التوزيـع المنـتظم المقدر فـإن التكـرار المتوقع يساوي (20 = (0.2) 100). كما هو موضح بالجدول (٧-١٨).

وبما أن الخطأ المعياري للتقدير يساوي $\hat{\gamma}$ حيث:-

$$\hat{\gamma} = \sqrt{\frac{\sum \varepsilon^2}{n-2}} = \sqrt{\frac{58}{100-2}} = 0.59$$

واضح هنا أن قيمة γ صغيرة أي أن التوفيق باستخدام التوزيع المنتظم توفيق جيد.

جدول (٧-١٨)

الفئات	التكرارات الفعلية f	التكرارات المتوقعة $\hat{f}$	$\varepsilon = (f - \hat{f})$	ε^2
1-	15	20	-5	25
2-	18	20	-2	4
4-	22	20	2	4
6	25	20	5	25
8-10	20	20	0	0
$\sum$	100	100		58

مثال (٧-٨)

الجدول التالي يوضح سعر بيع الوحدة الواحدة x بالجنيه من سلعة معينـة والكميـة المعروضة y بالألف وحدة من هذه السلعة.

جدول (٧-١٩)

سعر الوحدة بالجنيه x	3	5	6	8	9	11
الكمية المعروضة بالألف وحدة y	2	3	4	6	5	8

<u>المطلوب</u>

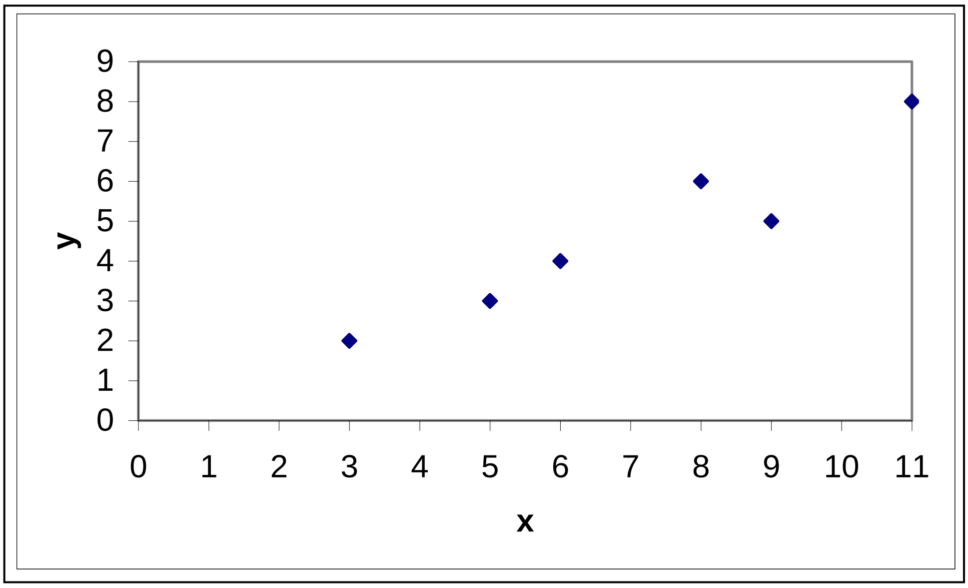
١- ارسم شكل الانتشار الذي يوضح العلاقة بين x , y.

٢- باستخدام طريقـة المربعـات الصـغرى قـدر منحنـى (أو ابنــي نمـوذج) انحـدار مناسب.

٣- احسب الخطأ في التقدير $\hat{\gamma}$.

<u>الحل</u>

١- الشكل التالي يوضح شكل الانتشار



شكل (٧-١٧)

٢- من الشكل يتضح أن اتجاه العلاقة هو الاتجاه الخطي وبالتالي إذا فرضنا أن العلاقة بين x , y على النحو التالي:-

$$y = a_0 + a_1 x + \varepsilon$$

ونموذج الانحدار الذي يمثل العلاقة السابقة هو:-

$$y = \hat{a}_0 + \hat{a}_1 x$$

حيث يمكن حساب $\hat{a}_0, \hat{a}_1$ باستخدام طريقة المربعات الصغرى ولحساب $\hat{a}_0, \hat{a}_1$ نكون الجدول (٧-٢٠).

بالتعويض في المعادلتين (7.52) , (7.51) من الجدول نحصل على:-

$$28 = 6\hat{a}_0 + 24\hat{a}_1 \qquad (7.81)$$

$$226 = 42\hat{a}_0 + 336\hat{a}_1 \qquad (7.82)$$

وبحل المعادلتين (7.82) , (7.81) نحصل على قيمة كل من $\hat{a}_0, \hat{a}_1$ على النحو التالي:-

$$\hat{a}_0 = -0.3 \;\;,\;\; \hat{a}_1 = 0.71$$

وبالتالي يصبح نموذج الانحدار على النحو التالي:-

$$\hat{y} = -0.3 + 0.71x \qquad (7.83)$$

جدول (٧-٢٠)

x	y	xy	x^2	$\hat{y}$	ε	ε^2
3	2	6	9	1.83	0.17	0.029
5	3	15	25	3.25	-0.25	0.063
6	4	24	36	3.96	0.04	0.002
8	6	48	64	5.38	0.62	0.384
9	5	45	81	6.09	-0.09	1.188
11	8	88	121	7.51	0.49	0.240
42	28	226	336			1.906

٣- وبالتعويض بقيم x المعطاة (من الجدول) في النموذج (7.83) نحصل على قيم $\hat{y}$ المناظرة لها كما هو موضح بالجدول السابق . وبالتالي يمكن حساب الخطأ المعياري للتقدير $\hat{\gamma}$ حيث:-

$$\hat{\gamma} = \sqrt{\frac{\sum \varepsilon^2}{n-2}} = \sqrt{\frac{1.906}{6-2}} = \sqrt{0.477} = 0.691$$

مثال (٨-٩)

ألقيت 6 وحدات من النقود 100 مرة وفي كل مرة سجلت المشاهدات والجدول التالي يعطي عدد مرات الرمي وفقاً لعدد الوحدات التي يظهر عليها الشعار في كل مرة:-

جدول (٧-٢١)

عدد الوحدات التي يظهر عليها الشعار (x)	0	1	2	3	4	5	6	$\sum$
عدد مرات الرمي (التكرارات) f	5	6	7	19	30	23	10	100

<u>المطلوب</u>

١- أوجد دالة التوزيع التكراري النسبي ووضح ذلك بيانياً.

٢- وفق دالة احتمال مناسبة باستخدام طريقة العزوم.

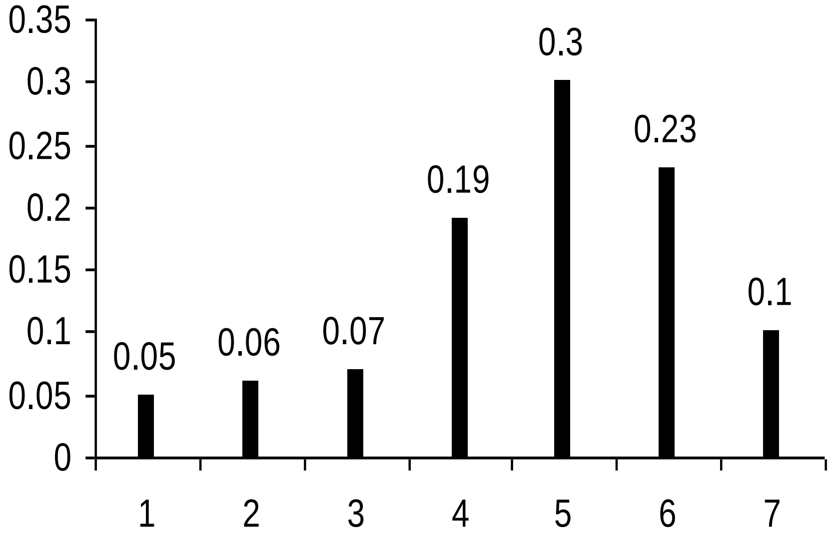
<u>الحل</u>

جدول (٧-٢٢)

x	0	1	2	3	4	5	6	$\sum$
التكرار النسبي $P_r(x)$	0.05	0.06	0.07	0.19	0.30	0.23	0.10	1

والشكل (٧-١٨) يوضح التكرار النسبي في الجدول السابق:-

٣- من الشكل يتضح أنه يمكن تقدير دالة الاحتمال المتغير x بدالة احتمال توزيع ذات الحدين:-



شكل (٧-١٨)

- **بما أن دالة احتمال المتغير ذات الحدين**

$$P_r(x) = C_x^6 (p)^x (1-p)^{6-x} \quad ,x =,1,2,...,6 \qquad (7.84)$$

- **بما أن** $E(x) = np = 6p$

ومن الجدول السابق نجد أن التوقع للعينة $\overline{x}$ حيث:-

$$\overline{x} = \frac{\sum x_i P_r(x)}{\sum f} = \frac{372}{100} = 3.72 \qquad (7.85)$$

وعندما

$$\mu = \overline{x} \to 6p = 3.72 \to p = 0.037$$

وبالتالي فإن:-

$$\hat{P}_r(x) = C_x^6 (0.037)(0.963)^{6-x} \quad ,x = 0,1,2,...,6$$

مثال (٧-١٠)

إذا كانتا العلاقة بين x , y على النحو التالي:

$$y = a\beta^{x}\varepsilon$$

والجدول التالي يعطي قيم x وقيم y المناظرة لها:

جدول (٧-٢٣)

x	0	1	2	3	4	5	6	7	8	9
y	3	6	11.5	23.5	48	96	192	384	766	1536

<u>المطلوب</u>

١- **باستخدام طريقة المربعات الصغرى أبني نموذج انحدار للمتغير y على المتغير x.**

٢- **أوجد الخطأ المعياري للتقدير $\hat{\gamma}$.**

<u>الحل</u>

١- **بما أن العلاقة بين x , y هي:-**

$$y = a\beta^{x}\varepsilon \qquad (7.86)$$

فبأخذ لوغاريتمات الطرفين في العلاقة السابقة نحصل على العلاقة التالية:-

$$\ln y = \ln a + x\ln\beta + \ln\varepsilon \qquad (7.87)$$

وبافتراض أن:-

$$y' = \ln y \quad , \quad a' = \ln a \quad , \beta' = \ln\beta \quad , \quad \varepsilon' = \ln\varepsilon$$

فإن العلاقة (7.87) سوف تصبح على النحو التالي:-

$$y' = a' + \beta' x + \varepsilon' \qquad (7.88)$$

ويكون النموذج الذي يمثل العلاقة (7.88) هو:-

$$y' = \hat{a}' + \hat{\beta}' x \qquad (7.89)$$

وتصبح المعادلات الطبيعية في هذه الحالة على النحو التالي:-

$$\sum y' = n\hat{a}' + \hat{\beta}'\sum x \qquad (7.90)$$

$$\sum xy' = \hat{a}'\sum x + \hat{\beta}'\sum x^2 \qquad (7.90)$$

وللحصول على التقديرات $\hat{a}', \hat{\beta}'$ نكون الجدول التالي:-

جدول (٧-٢٤)

x	y	$y' = \ln y$	xy'	x^2	$\hat{y}$	$\varepsilon^2 = (y - \hat{y})^2$
0	3	0.4771	0	0	2.9406	2.9406
1	6	0.7782	0.7782	1	5.8959	0.0108
2	11.5	1.0607	2.1219	4	11.8213	0.1032
3	23.5	2.371	4.1133	9	23.7017	0.0407
4	48	1.6812	1.8248	16	47.5219	0.2285
5	96	1.9823	9.9115	25	95.2813	0.5165
6	192	2.2833	13.6998	36	191.036	0.9224
7	384	2.5840	18.088	49	232.033	0.0010
8	688	2.8842	23.736	64	767.952	3.9283
9	1536	3.1864	28.6776	81	1539.804	14.4703
$\sum x$ 45	$\sum y$ 3066	$\sum y'$ 18.1878	$\sum xy'$ 107.1882	$\sum x^2$ 285		$\sum (y - \hat{y})$ 20.2264

وبالتعويض من الجدول في المعادلتين (7.91) , (7.90) نجد أن:-

$$18.1878 = 10\hat{a}' + 45\hat{\beta}'$$

$$107.1882 = 45\hat{a}' + 285\hat{\beta}'$$

وبحل المعادلتين نحصل على قيمة كل من $\hat{a}', \hat{\beta}'$ حيث:-

$$\hat{a}' = 0.4684 \rightarrow \hat{a} = (10)^{0.4684} = 2.9406$$

$$\hat{\beta}' = 0.3023 \rightarrow \hat{\beta} = (10)^{0.3023} = 2.005$$

وبالتالي يصبح نموذج الانحدار على النحو التالي:-

$$\hat{y} = \hat{a}\hat{\beta}^{x} = (2.9406)(2.005)^{x}$$

٢- وبما أن الخطأ المعياري للتقدير γ حيث:-

$$\hat{\gamma}^2 = \frac{\sum \varepsilon^2}{n-2} = \frac{20.2264}{10-2} = 2.5283$$

$$\hat{\gamma} = \sqrt{2.5283} = 1.59$$

الارتباط الخطي **Linear Correlation**

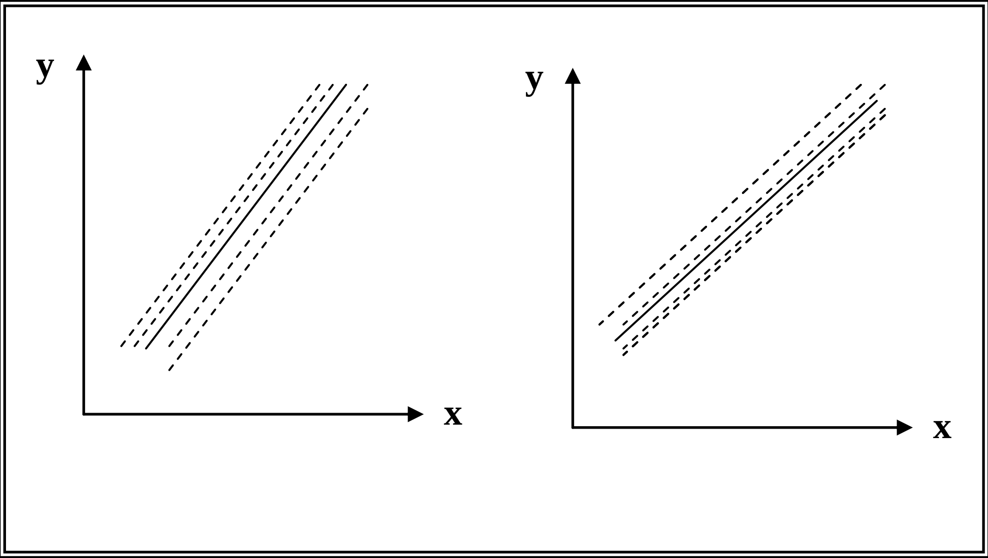
في الفصلين (٧-٣) ، (٧-٤) من هذا الباب تناولنا بالدراسة كيفية تقدير العلاقة الخطية بين متغير واحد تابع y ومتغير واحد مستقل x وهو ما يسمى بنموذج الانحدار الخطي **Linear Regression Model**.

وفي هذا الفصل سوف نتناول بالدراسة المقياس الذي يمكن باستخدامه قياس مدى الترابط بين الظاهرتين x , y (أي بين المتغير التابع والمتغير المستقل) وهو ما يسمى بمقياس الارتباط **correlation's Measure** .

فإذا كانت العلاقة بين المتغير التابع y والمتغير المستقل x هي:-

$$y = a_0 + a_1 x + \varepsilon$$

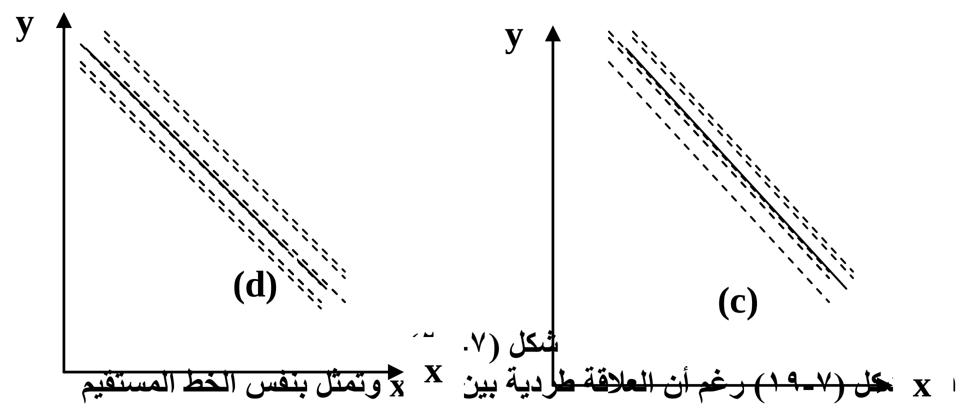
فإذا كانت العلاقة طردية كما في الشكل الانتشار (٧-١٩)



(b) (a)

شكل (٧-١٩)

أو عكسية كما في شكل الانتشار (٧-٢٠)



شكل (٧-...

شكل (٧-١٩) رغم أن العلاقة طردية بين x وتمثل بنفس الخط المستقيم

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x$$

إلا أن الترابط بين قيم x , y في الشكل (a) أكثر من الترابط بين قيمة x , y في شكل (b).

كذلك بالنسبة لشكل (٧-٢٠) فرغم أن العلاقة عكسية بين x , y وتمثل بنفس الخط المستقيم في شكلي (d) , (c)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

فإننا نلاحظ أن الترابط بين y, x في (c) أكثر من الترابط في (d) ، ولقياس شدة الترابط بين المتغير المستقل x والتابع y سوف نقدم بعض التعريفات الضرورية التالية:-

<u>تعريف (٧-١)</u>

التباين الكلي (الاختلاف الكلي) Total Variation

إذا كان المتغير التابع y يأخذ القيم $y_i \ , i = 1,2,.....,n$ في عينة حجمها n
$\bar{y}$ هو القيمة المتوسطة لهذا المتغير في العينة فإن المقدار:-

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad (7.92)$$

يسمى بالتباين الكلي للمتغير y.

تعريف (٧-٢)

التباين المفسر Explained Variation

إذا كانت $\hat{y}_i$ هي القيمة المقدرة لـ y_i فإن المقدار:-

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \qquad (7.93)$$

يسمى بالتباين المفسر.

تعريف (٧-٣)

التباين غير المفسر Unexplained Variation

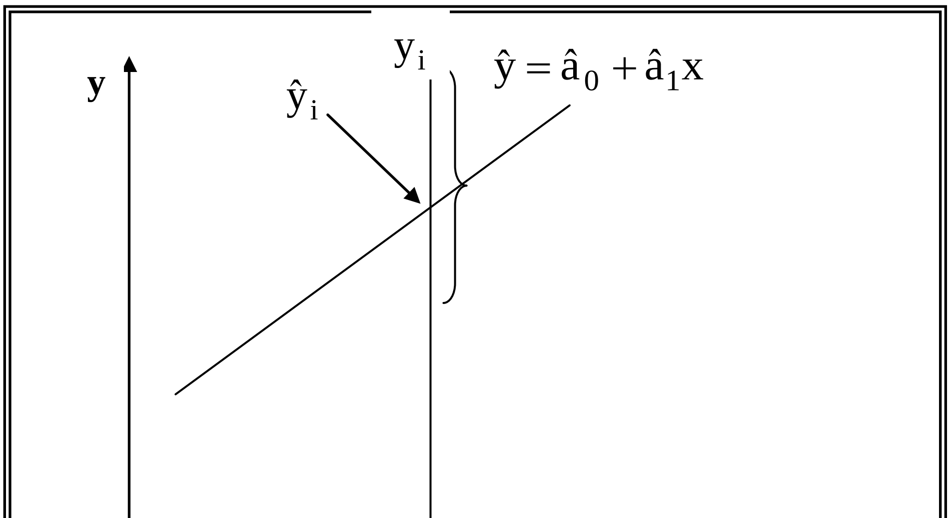
يسمى المقدار حيث:-

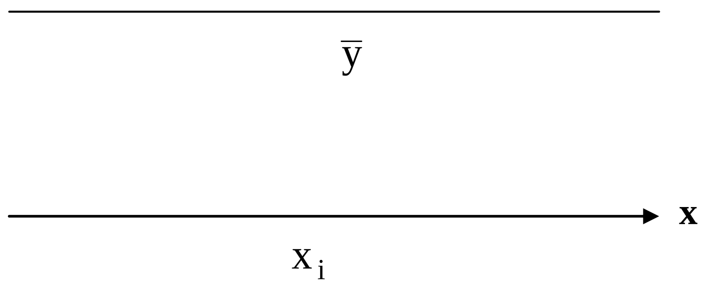
$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (7.94)$$

من التعريفات (7.94) : (7.92) نجد أن التباين الكلي هو مجموع التباين المفسر والتباين غير المفسر أي أن:-

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (7.95)$$

وشكل (٧-٢١) يوضح هذه العلاقة (7.95)





شكل (٧-٢١)

تعريف (٧-٤)

معامل التحديد The Coefficient of Determination

تسمى النسبة بين التباين المفسر والتباين الكلي بمعامل التحديد ويرمز لها بالرمز $\rho^2_{(x,y)}$ حيث :-

$$\rho^2_{(x,y)} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (7.96)$$

ومن العلاقة (7.95) نجد أن:-

$$\rho^2_{(x,y)} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (7.97)$$

أو

$$\rho^2_{(x,y)} = 1 - \frac{\sum_{i=1}^{n} \varepsilon_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (7.98)$$

<u>تعريف (٧-٥)</u>

معامل الارتباط Linear Correlation Coefficient

إذا كانت العلاقة بين المتغير التابع والمتغير المستقل x علاقة خطية كما في (7.2) فإن الجذر التربيعي لمعامل التحديد يسمى معامل الارتباط الخطي أو معامل ارتباط بيرسون Porson's Coefficient ويرمز له بالرمز $\rho_{(x,y)}$ أي أن:-

$$\rho_{(x,y)} = \pm \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} = \pm \sqrt{1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \qquad (7.99)$$

وباستخدام المعادلات الطبيعية (7.41) ، (7.40)نجد أن:-

$$\rho_{(x,y)} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}} \qquad (7.100)$$

$$\rho_{(x,y)} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{n\sigma_x \sigma_y} \qquad (7.101)$$

حيث σ_x , σ_y هما الانحراف المعياري للمتغير x , y على الترتيب.

ويأخذ معامل الارتباط ρ قيمة موجبة إذا كانت العلاقة بين x , y علاقة طردية ويأخذ قيمة سالبة إذا كانت العلاقة علاقة عكسية.

نظرية (٧-٧): **إذا كان $\rho_{(x,y)}$ هو معامل الارتباط الخطي بين x , y فإن:-**

$$\boxed{-1 \le \rho_{(x,y)} \le 1 \qquad (7.102)}$$

أي أن:-

$$\boxed{0 \le \left|\rho_{(x,y)}\right| \le 1 \qquad (7.103)}$$

الإثبات

بما أن:-

$$\rho_{(x,y)} = \pm\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

وبما أن التباين الكلي هو مجموع التباين المفسر والتباين غير المفسر فمن العلاقة (7.95) نجد أن:-

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

أي أن:

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \le \sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad (7.104)$$

وبالتعويض في (7.99) بـ(7.104) نجد أن:-

$$\rho_{(x,y)} = \pm\sqrt{1-\frac{\sum(y_i-\hat{y})^2}{\sum(y_i-\bar{y})^2}} \tag{7.105}$$

ومن (7.104) نجد أن:-

$$0 \le \frac{\sum(y_i-\hat{y})^2}{\sum(y_i-\bar{y})^2} \le 1 \tag{7.106}$$

وبالتالي من (7.105) : (7.106) نجد أن:-

$$-1 \le \rho_{(x,y)} \le 1$$

أي أن:-

$$0 \le \left|\rho_{(x,y)}\right| \le 1$$

ويلاحظ أن العلاقة بين x , y تكون خطية طردية قوية كلما اقتربت قيمة معامل الارتباط من 1 وتكون عكسية قوية كلما اقتربت قيمة معامل الارتباط من 1- وتكون العلاقة طردية أو عكسية ضعيفة كلما اقتربت ρ من الصفر.

نظرية (٧-٦): نجد أن:

$$\boxed{\rho_{(x,y)} = \rho_{(y,x)} \qquad (8.107)}$$

أي أن :-

العلاقة النسبية بين x , y لا تؤثر على قيمة معامل الارتباط الخطي أو بعبارة أخرى قيمة معامل الارتباط في الاعتبار أي المتغيرات مستقل وأيها تابع.

الإثبات

من العلاقة (7.101) نجد أن:-

$$\rho_{(x,y)} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{n\sigma_x\sigma_y} \tag{7.108}$$

ومن (7.107) : (7.108) نجد أن:-

$$\rho_{(x,y)} = \rho_{(y,x)}$$

نظرية (٧-٧): إذا كان x , y متغيران كل منهما يتبع التوزيع المعتاد (الطبيعي) والعلاقة بينهما علاقة خطية أي أن:-

$$y = a_0 + a_1 x + \varepsilon \qquad (7.109)$$

$$x = \beta_0 + \beta_1 y + \varepsilon \qquad (7.110)$$

فإذا أخذت عينة حجمها n من قيم x وقيم y المناظرة لها وتم حساب معامل الارتباط في العينـة ρ فـإن ρ يمثـل متغيـر عشـوائي(حيـث $\rho = \rho_{(x,y)} = \rho_{(y,x)}$) فـإن ρ

يمثل متغير بحيث المقدار t:-

$$\boxed{t = \frac{\hat{\rho}(n-2)}{\sqrt{1-\hat{\rho}^2}} \qquad (7.111)}$$

يمثل متغير عشوائي يتبع توزيع استيودنت بدرجات حرية (n-2).

مثال (٧-١١)

الجدول التـالي يوضـح عـدد الوحـدات المباعـة شـهرياً (y) مـن إحـدى المنتجـات الاستهلاكية (بالألف وحدة) والمبالغ المنفقة شـهرياً علـى الإعلان عن هذا المنـتج (x) بالألف جنيه لعدد الوحدات المباعة:-

جدول (٧-٢٥)

المنفق على الإعلان (x بالألف جنيه)	0	0.5	1	1.5	0.5	2	2.5	3	4	5
عدد الوحدات المباعة (y بالألف وحدة)	5	5.4	6	7	5.6	8.2	8	8.8	9	10

المطلوب

١- **احسب معامل الارتبـاط بين المبـالغ المنفقـة علـى الإعلان عن المنـتج وعـدد الوحدات المباعة من المنتج.**

٢- **من (١) حدد اتجاه وقوة العلاقة.**

الحل

لحساب معامل الارتباط نتبع الخطوات التالية:-

١- **نكون الجدول التالي:-**

جدول (٧-٢٦)

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
0	0	0	0	25

x_i	y_i	$x_i y_i$	x_i^2	y_i^2
5	**10**	**50**	**25**	**100**
0.5	**5.4**	**2.7**	**0.25**	**29.16**
1	**6**	**6**	**1**	**36**
1.5	**7**	**10.5**	**2.25**	**49**
0.5	**5.6**	**2.8**	**0.25**	**31.36**
2	**8.2**	**16.4**	**4**	**67**
2.5	**8**	**20**	**6.25**	**64**
3	**8.80**	**26.4**	**9**	**77.44**
4	**9**	**36**	**16**	**81**
$\sum x_i. = 20$	$\sum y_i = 73$	$\sum x_i y_i = 170.8$	$\sum x_i^2 = 64$	$\sum y_i^2 = 560.2$

من المعادلة (7.100) نجد أن:-

$$\rho_{(x,y)} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}}$$

بالتعويض من الجدول السابق في المعادلة السابقة نجد أن:-

$$\rho = \frac{10(170.8) - (20)(73)}{\sqrt{[10(64) - (20)^2][10(560.2) - (73)^2}}$$

$$= \frac{1708 - 1460}{\sqrt{240}(273)} = \frac{248}{255.97} = 0.97$$

٢- وبما أن $\rho = 0.97$ أي قيمة موجبة إذن العلاقة بين المنفق على الإعلان وعدد الوحدات المباعة علاقة طردية ، كذلك بما أن $\rho = 0.97$ أي قيمة تقترب من 1 إذن العلاقة علاقة طردية قوية.

(٧-٧) تمارين Exercises

(٧-١): الجدول التالي يعطي توزيع 200أسرة حسب عدد الأطفال في الأسرة الواحدة حيث تمثل الأسر عينة عشوائية بسيطة من الأسر بجمهورية مصر العربية:-

جدول (٧-٢٧)

عدد الأطفال في الأسرة الواحدة	0-	2-	4-	6-	8-10	$\sum$
عدد الأسر (التكرارات)	14	40	90	38	18	200

المطلوب

١- ارسم منحنى التكراري النسبي.

٢- أوجد متوسط عدد الأطفال في الأسرة الواحدة وكذلك الانحراف المعياري.

٣- استخدم طريقة العزوم في توفيق منحنى توزيع معتاد يمثل عدد الأطفال في الأسرة في جمهورية مصر العربية.

٤- احسب الخطأ المعياري للتقدير.

(٧-٢) الجدول التالي يوضح الدخل الشهري بالجنيه لعشرة أشخاص من طبقة المثقفين وعدد الأفراد في كل أسرة:-

جدول (٧-٢٨)

الدخل الشهري بالجنيه (x)	100	120	170	200	300	500	220	210	250	400
عدد الأفراد الأسرة (y)	2	2	3	4	4	3	4	3	3	3

المطلوب

١- ارسم شكل الانتشار الذي يوضح اتجاه العلاقة بين الدخل وعدد الأفراد في الأسرة.

٢- وفق منحنى (أو أبني نموذج) انحدار مناسب تقدير للعلاقة بين الدخل وعدد أفراد الأسرة.

٣- قدر عدد الأفراد في أسرة دخلها الشهري 500 جنيه.

٤- احسب الخطأ المعياري للتقدير.

٥- احسب معامل الارتباط بين الدخل الشهري وعدد أفراد الأسرة ثم عقب على الناتج.

(٧-٣): الجدول التالي يعطي قيم للمتغير المستقل x والقيم المناظرة للمتغير التابع y :-

جدول (٧-٢٩)

x	0	1	2	3	4
y	5	10	19	41	85

المطلوب

١- إذا كانت العلاقة بين x , y على النحو:-

$$y = a_0 a_1^x \varepsilon$$

حيث ε متغير عشوائي ، باستخدام طريقة المربعات الصغرى قدر كل من a_0, a_1 ثم أوجد نموذج الانحدار y على x.

٢- قدر قيمة y عندما x-4.

٣- احسب الخطأ المعياري للتقدير.

(٧-٤) الجدول التالي يوضح توزيع درجات 300 طالب وفقاً لدرجتهم في مادة الحاسب الآلي:-

جدول (٧-٣٠)

الدرجة	0-	5-	10-	15-	20-25	$\sum$
عدد الطلاب	12	88	100	90	10	300

المطلوب

١- كون جدول التوزيع التكراري النسبي ثم وضح ذلك بيانياً.

٢- وفق منحنى دالة كثافة الاحتمال للتوزيع المعتاد كتقدير لدالة كثافة احتمال درجة الطلاب في مادة الحاسب ووضح ذلك بيانياً.

(٧-٥): الجدول التالي يوضح توزيع 1000 عامل حسب مستوى الأداء بأحد المصانع:-

جدول (٧-٣١)

مستوى الأداء	أقل من متوسط	متوسط	جيد	ممتاز	$\sum$
عدد العمال	75	500	375	50	1000

المطلوب

١- حول مستوى الأداء من متغير وصفي إلى متغير كمي.

٢- أوجد التوزيع التكراري النسبي ثم وضح ذلك بيانياً.

٣- باستخدام طريقة المربعات العزوم قدر دالة احتمال مناسبة لمستوى الأداء.

(٧-٦): الجدول التالي يوضح قيم المتغير x والقيم المناظرة لها للمتغير y

جدول (٧-٣٢)

X	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

المطلوب

١- ارسم شكل الانتشار الذي يوضح اتجاه العلاقة بين المتغيرين.

٢- أبني نموذج انحدار y على x.

٣- احسب الخطأ المعياري.

٤- احسب معامل التحديد.

٥- باستخدام (٤) أوجد معامل الارتباط بين x , y ثم عقب على الناتج.

(٧-٧): الجدول التالي يوضح عدد الوحدات المنتجة x (بالألف وحدة) من منتج معين في إحدى الشركات كذلك الربح السنوي y (بالألف جنيه):-

جدول (٧-٣٣)

x	1	2	3	4	5	6	7	8
y	20	35	44	47	44	35	20	10

المطلوب

١- ارسم شكل الانتشار ومن الرسم حدد نوع العلاقة.

٢- باستخدام طريقة المربعات الصغرى أبني نموذج انحدار y على x.

٣- قدر قيمة y عندما x=4 , x=15.

٤- أوجد الخطأ المعياري للتقدير.

(٧-٨): قامت إحدى الشركات الإنتاجية بدراسة العلاقة بين مقدار التغير في إنتاجية العامل اليومية y ومقدار التغير في الدخل الشهري للعامل x_1 ومقدار التغير في مستوى الخدمات المقدمة للعامل x_2 والجدول التالي يعطي بيانات عن 10 عمال بهذه الشركة:-

جدول (٧-٣٤)

مقدار التغير في إنتاجية العامل y	**10**	**4**	**0**	**7**	**3**	**6**	**15**	**8**	**12**	**5**
مقدار التغير في دخل العامل x_1	**2**	**0.5**	**0.1**	**5**	**1**	**5**	**8**	**2**	**7**	**1.4**
مقدار التغير في مستوى الخدمات المقدمة للعامل x_2	**1**	**0**	**2**	**0**	**0.5**	**1**	**4.5**	**2**	**3.5**	**2.5**

المطلوب

بافتراض أن العلاقة بين مقدار التغير في إنتاجية العامل y و مقدار التغير في دخل العامل x_1 و مقدار التغير في مستوى الخدمات المقدمة للعامل x_2 :-

١- **حدد العلاقة بين** y, x_1, x_2.

٢- **قدر العلاقة بين** y, x_1, x_2.

٣- **قدر قسيمة y عندما** $x_1 = 4, x_2 = 3$.

(٧-٩): الجدول التالي يوضح درجات 10 طلاب في مادة الإحصاء y والرياضيات x_1 ومادة علوم الحاسب x_2 بافتراض أن النهاية العظمى للدرجة 100

جدول (٧-٣٥)

درجة الإحصاء y	**50**	**30**	**20**	**70**	**72**	**20**	**0**	**80**	**100**	**98**

درجة الرياضيات x_1	50	30	20	80	75	10	0	80	90	95
درجة علوم الحاسب x_2	50	10	10	70	60	0	10	55	70	95

المطلوب

١- بافتراض أن العلاقة خطية y, x_1, x_2 أوجد العلاقة بين y, x_1, x_2.

٢- قدر العلاقة y, x_1, x_2.

٣- قدر قيمة y هندما $x_1 = 65, x_2 = 60$.

الباب الثامن
نظرية المعاينة للانحدار والارتباط الخطي
Sampling Theory of Linear Regression and Correlation

(٨-١) العلاقات المقدرة
Estimated Relationships

(٨-٢) توزيعات المعاينة للانحدار الخطي
Sampling distributions of Linear Regression

(٨-٣) فترات الثقة لـ y_i, a_0, a_1
Confidence Intervals of y_i, a_0, a_1

(٨-٤) اختبار معنوية نموذج الانحدار
Test of Significance of a Regression Model

(٨-٥) اختبار معلمات الانحدار
Test of Regression Parameters

(٨-٦) اختبار معنوية معامل الارتباط الخطي
Test the Significance of the Linear Correlation Coefficient

(٨-٧) تمرينات
Exercises

(٨-١) العلاقة المقدرة Estimated Relationship

في الباب السابق تناولنا بالدراسة التفصيلية كيفية توفيق المنحنيات (أي الدوال أو النماذج) الإحصائية . حيث تناولنا بالتفصيل النماذج الخطية Linear Model أو النماذج التي تؤول إلى النماذج الخطية والتي تسمى بنماذج الانحدار الخطي Linear Regression Models.

فإذا تم بناء النموذج باستخدام بيانات المجتمع فإنه يسمى في هذه الحالة بنموذج الانحدار للمجتمع Population Regression Model أما إذا تم بناء نموذج الانحدار باستخدام بيانات العينة فإن النموذج (أو المنحنى) في هذه الحالة يمكن أن يمثل تقدير للعلاقة في المجتمع محل الدراسة.

وكما ذكرنا سابقاً أن المنحنيات (أي الدوال أو النماذج) التي تم توفيقها من بيانات عينة عشوائية مسحوبة من مجتمع الدراسة تختلف من عينة لأخرى وبالتالي تكون هذه المنحنيات منحنيات تقديرية.

ومن هنا يكون من الأهمية إجراء استدلال إحصائي لهذه النماذج التي يمكن أن تمثل علاقة تقديرية للعلاقة في المجتمع.

فإذا كانت العلاقة المقدرة من بيانات العينة للمتغير التابع y والمتغير (أو المتغيرات) المستقل (أو المستقلة) x على النحو :-

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x$$

فإنه يصبح من الضروري الإجابة على الأسئلة التالية:-

١- هل اتجاه العلاقة بين المتغير التابع y والمتغير المستقل x في المجتمع محل الدراسة اتجاه خطي أيضاً أم لا ؟

وللإجابة على هذا السؤال لابد من إجراء اختبار إحصائي يحدد مدى ملائمة المنحنى الذي تم توفيقه من بيانات العينة للمنحنى الذي يمثل العلاقة في المجتمع المسحوب منه العينة وتسمى هذه الاختبارات باختبارات جودة التوفيق **Goodness of Fitting Teses** وقد تكون العلاقة التقديرية التي تم إيجادها باستخدام بيانات العينة تمثل العلاقة في المجتمع محل الدراسة تمثيل جيد والاختلاف بين العلاقة النظرية[1] المفترضة في المجتمع والعلاقة التقديرية يرجع إلى عوامل المعاينة وهنا يصبح السؤال التالي:-

٢- ما هي فترات الثقة لكل مقدر من المقدرات $\hat{y}_i, \hat{a}_0, \hat{a}_1$ ؟

وبتطبيق نفس الشيء بالنسبة للتوزيعات الاحتمالية للمتغيرات العشوائية ، حيث دالة الاحتمال **Probability Function** أو دالة كثافة الاحتمال **Density Function** يمثلها المتغير التابع y (أي أن المتغير التابع هو السلوك الاحتمالي للمتغير x) والمتغير العشوائي يمثل المتغير المستقل x.

وقد يكون التوزيع المقدر من بيانات العينة بناء على افتراض توزيع نظري[2] معين تقدير جيد والاختلاف بين التوزيع المقدر من بيانات العينة والتوزيع النظري يرجع إلى عوامل المعاينة فقط . ولكن قد لا يمثل التوزيع الذي تم توفيقه من بيانات العينة التوزيع الاحتمالي الفعلي للمتغير محل الدراسة في المجتمع المسحوب منه العينة . وبالتالي يعتبر التوزيع الاحتمالي المقدر تقدير غير جيد ، وبصفة عامة التقدير غير الجيد للعلاقة بين متغير تابع y ومتغير (أو متغيرات) مستقلة يؤدي استخدامه إلى الوقوع في

[1]) العلاقة النظرية هي الصياغة الرياضية للدالة التي تمثل العلاقة بين المتغير التابع والمتغير المستقل مثل العلاقة الخطية ، العلاقة الأسية ،...، الخ.

[2]) التوزيع الاحتمالي النظري هو التوزيع المعروف بالصياغة الرياضية لدالة احتماله (أو دالة كثافة احتماله) كذلك خصائص خذه الدالة تكون محددة ومعروفة مثل توزيع ذات الحدين التوزيع المعتاد ، التوزيع الأسي ، ...الخ .

النوع الثاني من الخطأ عند اتخاذ القرار * أي قبول فرض خطأ (أنظر الفصل (٥-١)) وإجراء الاختبارات الإحصائية لتجنب الوقوع في النوع الثاني من الخطأ تعتبر اختبارات صعبة ومعقدة لغير المتخصصين لذلك صممت بعض الاختبارات يمكن أن يستخدمها غير المتخصصين والتي تجنب متخذ القرار في الوقوع في النوع الثاني من الخطأ.

وسوف نتناول في هذا الباب اختبارين أكثر استخداماً من هذه الاختبارات وهما اختبار χ^2 واختبار كولومجروف سيمرونوف في الباب التالي.

(٨-٢) توزيعات المعاينة للانحدار الخطي
Sampling Probabilities of Linear Regression

إذا أعتبر أن العلاقة بين x , y على النحو:-

$$y = a_0 + a_1 x + \varepsilon \qquad (8.1)$$

حيث ε متغير عشوائي يتبع التوزيع المعتاد بحيث

$$E(\varepsilon) = 0 \ , Var(\varepsilon) = \gamma^2 \ , Cov(\varepsilon_i, \varepsilon_j) = 0 \ , i \neq j \qquad (8.2)$$

فإذا كانت العلاقة التقديرية للنموذج الاحتمالي في (8.1) باستخدام عينة حجمها n تأخذ الشكل التالي:-

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x \qquad (8.3)$$

حيث تم الحصول على $\hat{a}_0, \hat{a}_1$ باستخدام طريقة المربعات الصغرى (أنظر الفصل (٧-٤)).

* Heinz Tohler (1994) : "Statistics for Business and Economics" , Harper Collins Collage Publishers , New york

وفيما يلي سوف نقدم التوزيعات الاحتمالية لكل من للتقديرات $\hat{y}_i, \hat{a}_0, \hat{a}_1$ **والتي تسمى بتوزيعات المعاينة لنموذج الانحدار وذلك لأهمية هذه التوزيعات لإيجاد فترات الثقة واجراء اختبارات الفروض لكل من** $\hat{y}_i, \hat{a}_0, \hat{a}_1$.

<u>نظرية (٨-١)</u>

إذا كان ε **متغير يتبع التوزيع المعتاد وبحيث:-**

$$E(\varepsilon) = 0 \ , \ Var(\varepsilon) = \gamma^2 \ , \ Cov(\varepsilon_i, \varepsilon_j) = 0 \quad , i \neq j$$

١- فإن $\hat{a}_0$ **في (8.3) متغير معتاد بحيث:-**

$$E(\hat{a}_0) = a_0 \ , Var(\hat{a}_0) = \frac{\gamma^2 (\sum_{i=1}^{n} x_i^2)}{n[\sum (x_i - \bar{x})^2]} \qquad (8.4)$$

٢- كذلك فإن $\hat{a}_1$ **متغير معتاد بحيث:-**

$$E(\hat{a}_1) = a_0 \ , Var(\hat{a}_0) = \frac{\gamma^2}{\sum (x_i - \bar{x})^2} \qquad (8.5)$$

٣- كذلك $\hat{y}$ **متغير معتاد بحيث:-**

$$E(\hat{y}_i) = y_i \ , Var(\hat{y}_i) = \gamma^2 [\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}] \qquad (8.6)$$

<u>نظرية (٨-٢)</u>

إذا كان γ^2 **يشير إلى تباين المتغير** ε **فإن** $\hat{\gamma}^2$ **حيث:-**

$$\hat{\gamma}^2 = \frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-2} = \frac{\sum_{i=1}^{n} (y_i - \hat{y})^2}{n-2} \qquad (8.7)$$

يمثل تقدير غير متحيز للمعلمة γ^2.

نظرية (٨-٣)

إذا كان المتغير العشوائي ε يتبع التوزيع المعتاد بحيث:-

$$E(\varepsilon)=0 \quad , \text{Var}(\varepsilon)=\gamma^2 \quad , \text{Cov}(\varepsilon_i,\varepsilon_j)=0 \quad , i \neq j \qquad (8.8)$$

فإن :-

١- المتغير

$$\left[\frac{\sum (y_i-\hat{y})^2 (n-2)}{\gamma^2}\right] \qquad (8.9)$$

متغير عشوائي يتبع توزيع χ^2 بدرجات حرية **(n-2)**.

٢- المتغير

$$\left[\frac{\sum (\hat{y}_i-\bar{y})^2}{\gamma^2}\right] \qquad (8.10)$$

متغير عشوائي يتبع χ^2 بدرجة حرية واحدة.

٣- وبالتالي فإن المتغير **f** حيث:

$$f=\frac{\left[\sum (\hat{y}-\bar{y})^2\right]}{\left[\sum (y_i-\hat{y})^2 (n-2)\right]} \qquad (8.11)$$

متغير عشوائي يتبع توزيع فيشر بدرجات حرية **(1,n-2)**.

نظرية (٨-٤)

إذا فرضنا أن ε متغير معتاد بحيث:-

$$E(\varepsilon)=0 \quad , \quad \text{Var}(\varepsilon)=\gamma^2 \quad , \quad \text{Cov}(\varepsilon_i,\varepsilon_j)=0 \quad , i \neq j$$

فإن:-

١- المتغير

$$\frac{\hat{a}_0 - a_0}{\sqrt{\dfrac{\gamma^2(\sum x_i^2)}{n[\sum(x_i-\bar{x})^2}}} = \frac{(\hat{a}_0 - a_0)\sqrt{n(\sum(x_i-\bar{x})^2)}}{\sqrt{\gamma^2(\sum x_i^2)}} \quad (8.12)$$

متغير يتبع التوزيع المعتاد القياسي.

٢- كذلك المتغير

$$\frac{\hat{a}_1 - a_1}{\sqrt{\dfrac{\gamma^2}{\sum(x_i-\bar{x})^2}}} = \frac{(\hat{a}_1 - a_1)\sqrt{(\sum(x_i-\bar{x})^2)}}{\sqrt{\gamma^2}} \quad (8.13)$$

متغير يتبع التوزيع المعتاد القياسي.

أما إذا استخدمنا $\hat{\gamma}^2$ كتقدير للمعلمة γ^2 فإنه في هذه الحالة:-

٣- المتغير:-

$$\frac{\hat{a}_0 - a_0\sqrt{n(\sum(x_i-\bar{x})^2)}}{\sqrt{\gamma^2(\sum x_i^2)}} \quad (8.14)$$

متغير يتبع توزيع استيودنت بدرجات حرية **(n-2)**.

٤- كذلك التغير:-

$$\frac{\hat{a}_1 - a_1\sqrt{(\sum(x_i-\bar{x})^2)}}{\sqrt{\hat{\gamma}^2}} \quad (8.15)$$

متغير يتبع استيودنت **(t)** بدرجات حرية **(n-2)** أيضاً.

<u>نظرية (٨-٥)</u>

إذا كانت ε متغير عشوائي يتبع التوزيع المعتاد بحيث:-

$$E(\varepsilon) = 0 \;,\; \mathrm{Var}(\varepsilon) = \gamma^2 \;,\; \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0 \;, i \neq j$$

فإن:-

١- المتغير $\hat{y}$ (عند المتغير x_i) متغير معتاد بحيث:-

$$E(\hat{y}_i) = y_i \ , Var(\hat{y}_i) = \gamma^2[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2} + 1] \quad (8.16)$$

٢- إذا كانت المعلمة γ^2 معلومة فإن المتغير:-

$$\frac{\hat{y}_i - y_i}{\gamma\sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2} + 1}} \quad (8.17)$$

متغير يتبع التوزيع المعتاد القياسي.

٣- إذا كانت المعلمة γ^2 غير معلومة فإن $\hat{\gamma}^2$ تستخدم تقدير لها وفي هذه الحالة فإن المتغير :-

$$\frac{\hat{y}_i - y_i}{\hat{\gamma}\sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2} + 1}} \quad (8.18)$$

متغير يتبع توزيع استيودنت بدرجات حرية (n-2).

ملحوظة

في حالة إذا كان حجم العينة كبير $(n \geq 30)$ والمعلمة γ^2 غير معلومة واستخدم $\hat{\gamma}^2$ تقدير لها فإن المتغير في (8.18) يؤول إلى التوزيع المعتاد القياسي.

(٨-٣) فترات الثقة لـ y_i, a_0, a_1

Confidence Intervals of y_i, a_0, a_1

باستخدام توزيعات المعاينة التي تم تقديمها في الفصل السابق سوف نقدم في هذا الباب فترات الثقة لكل من y_i , a_0 , a_1 على النحو التالي:-

أولاً : فترة الثقة للمعلمة (a_0)

١- إذا فرضنا أن درجة الثقة $(1-\alpha)$ وأن γ^2 معلومة فإن:-

$$P_r\{-Z_{\alpha/2} \le \frac{(\hat{a}_0 - a_0)\sqrt{n\sum(x_i-\bar{x})^2}}{\sqrt{\gamma^2\sum x_i^2}} \le Z_{\alpha/2}\} = 1-\alpha \qquad (8.19)$$

حيث يتم إيجاد $Z_{\alpha/2}$ من جدول التوزيع المعتاد القياسي،ومن العلاقة (8.19) نجد أن فترة الثقة للمعلمة a_0 على النحو التالي:-

$$\hat{a}_0 \pm Z_{\alpha/2}\sqrt{\frac{\gamma^2\sum x_i^2}{n\sum(x_i-\bar{x})^2}} \qquad (8.20)$$

٢- إذا كانت المعلمة γ^2 غير معلومة واستخدام $\hat{\gamma}^2$ تقدير لها فإنه عند درجة الثقة $(1-\alpha)$ نجد أن:-

$$P_r\{-t_{\alpha/2} \le \frac{(\hat{a}_0 - a_0)\sqrt{n\sum(x_i-\bar{x})^2}}{\sqrt{\gamma^2\sum x_i^2}} \le t_{\alpha/2}\} = 1-\alpha \qquad (8.21)$$

حيث يتم إيجاد $t_{\alpha/2}$ من جدول توزيع استيودنت عند درجة الثقة (n-2) وبالتالي فإن فترة الثقة للمعلمة a_0 تصبح على النحو التالي:-

$$\hat{a}_0 \pm t_{\alpha/2}\sqrt{\frac{\hat{\gamma}^2\sum x_i^2}{n\sum(x_i-\bar{x})^2}} \qquad (8.22)$$

ثانياً : فترة الثقة للمعلمة (a_1)

١- إذا كانت المعلمة γ^2 معلومة فإن:-

$$P_r\{-Z_{\alpha/2} \le \frac{(\hat{a}_1 - a_1)}{\sqrt{\frac{\gamma^2}{\sum(x_i-\bar{x})^2}}} \le Z_{\alpha/2}\} = 1-\alpha \qquad (8.23)$$

وبالتالي فإن فترة الثقة للمعلمة a_1 على النحو التالي:-

$$\hat{a}_1 \pm Z_{\alpha/2}\sqrt{\frac{\gamma^2}{\sum(x_i-\bar{x})^2}} \qquad (8.24)$$

٢- إذا كانت المعلمة γ^2 غير معلومة واستخدمت $\hat{\gamma}^2$ كتقدير لها فإن:-

$$P_r\{-t_{\alpha/2} \le \frac{(\hat{a}_1 - a_1)}{\sqrt{\frac{\hat{\gamma}^2}{\sum(x_i-\bar{x})^2}}} \le t_{\alpha/2}\} = 1-\alpha \qquad (8.25)$$

وبالتالي فإن فترة الثقة للمعلمة (a_1) في هذه الحالة تصبح على النحو التالي:-

$$\hat{a}_1 \pm t_{\alpha/2}\sqrt{\frac{\hat{\gamma}^2}{\sum(x_i-\bar{x})^2}} \qquad (8.26)$$

ثالثاً : فترة الثقة للقيمة y_i

١- إذا كانت γ^2 معلومة فإن فترة الثقة لـ y_i على النحو:-

$$\hat{y}_i \pm Z_{\alpha/2}\gamma\sqrt{\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum x_i^2 - n\bar{x}^2}+1} \qquad (8.27)$$

حيث $Z_{\alpha/2}$ يتم إيجادها من جدول التوزيع المعتاد القياسي.

٢- أما إذا كانت γ^2 غير معلومة فإن فترة الثقة لـ y_i تصبح على النحو التالي:-

$$\hat{y}_i \pm t_{\alpha/2}\hat{\gamma}\sqrt{\frac{1}{n}+\frac{(x_i-\bar{x})^2}{\sum x_i^2 - n\bar{x}^2}+1} \qquad (8.28)$$

يتم إيجاد $t_{\alpha/2}$ من جدول توزيع استيودنت بدرجات حرية **(n-2)**.

ملحوظة: في هذه الحالة إذا كان حجم العينة **n** كبير (أي $n \geq 30$) فإنه يمكن استخدام $Z_{\alpha/2}$ بدلاً من $t_{\alpha/2}$

مثال (٨-١)

الجدول التالي يوضح عينة من الكميات المطلوبة **(y)** بالألف وحدة وسعر بيع الوحدة المناظر **(x)** بالجنيه من إحدى السلع التموينية في الأسواق الاستهلاكية المختلفة

جدول (٨-١)

Y	**2**	**1**	**5**	**0**	**3**	**4**	**3**	**1**	**0**	**1**
x	**3**	**4**	**1**	**5**	**3**	**1**	**2**	**4**	**5**	**4**

باستخدام طريقة المربعات الصغرى فإذا تم تقدير العلاقة بين x , y فكانت على النحو التالي:-

$$\hat{y} = 5.59 - 1.121x$$

أي

$$\hat{a}_0 = 5.59, \hat{a}_1 = -1.121$$

المطلوب

إذا كان تباين المتغير العشوائي يساوي 0.04 (أي $\gamma^2 = 0.04$) أوجد بدرجة ثقة 95%:-

١- فترة الثقة للمعلمة a_0.

٢- فترة الثقة للمعلمة a_1.

٣- فترة الثقة للمعلمة y_i عندما $x_i = 3$.

الحل

نكون الجدول التالي (جدول (٨-٢)).

جدول (٨-٢)

x	y	x^2	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
3	2	9	3-3.2=0.2	0.004
4	1	16	4-3.2=0.8	0.064
1	5	1	1-3.2=-2.2	4.84
5	0	25	5-3.2=1.8	3.24
3	3	9	3-3.2=0.2	0.04
1	4	1	1-3.2=-2.2	4.84
2	3	4	2-3.2=-1.2	1.44
4	1	16	4-3.2=0.8	0.64
5	0	25	5-3.2=1.8	3.24

4	1	16	4-3.8=0.8	0.64
32	20	122		19.6

وبما أن γ^2 معلومة حيث $\gamma^2 = 0.04$ بالتالي عند درجة الثقة 95% نجد أن $\frac{\alpha}{2} = 0.025$ وبالتالي من جدول التوزيع المعتاد القياسي نجد أن:-

$$Z_{\alpha/2} = \pm 1.96$$

١- من العلاقة (8.20) نجد أن:

أ- الحد الأدنى لـ a_0 يساوي:-

$$\hat{a}_0 - Z_{\alpha/2}\gamma\sqrt{\frac{\sum x_i^2}{n\sum(x_i - \bar{x})^2}} = 5.59 - 1.96(0.2)\sqrt{\frac{122}{10(19.6)}}$$

$$= 5.59 - 3.416 = 2.174 \qquad (1)$$

ب- الحد الأعلى لـ a_0 يساوي:-

$$\hat{a}_0 + Z_{\alpha/2}\gamma\sqrt{\frac{\sum x_i^2}{n\sum(x_i - \bar{x})^2}} = 5.59 + 1.96(0.2)\sqrt{\frac{122}{10(19.6)}}$$

$$= 5.59 + 3.416 = 9.006 \qquad (2)$$

من (2) , (1) نجد أن:-

$$2.174 \le a_0 \le 9.006$$

وذلك باحتمال 95%.

٢- من العلاقة (8.26) نجد ان:-

أ- الحد الأدنى لـ a_1 يساوي:-

$$\hat{a}_1 - Z_{\alpha/2}\sqrt{\gamma^2 \sum (x_i - \bar{x})^2} = -1.121 - 1.96\sqrt{0.04(19.6)}$$

$$= -1.121 - 1.736 = -2.857$$

ب- الحد الأدنى لـ a_1 يساوي:-

$$\hat{a}_1 + Z_{\alpha/2}\sqrt{\gamma^2 \sum (x_i - \bar{x})^2} = -1.121 + 1.96\sqrt{0.04(19.6)}$$

$$= -1.121 + 1.736 = 0.65$$

وذلك بدرجة ثقة 95%.

٣- من العلاقة (8.28) نجد أن:-

أ – الحد الأدنى لـ y_i عندما $x_i = 3$ يساوي

$$\hat{y} - Z_{\alpha/2}\gamma\sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2} + 1}$$

$$= 2.227 - 1.96(0.2)\sqrt{\frac{1}{10} + \frac{(3-3.2)^2}{122 - 10(3.2)} + 1}$$

$$= 2.227 - 1.96(0.2)\sqrt{0.1 + \frac{0.04}{19.6} + 1}$$

$$= 2.227 - 1.96(0.2)(1.049)$$

$$= 2.227 - 1.96(0.2098)$$

$$= 2.227 - 0.411208 = 1.815792 \approx 1.82$$

ب-الحد الأعلى لـ y_i عندما $x_i = 3$ يساوي

$$\hat{y} + Z_{\alpha/2}\gamma\sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum x_i^2 - n\bar{x}^2} + 1}$$

$$= 2.227 + 1.96(0.2)\sqrt{\frac{1}{10} + \frac{(3-3.2)^2}{122-10(3.2)} + 1}$$

$$= 2.227 + 0.411208 = 2.6382 \approx 2.64$$

وبالتالي فإن:-

$$1.82 \le y_i \le 2.64$$

عندما x=3 وذلك بدرجة ثقة 95%.

مثال (٨-٢)

في المثال السابق إذا كانت γ^2 غير معلومة ، أوجد باستخدام طريقة المربعات الصغرى $\hat{\gamma}^2$ ثم أوجد:-

١- **فترة ثقة للمعلمة a_0 ثم قارن هذه الفترة بفترة الثقة لـa_0 عندما $\gamma^2 = 0.04$.**

٢- **أوجد فترق ثقة لـa_1 ثم قارن هذه الفترة بالفترة المناظرة لها عندما $\gamma^2 = 0.04$.**

<u>الحل</u>

نحسب من بيانات العينة في الجدول السابق قيمة التقدير $\hat{\gamma}^2$ على النحو التالي:-

$$\hat{\gamma}^2 = \frac{\sum \varepsilon_i^2}{n-2} = \frac{\sum (y_i - \hat{y}_i)^2}{n-2}$$

$$= \frac{\sum y_i^2 - \hat{a}_0 \sum y - \hat{a}_1 \sum x_i y_i}{n-2}$$

$$= \frac{66 - (5.59)(20) - (-1.121)(42)}{10-2}$$

$$= \frac{1.282}{8} = 0.16025$$

وبما أن درجة الثقة 95% فإنه عند درجات الحرية 8 فإنه يمكن إيجاد قيمة $t_{\alpha/2}$ **حيث:-**

$$t_{\alpha/2} = 2.306$$

وبالتالي فإن:-

١- حدود فترات الثقة لـ a_0 **على النحو التالي-**

$$\hat{a}_0 \pm t_{\alpha/2} \hat{\gamma} \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}$$

$$5.59 \pm 2.306 \frac{(0.4)(11.05)}{196}$$

$$5.59 \pm 0.052 = 5.438$$

$$\therefore 3.28 \le a_0 \le 5.64$$

ونلاحظ أن فترة الثقة هذه لـ a_0 **أقل من فترة الثقة عندما** $\gamma^2 = 0.04$.

٢- كذلك فإن حدود الثقة لـ a_1 **هي:-**

$$\hat{a}_1 \pm t_{\alpha/2}\hat{\gamma}\sqrt{\sum(x_i-\bar{x})^2}$$
$$-1.121 \pm 2.306(0.4)\sqrt{19.6}$$
$$-1.121 \pm 4.084$$
$$\therefore -5.205 \le a_1 \le 2.963$$

كذلك نلاحظ لـ a_1 في هذه الحالة أكبر من فترة الثقة لـ a_1 عندما $\gamma^2 = 0.04$.

(٨-٤) اختبار معنوية نموذج الانحدار الخطي
Test of Significance of a Linear Regression Model

وكما ذكرنا في الفصول السابقة أن نموذج الانحدار الخطي الذي يتم اشتقاقه باستخدام بيانات العينة يعتبر تقدير للعلاقة الفعلية في المجتمع محل الدراسة ولذا يسمى بالنموذج المقدر . وبالتالي يكون من الضروري إجراء اختبار إحصائي يمكننا من معرفة مدى ملائمة الصياغة الرياضية الخطية للنموذج المقدر للعلاقة في المجتمع محل الدراسة المسحوب منه العينة أو بعبارة أخرى هل العلاقة في المجتمع علاقة خطية أم لا؟ . وتسمى الاختبارات التي يمكن باستخدامها معرفة مدى ملائمة النموذج المقدر للعلاقة الفعلية في المجتمع باختبارات معنوية نماذج الانحدار **Testes of Significance of a Regression Model** وسوف تقتصر دراستنا في هذا الفصل على اختبار العلاقة الخطية.

من دراستنا في الفصل السابق نستنتج أن المتغير **f** حيث:-

$$f = \frac{\sum(\hat{y}_i-\bar{y})^2/1}{\sum(y_i-\hat{y})^2/n-2} \tag{8.29}$$

متغير عشوائي يتبع توزيع فيشر بدرجة حرية (n-1) . أي أن:-

$$f \sim F(1, n-2)$$

وسوف نستخدم استنتاج أن المتغير f يتبع توزيع فيشر في اختبار وجود علاقة انحدارية خطية بين المتغير التابع (y) والمتغير (أو المتغيرات) المستقل x أو عدم وجود هذه العلاقة الانحدارية الخطية وذلك بدرجة ثقة معينة ولتكن $(1-\alpha)$ وذلك باستخدام بيانات عينة حجمها n، ومما سبق يمكن تكوين جدول تحليل التباين Analysis of Table Variance وللاختصار يسمى بـ(ANOVA) على النحو التالي:-

جدول (٨-٣)

مصدر التباين	مجموع المربعات	عدد درجات الحرية	متوسط المربعات	f
الانحدار Regression	$SSR = \sum(\hat{y}-\bar{y})^2$	1	$SSR/1$	$\frac{MSR}{MSE}$
الخطأ Error	$SSE = \sum(y-\hat{y})^2$	n-2	$SSE/n-2$	
الكلي Total	$SST = \sum(y_i-\bar{y})^2$	n-1		

وباستخدام جدول تحليل التباين يمكن حساب قيمة f من بيانات العينة وبمقارنة f التي تم حسابه باستخدام جدول تحليل التباين بقيمة f من جدول توزيع فيشر عند درجات الحرية (1,n-2) ودرجة ثقة $(1-\alpha)$ يمكن معرفة الفرض القائل بوجود علاقة انحدار خطي بين المتغير التابع y والمتغير (أو المتغيرات) المستقل x وذلك بدرجة ثقة $(1-\alpha)$.
وسوف نوضح إجراء الاختبار في حالتين هما:-

الحالة الأولى

إذا اعتبرنا أن العلاقة المقدرة من العينة على النحو:-

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x$$

ويكون المطلوب هو اختبار الفرض القائل بعدم وجود علاقة خطية بين x , y في المجتمع المسحوب منه العينة وذلك من خلال بيانات العينة أو بعبارة أخرى اختبار الفرض القائل بأن المعلمة $a_1 = 0$ ضد الفرض البديل القائل بأن المعلمة $a_1 \neq 0$ (وهو وجود علاقة انحدار خطي بين x , y).

وفيما يلي خطوات إجراء الاختبار:-

١- الفرض العدمي

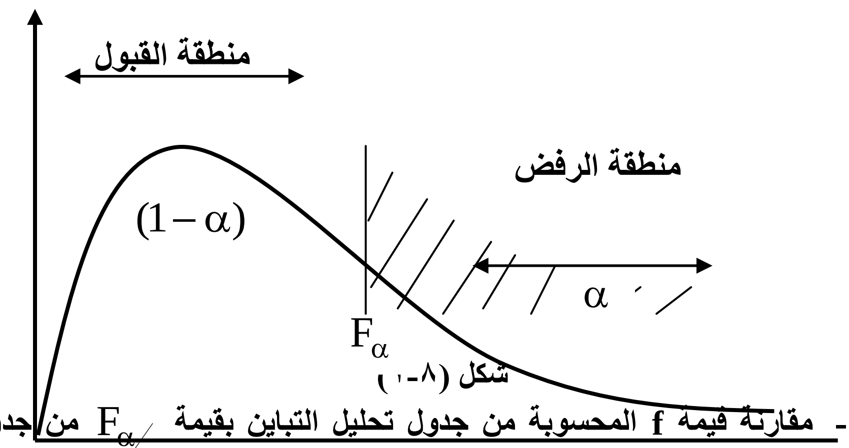
$H_0 : a_1 = 0$ (عدم وجود علاقة خطية)

الفرض البديل

$H_1 : a_1 \neq 0$ (وجود علاقة خطية)

٢- تكوين جدول تحليل التباين وإيجاد القيمة **f**.

٣- عند درجة ثقة $(1-\alpha)$ نوجد من جدول توزيع فيشر قيمة F_α (القيمة الجدولية) عند درجات حرية **(1,n-2)**.



شكل (٨-١٠)

٤- مقارنة قيمة **f** المحسوبة من جدول تحليل التباين بقيمة $F_{\alpha/2}$ من جدول توزيع فيشر فإذا كان:-

$$i)\ f < F_\alpha \qquad (8.30)$$

نقبل الفرض العدمي القائل بعدم وجود علاقة خطية بين **x , y** وذلك بدرجة ثقة $(1-\alpha)$.

$$ii)\ f \geq F_\alpha \qquad (8.31)$$

نرفض الفرض العدمي ونقبل الفرض البديل القائل بوجود علاقة خطية بين y , x وذلك بدرجة ثقة $(1-\alpha)$.

مثال (٨-٣)

الجدول التالي يوضح مقدار الزيادة في سعر اللتر الواحد من البنزين والزيادة المناظرة في تكلفة نقل الوحدة الواحدة من منتج معين من مراكز الإنتاج إلى مراكز الاستهلاك خلال فترة زمنية معينة.

جدول (٨-٤)

مقدار الزيادة في سعر اللتر الواحد من البنزين (x)	1	2	3	4	5
مقدار الزيادة في تكلفة نقل الوحدة من المنتج (y)	3	3	5	6	8

المطلوب

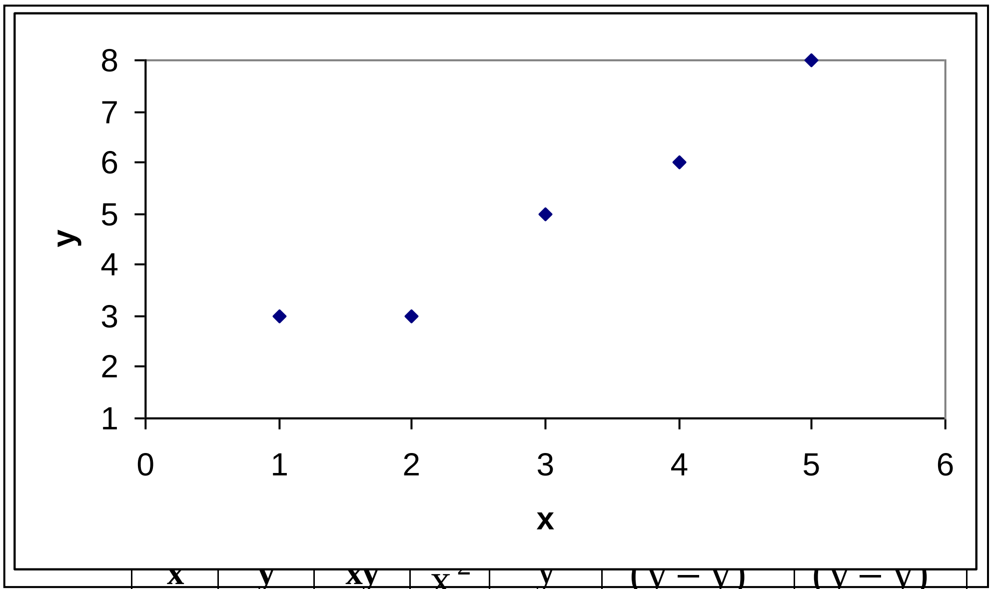
١- ارسم شكل الانتشار وحدد اتجاه العلاقة بين x , y.

٢- باستخدام طريقة المربعات الصغرى كون نموذج انحدار y على x.

٣- اختبر الفرض القائل بأنه يوجد علاقة خطية بين x , y في المجتمع المسحوب منه العينة بدرجة ثقة 95%.

<u>الحل</u>

١ - الشكل التالي يوضح شكل الانتشار بين x , y ويتضح من الشكل وجود علاقة طردية بين x , y.



x	y	xy	x^2	$\hat{y}$	$(y-\hat{y})^2$	$(y-\bar{y})^2$
1	3	3	1	2.4	0.36	4
2	3	6	4	3.7	0.49	4
3	5	15	9	5	0	0
4	6	24	16	6.3	0.09	1
5	8	40	25	7.6	0.16	9
15	25	88	55		1.1	18

من الجدول نكون المعادلات الطبيعية التالية:-

$$25 = 5\hat{a}_0 + 15\hat{a}_1$$

$$88 = 15\hat{a}_0 + 55\hat{a}_1$$

وبحل المعادلتين نجد أن:-

$$\hat{a}_0 = 1.1 \ , \ \hat{a}_1 = 1.3$$

وبالتالي فإن نموذج الانحدار باستخدام بيانات العينة يأخذ الشكل التالي:-

$$\hat{y} = 1.1 + 1.3x$$

٣- الفرض العدمي:

$H_0 : a_1 = 0$ (عدم وجود علاقة خطية في المجتمع المسحوب منه العينة)

الفرض البديل:

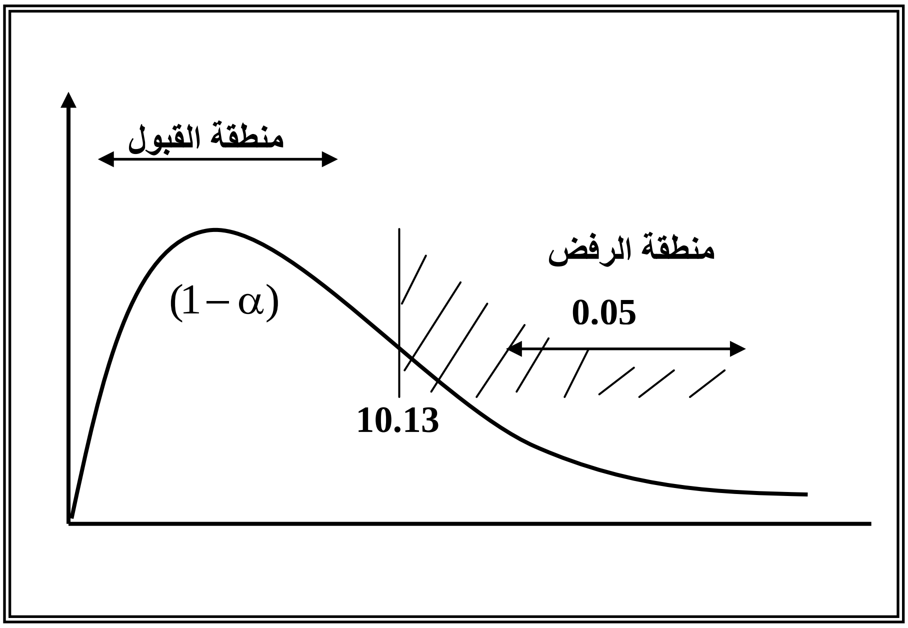
$H_1 : a_1 \neq 0$ (وجود علاقة خطية في المجتمع المسحوب منه العينة)

جدول تحليل التباين

مصدر التباين	مجموع المربعات	عدد درجات الحرية	متوسط المربعات	f
الانحدار	$SSR = 12.9$	1	$12.9/1 = 12.9$	$\frac{12.9}{0.367} = 35.15$
الخطأ	$SSE = 1.1$	5-2=3	$1.1/3 = 0.367$	
الكلي	$SST = 14$	5-1=4		

من جدول توزيع فيشر نجد أن F_α عند درجات حرية (1,3):-

$$F_\alpha = 10.13$$



شكل (٨-٣)

وبمقارنة قيمة f المحسوبة من جدول تحليل التبـاين بقيمـة F_α المحسوبة من جدول فيشر نجد أن:-

$$f > F_\alpha$$

وبالتالي نرفض الفرض العدمي القائل بعدم وجود علاقـة انحدار خطيـة بين x , y في المجتمع المسحوب منه العينة وذلك بدرجة ثقة 95%.

الحالة الثانية

إذا اعتبرنـا أن العلاقة المقدرة من بيانـات عينـة على النحو التالي:-

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + ... + \hat{a}_k x_k$$

ويكون المطلـوب هو اختبـار الفرض القائل بعدم وجـود علاقـة خطيـة بين المتغيـر التابع y والمتغيرات المستقلة $x_1, x_2..., x_n$ فـي المجتمـع المسـحوب منـه العينـة ، وذلـك من خـلال بيانـات العينـة. أو بعبـارة أخـرى اختبـار الفرض القائـل بـان المعلمـات $a_1 = a_2 = ... = a_n = 0$ (أي عدم وجود علاقة خطية في المجتمـع المسـحوب منه العينة) ضد الفرض البديل القائل بأنه يوجد على الأقل معلمة $a_j \neq 0$ بحيث

وفيما يلي خطوات إجراء الاختبار:-

١- الفرض العدمي

$$H_0 : a_j = 0 \qquad , j = 1,2,3,...,k$$

الفرض البديل

$$H_1 : a_j \neq 0 \qquad , j = 1,2,3,...,k$$

٢- نكون جدول تحليل التباين وحساب f (بنفس الأسلـوب فـي الحالـة الأولـى التـي سبق تقديمها على النحو التالي):-

جدول تحليل التباين (٨-٦)

مصدر التباين	مجموع المربعات	عدد درجات الحرية	متوسط المربعات	f
الانحدار	$SSR = \sum(\hat{y}-\bar{y})^2$	**k**	SSR/k	$\dfrac{SSR/k}{SSE/n-(k+1)}$
الخطأ	$SSE = \sum(y-\hat{y})^2$	**n-(k+1)**	$\dfrac{SSE}{n-(k+1)}$	
الكلي	$SST = \sum(y-\bar{y})^2$	**n-1**		

٣- مقارنة قيمة f المحسوبة من جدول تحليل التبـاين بقيمـة من جدول توزيـع فيشر بحيث نقبل الفرض أو نرفض الفرض العدمي بدرجة ثقة $(1-\alpha)$.

مثال (٨-٤)

في إحدى محالات بيع التحف الفنية بالمزاد وجد أن سعر بيع التحفة (y) يتوقف على عمر التحفة وكذلك على عدد العملاء الداخلين للشراء في المزاد والجدول التالي لعينة من أسعار التحف وعمر التحفة وعدد الداخلين في المزاد المناظرة لهذا السعر.

جدول (٨-٧)

y	x_1	x_2	y	x_1	x_2
1235	127	13	2131	170	14
1080	115	12	1550	182	8
845	127	7	1884	162	11
1522	150	9	2041	184	10
1047	156	6	845	143	6
1979	182	11	1483	159	9
1822	156	12	1055	108	14
1253	132	10	1545	175	8
1297	137	9	729	108	6
946	113	9	1792	179	9
1713	137	15	1175	111	15
1024	117	11	1593	187	8
1147	137	8	785	111	7
1092	153	6	744	115	7
1152	117	13	1356	194	5
1336	126	10	1262	168	7

من بيانات العينة في الجدول السابق وباستخدام طريقة المربعات الصغرى أمكن تكوين نموذج الانحدار التالي:-

$$\hat{y} = -1339 + 12.7x_1 + 86x_2$$

ومن الجدول السابق نحصل على التباينات التالية:-

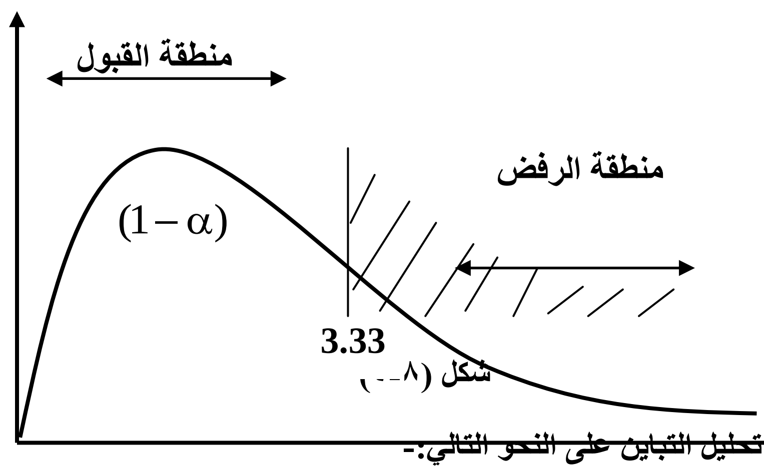
بدرجات حرية (2) SSR=4283063

بدرجات حرية (29) SSE=516727

وعند درجة الثقة 95% اختبر الفرض القائل بوجود علاقة انحدار خطية بين y, x_1, x_2 في المجتمع المسحوب منه العينة.

<u>الحل</u>

عند $(1-\alpha)$ يساوي 95% ، ودرجات حرية (2,29) نجد أن $F_\alpha = 3.33$



شكل (٨-..)

نكون جدول تحليل التباين على النحو التالي:-

جدول تحليل التباين (٨-٨)

مصدر التباين	مجموع المربعات	عدد درجات الحرية	متوسط المربعات	f

الانحدار	4283063	2	2141532	120.19
الخطأ	516727	29	17818	
الكلي	479989	31		

وبما أن:-

$$f > F_{\alpha}$$

وبمـا أن **f** تقـع فـي منطقـة الـرفض بالتـالي نـرفض الفـرض العـدمي القائـل بعـدم وجـود علاقة خطية بين y, x_1, x_2 وذلك بدرجة ثقة 95%.

(٨-٥) اختبارات معلمات الانحدار

Tests of Regression Parameters

في حالة إذا تم اختبار معنوية نموذج الانحدار الخطي وأثبتت أنها خطية عند درجـة ثقة معينة أي أن العلاقة بين المتغير التابع والمتغيرات المستقلة علاقة خطية.
بالنسبة للنموذج المقدر

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x$$

حيث تختلف القيم المقدرة من العينة $\hat{a}_0, \hat{a}_1$ من عينة إلى أخرى وفي الفصل (٨-٣) قمنـا بإنشـاء فتـرات الثقـة للمعلمـات فـي المجتمـع a_0, a_1 باسـتخدام $\hat{a}_0, \hat{a}_1$ ، وباسـتخدام توزيعـات المعاينـة فـي الفصـل (٩-٢) فإننـا فـي هـذا الفصـل سـوف نتنـاول بالدراسة الاختبارات الإحصائية الخاصة بهذه المعلمات على النحو التالي:-

أولاً : الاختبارات الخاصة بالمعلمة a_0

إذا فرضنا أن a'_0 قيمة معينة ويراد اختبار صحة الفرض القائل بـأن $a_0 = a'_0$ بدرجة ثقة $(1-\alpha)$ ويتم إجراء الاختبار على النحو التالي:-

١- الفرض العدمي:

$$H_0 : a_0 = a'_0$$

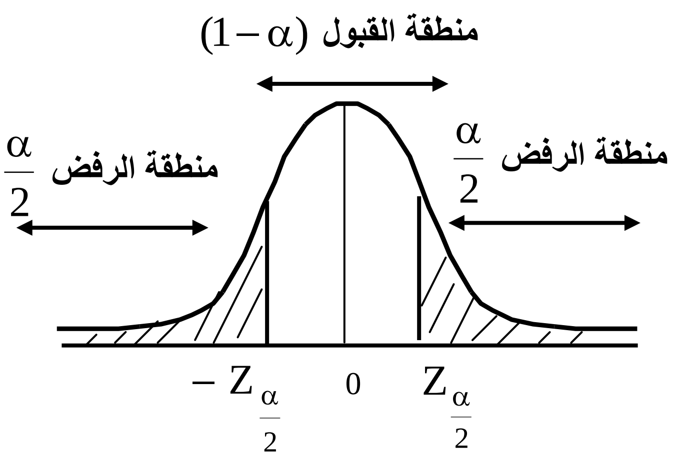
الفرض البديل:

$$H_1 : a_0 \neq a'_0$$

٢- إذا كان γ معلومة نحسب الإحصاء Z حيث:

$$Z = \frac{\hat{a}_0 - a'_0}{\gamma} \sqrt{\frac{n\sum (x_i - \bar{x})^2}{\sum x_i^2}}$$

٣- عند درجة الثقة $(1-\alpha)$ ومن جدول التوزيع المعتاد القياسي نحسب $Z_{\alpha/2}$



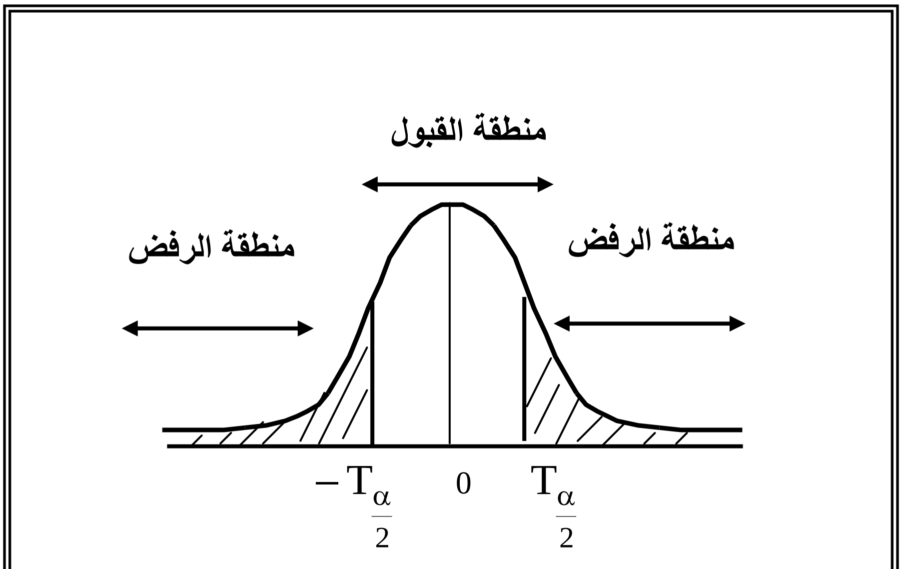
شكل (٨-٩)

نقارن بين القيمة المحسوبة والقيمة $Z_{\alpha/2}$ المحسوبة من جدول التوزيع المعتاد القياسي فإذا كانت Z تقع خارج حدي الثقة $\pm Z_{\alpha/2}$ فإننا نرفض الفرض العدمي بدرجة ثقة $(1-\alpha)$.

٤- إذا كانت γ غير معلومة فإننا نستخدم $\hat{\gamma}$ تقدير لها في هذه الحالة نحسب الإحصاء T حيث:

$$t = \frac{\hat{a}_0 - a'_0}{\hat{\gamma}} \sqrt{\frac{n\sum (x_i - \bar{x})^2}{\sum x_i^2}}$$

٥- وعند درجة الثقة $(1-\alpha)$ نوجد من جدول توزيع استيودنت وعند درجات حرية (n-2) القيمة $t_{\alpha/2}$.



شكل (٨-١٠)

ثم نقارن بين القيمة المحسوبة والقيمة $t_{\alpha/2}$ المحسوبة من جدول توزيع استيودنت.

فإذا كانت خارج حدود الثقة $\pm t_{\alpha/2}$ فغننا نرفض الفرض العدمي بدرجة ثقة $(1-\alpha)$.

مثال (٨-٥)

إذا كانت العلاقة التقديرية بين المتغير التابع y والمتغير المستقل x على النحو التالي:-

$$\hat{y} = 25 + 2x$$

بحيث تم بناء النموذج السابق من بيانات عينة مكونة من 27 مفردة عن المتغيرين x , y كذلك حسب المؤشرات التالية:-

$$\bar{x} = 7.5, \mathrm{Var}(x) = 2.25$$

<u>المطلوب</u>

١- اختبر الفرض القائل بأن $a_0 = 20$ في المجتمع المسحوب منه العينة بدرجة ثقة 95% . إذا فرضنا أن $\gamma^2 = 25$.

٢- اختبر الفرض القائل بأن $a_0 = 21$ بدرجة ثقة 95% إذا كان $\hat{\gamma}^2 = 2.25$.

الحل

١- من النموذج نجد ان $\hat{a}_0 = 25, \hat{a}_1 = 2$

وبما أن :-

$$\mathrm{Var}(x) = \frac{\sum (x_i - \bar{x})^2}{n}$$

$$2.25 = \frac{\sum (x_i - \bar{x})^2}{27}$$

$$\sum (x_i - \bar{x})^2 = (2.25)(27) = 60.75$$

وبما أن

$$\mathrm{Var}(\bar{x}) = \frac{\sum x_i^2}{n} - (\bar{x})^2$$

$$2.27 = \frac{\sum x_i^2}{27} - (7.5)^2$$

$$\sum x_i^2 = (2.27 + 56.25)(27) = 1579.5$$

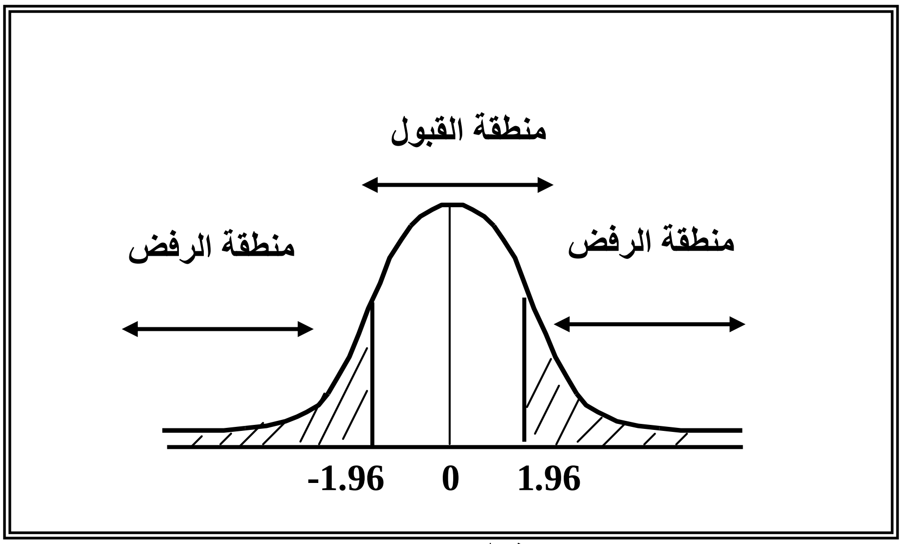
أ- الفرض العدمي:

$$H_0 : a_0 = 20$$

الفرض البديل:

$$H_1 : a_0 \neq 20$$

ب- بما أن γ معلومة حيث $\gamma = 5$ وعند درجة الثقة 95% فمن جدول التوزيع المعتاد القياسي نجد أن حدي الثقة $Z_{\alpha/2} = \pm 1.96$



شكل (٨-١١)

جـ- نحسب الإحصاء Z حيث:-

$$Z = \frac{\hat{a}_0 - a'_0}{\gamma}\sqrt{\frac{n\sum(x_i - \bar{x})^2}{\sum x_i^2}}$$

$$= \frac{25-20}{7}\sqrt{\frac{27(60-75)}{1579.5}} = 1.019$$

د- وبما أن Z تقع في منطقة القبول بالتالي نقبل الفرض القائل بأن $a_0 = 20$ وذلك بدرجة ثقة 95%.

٢-

أ- بما ان $\hat{\gamma} = 1.5$ بالتالي نستخدم الإحصاء t .

الفرض العدمي

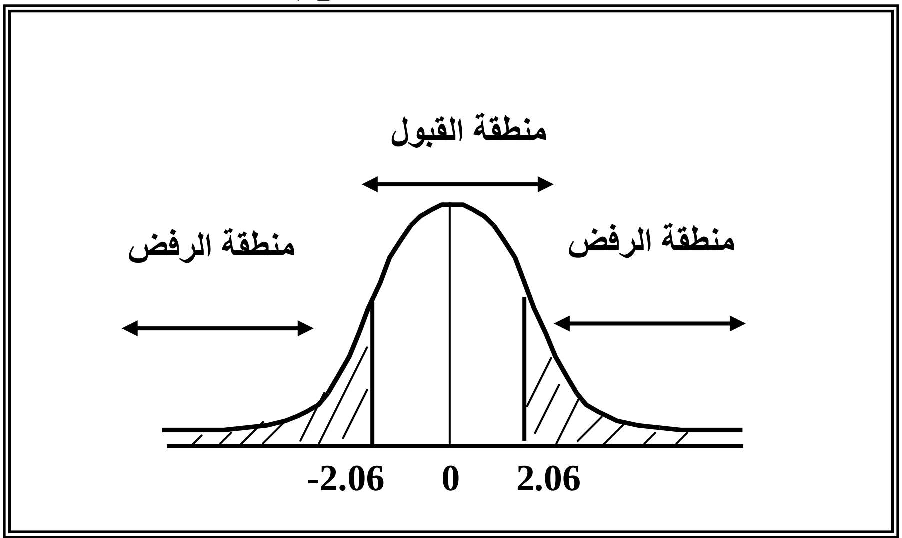
$$H_0 : a_0 = 21$$

الفرض البديل

$$H_1 : a_0 \neq 21$$

ب- من جدول توزيع استيودنت وعند درجات حرية 25 نجد أن:

$$t_{\alpha/2} = \pm 2.06$$



شكل (٨-١٢)

جـ- نحسب الإحصاء t حيث:-

$$t = \frac{\hat{a}_0 - a'_0}{\hat{\gamma}} \sqrt{\frac{n\sum(x_i - \bar{x})^2}{\sum x_i^2}}$$

$$= \frac{25-21}{1.5}\sqrt{\frac{27(60.75)}{1579.5}} = 2.718$$

وبما أن t=2.718 أي تقع في منطقة الرفض وبالتالي نرفض الفرض العدمي القائل بأن $a_0 = 21$ وذلك بدرجة ثقة 95%.

ثانياً : الاختبارات الخاصة بالمعلمة a_1

إذا فرضنا ان a'_1 قيمة معينة والمطلوب اختبار الفرض القائل بأن $a_1 = a'_1$ ففي هذه الحالةيكون الاختبار الإحصائي على النحو التالي:-

١- الفرض العدمي

$$H_0 : a_1 = a_1'$$

الفرض البديل

$$H_1 : a_1 \neq a_1'$$

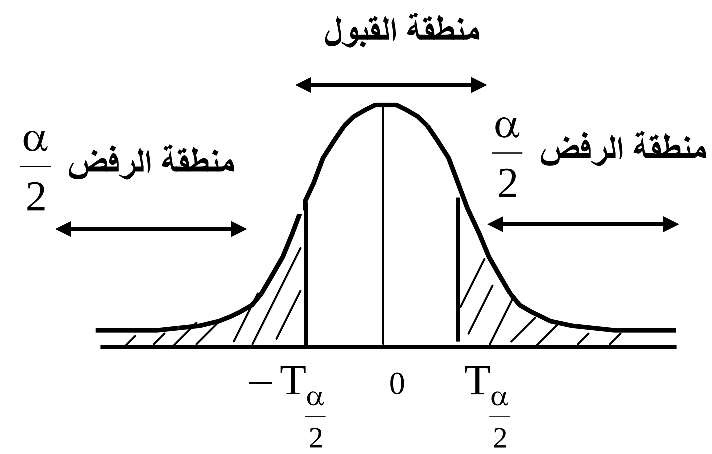
٢- إذا كانت γ معلومة نحسب الإحصاء Z حيث:-

$$Z = \frac{\hat{a}_1 - a_1'}{\gamma} \sqrt{\sum (x_i - \bar{x})^2}$$

٣- عند درجة الثقة $(1-\alpha)$ نوجد من جدول التوزيع المعتاد القياسي حدي الثقة $\pm Z_{\alpha/2}$ ثم نقارن بين Z المحسوبة وحدي الثقة $\pm Z_{\alpha/2}$

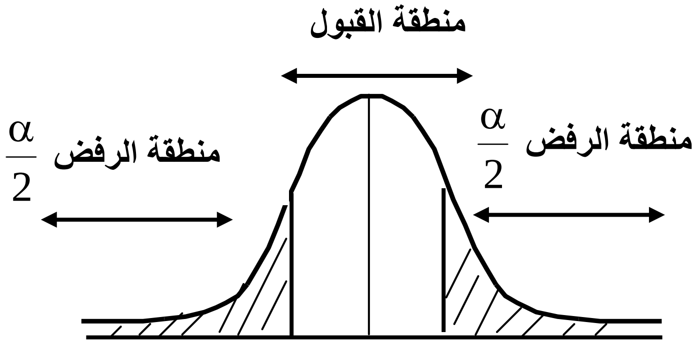
٤- أما إذا كانت γ غير معلومة واستخدم $\hat{\gamma}$ كتقدير لها فإننا نقوم بحساب الإحصاء t حيث:-

$$t = \frac{\hat{a}_1 - a_1'}{\hat{\gamma}} \sqrt{\sum (x_i - \bar{x})^2}$$



شكل (٨-١٣)

٥- عند درجة الثقة $(1-\alpha)$ نوجد من جدول توزيع استيودنت وعند درجات حرية (n-2) حدي الثقة $\pm t_{\alpha/2}$.



شكل (٨-١٤)

ثم نقارن t المحسوبة وحدي الثقة $\pm t_{\alpha/2}$ فإذا وقعت t خارج حدي الثقة فإننا نرفض الفرض العدمي بدرجة ثقة $(1-\alpha)$.

مثال (٨-٦)

في المثال السابق:-

١- اختبر الفرض القائل بأن $a_1 = 1.5$ بدرجة ثقة 95% إذا علمت أن $\gamma^2 = 25$.

٢- اختبر الفرض القائل بأن $a_1 = 2.4$ بدرجة ثقة 90% إذا كان $\hat{\gamma}^2 = 3.5$.

الحل

١- الفرض العدمي

$$H_0 : a_1 = 1.5$$

الفرض البديل

$$H_1 : a_1 \neq 1.5$$

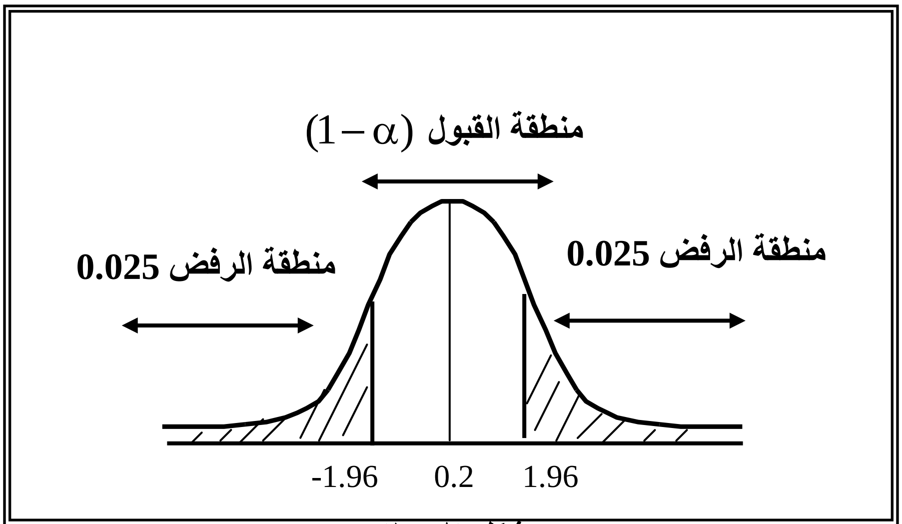
٢- عند درجة ثقة 95% أو من جدول التوزيع المعتاد القياسي نحسب $\pm Z_{\alpha/2}$

حيث $Z_{\alpha/2} = 1.96$.

٣- نحسب الإحصاء Z حيث:-

$$Z = \frac{\hat{a}_1 - a_1'}{\gamma}\sqrt{\sum(x_i - \bar{x})^2}$$

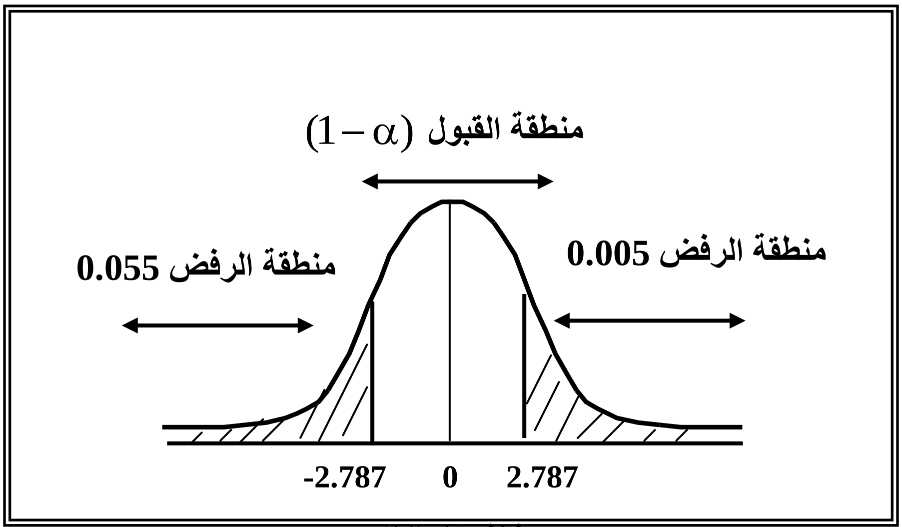
$$= \frac{2 - 1.5}{5}\sqrt{60.75} = 1.95$$



شكل (٨-١٥)

وبما أن Z تقع في منطقة القبول إذن نقبل الفرض العدمي القائل بأن $a_1 = 1.5$ بدرجة ثقة 95%.

٢- من جدول توزيع استيودنت عند درجات الحرية 25 ، درجة ثقة 90% نوجد $\pm t_{\alpha/2}$ حيث $t_{\alpha/2} = \pm 2.787$



شكل (٨-١٦)

٣- نحسب الإحصاء t حيث:-

$$t = \frac{\hat{a}_1 - a_1'}{\hat{\gamma}}\sqrt{\sum (x_i - \bar{x})^2}$$

$$= \frac{2-2.4}{3.5}\sqrt{60.75} = -0.891$$

وبما أن القيمة t=-0.891 تقع داخل حدي الثقة ± 2.787 بالتالي نقبل الفرض العدمي بأن $a_1 = 2.4$ وذلك بدرجة ثقة 95%.

(٨-٦) اختبار معنوية معامل الارتباط الخطي

Tests the significance of Linear Correlation Coefficient

إذا اعتبرنا أن (ρ) هو معامل الارتباط بين المتغير التابع y والمتغير المستقل x في المجتمع محل الدراسة ، $\hat{\rho}$ تقدير لمعامل الارتباط في المجتمع (ρ) تم حسابه باستخدام عينة عشوائية .

وكما ذكرنا في الباب السابع أن المتغير t حيث:-

$$t=\frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}} \qquad (8.32)$$

متغير عشوائي يتبع توزيع استيودنت بدرجات حرية (n-2) ولإجراء اختبار عن وجود ارتباط خطي بين x , y فإننا نتبع الخطوات التالية:-

١- الفرض العدمي:

$H_0 : \rho = 0$ **(وجود ارتباط خطي)**

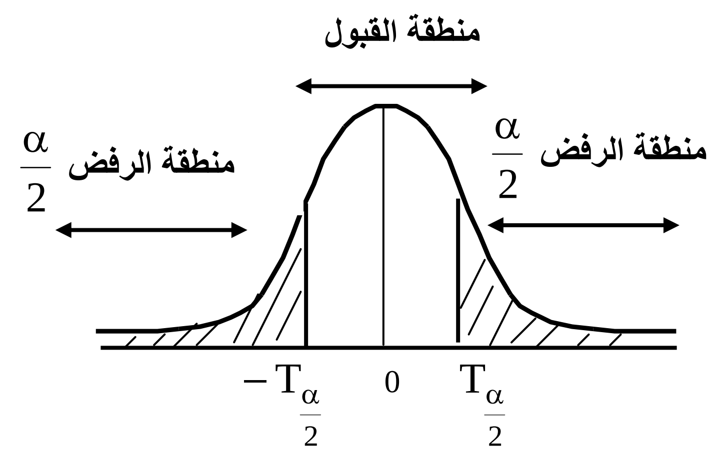
الفرض البديل:

$H_1 : \rho \neq 0$ **(عدم وجود ارتباط خطي)**

٢- يتم حساب الاحصاء t حيث

$$t=\frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}}$$

٣- عند درجة الثقة $(1-\alpha)$ يتم حساب حد (او حدي الثقة) $\pm t_{\alpha/2}$ عند درجات الحرية (n-2).



شكل (٨-١٧)

مثال (٨-٧)

في مثال (٨-١) نحسب $\hat{\rho}$ (معامل الارتباط من بيانات العينة) من العلاقة التالية:-

$$\hat{\rho} = \frac{\sum xy - n\bar{x}\bar{y}}{nS_x S_y}$$

من جدول (٨-٢) نجد أن:-

$$\bar{x} = \frac{\sum x}{n} = \frac{32}{10} = 3.2$$

$$\bar{y} = \frac{\sum y}{n} = \frac{20}{10} = 2$$

$$S_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{19.6}{9}} = \sqrt{2.18} = 1.48$$

$$S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{26}{9}} = \sqrt{2.89} = 1.7$$

$$\hat{\rho} = \frac{42 - 10(3.2)(2)}{10(1.48)(1.7)} = \frac{42-64}{25.16} = -0.87$$

وبما أن قيمة $\hat{\rho}$ تساوي (0.87-) فهذا يدل على أن العلاقة عكسية قوية

<u>المطلوب</u>

اختبر الفرض القائل بأنه لا يوجد ارتباط خطي بين x , y في المجتمع المسحوب منه العينة وذلك بدرجة ثقة ٩٥%.

<u>الحل</u>

١- الفرض العدمي

$$H_0 : \rho = 0$$

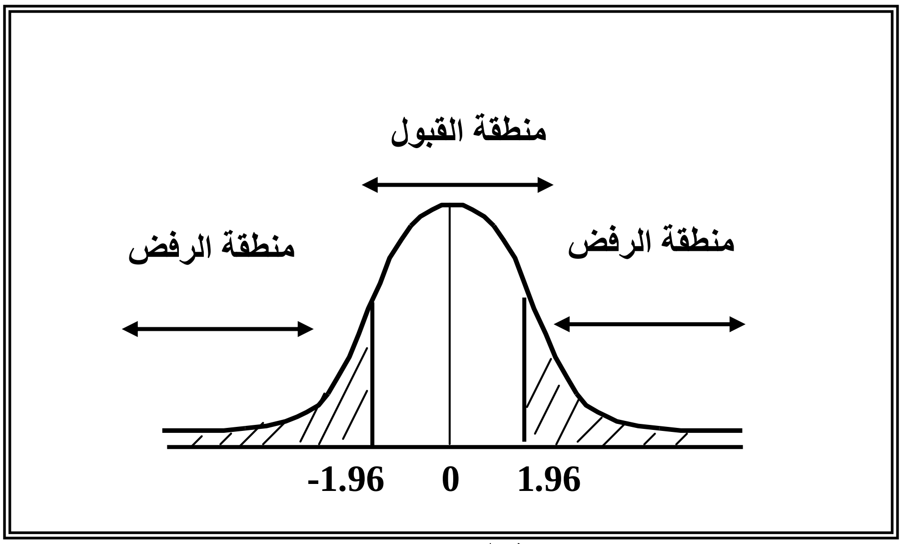
الفرض البديل

$$H_1 : \rho \neq 0$$

نحسب الإحصاء t حيث:-

$$t = \frac{\hat{\rho}\sqrt{n-2}}{\sqrt{1-\hat{\rho}^2}}$$

عند درجة الحرية (8=2-10) وعند درجة الثقة 90% نوجد حدي الثقة $\pm t_{\alpha/2}$



شكل (٨-١٨)

وبما أن قيمة $\hat{\rho}$ حيث $\hat{\rho} = -0.87$ تقع داخل حدي الثقة ± 1.86 إذن نقبل الفرض العدمي القائل بأنه لا يوجد ارتباط خطي بين المنغير التابع y والمتغير المستقل x في المجتمع المسحوب منه العينة.

(٨-٧) تمرينات Exercises

(٨-١): فيما يلي بيانات لعينة من الأسعار للوحدة الواحدة من منتج معين بالجنية، والكميات المباعة من السلعة (بالألف وحدة).

جدول (٨-٩)

الأسعار (بالجنية) X	2	5	7	10	12
الكميات (الألف وحدة) Y	10	13	15	25	20

المطلوب:

١- قدر العلاقة بين المتغير X والمتغير Y على النحو

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X$$

٢- أوجد $\hat{\gamma}^2$ ثم عقب على النتائج.

٣- أوجد فترات الثقة لكل من $\hat{Y}, \hat{a}_0, \hat{a}_1$.

٤- اختبر معنوية النموذج (اختبار جودة التوفيق) في (١).

(٨-٢): في التمرين السابق – اختبر الفرض القائل بأنه لا توجد علاقة خطية بين الأسعار والكميات المباعة.

(٨-٣): في تمرين (٨-١) أوجد معامل الارتباط بين الأسعار والكميات المباعة ثم عقب على النتائج.

(٨-٤): في التمرين السابق أوجد فترات الثقة لمعامل الارتباط. ثم اختبر الفرض القائل بأنه لا يوجد ارتباط خطي بين الأسعار والكميات المباعة.

(٨-٥): فيما يلي بيانات عن المتغيرين X، Y.

جدول (٨-١٠)

X	**0**	**1**	**2**	**3**	**4**	**5**
Y	**2**	**3.5**	**6**	**11**	**17**	**26**

المطلوب

١- ارسم شكل الانتشار بين x , y .

٢- باستخدام طريقة المربعات الصغري أوجد نموذج الانحدار.

٣- اختبر جودة العلاقة المقدرة في (٢).

الباب التاسع
اختبار جودة التوفيق
Statistical Estimation

(٩-١) التوزيعات المقدرة **Estimated Distributions**

(٩-٢) اختبـار χ^2 التوفيـق **Chi-Square Test**

(٩-٣) اختبار كولومجروف سميرونوف **Kolomogrov Simrnov Test**

(٩-٤) تطبيقات **Applications**

(٩-٥) تمرينات **Exercises**

(٩-١) التوزيعات المقدرة
Estimated Distributions

في الباب السابع تناولنا بالدراسة التفصيلية توفيق المنحنيات (أي الدوال أو النماذج) الإحصائية سواء لدوال الاحتمالات (أو دوال كثافة الاحتمال) التي تصف السلوك الاحتمالي للمتغير محل الدراسة أو العلاقة بين متغيرين أو أكثر. وكما ذكرنا سابقاً أن المنحنيات (أي الدوال أو النماذج) التي تم توفيقها من بيانات عينة عشوائية مسحوبة من مجتمع الدراسة تختلف من عينة لأخرى بالتالي تكون هذه المنحنيات منحنيات تقديرية.

وبالنسبة للاحتمالات التي تمثل التوزيع الاحتمالي ونظراً لأن هذه المنحنيات المبنية على بيانات عينة تمثل دوال (منحنيات) تقديرية، بالتالي لابد من إجراء اختبار يحدد جودة هذا التوفيق. والاختبار الإحصائي الذي يحدد مدي ملائمة المنحني الذي تم توفيقه باستخدام بيانات العينة للتوزيع الاحتمالي في المجتمع المسحوب منه العينة يسمي باختبار جودة التوفيق[١] **Goodness of Fitting Test**.

وقد يكون التوزيع الاحتمالي النظري[٢] للمتغير محل الدراسة على بيانات العينة المسحوبة من المجتمع، قد يمثل هذا التوزيع النظري للمتغير محل الدراسة في المجتمع تمثيل جيد والاختلاف بين التوزيع المقدر من بيانات العينة والتوزيع النظري يرجع إلى عوامل المعاينة فقط. ولكن قد لا يمثل التوزيع الذي تم توفيقه التوزيع الفعلي للمتغير محل الدراسة في المجتمع المسحوب منه العينة وبالتالي يعتبر التوزيع المقدر تقدير

(1) Heinz Kohler (1994): "Statistics for Business and Economics" , Harper Collins College Publishers, New York.

(٢) التوزيع الاحتمالي النظري هو التوزيع المعروف الصياغة الرياضية لدالة احتماله (أو كثافة احتماله)، كذلك خصائص هذه الدالة تكون محددة ومعروفة مثل توزيع ذات الحدين، التوزيع المعتاد، التوزيع الأسي، ... الخ.

غير جيد ويؤدي استخدامه إلى الوقوع في النوع الثاني من الخطأ عند اتخاذ القرار أي قبول فرض خطأ (أنظر الباب الخامس).

وإجراء الاختبارات الإحصائية لتجنب الوقوع في النوع الثاني من الخطأ تعتبر اختبارات صعبة ومعقدة لغير المتخصصين. لذلك صممت بعض الاختبارات التي يمكن أن يستخدمها غير المتخصصين والتي تجنب متخذ القرار من الوقوع في النوع الثاني من الخطأ. وسوف نتناول من هذه الاختبارات اختبارين أكثر استخداماً هما: اختبار مربع كا (χ^2) واختبار كولومجروف سيمرونوف في الفصول التالية.

مثال (٩-١)

إذا سحبت بيانات عن أحد المتغيرات من أحد المجتمعات وتم توفيق دالة ذات الحدين كتقدير لدالة الاحتمال للمتغير محل الدراسة في المجتمع المسحوب منه العينة كما سبق توضيح ذلك في الباب السابق. ويصبح السؤال هنا:

١- هل السلوك الاحتمالي للمتغير محل الدراسة في المجتمع المسحوب منه العينة يتفق مع دالة ذات الحدين أم لا؟

والذي يجيب عن هذا السؤال هو اختبارات جودة التوفيق.

٢- إذا كان السلوك الاحتمالي للمتغير محل الدراسة في المجتمع المسحوب منه العينة يتفق مع دالة ذات الحدين – فهل يتفق مع دالة ذات الحدين بنفس المعلمات المقدرة (n , p) أم ينفق مع ذات الحدين بمعلمات أخرى؟

ويجيب عن هذا السؤال أيضاً اختبارات جودة التوفيق.

وتوجد أنواع مختلفة لاختبارات جودة التوفيق سوف نتناول في هذا الباب اثنين من هذه الاختبارات الأكثر استخداماً من الناحية التطبيقية وهما:

أولاً: اختبار مربع كا (χ^2)

ثانياً: واختبار كولومجروف سيمرونوف

وعادة بالنسبة لاختبار جودة التوفيق يكون الفرض العدمي والفرض البديل على النحو التالي:

١- الفرض العدمي

H_0 : **لا يوجد فرق معنوي بين التوزيع الاحتمالي للمتغير في المجتمع والتوزيع الاحتمالي الذي تم توفيقه من بيانات العينة (أي التوفيق جيد).**

٢- الفرض البديل

H_1 : **يوجد فرق معنوي بين التوزيع الاحتمالي للمتغير في المجتمع والتوزيع الاحتمالي الذي تم توفيقه من بيانات العينة (أي التوفيق ردئ).**

(٩-٢) اختبار مربع كا **Chi-Square Test**

في هذا الفصل سوف نتناول بالدراسة كيفية إجراء اختبار مربع كا (χ^2) في الحالات التالية:

١- اختبار χ^2 لتوفيق توزيعات احتمالية نظرية.

٢- اختبار χ^2 لتوفيق أي توزيع معين Chi-Square Test To any Specified Distribution.

٣- اختبار χ^2 لاستقلال متغيرين (اختبار العلاقة بين متغيرين).

أولاً: اختيار χ^2 لتوفيق توزيعات احتمالية نظرية

إذا كان لدينا عينة عشوائية حجمها n تأخذ مفرداتها $x_1, x_2, x_3, \ldots, x_n$

حيث أن f_i **هو عدد المفردات التي تأخذ القيمة** x_i **(أي التكرار المناظر للقيمة** x_i**)**

كما هو موضح بالجدول التالي:

جدول (٩-١)

القيمة x_i	x_1	x_2	x_3	......	x_n	المجموع
التكرار f_i	f_1	f_2	f_3	......	f_n	$\sum f_i = n$

فإذا تم توفيق توزيع احتمالي نظري للمتغير X دالة احتماله (أو دالة كثافة احتماله) $f(X)$ بحيث أن التكرار المتوقع للقيمة x_i هو f_i حيث:

$$f_i = f(X = x_i)(\sum_{i=1}^{n} f_i) \qquad \textbf{(9.1)}$$

ويتطلب استخدام اختبار χ^2 أن تكون التكرار f_i أو التكرار المتوقع f_i' أكبر من 5مفردات، وللتغلب على ذلك إذا كان f_i أو f_i' أقل من أو يساوي 5 تدمج الفئة إلى الفئة (i-1) أو (i+1). وفي هذه الحالة نجد أن المتغير χ^2 حيث:

$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(f_i - f_i')^2}{f_i'}\right] \qquad \textbf{(9.2)}$$

أو

$$\chi^2 = \sum_{i=1} \frac{\text{التكرار المتوقع في الفئة (i) – التكرار المشاهد في الفئة (i)}}{\text{التكرار المتوقع في الفئة (i)}}$$

فإن المتغير χ^2 متغير عشوائي يتبع توزيع χ^2 بدرجات حرية h حيث:

$$h = \text{عدد الفئات} - 1 \qquad \textbf{(9.3)}$$

(أي h تساوي عدد النتائج الممكنة المستقلة).

وفيما يلي سوف نلخص خطوات إجراء اختبار جودة التوفيق على النحو التالي:

١- وضع الفرض العدمي والفرض البديل على النحو التالي:

$H_0:$ **التوزيع النظري المقدر باستخدام بيانات العينة يتفق مع التوزيع الاحتمالي للمتغير في المجتمع (التوفيق جيد).**

$H_1:$ **التوزيع النظري المقدر باستخدام بيانات العينة لا يتفق مع**

التوزيع الاحتمالي للمتغير في المجتمع (التوفيق ردئ).

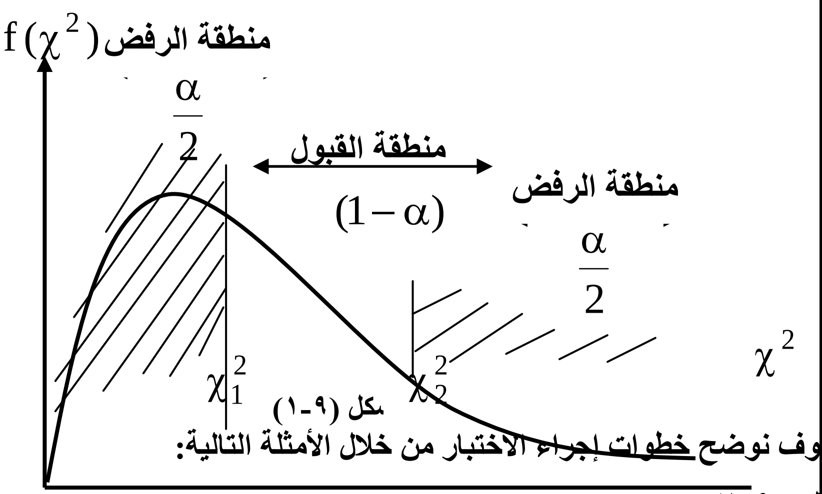
٢- **عند درجة ثقة $(1-\alpha)$ نوجد من جداول توزيع χ^2 بملحق رقم (٥) قيمتي χ_1^2، χ_2^2 عند درجات الحرية المساوية للقيمة h.**

٣- **نحسب قيم التكرارات المتوقعة f_i' من المعادلة (9.1).**

٤- **نحسب قيمة المقياس χ^2 من المعادلة (9.2).**

٥- **إذا كانت χ^2 بحيث $\chi_1^2 < \chi^2 < \chi_2^2$ نقبل الفرض العدمي بأن التوفيق جيد بدرجة ثقة $(1-\alpha)$. أما إذا كان $\chi_1^2 > \chi^2$ أو $\chi_2^2 < \chi^2$ فإننا نرفض الفرض العدمي ونقبل الفرض البديل بأن التوفيق ردئ.**

كما هو موضح بالشكل التالي:



شكل (٩-١)

وسوف نوضح خطوات إجراء الاختبار من خلال الأمثلة التالية:

مثال (٩-٢)

الجدول التالي يعطي التوزيع التكراري لعينة حجمها 100 عامل من العاملين بإحدى الشركات الإنتاجية وفقاً لكفاءة أداء العامل.

جدول (٩-٢)

مستوي الأداء	0 -	1 -	2 -	3 – 4	المجموع
عدد العمال f_i	14	40	34	12	100

المطلوب:

١- **توفيق منحني منتظم يوضح كفاءة العاملين بالشركة.**

٢- **أوجد التكرارات المتوقعة f_i'.**

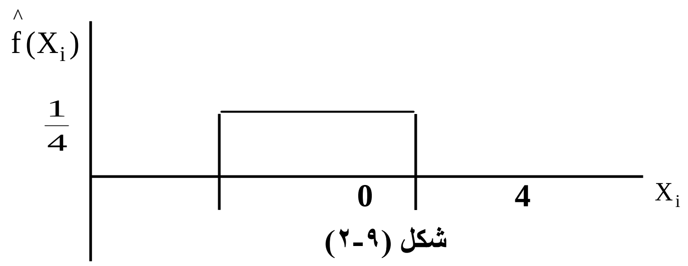
٣- اختبر جودة التوفيق بدرجة ثقة 95%.

الحل:

١- بما أنه يوجد 4 مستويات لكفاءة أداء العمال، فإذا أشرنا إلى مستوي كفاءة الأداء بالرمز X_i حيث ($0 \le X_i \le 4$) فإن دالة كثافة الاحتمال المقدرة (التوفيق) هي:

$$\hat{f}(X_i) = \frac{1}{4-0} = \frac{1}{4} \qquad 0 \le X_i \le 4$$

كما هو موضح بالشكل التالي:



شكل (٩-٢)

٢- إذا فرضنا أن f'_i هي التكرارات المتوقعة حيث:

$$f'_i = f(X_i)\sum_{i=1}^{n} f_i = \frac{1}{4}(100) = 25$$

وبما أن التوزيع النظري المفترض توزيع منتظم بالتالي فإن:

$$f'_1 = f'_2 = f'_3 = f'_4 = 25$$

٣- الفرض العدمي:

H_0: **كفاءة أداء العمال في المصنع تتفق مع التوزيع المنتظم (أي أن احتمال أن يكون العامل في المستوي الأول = احتمال أن يكون العامل في المستوي الثاني = احتمال أن يكون العامل في المستوي i ، (i=1,2,3,…n).**

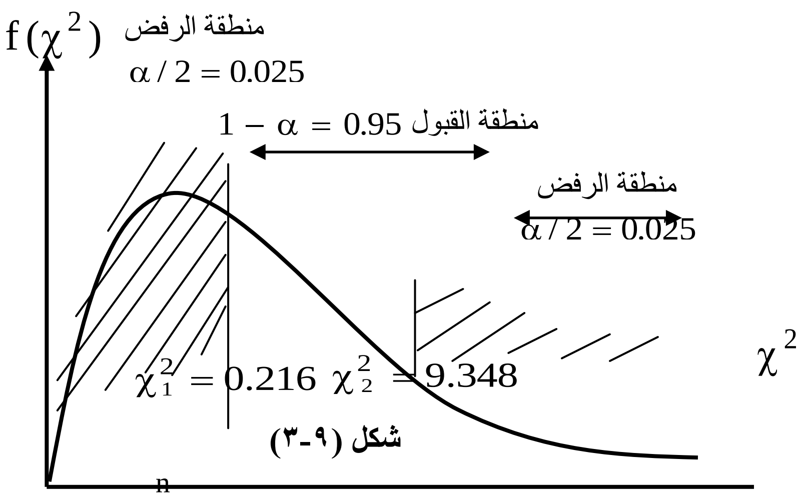
الفرض البديل:

H_1: **كفاءة أداء العمال في المصنع لا تتفق مع التوزيع المنتظم (أي أن احتمال أن يكون العامل في المستوي i يختلف عن احتمال أن يكون العامل في المستوي j).**

من ملحق رقم (٥) عند درجة الثقة 95%، ودرجات حرية تساوي (3=1-4) فإن $\chi^2_1 = 0.216$ ، $\chi^2_2 = 9.348$. كما هو موضح بالشكل (٩-٣). ونكون الجدول التالي:

جدول (٩-٣)

مستوي الأداء X_i	التكرارات المشاهدة f_i	التكرارات المتوقعة f_i'	$(f_i - f_i')^2$	$\frac{(f_i - f_i')^2}{f_i'}$
0 -	**14**	**25**	**121**	**4.84**
1 -	**40**	**25**	**225**	**9**
2 -	**34**	**25**	**81**	**3.24**
3 – 4	**12**	**25**	**169**	**6.67**
المجموع	**100**	**100**		**23.84**



شكل (٩-٣)

ملحوظة: لابد أن يساوي مجموع التكرارات المتوقعة ($\sum_{i=1}^{n} f_i'$) مجموع التكرارات الفعلية ($\sum_{i=1}^{n} f_i$) أي أن:

$$\sum_{i=1}^{n} f_i = \sum_{i=1}^{n} f_i' \qquad , \ i = 1,2,3,..., n \qquad \textbf{(9.4)}$$

ومن جدول (٩-٣) نجد أن:

$$\chi^2 = \sum_{i=1}^{n} \left[\frac{(f_i - f_i')^2}{f_i'} \right] = 23.84$$

وبما أن $\chi^2 < \chi_2^2$ أي قيمة χ^2 المحسوبة تقع في منطقة الرفض.

إذن نرفض الفرض العدمي القائل بأن التوزيع الاحتمالي لمستوي أداء العامل بالمصنع يتوافق مع التوزيع المنتظم. إذن يعتبر منحني التوزيع المنتظم توفيق ردئ لتوزيع مستوي كفاءة العاملين بالمصنع وذلك بدرجة ثقة 95%.

مثال (٩-٤)

الجدول التالي يعطي توزيع 200 أسرة حسب عدد الأطفال في الأسرة الواحدة في إحدى المجتمعات العمرانية الحديثة.

جدول (٩-٤)

عدد الأطفال في الأسرة الواحدة x_i	0	1	2	3	4 - 6	المجموع
عدد الأسر f_i	31	51	70	32	16	200

المطلوب:

١- باستخدام طريقة العزوم وفق دالة احتمال توزيع ذات الحدين لعدد الأطفال في الأسرة في هذا المجتمع العمراني الحديث.

٢- عند درجة ثقة 95% اختبر جودة التوفيق باستخدام اختبار χ^2.

الحل:

١- إذا فرضنا أن X تشير إلى عدد الأطفال في الأسرة الواحدة وأن احتمال انجاب طفل في الأسرة يساوي P فإن:

$$f(X) = C_x^6 (P)^x (1-P)^{6-x} \qquad , \; x = 0,1,2,3,...,6$$

ومن جدول (٩-٤) نجد أن متوسط عدد الأطفال في الأسرة الواحدة في العينة يساوي $\overline{X}$ حيث:

$$\overline{X} = \frac{0(31)+1(51)+2(70)+3(32)+4(16)}{200} = \frac{351}{200} = 1.755 \text{ طفل}$$

وباستخدام طريقة العزوم (أنظر الفصل (٨-٢)) نجد أن:

$$\overline{X} = n\hat{P} \rightarrow 1.755 = 6\hat{P} \rightarrow \hat{P} = \frac{1.755}{6} = 0.3$$

بالتعويض في الدالة $f(X)$ نجد أن:

$$\hat{f}(X) = C_x^6 (0.3)^x (0.7)^{6-x} \qquad , \; x = 0,1,2,3,..,6$$

٢- الفرض العدمي:

H_0 : **عدد الأطفال في الأسرة الواحدة في هذا المجتمع الحديث يتبع توزيع ذات الحدين بمعلمة (n = 6 , P = 0.3).**

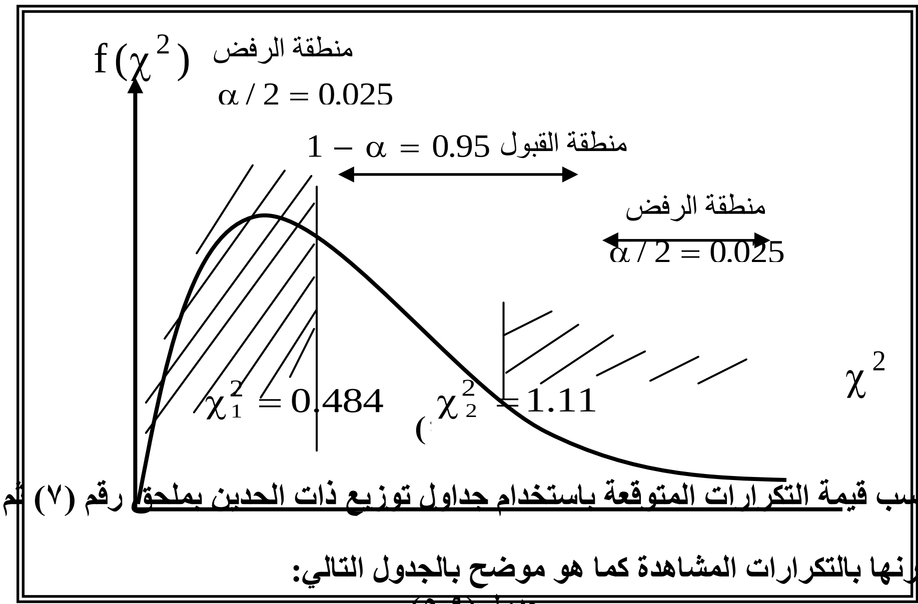
الفرض البديل:

عدد الأطفال في الأسرة الواحدة في هذا المجتمع الحديث يختلف عن توزيع ذات الحدين بمعلمة (n = 6 , P = 0.3). H_1 :

عند درجة الثقة 95% وعدد درجات الحرية يساوي (5-1=4) من جدول توزيع χ^2 **بملحق رقم (٥) نجد أن:**

$$\chi_1^2 = 0.484 \quad , \quad \chi_2^2 = 1.11$$

كما هو موضح بالشكل التالي:



نحسب قيمة التكرارات المتوقعة باستخدام جداول توزيع ذات الحدين بملحق رقم (٧) ثم نقارنها بالتكرارات المشاهدة كما هو موضح بالجدول التالي:

جدول (٩-٥)

مستوي الأداء X_i	التكرارات المشاهدة f_i	التكرارات المتوقعة f_i'	$(f_i - f_i')$	$\frac{(f_i - f_i')^2}{f_i'}$
0	**31**	**23.52**	**7.48**	**2.379**
1	**51**	**60.50**	**-9.50**	**1.492**
2	**70**	**64.82**	**5.18**	**0.414**
3	**32**	**37.04**	**-5.04**	**0.686**
4 – 6	**16**	**14.08**	**1.92**	**0.262**
المجموع	**100**		**0**	**5.233**

من جدول (٩-٥) نجد أن:

$$\chi^2 = 5.233$$

وبما أن: $\chi_1^2 < 5.233 < \chi_2^2$

أي أن قيمة χ^2 المحسوبة تقع في منطقة القبول. إذن نقبل الفرض القائل بأن التوزيع الاحتمالي لعدد الأطفال في الأسرة الواحدة في هذا المجتمع يتبع توزيع ذات الحدين بمعلمتين $(n = 6 , P = 0.3)$.وذلك بدرجة ثقة 95%.

مثال (٩-٥)

أخذت عينة مكونه من 100 فرد حسب الفئات العمرية بالسنوات، والجدول التالي يوضح توزيع الأفراد حسب الفئات العمرية المختلفة.

جدول (٩-٦)

الفئات العمرية بالسنوات	20 -	25 -	30 -	35 -	40 - 45	المجموع
عدد الأفراد	7	25	33	27	8	100

المطلوب:

١- وفق منحني التوزيع المعتاد باستخدام طريقة العزوم.
٢- اوجد التكرارات المتوقعة.
٣- اختبر جودة التوفيق بدرجة ثقة ٩٥%.

الحل:

نكون الجدول التالي لحساب توقع العينة $\overline{X}$ وتباينها S^2.

جدول (٩-٧)

الفئات	التكرارات المشاهدة f_i	مراكز الفئات X_i	$X_i f_i$	$X_i^2 f_i$
20 -	7	22.5	157.5	35453.75
25 -	25	27.5	687.5	18906.25
30 -	33	32.5	1072.5	34806.25
35 -	27	37.5	1012.5	37968.75
40 - 45	8	42	340	14450
المجموع	100		3270	109725

من الجدول نجد أن:

$$\overline{X} = \frac{\sum X_i f_i}{\sum f_i} = \frac{3270}{100} = 32.7$$

$$S^2 = \frac{\sum X_i^2 f_i}{\sum f_i} - (\overline{X})^2 = \frac{1097.25}{100} - (32.7)^2$$

$$= 1097.25 - 1096.29 = 27.96$$

$$\therefore S^2 = \sqrt{57.96} = 5.29$$

وباستخدام طريقة العزوم

$$\overline{X} = \mu \rightarrow \mu = 32.7$$

$$S = \sigma \rightarrow \sigma = 5.29$$

٢- ولحساب التكرارات المتوقعة f_i' نوجد أولاً الاحتمالات المناظرة لكل فئة $P_r(X)$ باستخدام التوزيع المعتاد القياسي ملحق رقم (٤) على النحو التالي:

$$P_r(20 < X < 25) = P_r(\frac{20-32.7}{5.29} < Z < \frac{25-32.7}{5.29})$$
$$= P_r(-1.46 < Z < 2.4) = 0.07895$$

بالمثل بالنسبة للاحتمالات المناظرة لباقي الفئات $P_r(X)$ كما هو موضح بالجدول التالي. ثم باستخدامها نحسب الاحتمالات المتوقعة f_i' بضرب كل احتمال في مجموع التكرارات (100).

جدول (٩-٨)

الفئات	التكرارات f_i	$P_r(X)$	التكرارات المتوقعة f_i'	$(f_i - f_i')^2$	$\frac{(f_i - f_i')^2}{f_i'}$
20 -	**7**	**0.07895**	**8.8**	**3.24**	**0.36**
25 -	**25**	**0.26609**	**26.6**	**2.56**	**0.09**
30 -	**33**	**0.39894**	**39.9**	**47.61**	**1.19**
35 -	**27**	**0.26609**	**26.6**	**0.16**	**0.006**
40 - 45	**8**	**0.08795**	**8.80**	**0.64**	**0.07**
المجموع	**100**				**1.716**

ومن الجدول نجد أن:

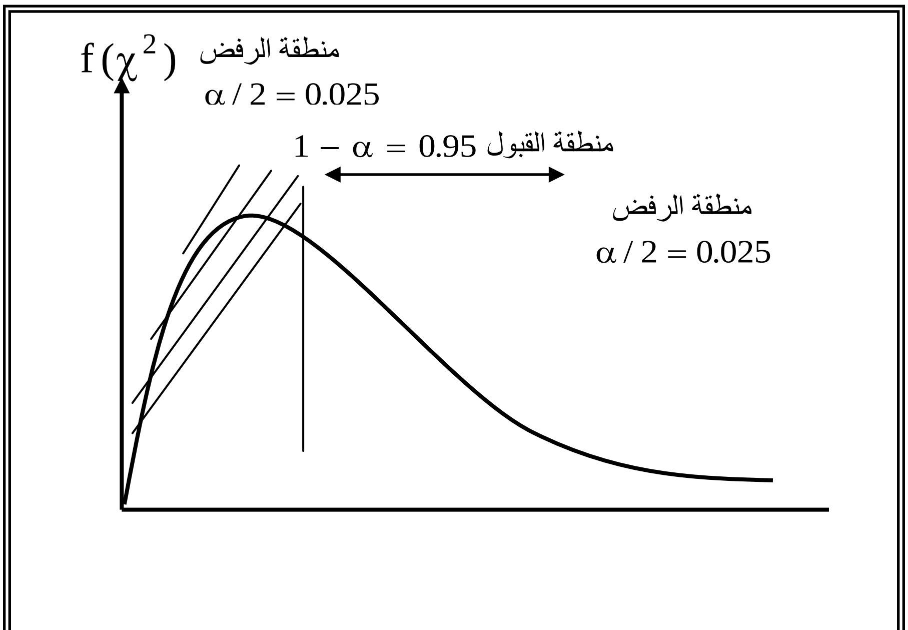
$$\chi^2 = \sum_{i=1}^{n}\left[\frac{(f_i - f_i')^2}{f_i'}\right] = 1.716$$

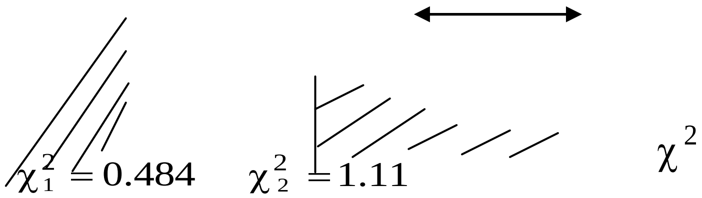
وعند درجة الثقة 95% ودرجات الحرية تساوي (5-1=4).

نجد من جدول χ^2 بملحق رقم (٥) أن:

$$\chi_1^2 = 0.484 \quad , \quad \chi_2^2 = 1.11$$

كما هو موضح بالشكل التالي:





شكل (٩-٥)

وبما أن:

$$\chi_1^2 < \chi^2 < \chi_2^2$$

أي تقع في منطقة القبول، إذن نقبل الفرض القائل أن أعمار الأفراد في المجتمع المسحوب منه العينة تتوزع توزيعاً طبيعياً بتوقع 32.7 سنة، وانحراف معياري 5.29 سنة وذلك بدرجة ثقة 95%.

ثانياً: اختبار χ^2 لتوفيق أي توزيع معين

إذا كان المتغير محل الدراسة في مجتمع ما يتبع توزيع احتمالي غير معروف (غير محدد الصياغة الرياضية)، وكان المطلوب اختبار ما إذا حدث تغير معنوي للسلوك الاحتمالي للمتغير أم لا – أو بعبارة أخرى التوزيع الاحتمالي للمتغير محل الدراسة تغير نتيجة حدوث تغير زمني، أو مكاني، ... أم ظل كما هو وذلك من خلال بيانات عينة. باستخدام بيانات العينة كما سوف يتضح من الأمثلة التالية.

مثال (٩-٤)

الجدول التالي يوضح نسبة الاستهلاك من الدخل لمجموعتين من الأسر في أحدى المحافظات بجمهورية مصر العربية في عامي 1985، 1995.

جدول(٩-٩)

نسبة الاستهلاك في الأسرة X_i	عدد الأسر في 1985 f_i	عدد الأسر في 1995 f_i
50% -	7	12
60% -	10	9
70% -	20	21
80% -	45	44
90% - 100%	18	14
المجموع	100	100

والمطلوب:

اختبار هل النمط الاستهلاكي للأسرة في هذه المحافظة سنة 1995 اختلف عن النمط الاستهلاكي للأسرة في هذه المحافظة سنة 1985 وذلك بدرجة ثقة 95%.

الحل:

نلاحظ هنا أن التوزيع الاحتمالي لنسبة الاستهلاك من الدخل للأسرة في المحافظة غير محدد ولكن باستخدام اختبار χ^2 يمكن اختبار هل هذا التوزيع للسلوك الاستهلاكي في هذه المحافظة اختلف في سنة 1995 عنه في سنة 1985 أم لا؟

١- في هذه الحالة يكون الفرض الإحصائي على النحو التالي:-

الفرض العدمي:

H_0 : **التوزيع الاحتمالي لنسبة الاستهلاك من الدخل للأسرة في هذه المحافظة سنة 1995 لا يختلف عن نسبة الاستهلاك في الأسرة سنة 1985.**

الفرض البديل:

H_1 : **التوزيع الاحتمالي لنسبة الاستهلاك من الدخل للأسرة في هذه المحافظة اختلف في سنة 1995 عنه في سنة 1985.**

٢- نحسب قيمة χ^2 حيث:

$$^{*}\chi^2 = \sum_{i=1}^{n}\left[\frac{(f_i - f_i')^2}{f_i'}\right] = 4.632$$

كما هو موضح بالجدول التالي:

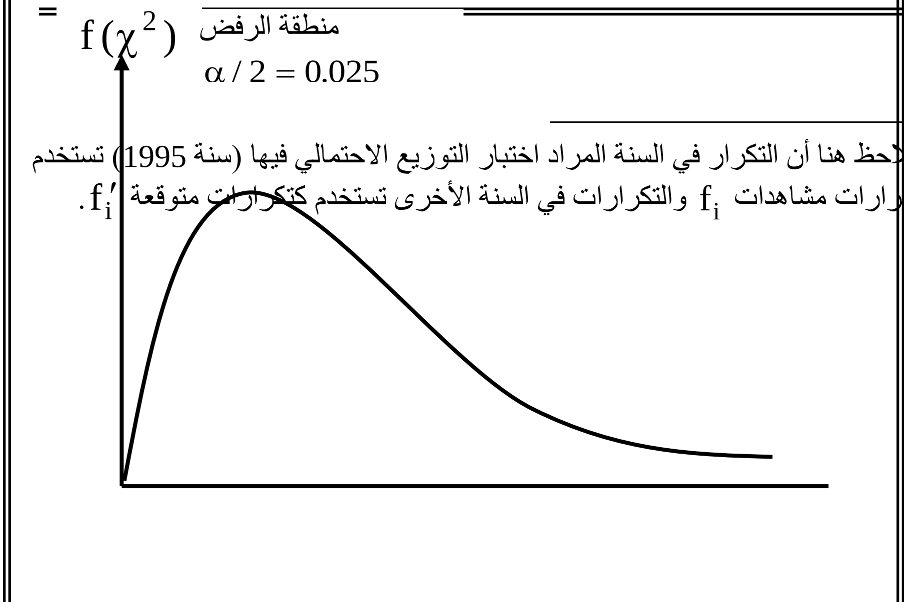
جدول (٩-٧)

X_i	f_i'	f_i	$(f_i - f_i')$	$\frac{(f_i - f_i')^2}{f_i'}$
50% -	**7**	**12**	**5**	**3.571**
60% -	**10**	**9**	**-1**	**0.100**
70% -	**20**	**21**	**1**	**0.050**
80% -	**45**	**44**	**-1**	**0.022**
90% - 100%	**18**	**14**	**-4**	**0.889**
المجموع	**100**	**100**		**4.632**

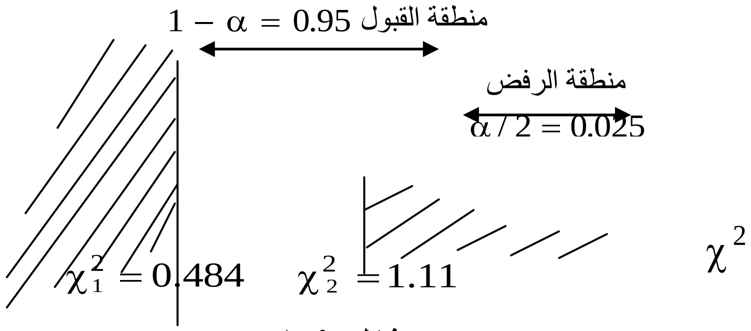
٣- وبما أن عند درجة الثقة 95% ودرجات الحرية (4=1-5) نجد أن:

$$\chi_1^2 = 0.484 \quad , \quad \chi_2^2 = 1.11$$

كما هو موضح بالشكل التالي:



* يلاحظ هنا أن التكرار في السنة المراد اختبار التوزيع الاحتمالي فيها (سنة 1995) تستخدم كتكرارات مشاهدات f_i والتكرارات في السنة الأخرى تستخدم كتكرارات متوقعة f_i'.



شكل (٩-٦)

وبما أن:

$$\chi_1^2<\chi^2<\chi_2^2$$

إذن نقبل الفرض القائل بأن التوزيع الاحتمالي للنمط الاستهلاكي للأسرة في هذه المحافظة لا يختلف في سنة 1995 عنه في سنة 1985 وذلك بدرجة ثقة 95%.

ثالثاً: اختبار χ^2 لاستقلال متغيرين (اختبار العلاقة بين متغيرين)

كذلك يمكن استخدام اختبار χ^2 في دراسة وجود علاقة بين متغيرين (ظاهرتين)

في المجتمع محل الدراسة بناءاً على بيانات عينة عشوائية عن قيم المتغير الأساسي X_i وقيمة المتغير التابع لها Y_j في جدول توزيع تكراري مزدوج*.

فإذا فرضنا أن X_i تشير إلى المستويات المختلفة للمتغير الأول (الأساسي) حيث $i=1,2,3,\dots,n$، Y_j هي المستويات المختلفة للمتغير التابع بحيث أن $j=1,2,3,\dots,n$.

فإذا فرضنا أن P_{ij} هو احتمال سحب مفردة من المجتمع وتكون في المستوي (i) للمتغير الأول، والمستوي (j) للمتغير الثاني. حيث:

$$P_{ij}=P_r(X_i \cap Y_j) \qquad \textbf{(9.4)}$$

فتجد في حالة استقلال المتغيرين X_i ، Y_j أن:

* انظر الجزء الأول من هذا المرجع.

$$P_{ij} = P_r(X_i \cap Y_j) = P_r(X_i)P_r(Y_j) \qquad \textbf{(9.5)}$$

وفي حالة عدم استقلال المتغيرين (أي في حالة وجود علاقة بينهما) نجد أن:

$$\boxed{P_{ij} = P_r(X_i \cap Y_j) \neq P_r(X_i)P_r(Y_j) \qquad \textbf{(9.6)}}$$

(انظر ملحق الاحتمالات الشرطية بملحق رقم (١))

فإذا فرضنا أن f_{ij} هو عدد المفردات (التكرارات) التي تنتمي إلى المستوي (i) للمتغير X، والمستوي (j) للمتغير Y. فإن التكرارات المتوقعة في هذه الحالة سوف نشير لها بالرمز f_{ij} حيث:

$$\boxed{f_{ij} = nP_{ij} \qquad \textbf{(9.7)}}$$

حيث n هو حجم العينة.

كذلك نجد أن المتغير χ^2 حيث:

$$\boxed{\chi^2 = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\left[\frac{(f_{ij} - f'_{ij})^2}{f'_{ij}}\right] \qquad \textbf{(9.8)}}$$

يتبع توزيع χ^2 بدرجات حرية $(n_1 - 1)(n_2 - 1)$.

ملحوظة: n_2, n_1 هما عدد مستويات المتغير X، وعدد مستويات المتغير Y على الترتيب.

ومن ثم يمكن تلخيص إجراء اختبار χ^2 لاختبار استقلال متغيرين على النحو التالي:

١- تكوين الفرض الإحصائي

الفرض العدمي:

$H_0 : P_{ij} = P_i P_j$ **(عدم وجود علاقة)**

الفرض البديل:

$H_1 : P_{ij} \neq P_i P_j$ **(وجود علاقة بين Y , X)**

حيث P_j, P_i **هما احتمال وقوع المفردة في المستوي (i) للمتغير الأول، واحتمال وقوع المفردة في المستوي (j) للمتغير الثاني على الترتيب. من بيانات العينة.**

٢- نحسب التكرارات المتوقعة f'_{ij} **حيث:**

$$f'_{ij} = nP'_iP'_j \qquad \textbf{(9.9)}$$

حيث:

$$P'_i = \frac{\sum_{j=1}^{n_2} f_{ij}}{n} \quad , \quad P'_j = \frac{\sum_{i=1}^{n_1} f_{ij}}{n} \qquad \textbf{(9.10)}$$

وبالتالي فإن:

$$f'_{ij} = \frac{\left(\sum_i f_{ij}\right)\left(\sum_j f_{ij}\right)}{n} \qquad \textbf{(9.11)}$$

أو بعبارة أخرى التكرار المتوقع في الصف (i) والعمود (j) =

(مجموع التكرارات في العمود j) (مجموع التكرارات في الصف i)

مجموع التكرارات

٣- نحسب قيمة الإحصاء χ^2 حيث:

$$\chi^2 = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \frac{[(f_{ij} - f'_{ij})^2]}{f'_{ij}}$$

٤- عند درجة الثقة $(1-\alpha)$ وعدد درجات حرية $(n_1 - 1)(n_2 - 1)$ نحسب قيمة χ^2_1, χ^2_2 من جدول توزيع χ^2 بملحق رقم (٥).

فإذا كان

$$\chi^2_1 < \chi^2 < \chi^2_2$$

فإننا نقبل الفرض القائل بعدم وجود علاقة بين المتغيرين في المجتمع المسحوب منه العينة بدرجة ثقة $(1-\alpha)$.

أما إذا كان

$$\chi^2_2 < \chi^2 \text{ أو } \chi^2_1 > \chi^2$$

فإننا نرفض الفرض العدمي ونقبل الفرض البديل القائل بوجود علاقة بين المتغيرين في المجتمع المسحوب منه العينة. والمثال التالي يوضح ذلك:

مثال (٩-٥)

الجدول التالي يوضح توزيع 500 عامل حسب مستوي تعليم العامل ومستوي الأداء للعامل.

جدول (٩-١١)

مستوي أداء العامل X / مستوي تعليم العامل X	أقل من المتوسط	متوسط	جيد	جيد جداً	ممتاز	المجموع
أمي	**50**	**40**	**15**	**5**	**15**	**125**
ابتدائي	**87**	**69**	**13**	**27**	**4**	**200**
إعدادي	**5**	**60**	**42**	**17**	**1**	**125**
ثانوي	**8**	**11**	**5**	**1**	**25**	**50**
المجموع	**150**	**180**	**75**	**50**	**45**	**500**

المطلوب:

اختبر الفرض القائل بوجود علاقة بين مستوي أداء العامل ومستوي تعليمة وذلك بدرجة ثقة 95%.

الحل:

١- الفرض العدمي

$$H_0 : P_{ij} = P_i P_j$$ **(عدم وجود علاقة)**

الفرض البديل:

$$H_1 : P_{ij} \neq P_i P_j$$ **(وجود علاقة بين X , Y)**

٢- نحسب التكرارات المتوقعة f'_{ij} باستخدام المعادلة (9.9) كما هو موضح بالجدول التالي:

جدول (٩-١٢)

j \ I	1	2	3	4	5
1	**37.5**	**45**	**18.75**	**12.5**	**11.25**
2	**60**	**72**	**30**	**20**	**18**
3	**37.5**	**45**	**18.75**	**12.5**	**11.25**
4	**15**	**18**	**7.5**	**5**	**4.5**

٣- ومن الجدول نحسب قيمة χ^2 حيث:

$$\chi^2 = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{[(f_{ij} - f'_{ij})^2]}{f'_{ij}}$$

$$= \frac{(50-37.5)^2}{37.5} + \frac{(40-45)^2}{45} + \dots + \frac{(25-4.5)^2}{4.5} = 217$$

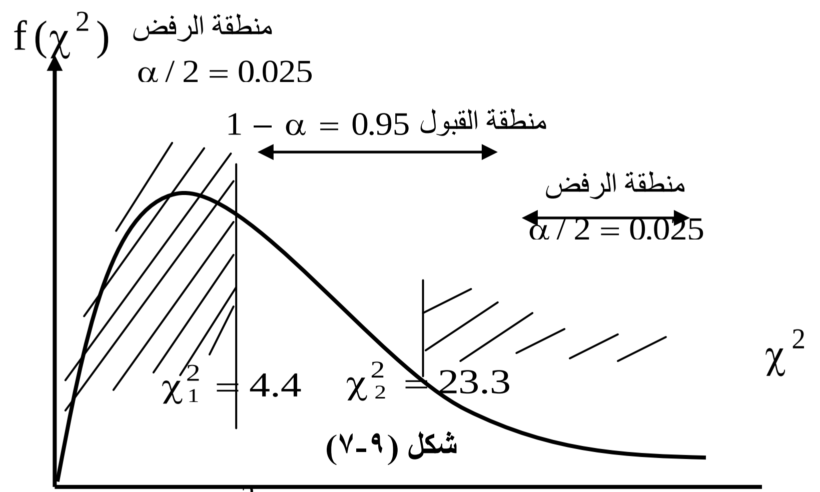
٤- ومـــن جـــدول توزيـــع χ^2 بملحـــق (٥) وعنـــد درجـــات حريـــة

$$(n_1 - 1)(n_2 - 1) = (4-1)(5-1) = 3 \times 4 = 12$$

نجد أن $\chi_1^2 = 4.4$ ، $\chi_2^2 = 23.3$

كما هو موضح بالشكل (٩-٧).

ويتضح أن $\chi^2 < \chi_2^2$ أي أن χ^2 تقع في منطقة الرفض. إذن نرفض الفرض العدمي القائل بعدم وجود علاقة ونقبل الفرض البديل القائل بوجود علاقة بين مستوي تعليم العامل ومستوي الأداء وذلك بدرجة ثقة 95%.



شكل (٩-٧)

ومما هو جدير بالذكر أنه لا يمكن استخدام اختبار χ^2 لجودة التوفيق أو اختبار وجود علاقة بين متغيرين إلا في حالة توافر الشرطين التاليين:

١- أن يكون مجموع التكرارات كبير (أزيد من 30 مفردة).

٢- أن تكون التكرارات الفعلية (f_i) أو المتوقعة (f_i') المناظرة لكل فئة أو لكل خلية في الجداول المزدوجة أكبر من أو تساوي 5 مفردات.

وفي حالة عدم توافر أحد هذين الشرطين أو الاثنين معاً فإنه يجب استخدام اختبار آخر لا يتطلب هذه الشروط مثل اختبار كولومجروف – سيمنروف والذي سوف نقدمه في الفصل التالي.

(٩-٣) اختبار كولومجروف سيمرونوف

Kolomogrov Simrnov Test

في الفصل السابق ذكرنا أن إجراء اختبار χ^2 لاختبار جودة التوفيق لا يصلح إلا في حالة توافر الشرطين:

١- أن يكون عدد التكرارات كبير (أزيد من 30 مفردة) أي عندما يكون حجم العينة كبير.

٢- أن يكون التكرار الفعلي أو المتوقع لأي خلية (أو لأي خلية في حالة الجداول المزدوجة) أكبر من أو تساوي 5 مفردات.

وفي حالة عدم توافر أحد الشرطين السابقين أو كلاهما فإنه لا يمكن استخدام اختبار χ^2 ويجب استخدام اختبارات أخرى. وأحد هذه الاختبارات لاختبار جودة توفيق دالة الاحتمال (أو دالة كثافة الاحتمال) اختبار كولومجروف سيمرنوف. أما بالنسبة للعينات الصغيرة التي لا يتوافر فيها الشرطين السابق ذكرهما أعلاه فقد قدم العالم كولومجروف هذا الاختبار عام ١٩٣٣ ميلادية ثم قام بتطويره في حالة وجود عينتين العالم سيمرنوف.

ويعتمد هذا الاختبار على دالة التوزيع التراكمية للتوزيع النظري المقترح للمتغير $F(X)$ ودالة التوزيع التراكمية للمتغير في العينة $\hat{F}(X)$. وحيث أن أكبر قيمة مطلقة للفرق بين الدالة التراكمية للتوزيع النظري والدالة التراكمية إلى العينة يمثل متغير عشوائي يتبع توزيع (K) بملحق رقم (٩). فإذا أشرنا إلى هذا المتغير العشوائي بالرمز K فإن:

$$K = Max.\left|F(X) - \hat{F}(X)\right| \qquad \textbf{(9.14)}$$

ولإجراء الاختبار نتبع الخطوات التالية:

١- الفرض العدمي

H_0 : المتغير محل الدراسة يتبع دالة التوزيع التراكمية $F(X)$ (أي التوفيق جيد)

الفرض البديل:

H_1 : المتغير محل الدراسة لا يتبع دالة التوزيع التراكمية $F(X)$ (أي التوفيق ردئ)

٢- عند درجة الثقة $(1-\alpha)$ وعدد درجات الحرية n (حيث n حجم العينة) نوجد القيمة الحرجة K_n^α من جدول توزيع K بملحق رقم (٩).

٣- نحسب قيمة المقياس K حيث:

$$K = Max.\left|F(X) - \hat{F}(X)\right|$$

٤- نقارن قيمة K المحسوبة بقيمة K_n^α من الجداول. فإذا كان $K > K_n^\alpha$ فإننا نرفض الفرض العدمي ونقبل الفرض البديل بدرجة ثقة $(1-\alpha)$.

ومما سبق يتضح أن إجراء اختبار كومجروف سيمرنوف يعتبر أكثر قوة من اختبار χ^2 في حالة عدم توافر الشرطين السابق ذكرهما.

مثال (٩-٦)

إذا ألقيت زهرة طاولة عشوائياً 18 مرة وسجلت النتائج في الجدول التالي:

جدول (٩-١٣)

نتائج الرمي (X)	1	2	3	4	5	6	المجموع
التكرارات (f)	2	1	3	2	6	4	18

المطلوب:

اختبر الفرض القائل بان نتائج الرمي متغير يتبع التوزيع المنتظم باستخدام اختبار كولومجروف سيمرنوف بدرجة ثقة 95%.

الحل:

بما أن دالة الاحتمال للتوزيع المنتظم في هذه الحالة $f(X)$ حيث:

$$f(X) = \frac{1}{6} \qquad , \qquad X=1,2,\dots,6$$

١- إذن دالة التوزيع التراكمية للمتغير المنتظم $F(X)$ حيث:

$$F(X) = \frac{X}{6} \qquad , \qquad X=1,2,\dots,6$$

وبالتالي فإن:-

الفرض العدمي:

$$H_0 : F(X) = \frac{X}{6}$$

الفرض البديل:

$$H_1 : F(X) \neq \frac{X}{6}$$

٢- عند درجة الثقة 95% أي مستوي المعنوية $\alpha = 0.05$ وعدد درجات الحرية (n = 18) من ملحق رقم (٩) نجد ان:

$$K_n^{\alpha} = K_{18}^{0.05} = 0.29$$

٣- ولحساب قيمة K نكون الجدول التالي:

جدول (٩-١٤)

X	f	$\hat{F}(X)$	$F(X)$	$\left\|F(X)-\hat{F}(X)\right\|$
1	**2**	**2/18**	**3/18**	**1/18**
2	**1**	**3/18**	**6/18**	**3/18**
3	**3**	**6/18**	**9/18**	**3/18**
4	**2**	**8/18**	**12/18**	**4/18**
5	**6**	**14/18**	**15/18**	**1/18**
6	**4**	**18/18**	**18/18**	**0**
المجموع	**18**			

حيث $\hat{F}(X)$ **هي عبارة عن الدالة التراكمية من العينة أي:**

$$\hat{F}(X_j)=\frac{\sum_{i=1}^{j} f_i}{\sum_{i=1}^{6} f_i}$$

ومن الجدول السابق (٩-١٤) نجد أن:

$$K=\frac{4}{18}=0.22$$

أي أن: $K<K_n^{\alpha}$

إذن نقبل الفرض القائل بأن نتائج الرمي تتبع التوزيع المتنظم بدرجة ثقة 95%.

(٩-٤) تطبيقـــــــــــــــــــــــــــــــــات Applications

تطبيق (٩-١): بإستخدم اختبار كولومجروف سيمرنوف اختبر الفرض القائل بأن درجات الحرارة في خلال 20 يوم في إحدى المناطق الشمالية ستتبع التوزيع المعتاد بتوقع 0.5 وتباين 1، وذلك بدرجة ثقة 95%.

-0.74	**-0.15**	**0.24**	**-0.95**	**0.56**	**1.74**	**1.86**	**-0.56**	**0.92**	**0.36**
-0.49	**-0.48**	**0.15**	**-1.06**	**1**	**-1.26**	**-0.10**	**0.70**	**0.82**	**0.32**

الحل:

١- **الفرض العدمي**

H_0 : **المتغير محل الدراسة يتبع دالة التوزيع المعتاد بتوقع 0.5 وتباين 1.**

الفرض البديل:

H_1 : **المتغير محل الدراسة لا يتبع دالة التوزيع المعتاد بتوقع 0.5 وتباين 1.**

٢- **نكون الجدول التالي:**

جدول (٩-١٥)

x	F(X)	$\hat{F}(X)$	$\left\|F(X)-\hat{F}(X)\right\|$
-1.26	**1/20 = 0.05**	**0.04**	**0.01**
-1.06	**2/20 = 0.10**	**0.06**	**0.04**
-0.95	**3/20 = 0.15**	**0.07**	**0.08**
-0.74	**4/20 = 0.20**	**0.11**	**0.09**
-0.56	**5/20 = 0.25**	**0.14**	**0.11**
-0.49	**6/20 = 0.30**	**0.16**	**0.14**
-0.48	**7/20 = 0.35**	**0.16**	**0.19**

x	$F(X)$	$\hat{F}(X)$	$\left\lvert F(X)-\hat{F}(X) \right\rvert$
-0.15	**8/20 = 0.40**	**0.26**	**0.14**
-0.10	**9/20 = 0.45**	**0.27**	**0.18**
0.15	**10/20 = 0.50**	**0.36**	**0.14**
0.24	**11/20 = 0.55**	**0.40**	**0.15**
0.32	**12/20 = 0.60**	**0.43**	**0.17**
0.36	**13/20 = 0.65**	**0.44**	**0.21**
0.55	**14/20 = 0.70**	**0.52**	**0.18**
0.56	**15/20 = 0.75**	**0.52**	**0.23**
0.70	**16/20 = 0.80**	**0.58**	**0.22**
0.82	**17/20 = 0.85**	**0.63**	**0.22**
0.92	**18/20 = 0.90**	**0.66**	**0.24**
1.74	**19/20 = 0.95**	**0.89**	**0.06**
1.86	**20/20 = 1**	**0.91**	**0.09**

حيث قيمة $\hat{F}(X)$ تم حسابها باستخدام جداول التوزيع المعتاد القياسي بملحق (٤) فمثلاً:-

$$K(X=-1.26)=P_r(X\le -1.26)$$
$$=P_r(Y\le \frac{-1.26-0.5}{1})$$
$$=P_r(Y\le \frac{-1.76}{1})$$
$$=0.5-P_r(0\le Y\le 1.76)$$
$$=0.5-0.46=0.04$$

ومن الجدول السابق نجد أن:

$$K^{\alpha}=0.24$$

٣- بما أن $K_n^{\alpha}=D_{20}^{0.05}=0.29$ **من ملحق (٩)**

وبمقارنة قيمة D لقيمة $K_{20}^{0.05}$ **نجد ان:**

$$D<D_{20}^{0.05}$$

وبالتالي نقبل الفرض العدمي القائل بأن درجات الحرارة في هذه المنطقة تتوزع توزيعاً معتاداً في هذه الفترة الزمنية بتوقع 0.5 وتباين 1 وذلك بدرجات ثقة 95%.

تطبيق (٩-٢): **قامت إحدى الشركات بعمل دورات تدريبية للعاملين بها لرفع كفاءة الأداء فإخذت عينة مكونة من 200 عامل وتم توزيعهم وفقاً لكفاءتهم قبل وبعد عمل الدورات وسجلت النتائج في الجدول التالي:**

جدول (٩-١٦)

مستوي الأداء بعد التدريب / مستوي الأداء قبل التدريب	غير ماهر	ماهر	المجموع
غير ماهر	20	80	100
ماهر	7	93	100
المجموع	27	173	200

اختبر الفرض القائل بأنه توجد علاقة بين كفاءة العامل والدورات التدريبية (أي يوجد تأثير للدورات على كفاءة العامل) وذلك بدرجة ثقة 95%.

الحل:

١- الفرض العدمي:

$$H_0 : P_r(A \cap B) = P_r(A) \times P_r(B)$$ **(عدم وجود علاقة)**

الفرض البديل

$$H_1 : P_r(A \cap B) \neq P_r(A) \times P_r(B)$$ **(أي يوجد علاقة)**

٢- نحسب التكرارات المتوقعة f' المناظرة للتكرارات الفعلية كما هو موضح بالجدول التالي:

جدول (٩-١٧)

(B) / (A)	ماهر		غير ماهر		المجموع	
ماهر	20	13.5	80	86.5	100	100
غير ماهر	7	13.5	93	86.5	100	100
المجموع	27	27	173	173	200	200

حيث يتم حساب f من العلاقة:

$$f_{ij} = \frac{\text{(مجموع التكرارات في العمود j) (مجموع التكرارات في الصف i)}}{\text{مجموع التكرارات}}$$

٣- يتم حساب χ^2 من الجدول السابق حيث:

$$\chi^2 = \sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\frac{[(f_{ij} - f'_{ij})^2]}{f'_{ij}}$$

$$= \frac{(20-13.5)^2}{13.5} + \frac{(80-86.5)^2}{86.5} + \frac{(7-13.5)^2}{13.5} + \frac{(93-86.5)^2}{86.5}$$

$$= 3.22 + 0.49 + 3.13 + 0.49 = 7.33$$

٤- بما أن $i = 1,2$ ، $j = 1,2$ **فإن:**

1 = (2-1)(2-1) = (1 – عدد الأعمدة)(1 – عدد الصفوف) = عدد درجات الحرية

وبالتالي فإن عند درجة ثقة 95% وعدد درجات حرية 1 فإنه من جداول χ^2 **نجد أن:** $\chi^2_{(1,0.05)} = 5.02$

وبما أن $\chi^2 = 7.33$ **أي تقع في منطقة الرفض وبالتالي نقبل الفرض البديل القائل بأنه توجد علاقة بين الدورات التدريبية وكفاءة العاملين بدرجة ثقة 95%.**

تطبيق (٩-٣): فرض دالة توزيع بواسون للبيانات التالية:

3 , 0 , 1 , 2 , 4 , 5 , 6 , 1 , 0 , 1 , 3 , 2 , 0 , 1 , 2 , 0 , 1 , 3 , 2 , 2 , 4 , 5 , 6 , 3 , 0 , 1 , 2 , 3 , 4 , 0

ثم اختبر جودة التوفيق باختبار χ^2 **بدرجة ثقة 95%.**

الحل:

من البيانات السابقة يمكن تكوين الجدول التكراري التالي:

جدول (٩-١٨)

القيمة X_i	0	1	2	3	4	5	6	المجموع
التكرارات f_i	6	6	6	5	3	2	2	30

القيمة المتوقعة $\overline{X} = \frac{\sum X_i f_i}{\sum f_i} = \frac{67}{30} = 2.23$

لتوفيق توزيع بواسون سوف نعتبر $\lambda = 2.23$

$$f(X) = \frac{e^{-\lambda}\lambda^X}{X!} = \frac{e^{-2.23}(2.23)^X}{X!} \qquad , \quad X = 0,1,2,....$$

ولاختبار جودة التوفيق باستخدام χ^2 **نتبع الخطوات التالية:**

باستخدام دالة الاحتمال f(X) نحسب التكرارات المتوقعة f_i' **حيث:**

$$f_i' = f(X_i)\sum f_i$$

ثم نكون الجدول التالي:

جدول (٩-١٨)

X_i	f_i	$f(X_i)$	f_i'	$\frac{(f_i - f_i')^2}{f_i'}$
0	6	0.1076	3.228	2.3804
1	6	0.2398	7.194	0.1982
2	6	0.2674	8.022	0.5097
3	5	0.1987	5.961	0.1549
4	3	0.1108	3.324	0.0316
5	2	0.0494	1.482	0.1881
6	2	0.0181	0.543	3.9095
المجموع	30			7.3654

من الجدول السابق نجد أن: $\chi^2 = 7.3654$

وعند درجات حرية 6 ومستوي معنوية 5% نجد أن:

$$\chi^2_{(2,0.05)} = 12.6$$

وبما أن χ^2 المحسوبة < χ^2 الجدولية

إذن نقبل الفرض العدمي القائل بان توزيع بواسون توفيق جيد في هذه الحالة.

تطبيق (٩-٤): قامت هيئة التليفزيون باستطلاع الرأي بالنسبة لمشاهدة الجمهور لأحد البرامج الثقافية وعلاقته بالإعلان عند هذا البرنامج وأهميته قبل عرضه. فأخذت عينة مكونه من 2000 مشاهد وتم سؤال كل منهم عن مدي متابعة البرنامج قبل وبعد الإعلان عنه. والجدول التالي يوضح توزيع المشاهدين في العينة وفقاً لمتابعتهم للبرنامج قبل وبعد الإعلان.

جدول (٩-٢٠)

مستوي المتابعة / الإعلان	متابع دائماً	متابع أحياناً	غير متابع	المجموع
قبل الإعلان	400	190	410	1000
بعد الإعلان	420	170	410	1000
المجموع	820	360	820	2000

اختبر الفرض القائل بأنه يوجد علاقة بين الإعلان عن البرنامج ومستوي المتابعة وذلك بدرجة ثقة 95%.

الحل:

١- إذا فرضنا أن P_{ij} هو احتمال أن تكون حالة المشاهد بالنسبة للإعلان (i) والحالة لمتابعة البرنامج (j) كذلك P_i' ، P_j' هي احتمال الحالة (i) واحتمال الحالة (j) على الترتيب. وبالتالي يصبح الفرض الإحصائي:

الفرض العدمي:

$$H_0 : P_{ij} = P_i' \times P_j' \qquad \text{(عدم وجود علاقة)}$$

الفرض البديل:

$$H_1 : P_{ij} \neq P_i' \times P_j' \qquad \text{(وجود علاقة)}$$

٢- من الجدول السابق نحسب التكرارات f_{ij}' حيث:

$$f_{ij}' = \frac{\left(\sum_i f_{ij}\right)\left(\sum_j f_{ij}\right)}{\sum_i \sum_j f_{ij}}$$

كما هو موضح بالجدول التالي:

جدول (٩-٢١)

i \ J	متابع دائماً		متابع أحياناً		غير متابع		المجموع	
قبل الإعلان	410	0.24	180	0.56	410	0	1000	0.80
بعد الإعلان	410	0.24	180	0.56	410	0	1000	0.80
المجموع	820	0.48	360	1.12	820	0	2000	1.6

حيث تمثل أي خلية بالجدول السابق:

f'_{ij}
$\dfrac{(f_i - f'_i)^2}{f'_i}$

وبالتالي فإن:

$$\chi^2 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left[\frac{(f_{ij} - f'_{ij})^2}{f'_{ij}} \right] = 1.6$$

٣- عند درجات الحرية [2 = (1-3)(1-2)] ودرجة ثقة 95% نجد أن:

$$\chi_1^2 = 0.0506 \quad , \quad \chi_2^2 = 7.38$$

وبما ان χ^2 في منطقة القبول.

∴ نقبل الفرض العدمي القائل باستقلال الحالة الإعلانية عن مدي متابعة المشاهد للبرنامج وذلك بدرجة ثقة 95%.

(٩-٥) تمرينـــــــــــــــــــــــــــــــــــــات Exercises

(٩-١): وفق دالة توزيع بواسون للبيانات التالية:

8, 7, 6, 7, 8, 7, 5, 9, 6, 8, 7, 6, 6, 9, 5

ثم اختبر جودة التوفيق باختبار مناسب بدرجة ثقة 95%، 85%، 99%.

(٩-٢): قامت هيئة التليفزيون باستطلاع الرأي بالنسبة لمشاهدة الجمهور لأحد البرامج الثقافية وعلاقته بالإعلان عن هذا البرنـامج وأهميتـه قبل عرضـه. فأخذت عينـة مكونـه من 2000 مشاهد وتم سؤال كل مـنهم عن مـدي متابعـة البرنـامج قبـل وبعد الإعلان عنه.

والجدول التالي يوضـح توزيـع المشـاهدين فـي العينـة وفقـاً لمتـابعتهم للبرنـامج قبل وبعد الإعلان.

جدول (٩-٢٢)

الإعلان \ مستوي المتابعة	متابع دائماً	متابع أحياناً	غير متابع	المجموع
قبل الإعلان	400	190	410	1000
بعد الإعلان	420	170	410	1000
المجموع	820	360	820	2000

المطلوب:

اختبـر الفـرض القائـل بأنـه يوجـد علاقـة بـين الإعـلان عن البرنـامج ومسـتوي متابعة المشاهدين لها وذلك بدرجة ثقة 95%، 99%.

(٩-٣): الجدول التالي يوضح التوزيع التكراري للدخل اليومي بالجنيـة لـ 100 عامل من العمال المهرة.

جدول (٩-٢٣)

الدخل بالجنية (المتغير)	30 -	40 -	50 -	60 -	70 - 80	المجموع
عدد العمال	14	24	32	20	10	100

(التكرارات)						

المطلوب:

١- وفق المنحني الطبيعي للدخل في المجتمع المسحوب منه العينة.
٢- اختبر جودة التوفيق بدرجة ثقة 90%، 99% وذلك باختبار مناسب.

(٩-٤): تقوم إحدى الشركات بإنتاج 3 أنواع من منتج معين A, B, C فإذا أخذت عينة مكونه من 500 وحدة من المنتج من الأنواع المختلفة وتم تحديد عدد الوحدات المعيبة والسليمة في الجدول التالي:

جدول (٩-٢٤)

مستوي الجودة / نوع المنتج	سليم	معيب	المجموع
A	150	10	160
B	200	25	225
C	100	15	115
المجموع	450	50	500

المطلوب:

اختبر الفرض القائل بأنه يوجد علاقة بين نوع المنتج ومستوي جودته وذلك بدرجة ثقة 95%.

الباب العاشر
السلاسل الزمنية
Time Series

(١٠-١) السلاسل الزمنية
Time Series

(١٠-٢) مركبات السلسلة الزمنية
Components of Time Series

(١٠-٣) نماذج الاتجاه العام
General Trend Models

(١٠-٤) الارتباط الذاتي
Auto Correlation

(١٠-٥) اختبار ديربن واتسون
Durbin Waston Test

(١٠-٦) تمرينات
Exercises

(١٠-١)السلاسل الزمنية
Introduction

في كثير من الدراسات يتطلب البحث دراسة التطور التاريخي للظاهرة (المتغير) محل البحث، حتى يمكن باستخدام هذه الدراسة التنبؤ بسلوك الظاهرة في الفترات المستقبلية، وتتطلب هذه الدراسة توافر بيانات عن الظاهرة في وحدات زمنية متتالية تسمي بالسلسلة الزمنية **Time Series** فباستخدام هذه البيانات يمكن التنبؤ بسلوك الظاهرة في الفترات المستقبلية وبالتالي تقدير قيمتها.

ويهدف هذا الباب إلى دراسة سلوك الظاهرة من واقع البيانات التاريخية عنها وباستخدام هذه البيانات يتم بناء نموذج إحصائي يعطي العلاقة بين المتغير الأساسي (الزمن) والمتغير التابع (الظاهرة محل الدراسة) ويسمي هذا النموذج بالسلسلة الزمنية. وباستخدام هذا النموذج يمكن التنبؤ بقيمة الظاهرة (أي تقدير قيمة الظاهرة) في الفترات المستقبلية كذلك دراسة الارتباط بين قيم الظاهرة وهو ما يسمي بالارتباط الذاتي **Auto-Correlation**.

ففي الفصل (١٠-٢) سوف نتناول العناصر (المركبات) المكونة للسلسلة الزمنية. وفي الفصل (١٠-٣) نقدم ثلاثة نماذج انحدار لتقدير القيم الاتجاهية للظاهرة. وفي الفصليين (١٠-٤)، (١٠-٥) نتناول الارتباط الذاتي واختبار ديربن واتسون للارتباط بين قيم الظاهرة في الفترات الزمنية المتتالية.

(١٠-٢) مركبات السلسلة الزمنية

Components of Time Series

مما سبق يتضح أن السلسلة الزمنية هي عبارة عن قيم الظاهرة في فترات زمنية متتالية. حيث يمكن باستخدام هذه البيانات بناء نموذج إحصائي يعطي العلاقة بين الزمن وقيم المتغير (الظاهرة) محل الدراسة.

فعلى سبيل المثال يوضح الجدول التالي الدخل القومي بالنسبة ارأس المال خلال الفترة الزمنية [1986/1987 – 1980/1981] في جمهورية مصر العربية.

جدول (١٠-١)

السنة	1980/ 1981	1981/ 1982	1982/ 1983	1983/ 1984	1984/ 1985	1985/ 1956	1986/ 1987
الدخل القومي بالنسبة لرأس	407.2	450.2	621.8	745.7	836	892	1044.5

المال							

المصدر: الجهاز المركزي للتعبئة العامة والإحصاء.

فمن الجدول يتضح زيادة الدخل القومي بالنسبة لرأس المال مع تزايد الزمن، أي يوجد علاقة طردية بين الزمن والدخل القومي بالنسبة لرأس المال.

كذلك يوضح جدول (١٠-٢) قيمة الصادرات والواردات لجمهورية مصر العربية في الفترة [1989-1984] بالمليون جنية.

جدول (١٠-٢)

السنة	قيمة الصادرات (بالألف جنية)	قيمة الواردات (بالألف جنية)	الفائض أو العجز
1984	2197933	7536068	-5338135
1985	2599941	6973061	-4373120
1986	2053659	4051432	-5997473
1987	3046019	11357837	-4311827
1988	3994436	16308572	-12314136
1989	5734726	16623623	-10888897

المصدر: الجهاز المركزي للتعبئة العامة والإحصاء.

ومن جدول (١٠-٢)يتضح تزايد قيمة الصادرات كذلك تزايد قيمة الواردات مع تزايد الزمن ولكن من العمود الأخير بالجدول يتضح أن قيمة الزيادة في الواردات تفوق قيمة الزيادة في الصادرات.

وتعتبر قيمة الظاهرة (المتغير) محل الدراسة عند أي نقطة زمنية في السلسلة محصلة تفاعل عدة عوامل تسمي بمركبات السلسلة الزمنية وهي*:

١- الاتجاه العام General Trend.

٢- التغيرات الموسمية Seasonal Movements.

* أ.د. حسن عبد العزيز ، د. طارق عميرة (١٩٨٧) : "مبادئ في الإحصاء واستخداماته" – دار النهضة العربية – القاهرة.

٣- التغيرات الدورية Cyclical Movements.

٤- التغيرات العرضية Irregular Movements.

ويختلف تأثير كل عامل (أو عنصر) من العوامل المذكورة أعلاه على قيمة الظاهرة محل الدراسة. وفيما يلي سوف نتناول هذه العوامل (العناصر) باختصار

أولاً: الاتجاه العام

الاتجاه العام هو أثر يظهر ويستمر لمدة طويلة ويمكن تمثيل العلاقة بين هذا الأثر والزمن بنموذج إحصائي خطي أو غير خطي كما سوف يتضح في الفصل التالي:

ثانياً: التغيرات الموسمية

هي تقلبات تتكرر بانتظام خلال الفترات الزمنية الصغيرة (المواسم) وعادة يكون أثرها داخل الموسم (وعادة يكون الموسم هو الوحدة الزمنية المستخدمة خلال الفترة محل الدراسة) هذا بالإضافة إلى أن مجموعة التغيرات الموسمية خلال سنة يقترب من الصفر.

~~ثالثاً: التغيرات الدورية~~

هي تقلبات تتكرر كل دورة وعادة ما يكون طول الدورة يزيد عن الوحدة الزمنية المستخدمة في الدراسة. ومن أمثلة ذلك الدورات الاقتصادية في حالة حدوث تضخم وانكماش على فترات متتالية. وعادة يقترب مجموع التغيرات الدورية خلال دورة كاملة من الصفر.

رابعاً: التغيرات العرضية

التغيرات العرضية هي تقلبات فجائية تحدث عادة نتيجة أحداث نادرة مثل الحروب، الأوبئة، وتكون هذه التقلبات غير منتظمة في سلوكها مع الزمن وفي كثير من الحالات يفترض أن هذه التغيرات تخضع لقانون احتمالي معين.

وتحليل السلسلة الزمنية يعني قياس حجم وتحديد اتجاه كل عنصر من العناصر المذكورة أعلاه المؤثرة على قيمة الظاهرة وبالتالي يمكن استبعاد أثره للتنبؤ بقيمة

الظاهرة في المستقبل. وذلك اعتماداً على البيانات التاريخية للظاهرة (المتغير) محل الدراسة.

وكما ذكرنا سابقاً أن السلسلة الزمنية هي علاقة بين متغير مستقل (الزمن) X ومتغير تابع (قيمة الظاهرة) Y فإنه يمكن صياغة العلاقة بين قيمة الظاهرة (Y) والزمن (X) في شكل نموذج انحدار (أنظر الباب) وعادة تسمي نماذج الانحدار لتقدير قيمة الظاهرة بنماذج التنبؤ Forcasting Models (حيث يمثل المتغير المستقل في هذه النماذج الزمن) وذلك لأنه باستخدام هذه النماذج يمكن تقدير قيمة الظاهرة في فترة مستقبلية.

وفيما يلي نتناول نوعين* من نماذج الانحدار لتقدير قيمة الظاهرة في المستقبل وهو ما يسمي بنماذج التنبؤ Forcasting Models .

---

* أ.د. نادية مكاري جرجس (١٩٧٣): "الإحصاء الاقتصادي" كلية الاقتصاد والعلوم السياسية – جامعة القاهرة.

(١) النموذج التجميعي Additional Model

في هذه الحالة يفترض أن عناصر (مركبات) السلسلة الزمنية تؤثر بطريقة تجميعية على الظاهرة محل البحث، بمعني أن تأثيرها مطلق ولا يعتمد على المستوي العام للظاهرة.

فإذا فرضنا أن:

- القيمة الفعلية للظاهرة في الزمن X تساوي Y.
- القيمة الاتجاهية للظاهرة في الزمن X تساوي $f_1(X)$.
- قيمة التغير الموسمي في الزمن X تساوي $f_2(X)$.
- قيمة التغيرات الدورية في الزمن X تساوي $f_3(X)$.
- قيمة التغيرات العرضية في الزمن X تساوي $\varepsilon(X)$.

فإن:

$$Y = f(X) = f_1(X) + f_2(X) + f_3(X) + \varepsilon(X) \quad \textbf{(10.1)}$$

(٢) النموذج النسبي Relative Model

وفي هذه الحالة يفترض أن تأثير العوامل الموسمية والدورية والعشوائية تأثير نسبي يتوقف على مستوي الاتجاه العام للظاهرة فإذا فرضنا أن:

- أثر التغيرات الدورية في الزمن X يساوي $h_1(X)$.
- أثر التغيرات الموسمية في الزمن X يساوي $h_2(X)$.

فإن:

$$Y = f(X) = f_1(X) \times h_1(X) \times h_2(X) \times \varepsilon(X) \qquad \textbf{(10.2)}$$

وتعتبر العلاقة (10.2) أقرب للواقع بالنسبة لكثير من الظواهر الفعلية.

وفي هذا الباب سوف نقتصر دراستنا على بعض الظواهر التي لا تخضع للمتغيرات الموسمية والدورية. أو بعبارة أخرى سوف نكتفي بدراسة بعض الظواهر التي يكون فيها القيمة الفعلية للظاهرة هي محصلة التفاعل بين الاتجاه العام ($f_1(X)$) والعوامل العرضية ($\varepsilon(X)$) كما سوف يتضح في الفصل التالي.

وفي هذا الباب سوف نقتصر دراستنا على تقدير القيمة الاتجاهية **Trend Value** للظاهرة أي تقدير الدالة $f_1(X)$ سواء كان تأثير باقي مركبات السلسلة الزمنية (التغيرات الدورية، التغيرات العرضية، التغيرات الموسمية) على قيمة الظاهرة تأثير جمعي كما في العلاقة (10.1) أو تأثير نسبي كما في العلاقة (10.2). وسوف نوضح ذلك بالتفصيل في الفصل التالي.

(١٠-٣) نماذج الاتجاه العام

General Trend Model

في هذا الفصل سوف نتناول العلاقة بين القيمة الاتجاهية للظاهرة $f_1(X)$ والزمن (X) كنماذج انحدار في الحالات الثلاثة التالية:

١- إذا كان الاتجاه العام للظاهرة اعتماداً على الزمن يأخذ الصورة الخطية.

٢- إذا كان الاتجاه العام للظاهرة اعتماداً على الزمن يأخذ صورة منحني من الدرجة الثانية.

٣- إذا كان الاتجاه العام للظاهرة اعتماداً على الزمن يأخذ الصورة الأسية.

الحالة الأولي

إذا كانت العلاقة بين القيمة الاتجاهية $f_1(X)$ والزمن X على الصورة:

$$f_1(X) = a_0 + a_1X + \varepsilon \qquad (10.3)$$

حيث ε مجموعة من المتغيرات العشوائية **Random Variables**.

ومن العلاقة (10.3) نجد أن تقدير الاتجاه العام $f_1(X)$ أي نموذج الاتجاه العام هو:

$$\hat{f}_1(X) = \hat{a}_0 + \hat{a}_1X \qquad (10.4)$$

حيث $f_1(X)$ هي القيمة الاتجاهية المقدرة للظاهرة في الزمن X (أي قيمة الظاهرة وهي متأثرة بالاتجاه العام فقط في الزمن X) كذلك $\hat{a}_0$ ، $\hat{a}_1$ هي تقديرات المعلمات a_0 ، a_1.

وفي حالة افتراض أن ε مجموعة من المتغيرات العشوائية المستقلة المتجانسة (أي لها نفس التباين) وتتبع التوزيع الطبيعي أي أن:

$$E(\varepsilon) = 0 \;,\; V(\varepsilon) = \gamma^2 \;,\; \mathrm{cov}(\varepsilon_j, \varepsilon_k) = 0 \qquad j \neq k \quad \textbf{(10.5)}$$

وباستخدام نظرية (٧-٣) بالباب السابع، فإنه يمكن استخدام طريقة المربعات الصغرى لحساب قيم التقديرات $\hat{a}_0$ ، $\hat{a}_1$ من المعادلتين التاليتين:

$$\sum_{j=1}^{n} Y_j = n\,\hat{a}_0 + \hat{a}_1 \sum_{j=1}^{n} X_j \qquad \textbf{(10.6)}$$

$$\sum_{j=1}^{n} Y_j X_j = \hat{a}_0 \sum_{j=1}^{n} X_j + \hat{a}_1 \sum_{j=1}^{n} X_j^2 \qquad \textbf{(10.7)}$$

وبحساب قيم $\hat{a}_0$ ، $\hat{a}_1$ فإنه يمكن تقدير القيمة الاتجاهية للظاهرة في الزمن X حيث:

$$\hat{f}_1(X) = \hat{a}_0 + \hat{a}_1 X \qquad \textbf{(10.8)}$$

قياس التغيرات الدورية والعشوائية

إذا كانت بيانات السلسلة الزمنية بيانات سنوية فإن قيمة الظاهرة (المتغير) في هذه الحالة تعتمد على: الاتجاه العام Trend، وكل من التغيرات الدورية Cyclical، والتغيرات العرضية Irregular. أو بعبارة أخري لا تعتمد قيمة الظاهرة على التغيرات الموسمية (حيث أن الآثار الموسمية خلال العام تساوي صفر).

فإذا استخدمنا النموذج النسبي كما في العلاقة (10.2) حيث:

$$Y = f(X) = f_1(X) \times h_1(X) \times \varepsilon(X) \qquad \textbf{(10.9)}$$

ومن العلاقة (10.9) نجد أن:

$$\frac{Y}{f_1(X)} = \left[h_1(X) \times \varepsilon(X)\right] \qquad \textbf{(10.10)}$$

أي أن نسبة تأثير التغيرات الدورية والعشوائية معاً $\left[h_1(X) \times \varepsilon(X)\right]$ يساوي النسبة بين القيمة الفعلية للظاهرة (Y) والقيمة الاتجاهية لها ($f_1(X)$).

كذلك إذا كان $\hat{f_1}(X)$ هي تقدير $f_1(X)$ فإن:

$$\frac{Y}{\hat{f_1}(X)} = \text{تقدير نسبة تأثير التغيرات الدورية والعشوائية} \qquad \textbf{(10.11)}$$

ومن العلاقتين (10.10)، (10.11) نجد أن:

١- إذا كان $\frac{Y}{\hat{f_1}(X)} = 1$ فإنه لا يوجد تأثير للتغيرات الدورية والعرضية على الظاهرة.

٢- إذا كان $\frac{Y}{\hat{f_1}(X)} < 1$ فإن القيمة الفعلية أقل من القيمة الاتجاهية للظاهرة – أي ان القيمة الفعلية Y تقع أسفل خط الاتجاه العام $\hat{f_1}(X)$ - وهذا يعني أن تأثير التغيرات الدورية والعرضية تأثير سلبي على الظاهرة.

٣- إذا كان $\frac{Y}{\hat{f}_1(X)} > 1$ فإن القيمة الفعلية أكبر من القيمة الاتجاهية للظاهرة

– أي أن القيمة الفعلية Y تقع أعلى خط الاتجاه العام $\hat{f}_1(X)$ - وهذا يعني أن تأثير التغيرات الدورية والعرضية تأثير إيجابي على الظاهرة.

وسوف نوضح النقاط المذكورة أعلاه من خلال المثال التالي:

مثال (١٠-١)

الجدول التالي يوضح قيمة إيرادات إحدى الشركات التجارية بالألف جنية وذلك خلال الفترة الموضحة.

جدول (١٠-٣)

السنة	1990	1991	1992	1993	1994	1995	1996
قيمة الإيرادات بالألف جنية	10	12	14	17	22	25	27

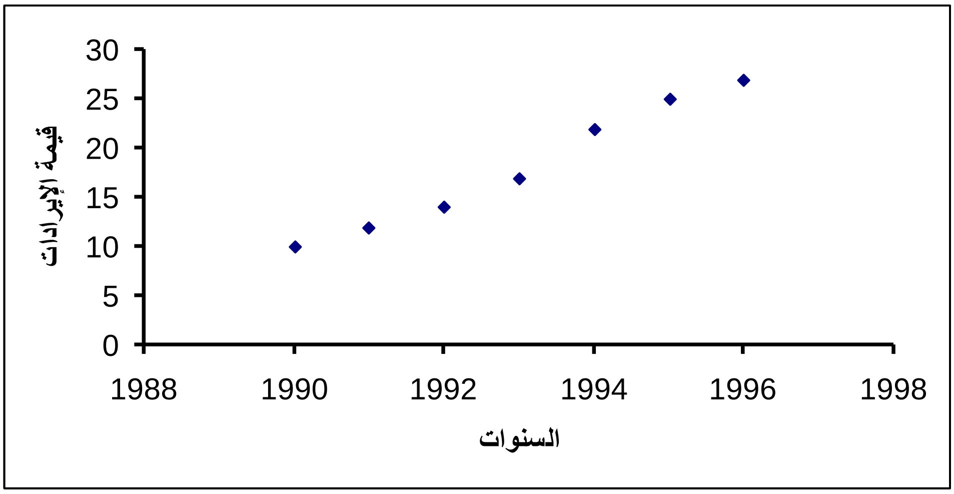
والمطلوب:

أ) وضح بيانياً التطور التاريخي لقيمة الإيرادات.
ب) أبني نموذج الاتجاه العام للظاهرة.
ج) باستخدام النموذج في (ب) قدر القيم الاتجاهية للإيرادات في السنوات المعطاة.
د) قدر القيم الاتجاهية للظاهرة ثم أوجد نسبة واتجاه تأثير التغيرات الدورية والعشوائية في السنوات المعطاة.
ه) باستخدام النموذج قدر القيم الاتجاهية للظاهرة في السنوات ١٩٩٣، ١٩٩٦.

الحل:

من الشكل يتضح انه يمكن افتراض أن دالة الاتجاه العام دالة خطية على النحو: $f_1(X) = a_0 + a_1X + \varepsilon$

أ) الشكل التالي يوضح الاتجاه العام للظاهرة (قيمة الإيرادات بالألف جنية).



شكل (١٠-١)

وأن تأثير التغيرات الدورية والعشوائية تأثير نسبي أي أن:

$$Y = f(X) = f_1(X) \times h_1(X) \times \varepsilon(X)$$

وبالتـالي يمكـن بنـاء نمـوذج خطـي يمكـن باسـتخدامه تقـدير القيم الاتجاهيـة للظـاهرة $\hat{f}_1(X)$ حيث:

$$\hat{f}_1(X) = \hat{a}_0 + \hat{a}_1 X$$

حيث يتم حساب $\hat{a}_0$ ، $\hat{a}_1$ من المعادلتين (10.6)، (10.7) لذلك نكون الجدول التالي:

جدول (١٠-٤)

السنة	قيمة الإيرادات (Y)	X	XY	X^2
1990	**10**	**-3**	**-3**	**9**
1991	**12**	**-2**	**23**	**4**
1992	**14**	**-1**	**-14**	**1**
1993	**17**	**0**	**0**	**0**
1994	**22**	**1**	**22**	**1**

1995	25	2	50	4
1996	27	3	81	9
المجموع	127	0	85	82

بالتعويض في المعادلات (10.6)، (10.7) من الجدول نجد:

$$\sum_{j=1}^{n} Y_j = n\hat{a}_0 + \hat{a}_1 \sum_{j=1}^{n} X_j$$

$$127 = 7\hat{a}_0 + (0)\hat{a}_1 \rightarrow \hat{a}_0 = 18.14$$

وكذلك:

$$\sum_{j=1}^{n} Y_j X_j = \hat{a}_0 \sum_{j=1}^{n} X_j + \hat{a}_1 \sum_{j=1}^{n} X_j^2$$

$$85 = (0)\hat{a}_0 + 28\hat{a}_1 \rightarrow \hat{a}_1 = 3.04$$

وبالتالي فإن نموذج الاتجاه العام هو:

$$\hat{f}_1(X) = 18.14 + 3.04X$$

ج) وبالتعويض بقيم X المعطاة في النموذج السابق يمكن حساب $\hat{f}_1(X)$ في هذه الفترة كما هو موضح بالعمود رقم (3) بجدول (١٠-٥).

جدول (١٠-٥)

(1) السنوات	(2) القيم الفعلية Y	(3) القيم الاتجاهية المقدرة $\hat{f}_1(X)$	(4) $\dfrac{Y}{\hat{f}_1(X)}$	(5) تأثير التغيرات الدورية والعشوائية
1990	10	9.20	1<1.11	تأثير موجب
1991	12	12.06	1>0.995	تأثير سالب
1992	14	15.01	1>0.933	تأثير سالب
1993	17	18.14	1>0.937	تأثير سالب
1994	22	21.18	1<1.039	تأثير موجب

تأثير موجب	1<1.032	24.22	25	1995
تأثير سالب	1>0.998	27.06	27	1996

د) العمود رقم (4) بالجدول يعطي نسبة تأثير المتغيرات الدورية والعشوائية على الظاهرة. فعلي سبيل المثال في سنة 1990 نجد أن:

$$\frac{Y}{\hat{f}_1(X)} = 1.11 \approx 111\%$$

أي أن القيمة الفعلية (Y) للظاهرة تزيد عن القيمة الاتجاهية التقديرية بمقدار 11% وهذا يرجع إلى أن تأثير التغيرات الدورية والعرضية إيجابي أدي إلى زيادة القيمة الفعلية للظاهرة بمقدار 11%.

كذلك في سنة 1993 نجد أن:

$$\frac{Y}{\hat{f}_1(X)} = 0.937 \approx 93.7\%$$

أي أن القيمة الفعلية (Y) للظاهرة تقل عن القيمة الاتجاهية التقديرية بمقدار 6.3% (6.3 = 93.7 - 100) وهذا يرجع إلى أن تأثير التغيرات الدورية والعرضية سلبي أدي إلى نقص القيمة الفعلية للظاهرة بمقدار 6.3%.

هـ) وبما أن قيمة X المناظرة لسنة 1993 تساوي صفر. بالتالي يمكن تقدير القيمة الاتجاهية للإيرادات لسنة 1993 على النحو التالي:

$$\hat{f}_1(X=0) = 18.14 + 3.04(0) = 18.14$$ **ألف جنية**

وبالمثل يمكن حساب سنة 1996.

$$\hat{f}_1(X=3) = 18.14 + 3.04(3) = 27.26$$ **ألف جنية**

الحالة الثانية

إذا كانت العلاقة بين دالة الاتجاه العام الاتجاهية $f_1(X)$ والزمن X على الصورة:

$$f_1(X) = a_0 + a_1 X + a_2 X^2 + \varepsilon \qquad \textbf{(10.12)}$$

حيث ε مجموعة من المتغيرات العشوائية Random Variables، و a_0, a_1, a_2 هي معلمات العلاقة. ومن العلاقة (10.12) نجد أن نموذج الاتجاه العام يصبح على النحو التالي:

$$\hat{f}_1(X) = \hat{a}_0 + \hat{a}_1 X + \hat{a}_2 X^2 \qquad \textbf{(10.13)}$$

حيث $\hat{f}_1(X)$ هي القيمة الاتجاهية المقدرة للظاهرة في الزمن X كذلك $\hat{a}_0, \hat{a}_1, \hat{a}_2$ هي تقديرات للمعلمات a_0, a_1, a_2. وباستخدام طريقة المربعات الصغرى أيضاً يمكن حساب $\hat{a}_0, \hat{a}_1, \hat{a}_2$ من المعادلات الطبيعية التالية:

$$\sum_{j=1}^{n} Y_j = n\hat{a}_0 + \hat{a}_1 \sum_{j=1}^{n} X_j + \hat{a}_2 \sum_{j=1}^{n} X_j^2 \qquad \textbf{(10.14)}$$

$$\sum_{j=1}^{n} Y_j X_j = \hat{a}_0 \sum_{j=1}^{n} X_j + \hat{a}_1 \sum_{j=1}^{n} X_j^2 + \hat{a}_2 \sum_{j=1}^{n} X_j^3 \qquad \textbf{(10.15)}$$

$$\sum_{j=1}^{n} Y_j X_j^2 = \hat{a}_0 \sum_{j=1}^{n} X_j^2 + \hat{a}_1 \sum_{j=1}^{n} X_j^3 + \hat{a}_2 \sum_{j=1}^{n} X_j^4 \qquad \textbf{(10.16)}$$

وبحل المعادلات (10.14) – (10.16) نحصل على قيم $\hat{a}_0, \hat{a}_1, \hat{a}_2$.

مثال (١٠-٢)

الجدول التـالي يوضـح صـافي الـربح المتحصل مـن إنتـاج سـلعة معينـة فـي الفتـرة [1990-1996] بالألف جنية بالنسبة لإحدى الشركات المنتجة لهذه السلعة.

جدول (١٠-٦)

السنة	1990	1991	1992	1993	1994	1995	1996
صافي الربح (بالألف جنية)	10	45	70	100	75	50	30

والمطلوب:

أ) وضح بيانياً العلاقة بين الزمن وصافي الربح المتحصل من هذه السلعة.

ب) كون نموذج انحدار للاتجاه العام للظاهرة.

ج) أوجد القيمة الاتجاهية المقدرة في سنة 1995، 2000.

الحل:

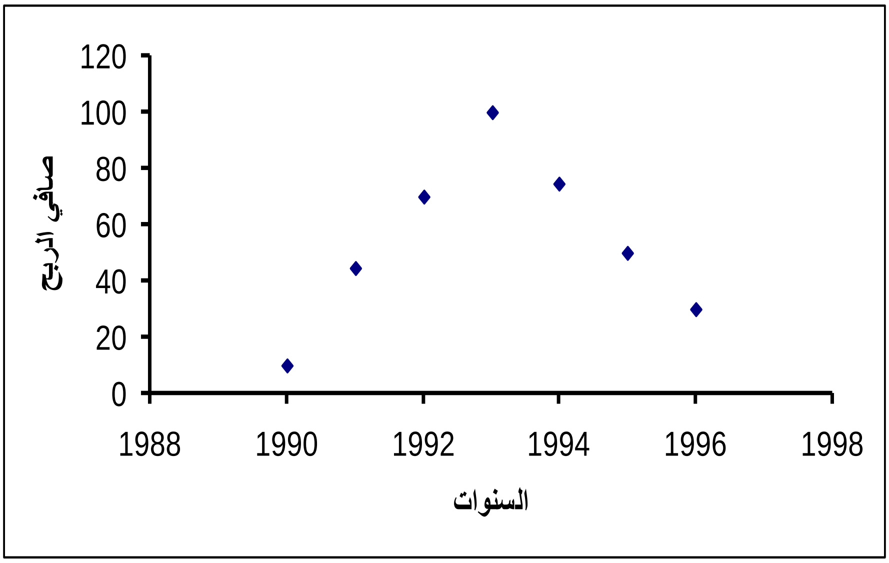
أ) الشكل التـالي يوضـح العلاقـة بـين قيمـة الظـاهرة Y والـزمن X فـي الفتـرة [1990-1996].

ب) من الشكل يتضح أن علاقة الاتجاه العام للظاهرة تأخذ الصورة التالية:

$$f_1(X) = a_0 + a_1 X + a_2 X^2 + \varepsilon$$

حيث ε **مجموعة من المتغيرات العشوائية المستقلة.**

$$\hat{f}_1(X) = \hat{a}_0 + \hat{a}_1 X + \hat{a}_2 X^2$$



شكل (١٠-٢)

ولحساب قيم التقديرات $\hat{a}_0, \hat{a}_1, \hat{a}_2$ نكون الجدول التالي:

جدول (١٠-٧)

السنة	صافي الربح (بالألف جنية)	X	X^2	X^3	X^4	YX	YX^2
1990	**10**	**-3**	**9**	**-27**	**81**	**-30**	**90**
1991	**45**	**-2**	**4**	**-8**	**16**	**-90**	**180**
1992	**70**	**-1**	**1**	**-1**	**1**	**-70**	**70**
1993	**100**	**0**	**0**	**0**	**0**	**0**	**0**
1994	**75**	**1**	**1**	**1**	**1**	**75**	**75**
1995	**50**	**2**	**4**	**8**	**16**	**100**	**200**
1996	**30**	**3**	**9**	**27**	**81**	**90**	**270**
المجموع	**380**	**0**	**28**	**0**	**196**	**75**	**885**

وبالتعويض من الجدول في المعادلات الطبيعية (10.14) – (10.16) نحصل على المعادلات التالية:

$$380 = (7)\hat{a}_0 + (0)\hat{a}_1 + (28)\hat{a}_2$$

$$75 = (0)\hat{a}_0 + (28)\hat{a}_1 + (0)\hat{a}_2$$

$$75 = (28)\hat{a}_0 + (0)\hat{a}_1 + (196)\hat{a}_2$$

وبحل المعادلات نحصل على قيم $\hat{a}_0, \hat{a}_1, \hat{a}_2$ **على النحو التالي:**

$$\hat{a}_0 = 84.53 \quad , \quad \hat{a}_1 = 2.68 \quad , \quad \hat{a}_2 = -7.56$$

$$\boxed{\hat{f}_1(X) = 48.53 + 2.68X - 7.56X^2 \qquad \textbf{(10.17)}}$$

ج) وبما أن سنه الأساس هي 1993 وبالتالي فإنه يمكن حساب X سنة 1995 من المعادلة:

X = 1995 – 1993 = 2

وبالتعويض بقيمة X في النموذج (11.17) نحصل على القيمة الاتجاهية المقدرة للربح سنة 1995 كما يلي:

$$\hat{f}_1(X) = 48.53 + 2.68(2) - 7.56(2)^2 = 59.65$$ **ألف جنية**

الحالة الثالثة

إذا كانت دالة الاتجاه العام على النحو التالي:

$$\boxed{f_1(X) = a_0(1 + a_1)^X \varepsilon \qquad \textbf{(10.18)}}$$

أي أن قيمة الظاهرة تتزايد (أو تتناقص) مع الزمن بمعدل ثابت هو a_1 ـ وهي علاقة غير خطية تأخذ الشكل الأسى.

وبالتالي فإن نموذج الاتجاه العام المنظر للدالة (10.18) هو

$$\boxed{\hat{f}_1(X) = \hat{a}_0(1 + \hat{a}_1)^X \qquad \textbf{(10.19)}}$$

والعلاقة (10.18) علاقة غير خطية يمكن تحويلها إلى الصورة الخطية بأخذ لوغاريتمات الطرفين على النحو التالي:

$$\log f_1(X) = \log a_0 + X\log(1+a_1) + \log\varepsilon \qquad \textbf{(10.20)}$$

فإذا فرضنا أن:

$$\log f_1(X) = f_1'(X)$$

$$\log a_0 = a_0'$$

$$\log(1+a_1) = a_1'$$

$$\log\varepsilon = \varepsilon'$$

وبالتالي فإن المعادلة (10.18) يمكن أن تكتب على الصورة:

$$\boxed{f_1'(X) = a_0' + a_1'X + \varepsilon' \qquad \textbf{(10.21)}}$$

والعلاقة (10.21) علاقة خطية يمكن استخدام طريقة المربعات الصغرى لحساب تقديراتها $\hat{a}_1'$, $\hat{a}_0'$ من المعادلات الطبيعية. وسوف نوضح ذلك من خلال المثال التالي:

مثال (١٠-٣)

الجدول التالي يوضح عدد السكان في إحدى المدن بالألف نسمة خلال الفترة [1972-1978].

جدول (١٠-٨)

السنوات	1972	1973	1974	1975	1976	1977	1978
عدد السكان (بالألف نسمة)	2.28	3	3.85	5	6.5	8.46	11

المطلوب:

أ) إذا كانت دالة الاتجاه العام لعدد السكان بالسنوات على النحو التالي:

$$Y = f_1(X) = a_0(1+a_1)^X\varepsilon$$

حيث Y تشير إلى عدد السكان بالألف نسمة، وX تشير إلى انحرافات السنوات عن سنة الأساس التي تم اختيارها، ε تشير إلى العوامل العشوائية. كون نموذج انحدار مناسب للدالة $f_1(X)$.

ب) قدر عدد السكان المتوقع عام 1991.

الحل:

أ) بما أن:

$$Y = f_1(X) = a_0(1+a_1)^X \varepsilon$$

فإذا فرضنا ان:

$$\log Y = Y'$$

$$\log a_0 = a_0'$$

$$\log(1+a_1) = a_1'$$

$$\log \varepsilon = \varepsilon'$$

فإن:

$$Y' = a_0' + a_1' X + \varepsilon'$$

$$\hat{Y}' = \hat{a}_0' + \hat{a}_1' X$$

وبحساب قيم $\hat{a}_0' \, , \; \hat{a}_1'$ **وبالتالي** a_0', a_1' **نكون الجدول التالي:**

جدول (١٠-٩)

السنوات	Y	$\log Y = Y'$	X	X^2	$Y'X^2$
1972	**2.28**	**0.825**	**-3**	**9**	**-2.475**
1973	**3.00**	**1.099**	**-2**	**4**	**-2.198**
1974	**3.85**	**1.348**	**-1**	**1**	**-1.348**
1975	**5.000**	**1.609**	**0**	**0**	**0**
1976	**6.5**	**1.872**	**1**	**1**	**1.872**
1977	**8.46**	**2.135**	**2**	**4**	**4.27**
1978	**11**	**2.398**	**3**	**9**	**7.194**
المجموع		**11.286**	**0**		**7.315**

وبما أن المعادلات الطبيعية هي:

$$\sum_{j=1}^{n} Y_j' = n\hat{a}_0' + \hat{a}_1' \sum_{j=1}^{n} X_j$$

$$\sum_{j=1}^{n} Y_j' X_j = \hat{a}_0' \sum_{j=1}^{n} X_j + \hat{a}_1' \sum_{j=1}^{n} X_j^2$$

بالتعويض من الجدول السابق في المعادلتين المذكورتين أعلاه نحصل على المعادلتين التاليتين:

$$11.286 = 7\hat{a}_0' + 0 \rightarrow \hat{a}_0' = \frac{11.286}{7} = 1.61$$

$$7.315 = 0 + 28\hat{a}_1' \rightarrow \hat{a}_1' = \frac{7.315}{28} = 0.26$$

ونجد أن:

$$\hat{a}_0 = e^{\hat{a}_0'} \rightarrow \hat{a}_0 = e^{1.16} = 5.003$$

وكذلك

$$\hat{a}_1 = e^{\hat{a}_1'} - 1 \rightarrow \hat{a}_1 = e^{0.26} - 1 = 1.297 - 1 = 0.297$$

وبالتعويض في النموذج:

$$f_1(\hat{X}) = \hat{a}_0(1 + \hat{a}_1)^X$$

بقيم $\hat{a}_0$ ، $\hat{a}_1$ المحسوبة أعلاه نجد أن نموذج الاتجاه العام يصبح على النحو التالي:

$$f_1(\hat{X}) = 5.003(1 + 1.297)^X$$

وباستخدام هذا النموذج يمكن تقدير القيمة الاتجاهية لعدد السكان في هذه المدينـة في أي سنة.

ب) في سنة 1991 نجد أن:

X = 1991 – 1975 = 16

وبالتعويض بقيمة X في النموذج نجد أن:

القيمة الاتجاهية لعدد السكان في سنة 1991 هي:

$$\hat{f}_1(X) = 5.003(1+1.297)^X$$
$$= 5.003(1+1.297)^{16} = 320.83$$ ألف نسمة

(١٠-٤) الارتباط الذاتي Auto-Correlation

في الباب السابع تناولنا الفروض الأساسية التي تعتمد عليها طريقة المربعات الصغرى لحساب أفضل تقديرات (التقديرات المثلي) للعلاقة الخطية بين المتغير الأساسي X والتابع Y حيث:

$$Y_j = a_0 + a_1 X_j + \varepsilon_j \qquad , j = 1,2,\ldots,n$$

وكان أهم هذه الفروض هو افتراض أن مجموعة المتغيرات العشوائية ε متغيرات مستقلة ومتجانسة (أي متساوية التباين)، وتتبع التوزيع الطبيعي بتوقع صفر وتباين γ^2 أي أن:

$$E(\varepsilon) = 0 \;,\; V(\varepsilon) = \gamma^2 \;,\; \text{cov}(\varepsilon_j, \varepsilon_k) = 0 \qquad j \neq k$$

أو بعبارة أخري أنه ليس هناك ارتباط بين المتغيرات العشوائية ($\varepsilon_j \to j = 1,2,\ldots,n$).

وبالتالي إذا لم يتحقق هذا الفرض لم تصبح التقديرات المحسوبة باستخدام طريقة المربعات الصغرى من النموذج التالي:

$$\hat{Y}_j = \hat{a}_0 + \hat{a}_1 X_j \qquad , j = 1,2,3,\ldots,n$$

تقديرات مثلي.

لذلك يجب دراسة المتغيرات العشوائية ε_j (أو ما تسمي أحياناً بالأخطاء Errors) باستخدام النموذج من العلاقة التالية:

$$\varepsilon_j = (Y_j - \hat{Y}_j) \qquad , j = 1,2,3,\ldots,n$$

لتحديد هل المتغيرات $\varepsilon_j \to j = 1,2,\ldots,n$ مرتبطة ببعضها أم لا؟

وبالنسبة للسلاسل الزمنية يعتبر دراسة الارتباط بين المتغيرات $\varepsilon_j \to j = 1,2,\ldots,n$ هام، حيث وجود أخطاء في أحد السنوات **j** مثلاً ممكن أن يترتب عليه أخطاء أخرى تتراكم في السنوات التالية للسنة **j** طوال السلسلة الزمنية. وبالتالي التنبؤ بالقيم التقديرية للظاهرة في السنوات التالية للسنة **j**، وباستخدام النماذج التي تم بناءها باستخدام طريقة المربعات الصغرى تكون تقديرات متفائلة أكثر من اللازم.

فعلى سبيل المثال حدوث تضخم أو انكماش في الاقتصاد في أحد السنوات سوف يؤثر على الاقتصاد في السنوات التالية لهذه السنة طوال السلسلة الزمنية.

كذلك حدوث جفاف في أحد السنوات يؤثر على المنتج الزراعي في هذه السنة ويترتب عليه أخطاء تتراكم في السنوات التالية لهذه السنة.

ومن هنا كان من الأهمية دراسة المتغيرات العشوائية (الأخطاء) باستخدام النموذج لمعرفة هل هناك ارتباط بين قيم الظاهرة (المشاهدات) في الفترات الزمنية المتتالية أم لا.

ويسمي الارتباط بين المشاهدات لنفس المتغير بالارتباط الذاتي **Auto-Correlation**.

معامل الارتباط الذاتي

هو مؤشر باستخدامه يمكن معرفة هل هناك ارتباط بين المشاهدات لنفس المتغير أم لا.
فإذا كان ρ هو معامل ارتباط الذاتي فإن:

$$\hat{\rho} = 1 - \frac{(DW)}{2} \qquad \textbf{(10.22)}$$

حيث

$$(DW) = \frac{\sum_{j=1}^{n}(\varepsilon_j - \varepsilon_{j-1})^2}{\sum_{j=1}^{n}\varepsilon_j^2} \qquad \textbf{(10.23)}$$

ويمثل (DW) متغير عشوائي يتبع توزيع ديربن واتسون بدرجات حرية (1 , (n-1)) وبملحق رقم (١٠) التوزيع الاحتمالي لمتغير ديربن واتسون والذي يرمز له بتوزيع (D) حيث تنحصر قيم المتغير (DW) بين الصفر والأربعة. أي أن:

$$0 \le (DW) \le 4 \qquad \textbf{(10.24)}$$

كذلك

$$-1 \le \hat{\rho} \le +1 \qquad \textbf{(10.25)}$$

وبما ان (DW) تتراوح بين الصفر والأربعة. فبالتالي:

$$(DW) = 0 \rightarrow \hat{\rho} = 1 \qquad \textbf{(10.26)}$$

وهذا يعني أنه كلما اقتربت قيمة (DW) من الصفر يقترب معامل الارتباط الذاتي من (+1) أي أن هناك ارتباط ذاتي موجب.
كذلك إذا كان:

$$(DW) = 4 \rightarrow \hat{\rho} = -1 \qquad \textbf{(11.27)}$$

أي أنه كلما اقتربت قيمة (DW) من 4 يقترب معامل الارتباط الذاتي من (-1) أي أن هناك ارتباط ذاتي سالب.

مثال (١٠-٣)

الجدول التالي يوضح عدد الوحدات المباعة (بالألف وحدة) في الفترة الزمنية [1980-1990] في إحدى الشركات.

جدول (١٠-١٠)

السنوات	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990
عدد الوحدات	140	140	142	145	147	150	152	160	162	165	170

المطلوب:

أ) أوجد نموذج الاتجاه العام للوحدات المباعة.

ب) أوجد الخطأ المعياري للتقدير.

ج) أوجد معامل الارتباط الذاتي بين عدد الوحدات المباعة في الفترة محل الدراسة.

الحل:

أ) كون الجدول التالي:

جدول (١٠-١١)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
السنوات	Y	X	X^2	XY	$\hat{Y}$	ε_j	$(\varepsilon_j - \varepsilon_{j-1})$	$(\varepsilon_j - \varepsilon_{j-1})^2$	ε_j^2
1980	140	-5	25	-700	136.36	3.16	-	-	13.09
1981	140	-4	16	-560	139.53	0.47	-3.14	9.86	0.22
1982	142	-3	9	-546	142.67	-0.67	-1.14	1.29	0.45

1983	**145**	**-2**	**4**	**-290**	**145.81**	**-0.18**	**-1.4**	**1.96**	**0.66**
1984	**147**	**-1**	**1**	**-147**	**148.95**	**-1.95**	**-1.14**	**1.29**	**3.80**
1985	**150**	**0**	**0**	**0**	**152.09**	**2.09**	**-0.14**	**0.02**	**4.37**
1986	**152**	**+1**	**1**	**152**	**155.23**	**-3.23**	**-1.14**	**1.29**	**10.43**
1987	**160**	**+2**	**4**	**320**	**158.27**	**1.63**	**4.86**	**23.62**	**2.66**
1988	**162**	**+3**	**9**	**486**	**161.51**	**0.49**	**1.140**	**1.29**	**0.24**
1989	**165**	**+4**	**16**	**660**	**164.65**	**0.35**	**0.14**	**0.02**	**0.12**
1990	**170**	**+5**	**25**	**850**	**167.79**	**2.21**	**1.86**	**3.46**	**4.88**
المجموع	**1673**	**0**	**110**	**345**				**44.1**	**40.86**

بما أن المعادلات الطبيعية على النحو التالي:

$$\sum_{j=1}^{n} Y_j = n\hat{a} + \hat{b}\sum_{j=1}^{n} X_j$$

$$\sum_{j=1}^{n} Y_j X_j = \hat{a}\sum_{j=1}^{n} X_j + \hat{b}\sum_{j=1}^{n} X_j^2$$

بالتعويض من الجدول في المعادلات الطبيعية نجد أن:

$$1673 = 11\hat{a} \rightarrow \hat{a} = 152.09$$

$$345 = 110\hat{b} \rightarrow \hat{b} = 3.14$$

وبالتالي فإن نموذج الاتجاه العام يصبح على النحو التالي:

$$\hat{Y} = 152.09 + 3.14X$$

ب) ولحساب الخطأ المعياري للتقدير $\hat{\gamma}^2$ نعوض بقيم X في العمود رقم (3) بالجدول السابق في النموذج فنحصل على قيم $\hat{Y}$ المناظرة لها كما هو موضح بالعمود رقم (6) بالجدول.

وبما أن الخطأ المعياري للتقدير $\hat{\gamma}^2$ حيث:

$$\hat{\gamma}^2 = \frac{1}{n}\left[\left(\sum_j Y_j - \hat{Y}_j\right)^2\right] = \frac{1}{n}\left[\varepsilon_j^2\right] = \frac{1}{11}(40.86) = 3.72$$

ج) معامل الارتباط الذاتي $\hat{\rho}$ حيث:

$$\hat{\rho} = 1 - \frac{(DW)}{2}$$

وبما أن:

$$(DW) = \frac{\sum_{j=1}^{n}(\varepsilon_j - \varepsilon_{j-1})^2}{\sum_{j=1}^{n}\varepsilon_j^2} = \frac{44.1}{27.83} = 1.59$$

فإن:

$$\hat{\rho} = 1 - \frac{(DW)}{2} = 1 - \frac{1.59}{2} = 0.205$$

أي انه يوجد ارتباط ذاتي طردي ضعيف حيث القيمة (0.205) أقل للصفر من (+1)

(١٠-٥) اختبار ديربن واتسون

Durbin and Watson Test

إذا كان لدينا بيانات لسلسلة زمنية والمطلوب اختبار وجود ارتباط ذاتي بين مشاهدات الظاهرة أو لا. فيوجد العديد من الاختبارات لتحديد ما إذا كان يوجد ارتباط ذاتي بين قيم المفردات أم لا. ويعتبر اختبار ديربن-واتسون أكثر استخداماً. ويتم الاختبار على النحو التالي:

١- الفرض العدمي $\hat{\rho} = 0$

الفرض البديل $\hat{\rho} \neq 0$

٢- نحسب قيمة المقياس **(DW)** حيث:

$$(DW) = \frac{\sum_{j=1}^{n} (\varepsilon_j - \varepsilon_{j-1})^2}{\sum_{j=1}^{n} \varepsilon_j^2}$$

٣- نحدد قيمة كل من f_1 ، f_2 من جدول (ديربن - واتسون) عند درجة الثقة $(1-\alpha)$ وعدد درجات الحرية **(n-2 , 1)** بملحق رقم (١٠).

٤- وبحساب القيمة **(DW)** يكون أمامنا إحدى الحالات التالية:

أ) إذا كانت القيمة (**DW**) بحيث:

$$\boxed{4 - f_1 < DW < 4 \qquad \textbf{(10.28)}}$$

وهذا يعني وجود ارتباط ذاتي سالب (وبالتالي يرفض الفرض العدمي ويقبل البديل).

ب) إذا كانت القيمة (**DW**) بحيث:

$$\boxed{0 < DW < f_1 \qquad \textbf{(10.29)}}$$

وهذا يعني وجود ارتباط ذاتي موجب (وبالتالي يرفض الفرض العدمي ويقبل الفرض البديل).

ج) أما إذا كانت القيمة (**DW**) بحيث:

$$\boxed{2 < DW < 4 - f_2 \qquad \textbf{(10.30)}}$$

وهذا يعني عدم وجود ارتباط ذاتي (أي قبول الفرض العدمي ورفض الفرض البديل).

د) أما إذا كانت القيمة (**DW**) بحيث:

$$f_2 < DW < 2 \qquad \textbf{(10.31)}$$

وهذا يعني عدم وجود ارتباط ذاتي (أي قبول الفرض العدمي ورفض الفرض البديل).

هـ) أما إذا كانت القيمة (**DW**) بحيث:

$$4 - f_2 < DW < 4 - f_1 \qquad \textbf{(10.32)}$$

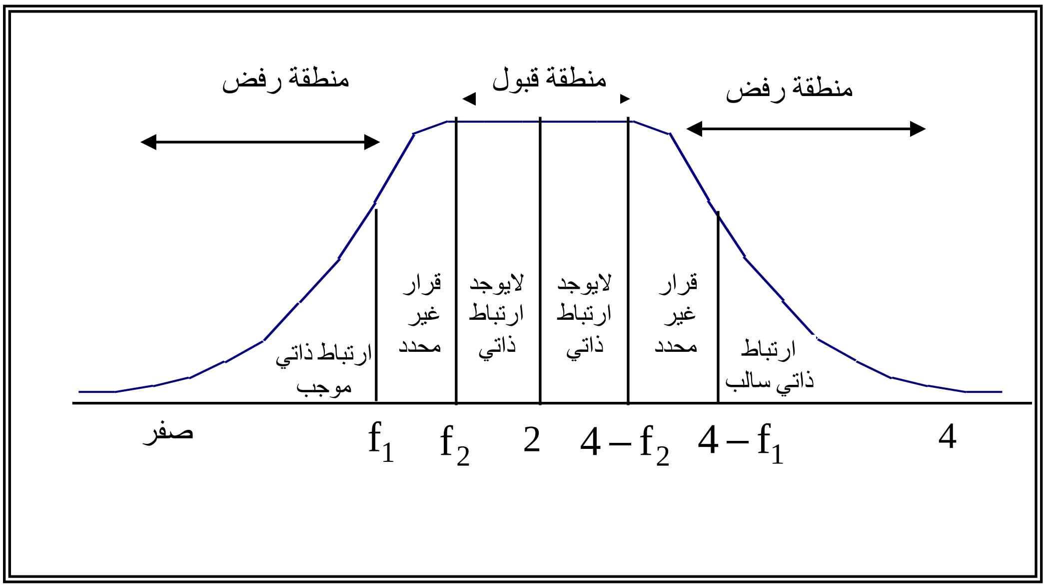
وهذا يعني عدم كفاية البيانات لتحديد هل هناك ارتباط ذاتي أم لا. وبالتالي تتطلب الدراسة أخذ سلسلة زمنية أطول وفي هذه الحالة لا نستطيع أخذ قرار محدد.

و) كذلك إذا كان:

$$f_1 < DW < f_2 \qquad \textbf{(10.33)}$$

فهذا يعني عدم كفاية البيانات لتحديد هل هناك ارتباط أم لا. وبالتالي تتطلب الدراسة أخذ سلسلة زمنية أطول وفي هذه الحالة أيضاً لا نستطيع أخذ قرار محدد.

والشكل التالي يوضح الحالات الست المذكورة أعلاه.



شكل (١٠-٣)

مثال (١١-٤)

الجدول التالي يوضح الأرباح المحققة (بالألف جنية) لإحدى الشركات خلال الفترة [1973-1989].

جدول (١٠-١٢)

السنوات	1973	1974	1975	1976	1977	1978	1979	1980
الأرباح بالألف جنية	1	2	3.5	4.5	5.5	6	8.2	8

تابع جدول (١٠-١٢)

1981	1982	1983	1984	1985	1986	1987	1988	1989
10	10.5	11	11.5	13.5	14.5	15	16	18

المطلوب:

- **هل هذه البيانات تكفي لاتخاذ قرار بوجود ارتباط ذاتي أم لا. بدرجة ثقز 95%.**

الحل:

١- إذا فرضنا أن العلاقة الاتجاهية للظاهرة مع الزمن على النحو التالي:

$$Y = a + bX + \varepsilon$$

فباستخدام طريقة المربعات الصغرى نجد أن:

$$\hat{Y} = \hat{a} + \hat{b}X$$

حيث يمكن حساب قيم $\hat{a}$ ، $\hat{b}$ **من المعادلات الطبيعية ولتكوين المعادلات الطبيعية نكون الجدول التالي:**

جدول (١٠-١٣)

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
السنوات	Y	X	XY	X^2	$\hat{Y}$	ε_j	$(\varepsilon_j - \varepsilon_{j-1})$	$(\varepsilon_j - \varepsilon_{j-1})^2$	ε_j^2

1973	1	-8	-8	64	1.23	-0.23	-	-	0.0529
1974	2	-7	-14	49	2.24	-0.4	-0.01	0.0001	0.0579
1975	3.5	-6	-21	36	3.25	0.25	0.49	0.2041	0.0625
1976	4.2	-5	-21	25	4.26	-0.06	-0.31	0.0961	0.0036
1977	5.50	-4	-22	16	5.27	0.23	0.29	0.0841	0.0529
1978	5	-3	-18	9	6.28	0.28	-0.51	0.2691	0.0784
1979	8.2	-2	-16.4	4	7.29	0.91	1.19	1.4161	0.8281
1980	8	-1	-8	1	8.3	-0.3	-1.21	1.4641	0.09
1981	10	0	0	0	9.31	0.69	0.99	0.9801	0.4761
1982	10.5	1	10.5	1	10.32	0.18	-0.51	0.2601	0.0324
1983	11	2	22	4	11.33	-0.33	-0.51	0.2601	0.1089
1984	11.5	3	34.5	9	12.34	-0.84	-0.51	0.2601	0.7056
1985	13.5	4	54	16	13.35	-0.15	0.99	0.9801	0.0225
1986	14.3	5	71.5	26	14.36	-0.06	-0.09	0.0081	0.0036
1987	15	6	90	36	15.37	-0.37	-0.31	0.0961	0.1369
1988	16	7	1122	49	16.38	-0.38	-0.01	0.0001	0.1444
1989	18	8	1144	64	17.39	0.16	0.99	0.9801	0.3721
المجموع	158.2	0	410.1	408				7.39	3.23

وبالتعويض من الجدول في المعادلات الطبيعية نحصل على المعادلتين التاليتين:

$$158.2 = 17\hat{a} \rightarrow \hat{a} = 9.31$$

$$410.1 = 408\hat{b} \rightarrow \hat{b} = 1.01$$

وبالتالي يصبح نموذج الاتجاه العام على النحو التالي:

$$\hat{Y} = 9.311 + 1.01X$$

٢- وبالتعويض المتتالي بقيم X في النموذج نحصل على قيم Y المناظرة كما هو موضح بجدول (١١-١٣).

٣- نحسب قيم ε_j ، $(\varepsilon_j - \varepsilon_{j-1})$ كما هو موضح بالجدول.

٤- نحسب القيمة (DW) على النحو التالي:

$$(DW) = \frac{\sum_{j=1}^{n}(\varepsilon_j - \varepsilon_{j-1})^2}{\sum_{j=1}^{n}\varepsilon_j^2} = \frac{7.39}{3.27} = 2.26$$

٥- عند درجة الثقة 95% ودرجات الحرية (1 , 15) نجد أن $f_1 = 1.08$ ، $f_2 = 1.36$ من جدول (ديربن - واتسون).

وبما ان قيمة (DW) تساوي 2.26 أي أن:

$$2 < DW < 4 - f_1$$
$$2 < 2026 < 4 - 1.08$$

وهذا يعني أن قيمة (DW) تقع في منطقة الرفض وبالتالي لا يوجد ارتباط ذاتي بين قيم المشاهدات وذلك بدرجة ثقة 95%.

(١٠-٦) تمرينات
Exercises

(١٠-١) : الجدول التالي يوضح قيمة الصادرات لإحدى الشركات بالألف جنية سنوياً خلال الفترة [1990-1994].

جدول (١٠-١٣)

السنوات	1990	1991	1992	1993	1994
قيمة الصادرات بالألف جنية	120	170	200	250	350

المطلوب:

١- ارسم شكل الانتشار.

٢- بافتراض العلاقة الخطية للاتجاه العام لتطور قيمة الصادرات خلال الفترة [1990-1994] – كون نموذج انحدار مناسب للاتجاه العام لقيمة الصادرات.

٣- هل يوجد ارتباط ذاتي بين المشاهدات – وإذا وجد حدد نوع العلاقة بين الزمن وقيمة الصادرات.

٤- قدر قيمة الصادرات في السنوات التالية: 1995، 2000، 2005.

(١٠-٢) : الجدول التالي يوضح معدلات النمو في الألف السنوي للسكان خلال الفترة من [1990-1995] لإحدى المدن في إحدى الدول.

جدول (١٠-١٤)

السنوات	1990	1991	1992	1993	1994	1995
معدل النمو	1	1.5	4	5	6	9

بافتراض أن العلاقة الاتجاهية بين معدل النمو في الألف والزمن علاقة أسية على النحو التالي:

$$Y = ab^{X}\varepsilon$$

حيث ε متغير عشوائي وX تشير انحرافات السنوات عن سنة الأساس، وكل من b , a معلمات.

المطلوب:

أ) كون نموذج انحدار مناسب.

ب) قدر معدل النمو في 1998، 2005.

ج) اجري اختبار ديربن واتسون لاختبار الفرض القائل بأنه لا يوجد ارتباط ذاتي بين مشاهدات السلسلة الزمنية.

(١٠-٣) : في المثال (١٠-١) هل يوجد ارتباط ذاتي بين قيم الإيرادات للشركة في السنوات المتتالية. وذلك بدرجة ثقة 90%.

(١٠-٤) : في المثال (١٠-٣) هل بيانات السلسلة الزمنية خلال الفترة [1972- 1978] كافية لتمكن متخذ القرار من تحديد هل يوجد ارتباط ذاتي أم لا. وذلك بدرجة ثقة 95%.

الملاحق
Appendices

ملحق (١): مقدمة في الاحتمالات

ملحق (٢): ملخص لبعض التوزيعات الاحتمالية

ملحق (٣): الأعداد العشوائية

ملحق (٤): التوزيع المعتاد القياسي

ملحق (٥): توزيع مربع كا (χ^2)

ملحق (٦): توزيع استيودنت (**T**)

ملحق (٧): توزيع ذات الحدين

ملحق (٨): توزيع فيشر (**F**)

ملحق (٩): توزيع كولومجروف سيمزوف (**K**)

ملحق (١٠): توزيع ديربن واتسون (**D**)

ملحق (١١): التوزيع الأسي

ملحق (١٢): توزيع ذات الحدين

ملحق (١)

مقدمة في الاحتمالات

فـي هـذا البـاب سـوف نتنـاول بالدراسـة المفـاهيم الأساسـية والضـرورية لنظريـة الاحتمالات Probability Theory ومجموعة القواعد Rules المرتبطة بها والتي يجب الإلمام بها عند دراسة الظواهر التي لا تتوافر معرفة كاملة عن خصائصها هذا من ناحية. ومن ناحية أخرى فإن محتويات هذا الباب تمثل عنصر أساسي في دراسة الأبواب التالية سواء من الناحية النظرية أو الناحية التطبيقية. كذلك يتضمن هذا الباب مجموعة من الأمثلة التطبيقية التي توضح استخدام نظرية الاحتمالات في حل العديد من المشاكل الفعلية وكيفية الوصول إلى قرار صحيح بشأنها.

وفيما يلي سوف نقدم بعض المفاهيم Concepts الضرورية لدراسة الاحتمالات. ومما هو جدير بالذكر أن دراسة نظرية الاحتمالات مرتبطة بنظرية الفئات* Set's Theory.

(١) المتغير **Variable : هو الشيء الذي يمكن أن يأخذ قيم مختلفة في الظروف المختلفة (زمنية، مكانية، سياسية، اقتصادية، ... الخ) فمثلاً سعر كيلو الدقيق يختلف من سوق إلى أخر، ويختلف في نفس السوق من عام إلى عام أخر، ...، كذلك وزن الفرد يمثل متغير يتوقف على عمر الفرد، وطوله، ... الخ. وعادة يرمز للمتغير الذي يأخذ عدد n من القيم بالرمز X حيث X تأخذ القيم X_i حيث $i = 1,2,3,...,n$.**

كذلك الشيء الثابت Constant هو مقدار لا يتغير بتغير الظروف وعادة يرمز له بالرمز C وإذا كان لدينا متغيران Y , X تربط بينهما العلاقة (الدالة) f بحيث:

$$Y = f(X) \qquad \textbf{(A.1)}$$

* أ.د عفاف الدش (١٩٩٤): " الرياضيات وصناعة القرار" – مكتبة عين شمس – القاهرة برقم إيداع ٩٤/٨٦٨٢.

فـإن المتغيـر X يمثـل نطـاق الدالــة f ويسـمي بـالمتغير المسـتقل Independent Variable ويســمي المتغيــر Y بمــدي الدالـــة أو المتغيــر التـــابع Dependent Variable.

والمتغير الـذي يمكن أن يأخذ أي قيمـة بـين قيمتـين معينتـين يسمي متغيـر متصـل Continuous Variable، وخلاف ذلك يسمي المتغير بالمتغير المتقطـع Discrete Variable.

مثال (١)

إذا كان X متغير يمثل عدد الأطفال الممكن إنجـابهم فـي الأسـرة، فـإن X ممكـن أن تأخذ القيم 0 ، 1 ، 2 ، 3 ، 4 ، ..... ، وتكتب على النحو:

$$X = 0 , 1 , 2 , 3 , 4 , \dots , 10$$

ولا يمكن أن تأخذ X القيمـة 1.5 ، 2.31 وبالتـالي فـإن X فـي هـذه الحالـة يمثـل متغيـر متقطع.

مثال (٢)

إذا كان المتغير Y يمثل عمر الفرد فمن الممكن أن تأخذ Y القيمـة 15 سـنة، 15.2 سنـة، 15.25 سنـة. وبالتـالي فـإن المتغيـر Y يمكـن أن يأخـذ أي قيمـة فـي فتـرة زمنيـة معينـة.

وبالتالي فإن Y يمثل متغير متصل.

مثال (٣)

إذا كان X حيث:

$$X = \{X \mid X \in I \quad , \quad 1 \le X \le 10\} \qquad \textbf{(A.2)}$$

حيث I تمثل فئـة الأعداد الصحيحة فإن: $X = 1 , 2 , 3 , 4 , \dots , 10$

أي أن X متغير متقطع.

أما إذا كان المتغير Z بحيث:

$$Z = \{Z \mid Z \in R \quad , 1 \le Z \le 10\} \qquad \textbf{(A.3)}$$

حيث R تمثل فئـة الأعداد الحقيقية، وبالتالي فإن المتغير Z يمكن أن يأخـذ أي قيمـة بين 1 ، 10 مثل 1.5 ، 2.4 ، ... الخ، وبالتالي فإن Z يمثل متغير متصل.

(٢) التجربـة العشـوائية **Random Experiment : هـي التجربـة التـي يعرف مقدماً جميع النتـائج الممكنـة لهـا، ولكـن لا يمكـن أن يعرف مقدماً ترتيب**

حدوث هذه النتائج. ويمثل نتائج Outcomes التجربة العشوائية متغير يسمي بالمتغير العشوائي Random Variable حيث أن الصدفة (الصدفة هي عبارة عن مجموعة من العوامل التي لا يمكن التحكم فيها) هي التي تحدد نتائج التجربة.

فمثلاً عند إلقاء قطعة عملة متوازنة فإننا نعرف قبل الرمي أن نتيجة هذه التجربة إما صورة ونرمز لها بالرمز X_1 وإما كتابة ونرمز لها بالرمز X_2. ولكن أيهما سيقع أولاً لا يمكن معرفة ذلك قبل الرمي. وإذا رمزنا لنتائج التجربة السابقة بالرمز X فإن X تمثل متغير عشوائي يأخذ القيم X_1 ، X_2. أي أن:

$$X = X_1 , X_2 \qquad \textbf{(A.4)}$$

كذلك عند إلقاء زهرة نرد (طاولة) متوازنة فإننا نعرف قبل الرمي أن نتائج هذه التجربة إما أن تكون النتيجة الرقم 1 ، 2 ، 3 ، 4 ، 5 ، 6 وبذلك يعتبر إلقاء هذه الزهرة تجربة عشوائية. فإذا رمزنا لنتائج هذه التجربة بالرمز X فإن X يمثل متغير عشوائي بحيث:

$$\mathbf{X = 1 , 2 , 3 , 4 , 5 , 6} \qquad \textbf{(A.5)}$$

(٣) فراغ المعاينة **Sampling Space : هو الفئة التي تتكون من جميع النتائج الممكنة للتجربة العشوائية. حيث يمثل كل عنصر في هذه الفئة نتيجة من نتائج التجربة العشوائية. وأحياناً تسمي هذه الفئة بفراغ النتائج* Outcome Space أو فراغ الاحتمال Probability Space. ويرمز عادة لفراغ المعاينة بالرمز S فمثلاً فراغ المعاينة لرمي قطعة عملة في (A.4) هي:**

$$S = \{X_1 , X_2\} \qquad \textbf{(A.6)}$$

كذلك فراغ المعاينة لرمي زهرة النرد في (A.5) هي:

$$S = \{1 , 2 , 3 , 4 , 5 , 6\} \qquad \textbf{(A.7)}$$

وبالتالي فإن فراغ المعاينة هو الفئة الشاملة Universal Set لنتائج التجربة العشوائية.

* Heinz Kohler (1994): " Statistics for Business and Economics ", Harper Collins Publishers, New York.

(٤) الأحداث العشوائية **Random Events : الحدث العشوائي هو فئة جزئية Subset من فراغ المعاينة. فقد يكون الحدث فئة مكونة من عنصر واحد (أي نتيجة واحدة) أو أكثر من عنصر.**
فعند رمي قطعة العملة، إذا كان الحدث A يمثل ظهور صورة فإن:

$$A = \{X_1\} \qquad \textbf{(A.8)}$$

كذلك عند رمي زهرة النرد، فإذا كان الحدث B يمثل ظهور رقم فردي فإن:

$$B = \{1, 3, 5\} \qquad \textbf{(A.9)}$$

فنجد أن الحدث A يمثل فئة جزئية من فراغ المعاينة S في (A.6) هو فئة تحتوي على عنصر واحد. كذلك الحدث B في (A.9) هو فئة تحتوي على ثلاثة عناصر وهو فئة جزئية من فراغ المعاينة S في (A.7). ويسمي كل من الحدث A، والحدث B بالأحداث البسيطة Simple Events حيث العناصر (النتائج) الموجودة في كل حدث تعتبر عناصر (نتائج) بسيطة. ويوجد نوع اخر من الأحداث العشوائية تسمي بالأحداث المركبة Compound Events حيث تتكون هذه الأحداث من عناصر (أو نتائج) مركبة كما سوف يتضح من المثال التالي.

مثال (٤)
إذا رميت قطعتي عملة متوازنتين بحيث تشير X_1 للصورة، وتشير X_2 للكتابة فإن فراغ المعاينة لهذه التجربة S يصبح:

$$S = \{(X_1, X_1), (X_1, X_2), (X_2, X_1), (X_2, X_2)\} \qquad \textbf{(A.10)}$$

فنجد أن كل عنصر في الفئة S في (A.10) عنصر (نتيجة) مركبة. فالعنصر الواحد مكون من النتيجتين نتيجة على القطعة الأولي وأخرى على القطعة الثانية. فمثلاً العنصر (X_1, X_1) يشير إلى أن النتيجة على القطعة الأولي صورة والنتيجة على القطعة الثانية صورة أيضاً، في حين أن العنصر (X_2, X_1) يشير إلى ظهور الكتابة على القطعة الأولي والصورة على القطعة الثانية. ويمكن أن يتكون العنصر الواحد (أو النتيجة الواحدة) المركب من أكثر من نتيجتين.

مثال (٥)
في إحدى شركات إنتاج المعلبات الغذائية تم فحص 3 وحدات لتحديد أي عبوة صالحة للاستخدام وأي عبوة غير صالحة. والمطلوب:

١- حدد فراغ المعاينة ووضح ذلك من خلال شكل فإن.
٢- أوجد الأحداث التالية:
- **جميع الوحدات صالحة.**
- **وحدة واحدة على الأكثر فاسدة.**
- **جميع الوحدات فاسدة.**

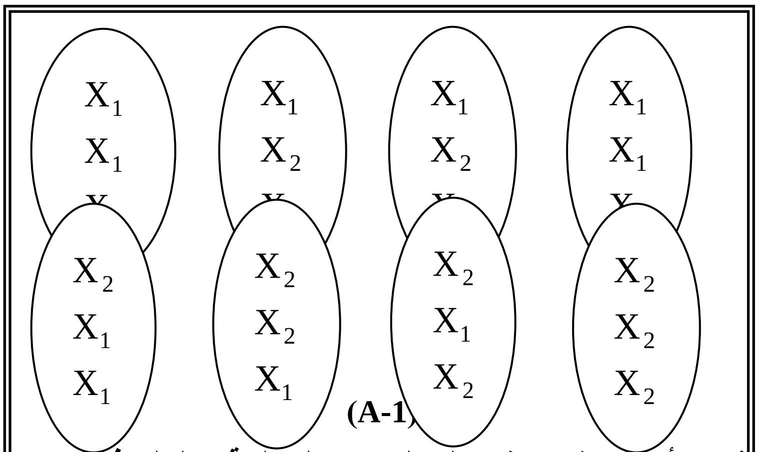
الحل:

(١) إذا افترضنا أن X_1 ، X_2 تشير إلى العبوة الصالحة والعبوة غير الصالحة على الترتيب، كذلك S تشير إلى فراغ المعاينة. وفي هذه الحالة نجد أن عدد العناصر في فراغ المعاينة يساوي ($2^3 = 8$) عناصر مركبة على النحو التالي:

$$S = \begin{Bmatrix} (X_1, X_1, X_1), (X_1, X_1, X_2), (X_1, X_2, X_1) \\ (X_1, X_2, X_2), (X_2, X_1, X_1), (X_2, X_1, X_2) \\ (X_2, X_2, X_1), (X_2, X_2, X_2) \end{Bmatrix} \qquad \textbf{(A.11)}$$

وشكل فإن التالي يوضح فراغ المعاينة S.

S



(A-1)

(٢) – نفرض أن الحدث A يشير إلى الوحدات الصالحة وبالتالي فإن:

$$A = \{(X_1, X_1, X_1)\} \qquad \textbf{(A.12)}$$

فنجد أن الفئة A تتكون من عنصر مركب واحد.

- **كذلك إذا فرضنا أن B تشير إلى الحدث وجود وحدة واحدة على الأكثر فاسدة فإن الحدث B على النحو التالي:**

$$B = \{(X_1, X_1, X_1), (X_2, X_1, X_1), (X_1, X_2, X_1), (X_1, X_1, X_2)\} \qquad \textbf{(A.13)}$$

فنجد أن الحدث B يتكون من 4 عناصر مركبة كل عنصر مركب من ثلاثة عناصر بسيطة.

- **بالمثل إذا فرضنا أن C تشير إلى الحدث "جميع الوحدات فاسدة" فإن:**

$$C = \{(X_2, X_2, X_2)\} \qquad \textbf{(A.14)}$$

ووقوع الحدث العشوائي يعني ظهور أحد عناصر فئة الحدث كناتج للتجربة العشوائية، كما أن عدم وقوع الحدث يعني عدم ظهور أي عنصر من عناصر فئة الحدث كنتيجة للتجربة وبالتالي فهو يعني وقوع الحدث المكمل له Complement Event. فمثلاً عدم وقوع الحدث B يعني وقوع الحدث B' (أي وجود وحدة صالحة على الأكثر) حيث:

$$B' = \{(X_2, X_2, X_2), (X_1, X_2, X_2), (X_2, X_1, X_2), (X_2, X_2, X_1)\} \qquad \textbf{(A.15)}$$

وممن سبق يتضح أنه إذا كانت a_1، a_2، ...، a_n هي عناصر فئة الحدث A. فإن وقوع الحدث A يعني وقوع واحد على الأقل من الأحداث الجزئية $\{a_1\}$، $\{a_2\}$، ...، $\{a_n\}$ أو بعبارة أخري

$$A = \{a_1\} \cup \{a_2\} \cup ..... \cup \{a_n\} = \bigcup_{i=1}^{n} \{a_i\} \qquad \textbf{(A.16)}$$

(٥) الحوادث المتنافية **Exclusive Events : يقال أن الحدثين A، B حدثين متنافيين إذا كان وقوع إحداهما يمنع وقوع الأخر. فمثلاً عند رمي قطعة عملة متوازنة، إذا ظهر الوجه الذي علية الصورة فإن هذا يؤدي إلى منع ظهور الوجه الذي علية الكتابة وبالتالي فإن حدث ظهور الصورة وحدث ظهور الكتابة يمثلان حدثين متنافيين. كذلك وجود شخص ما في مكان معين في لحظة معينة يمنع وجوده في مكان أخر في نفس اللحظة الزمنية، وبالتالي فوجود الشخص في مكانيين مختلفين في نفس اللحظة الزمنية يمثل حدثين متنافيين.**

وبالتالي يكون الحدثين A، B حدثين متنافيين إذا كان:

$$A \cap B = \phi \qquad \textbf{(A.17)}$$

حيث ϕ تمثل الفئة الخالية.

(٦) الحوادث المستقلة: **وقوع الحدث A لا يمنع ولا يؤثر على وقوع الحدث B، كذلك وقوع الحدث B لا يمنع ولا يؤثر على وقوع الحدث A. فإنه يقال أن الحدثين A، B حدثين مستقلين.**

ويمكن تعميم ذلك، بمعني بمعنى إذا كان لدينا الأحداث A_i، i=1,2,3,…,n بحيث وقوع أي حدث A_i لا يمنع أو يؤثر على وقوع أي حدث أخر A_j بحيث $i \neq j$ فإنه يقال أن الحدثين A_i، A_j حدثين مستقلين.

ومن (٥) ، (٦) يتضح أن مفهوم الحوادث المتنافية يختلف عن مفهوم الحوادث المستقلة.

(٧) الاحتمال Probability: إذا كانت m تمثل عدد مرات وقوع الحدث A في N من المحاولات (التجارب) المتماثلة، وعرفنا أن الدالة $P_r(A)$ حيث $P_r(A)$ تشير إلى احتمال وقوع الحدث A فإن:

$$P_r(A) = \lim_{N \to \infty} \frac{m}{N} \qquad \textbf{(A.18)}$$

والمقصود * بكلمة $\lim_{N \to \infty}$ ليس المعني الرياضي المعروف ولكن المقصود هو أن قيمة هذا الاحتمال سوف تقترب من الصحة إذا كان عدد المحاولات كبيراً جداً. ويسمي الاحتمال المعروف في (A.18) بالاحتمال التجريبي Experimental Probability.

مثال (٦)

إذا قذفت قطعة عملة متوازنة عشوائياً 100000 مرة، وكانت الأحداث X_1 ، X_2 تشير إلى ظهور الصورة والكتابة على الترتيب. فإذا كانت نتيجة القذف 5720 مرة ظهور الصورة، 4280 مرة ظهور الكتابة. أوجد احتمال ظهور الصورة $P_r(X_1)$ واحتمال ظهور الكتابة $P_r(X_2)$.

$$P_r(X_1) = \frac{5720}{100000} = 0.572 \qquad \textbf{(A.19)}$$

$$P_r(X_2) = \frac{4280}{100000} = 0.428 \qquad \textbf{(A.20)}$$

كذلك إذا أجريت تجربة عشوائية وكان عدد النتائج الممكنة لهذه التجربة N (عدد عناصر فراغ المعاينة) بحيث كان لكل نتيجة للتجربة نفس الفرصة في الظهور. فإذا كان m من هذه النتائج تؤدي إلى وقوع الحدث A فإنه يمكن حساب احتمال وقوع الحدث A كما يلي:

$$P_r(A) = \frac{m}{N} \qquad \textbf{(A.21)}$$

وبعبارة أخري فإن احتمال وقوع الحدث A هو $P_r(A)$ حيث:

عدد الحالات المواتية

---

* أ.د سمير كامل عاشور (١٩٨٦): " الإحصاء التحليلي "، معهد الدراسات والبحوث الإحصائية – جامعة القاهرة.

$P_r(A) =$ ——————— **(A.22)**

حيث أن الحالات المواتية هي النتائج المؤدية إلى الـ عدد الحالات الممكنة الممكنـة هي ~~عناصر فراغ المعاينة.~~

مثال (٧)

إذا رميت زهرة نرد (طاولة) متوازنـة عشـوائياً، احسب احتمـال ظهور رقم فردي، وكذلك احتمال ظهور رقم زوجي.

بما أن فراغ المعاينة في هذه الحالة هو: $S = \{1 , 2 , 3 , 4 , 5 , 6\}$

وبالتالي فإن $N = 6$. فإذا كان A، B تشير إلى حدث ظهور رقم فردي وحدث ظهور رقم زوجي على الترتيب. وبالتالي فإن:

$$A = \{1 , 3 , 5\} \quad , \quad B = \{2 , 4 , 6\}$$

وبالتالي فإن:

$$P_r(A) = \frac{3}{6} = \frac{1}{2} \quad , \quad P_r(B) = \frac{3}{6} = \frac{1}{2}$$

مثال (٨)

الجدول التالي يوضح توزيع 5000 أسرة من ذو الدخل المتوسط حسب عدد الغرف التي تسكنها الأسرة الواحدة.

جدول (١)

عدد الغرف التي تسكنها الأسرة الواحدة	1	2	3	4	المجموع
عدد الأسر	500	3000	1000	500	5000

إذا فرضنا أن X تشير إلى عدد الغرف التي تسنها الأسرة الواحدة. أي أن
$X = 1 , 2 , 3 , 4$

وبما أن عدد الأسر الكلي يسـاوي 5000 حالـة وبالتـالي فـإن عدد الحـالات الممكنـة يساوي 5000 حالة. وبالتالي فإن احتمال أن تسكن الأسرة غرفة واحدة هو:

$$P_r(X = 1) = \frac{\text{عدد الحالات المواتية}}{\text{عدد الحالات الممكنة}} = \frac{500}{5000} = 0.1$$

وبالمثل احتمال أن تسكن الأسرة غرفتين هو $P_r(X = 2)$ حيث:

$$P_r(X = 2) = \frac{3000}{5000} = 0.6$$

واحتمال أن تسكن الأسرة ثلاثة غرف هو $P_r(X=3)$ حيث:

$$P_r(X=3)=\frac{1000}{5000}=0.2$$

واحتمال أن تسكن الأسرة أربعة غرف هو $P_r(X=4)$ حيث:

$$P_r(X=4)=\frac{500}{5000}=0.1$$

مثال (٩)

إذا ألقي زهرتي نرد عشوائياً، أوجد ما يلي:

١- فراغ المعاينة لتجربة الرمي.

٢- احتمال أن يكون الرقم على الزهرة الأولي يساوي الرقم على الزهرة الثانية.

٣- احتمال أن يكون مجموع الرقمين على الزهرتين مساوي 4، كذلك احتمال أن يكون المجموع أقل من 4.

٤- احتمال أن يكون مجموع الرقمين على الزهرتين أكبر من 9.

الحل:

الجدول التالي يوضح جميع النتائج الممكنة لرمي زهرتي النرد.

جدول (٢)

نتائج الزهرة الأولي نتائج الزهرة الثانية	1	2	3	4	5	6
1	(1 , 1)	2 ,) (1	(3 , 1)	(4 , 1)	5 ,) (1	(6 , 1)
2	(1 , 2)	2 ,) (2	(3 , 2)	(4 , 2)	5 ,) (2	(6 , 2)
3	(1 , 3)	2 ,) (3	(3 , 3)	(4 , 3)	5 ,) (3	(6 , 3)
4	(1 , 4)	2 ,) (4	(3 , 4)	(4 , 4)	5 ,) (4	(6 , 4)
5	(1 , 5)	2 ,) (5	(3 , 5)	(4 , 5)	5 ,) (5	(6 , 5)
6	(1 , 6)	2 ,) (6	(3 , 6)	(4 , 6)	5 ,) (6	(6 , 6)

من الجدول يتضح أن عدد النتائج الممكنة $N = 36$ نتيجة (أو حالة) وهي عبارة عن نتائج (أو حالات) مركبة كل نتيجة مكونة من نتيجتين (النتيجة على الزهرة الأولي، والنتيجة على الزهرة الثانية). وبالتالي فإن فراغ المعاينة S يأخذ الشكل التالي:

$$S = \{ (1 , 1) , (1 , 2) , (1 , 3) , ... ,(1 , 6) , (2 , 1) , ... , (6 , 6)\}$$

٢- إذا فرضنا أن الحدث A يمثل أن تكون نتيجة الرمي "الرقم على الزهرة الأولي يساوي الرقم على الزهرة الثانية" فإن:

$$A = \{ (1 , 1) , (2 , 2) , (3 , 3) , (4 , 4) , (5 , 5) , (6 , 6) \}$$

ونلاحظ أن عدد العناصر في الفئة (A) يساوي عدد الحالات المواتية (m) يساوي 6 عناصر، وبالتالي فإن احتمال أن يكون الرقم على الزهرة الأولي مساوي للرقم على الزهرة الثانية يساوي $P_r(A)$ حيث:

$$P_r(A) = \frac{m}{N} = \frac{6}{36} = \frac{1}{6}$$

٣- إذا فرضنا أن الحدث B يمثل "أن يكون مجموع الرقمين على الزهرتين مساوي4" فإن:

$$B = \{ (1 , 3) , (3 , 1) , (2 , 2) \}$$

وبالتالي فإن احتمال أن يكون مجموع الرقمين على الزهرتين مساوي 4 هو $P_r(B)$ حيث:

$$P_r(B) = \frac{m}{N} = \frac{3}{36} = \frac{1}{12}$$

كذلك إذا فرضنا أن الحدث (C) هو "أن يكون مجموع الرقمين على الزهرتين أقل من او يساوي 4" فإن:

C = { (1 , 1) , (2 , 1) , (1 , 2) , (2 , 2) , (1 , 3) , (3 , 1) }

وبالتالي فإن احتمال أن يكون مجموع الرقمين على الزهرتين أقل من أو يساوي 4 هو $P_r(C)$ حيث:

$$P_r(C) = \frac{m}{N} = \frac{6}{36} = \frac{1}{6}$$

٤- بالمثل إذا كان الحدث H هو "أن يكون مجموع الرقمين على الزهرتين أكبر من9 " فإن:

H = { (5 , 5) , (6 , 5) , (5 , 6) , (6 , 6) }

وبالتالي فإن احتمال أن يكون مجموع الرقمين على الزهرتين أكبر من 9 هو $P_r(H)$ حيث:

$$P_r(H) = \frac{m}{N} = \frac{4}{36} = \frac{1}{9}$$

٥- إذا فرضنا أن الحدث (T) هو "أن يكون حاصل ضرب الرقمين على الزهرتين رقم فردي أقل من 10" فإن:

T = { (1 , 1) , (3 , 1) , (1 , 3) , (5 , 1) , (1 , 5) ,
(1 , 7) , (7 , 1) , (1 , 9) , (9, 1) }

وبالتالي فإن احتمال ان يكون حاصل ضرب الرقمين على الزهرتين رقم فردي أقل من 10 هو $P_r(T)$ حيث:

$$P_r(T) = \frac{m}{N} = \frac{9}{26} = \frac{1}{4}$$

قوانين الاحتمالات **Laws of Probabilities**

من التعريف السابق للاحتمال اتضح أنه يمكن إيجاد قيمة احتمال وقوع الحدث بقسمة عدد العناصر في فئة الحدث محل الدراسة (m) على عدد العناصر في فئة فراغ المعاينة (N).

ومن العمليات على الفئات Set’s Operations فقد يكون الحدث محل الاعتبار ناتج من عملية على حدثين أو أكثر، فعلي سبيل المثال قد يكون الحدث محل الاعتبار **ناتج من عملية على حدثين أو أكثر، فعلي سبيل المثال قد يكون الحدث محل الاعتبار** **هو A عبارة عن اتحاد Union الحدثين B، C أي أن:**

$$A = (B \cup C)$$

أو قد يكون الحدث A هو حدث ناتج من تقاطع Intersection الحدثين B، C أو أن:

$$A = (B \cap C)$$

أو قد يكون الحدث A هو حدث ناتج من الفرق Difference بين الحدثين B، C أو أن:

$$A = (B - C)$$

وفي أواخر القرن السابع عشر والنصف الأول من القرن الثامن عشر الميلادي اشتق عالم الرياضيات ابراهام دو موفر Abraham De Moiver (١٦٦٧-١٧٥٤) الصياغات الأولي لقوانين الإضافة والضرب Addition and Multiplication Laws. وكان لاشتقاق هذه القوانين وتطويرها أهمية بالغة في تطوير وتنمية علم الاحتمالات بصفة خاصة وعلوم الإحصاء والرياضيات بصفة عامة.

وفي هذا الفصل سوف نتناول أهم هذه القوانين التي يمكن باستخدامها الحصول على احتمال لحدث ناتج من عملية أو أكثر على حدثين أو أكثر من خلال مجموعة من النظريات الأساسية لعلم الاحتمالات بالإضافة إلى بعض المسلمات Axioms المبنية عليها هذه النظريات بالإضافة إلى العديد من الأمثلة على هذه القوانين.

مسلمة (١): **إذا كان الحدث A فئة جزئية من فراغ المعاينة S أي أن:**

$$A \subseteq S$$

فإن:

$$P_r(A) \geq 0 \qquad \textbf{(A.24)}$$

مسلمة (٢): **إذا كان S هي فراغ المعاينة فإن:**

$$P_r(S) = 1 \qquad \textbf{(A.25)}$$

مسلمة (٣): **إذا كانت الأحداث ($A_1, A_2, ..., A_n$) سلسلة من الأحداث المتنافية Sequence of Mutually Exclusive Events في فراغ المعاينة S بمعني:**

$$(A_i \cap A_j) = \{\ \} = \phi \qquad i \neq j \qquad \textbf{(A.26)}$$

فإن:

$$P_r(A_1 \cup A_2 \cup A_3 \cup \ldots\ldots \cup A_n) = P_r(A_1) + P_r(A_2) + \ldots\ldots + P_r(A_n)$$
$$= P_r(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P_r(A_i)$$

(A.27)

ومن المسلمات الثلاثة السابقة أمكن استنتاج النظريات التالية:

نظرية (١):

إذا كان الحدث A' **هو مكمل الحدث** A **في فراغ المعاينة S فإن:**

$$\boxed{P_r(A') = 1 - P_r(A) \qquad \textbf{(A.28)}}$$

الإثبات: من نظرية الفئات نجد أن:

$$(A \cap A') = \phi \quad , \quad (A \cup A') = S \qquad \textbf{(A.29)}$$

وبالتالي فإن:

$$P_r(A \cup A') = P_r(S) = 1 \qquad \textbf{(A.30)}$$

وبما أن الحدثين A' ، A **حدثين متنافيين فمن المسلمة (٣) نجد أن:**

$$P_r(A \cup A') = P_r(A) + P_r(A') \qquad \textbf{(A.31)}$$

وبالتعويض في (A.31) بقيمة $P_r(A \cup A')$ **في (A.30) نجد أن:**

$$P_r(A) + P_r(A') = 1 \rightarrow P_r(A') = 1 - P_r(A)$$

نظرية (٢):

إذا كان الحدث A بحيث $A \subseteq S$ **فإن:**

$$\boxed{0 \le P_r(A) \le 1 \qquad \textbf{(A.32)}}$$

الإثبات: من المسلمة (١) نجد أن:

$$P_r(A) \ge 0 \quad , \quad P_r(A') \ge 0 \qquad \textbf{(A.33)}$$

ومن النظرية السابقة نجد أن:

$$P_r(A) = 1 - P_r(A') \qquad \textbf{(A.34)}$$

وبما أن $P_r(A) \ge 0$ **فإن:**

$$P_r(A) \le 1 \quad \rightarrow \quad 0 \le P_r(A) \le 1 \qquad \textbf{(A.25)}$$

نظرية (٣):

إذا كان الحدث A فئة خالية Null Set أي أن $A=\phi$ **فإن:**

$$P_r(A)=0 \qquad \textbf{(A.36)}$$

الإثبات: بما أن $A=\phi$ **إذن** $A=S'=\phi$ **وبالتالي فإن:**

$$P_r(A)=P_r(S') \qquad \textbf{(A.37)}$$

$$P_r(S\cup S')=P_r(S)+P_r(S')=1+P_r(A) \qquad \textbf{(A.38)}$$

وبما أن $S=(S\cup S')$ **وبالتالي فإن:**

$$P_r(S\cup S')=P_r(S) \qquad \textbf{(A.39)}$$

بالتعويض في الطرف الأيمن من المعادلة (A.38) بـ $P_r(S\cup S')$ **من المعادلة (A.39) نجد أن:**

$$P_r(S)=1+P_r(A)$$

$$1=1+P_r(A)$$

$$P_r(A)=0$$

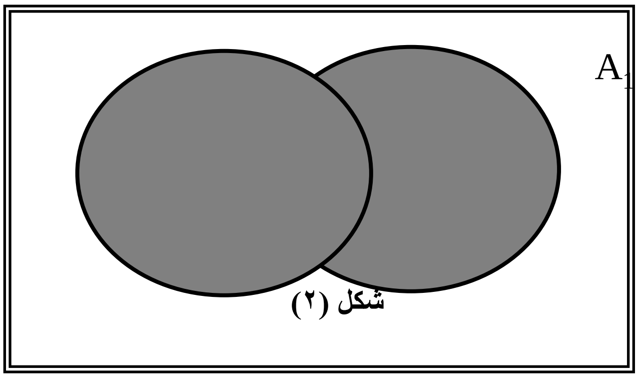
نظرية (٤):

إذا كان الحدثين A_1 **،** A_2 **بحيث** $A_1,A_2\subseteq S$ **فإن:**

$$P_r(A_1\cup A_2)=P_r(A_1)+P_r(A_2)-P_r(A_1\cap A_2) \qquad \textbf{(A.40)}$$

الإثبات: بما أن $(A_1\cup A_2)=(A_1)\cup(A_1'\cap A_2)$ **كما هو موضح بالشكل التالي:**

S



شكل (٢)

وبما أن الحدثين (A_1) ، $(A'_1 \cap A_2)$ **حدثين متنافيين، فمن المسلمة (٢) نجد أن:**

$$\begin{aligned} P_r(A_1 \cup A_2) &= P_r\left[(A_1) \cup (A'_1 \cap A_2)\right] \\ &= P_r(A_1) + P_r(A'_1 \cap A_2) \end{aligned} \qquad \textbf{(A.41)}$$

وبما أن:

$$A_2 = (A_1 \cap A_2) \cup (A'_1 \cap A_2) \qquad \textbf{(A.42)}$$

وحيث أن الحدثين $(A_1 \cap A_2)$ ، $(A'_1 \cap A_2)$ **حدثين متنافيين فإن:**

$$\begin{aligned} & P_r(A_2) = P_r(A_1 \cap A_2) + P_r(A'_1 \cap A_2) \\ & \therefore P_r(A'_1 \cap A_2) = P_r(A_2) - P_r(A_1 \cap A_2) \end{aligned} \qquad \textbf{(A.43)}$$

بالتعويض في الطرف الأيسر من المعادلة (A.41) بالطرف الأيمن في المعادلة (A.43) نجد أن:

$$P_r(A_1 \cup A_2) = P_r(A_1) + P_r(A_2) - P_r(A_1 \cap A_2) \qquad \textbf{(A.44)}$$

نتيجة: **إذا كان الحدثين** A_1 ، A_2 **حدثين متنافيين فإن** $(A_1 \cap A_2) = \phi$ **ومن نظرية (٣) نجد أن**

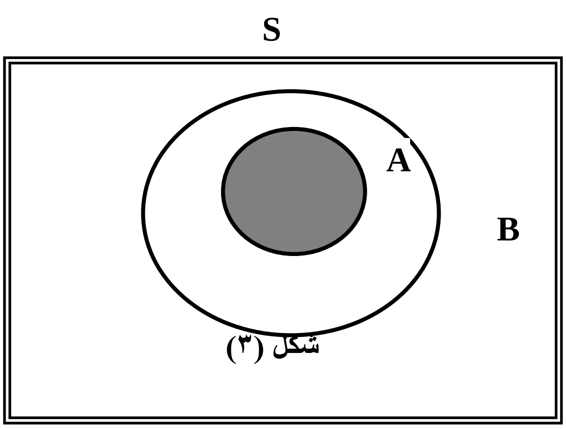
$$P_r(A_1 \cap A_2) = P_r(\phi) = 0 \qquad \textbf{(A.45)}$$

بالتعويض بقيمة $P_r(A_1 \cap A_2)$ **من المعادلة (A.45) في الطرف الأيمن من المعادلة (A.44) نجد أن:**

$$P_r(A_1 \cup A_2) = P_r(A_1) + P_r(A_2) \qquad \textbf{(A.46)}$$

نظرية (٥)

إذا كان الحدثين $A, B \subseteq S$ **بحيث** $A \subseteq B$ **كما هو موضح بالشكل التالي:**



شكل (٣)

فإن:

$$P_r(A) \le P_r(B) \qquad \textbf{(A.47)}$$

الإثبات: بما أن $A \subseteq B$ **إذن:**

$$(B) = (A) \cup (A' \cap B) \qquad \textbf{(A.48)}$$

أي أن الفئة (A)، $(A' \cap B)$ **فئتين متنافيتين. فمن النتيجة السابقة نجد أن:**

$$P_r(B) = P_r(A) + P_r(A' \cap B) \qquad \textbf{(A.49)}$$

ومن المسلمة (١) نجد أن:

$$P_r(A' \cap B) \ge 0 \qquad \textbf{(A.50)}$$

ومن المعادلة (A.49) والمتباينة (A.50) نجد أن:

$$P_r(A) \le P_r(B)$$

مثال (١٠)

إذا كانت S هي فراغ المعاينة، والحدثين A ، B بحيث $(A, B \subseteq S)$**. بحيث:**

$$P_r(A) = 0.3 \quad , \quad P_r(B) = 0.5 \quad , \quad P_r(A \cap B) = 0.1$$

احسب الاحتمالات التالية:

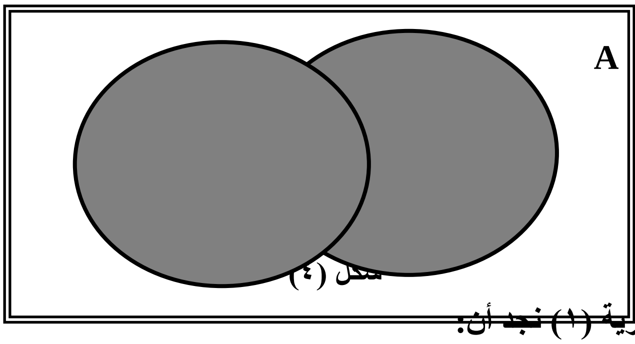
(١) $P_r(A')$ ، $P_r(B')$.

(٢) $P_r(A' \cap B)$ ، $P_r(A' \cap B')$.

الحل:

الشكل التالي يوضح الحدثين A ، B.

S



شكل (٤)

(١) من نظرية (١) نجد أن:

$$P_r(A') = 1 - P_r(A) = 1 - 0.3 = 0.7$$

$$P_r(B') = 1 - P_r(B) = 1 - 0.5 = 0.5$$

(٢) من الشكل نجد أن:

$$(B) = (A' \cap B) \cup (A \cap B)$$
$$P_r(B) = P_r(A' \cap B) + P_r(A \cap B)$$
$$P_r(A' \cap B) = P_r(B) - P_r(A \cap B) = 0.5 - 0.1 = 0.4$$

كذلك:

$$(A' \cap B') = (A \cup B)' = S - (A \cup B)$$
$$P_r(A' \cap B') = P_r(S) - P_r(A \cup B)$$
$$= 1 - [0.3 + (0.5 - 0.1)] = 1 - 0.7 = 0.3$$

ملحوظة: إذا كان وقوع الحدثين A ، B معاً هو $(A \cap B)$ فإن احتمال وقوع الحدثين A ، B معاً $(A \cap B)$ يسمي بالاحتمال المشترك Joint Probability.

الاحتمالات الشرطية Conditional Probabilities

إذا كان A ، B حدثين في فراغ المعاينة S أي $A, B \leq S$. وإذا كان حدوث الحدث A يتوقف على حدوث الحدث B فإن احتمال وقوع الحدث A بشرط وقوع الحدث B يسمي بالاحتمال الشرطي Conditional Probability ويرمز له بالرمز $P_r(A \mid B)$ ويقرأ "احتمال وقوع الحدث A بشرط وقوع الحدث B". ويسمي الحدث B بالحدث القبلي Prior Event ويسمي الحدث $(A \mid B)$ بالحدث البعدي Posterior Event.

ويتم حساب قيمة $P_r(A \mid B)$ من المعادلة التالية:

$$P_r(A \mid B) = \frac{P_r(A \cap B)}{P_r(B)} \quad , P_r(B) > 0 \qquad \textbf{(A.51)}$$

بالمثل:

$$P_r(B \mid A) = \frac{P_r(A \cap B)}{P_r(A)} \quad , P_r(A) > 0 \qquad \textbf{(A.52)}$$

من المعادلتين (A.51) ، (A.52) نجد أن:

$$P_r(A \cap B) = P_r(A \mid B) P_r(B) \qquad (A.53)$$
$$= P_r(B \mid A) P_r(A) \qquad (A.54)$$

والمثال التالي يوضح مفهوم الاحتمال الشرطي الذي يمكن حسابه من المعادلتين(A.51) ، (A.52).

مثال (١١)

الجدول التالي يوضح عدد العمال المهرة وغير المهرة في ثلاثة أقسام بأحد المصانع.

جدول (٣)

القسم / نوع العامل	الأول	الثاني	الثالث	المجموع
ماهر	370	410	55	835
غير ماهر	30	90	45	165
المجموع	400	500	100	1000

احسب الاحتمالات التالية:

١- **احتمال اختيار عامل من القسم الأول، واحتمال اختيار عامل من القسم الثاني، احتمال اختيار عامل من القسم الثالث.**

٢- **احتمال اختيار عامل ماهر بشرط أن يكون من القسم الأول.**

٣- **احتمال اختيار عامل من القسم الأول بشرط أن يكون ماهر.**

٤- **احتمال اختيار عامل غير ماهر بشرط أن يكون من القسم الثالث.**

الحل:

١- **إذا فرضنا أن الأحداث** A_1 ، A_2 ، A_3 **تمثل اختيار عامل من القسم الأول، الثاني، الثالث على الترتيب. فإن:**

$$P_r(A_1) = \frac{\text{عدد الحالات المواتية}}{\text{عدد الحالات الممكنة}} = \frac{400}{1000} = 0.4$$

$$P_r(A_2) = \frac{\text{عدد الحالات المواتية}}{\text{عدد الحالات الممكنة}} = \frac{500}{1000} = 0.5$$

$$P_r(A_3) = \frac{\text{عدد الحالات المواتية}}{\text{عدد الحالات الممكنة}} = \frac{100}{1000} = 0.1$$

٢- **إذا فرضنا أن الحدث B يمثل اختيار عامل ماهر، فإن الحدث** $(B \mid A_1)$ **يمثل اختيار عامل ماهر بشرط أن يكون من القسم الأول. ففي هذه الحالة يكون الحدث القبلي هو** A_1 **أي ان اختيار عامل من القسم الأول، والحدث البعدي هو** $(B \mid A_1)$ **أي اختيار عامل ماهر من عمال القسم الأول. أي عند اختيار**

العامل يكون الاختيار من عمال القسم الأول فقط وليس من عمال المصنع ككل، وبالتالي فإن:

$$P_r(B \mid A_1) = \frac{P_r(B \cap A_1)}{P_r(A_1)} = \frac{370/1000}{400/1000} = \frac{370}{400} = \frac{37}{40}$$

٣- كذلك احتمال اختيار عامل من القسم الأول بشرط أن يكون ماهر هو $P_r(A_1 \mid B)$. أي أن الحدث القبلي هو (B) أي اختيار عامل ماهر، والحدث البعدي هو $(A_1 \mid B)$ أي اختيار عامل من العمال المهرة بشرط أن يكون من القسم الأول. أي عند اختيار العامل يكون الاختيار من العمال المهرة فقط وليس من عمال المصنع ككل. وبالتالي فإن:

$$P_r(A_1 \mid B) = \frac{P_r(A_1 \cap B)}{P_r(B)} = \frac{370/1000}{835/1000} = \frac{370}{835} = \frac{37}{167}$$

ونلاحظ أن $P_r(A_1 \mid B) \neq P_r(B \mid A_1)$

٤- بالمثل احتمال اختيار عامل غير ماهر بشرط أن يكون من القسم الثالث هو $P_r(B' \mid A_3)$ حيث:

$$P_r(B' \mid A_3) = \frac{P_r(B' \cap A_3)}{P_r(A_3)} = \frac{45/1000}{100/1000} = \frac{45}{100}0.45$$

مما سبق يتضح أن دراسة الاحتمال الشرطي لحدث معين في تجربة ما يمكننا من دراسة مفردات الحدث القبلي فقط بدلاً من دراسة جميع مفردات فراغ المعاينة أو بعبارة أخرى يؤدي إلى التعامل مع جزء من عناصر فراغ المعاينة وليس الكل.

مثال (١٢)

إذا ألقيت قطعتي عملة متوازنتين عشوائياً ما هو احتمال ظهور صورة على القطعة الأولي بشرط ظهور كتابة على القطعة الثانية؟

الحل:

الجدول التالي يوضح النتائج الممكنة للتجربة حيث H تشير إلى الصورة، W تشير إلى الكتابة.

جدول(٤)

نتائج القطعة الثانية / نتائج القطعة الأولي	H	W
H	**(H , H)**	**(H , W)**
W	**(W , H)**	**(W , W)**

من الجدول يتضح أن عدد الحالات الممكنة N = 4 حيث:

S = {(H , H) , (H , W) , (W , H) , (W , W)} فراغ المعاينة

فإذا فرضنا أن الحدث A يشير إلى ظهور صورة على القطعة الأولي والحدث B يشير إلى ظهور الكتابة على القطعة الثانية. فإن:

$$(A) = \{(H,H),(H,W)\} \rightarrow P_r(A) = \frac{2}{4} = \frac{1}{2} \qquad \textbf{(A.53)}$$

$$(B) = \{(H,W),(W,W)\} \rightarrow P_r(B) = \frac{2}{4} = \frac{1}{2} \qquad \textbf{(A.54)}$$

كذلك:

$$(A \cap B) = \{(H,W)\} \rightarrow P_r(A \cap B) = \frac{1}{4}$$

كذلك إذا فرضنا أن الحدث $(A|B)$ هو ظهور الصورة على القطعة الأولي بشرط ظهور الكتابة على الثانية. فنجد أن:

$$(A|B) = \{(H,W)\} \qquad \textbf{(A.55)}$$

ومن تعريف الاحتمالات نجد أن:

$$P_r(A|B) = \frac{\text{عدد الحالات المواتية}}{\text{عدد الحالات الممكنة}} = \frac{1}{2} \qquad \textbf{(A.56)}$$

حيث أن عدد الحالات المواتية حالة واحدة الموجودة في المعادلة (A.55) وعدد الحالات الممكنة في هذه الحالة يساوي حالتين كما هو واضح في (A.51). وفي (A.56) تم حساب الاحتمال الشرطي عن طريق تعريف الاحتمال، والآن سوف نقوم بحساب $P_r(A|B)$ باستخدام قاعدة الاحتمال الشرطي في (A.51) على النحو التالي:

$$P_r(A|B) = \frac{P_r(A \cap B)}{P_r(B)} = \frac{1/4}{2/4} = \frac{1}{2} \qquad \textbf{(A.57)}$$

ونلاحظ أن قيمة $P_r(A|B)$ باستخدام قاعدة الاحتمالات الشرطية في (A.57) هي نفس القيمة باستخدام تعريف الاحتمال في (A.56).

مثال (١٣)

أجريت دراسة على 1000 أسرة من ذو الدخل المحدود لمعرفة العلاقة بين مستوي تعليم الأم وعدد الأطفال في الأسرة. والجدول التالي يوضح توزيع 1000 حسب مستوي تعليم الأم وعدد الأطفال في الأسرة الواحدة.

جدول(٥)

عدد الأطفال في الأسرة / مستوي تعليم الأم	1	2	3	4	المجموع
أقل من المتوسط	1	7	70	300	379
متوسط	3	23	60	205	291
عالي	125	30	174	1	330
المجموع	130	60	304	306	1000

المطلوب:

١- **احسب احتمال وجود طفلين في الأسرة من ذو الدخل المحدود وفي نفس الوقت يكون مستوي تعليم الأم أقل من المتوسط.**

٢- **احسب احتمال وجود أكثر من طفلين في الأسرة وفي نفس الوقت مستوي تعليم الأم عالي.**

٣- **احسب احتمال وجود 3 اطفال في الأسرة وفي نفس الوقت مستوي تعليم الأم متوسط.**

٤- **احسب احتمال أن يكون عدد الأطفال في الأسرة 3 أطفال بشرط أن يكون مستوي تعليم الأم متوسط.**

٥- **احسب احتمال أن يكون عدد الأطفال في الأسرة أكثر من طفلين بشرط أن يكون مستوي تعليم الأم عالي.**

٦- **احسب احتمال أن يكون عدد الأطفال في الأسرة أقل من 3 أطفال بشرط أن يكون مستوي تعليم الأم أقل من العالي.**

الحل:

١- **نفرض أن الحدث** A_i **يشير إلى مستوي تعليم الأم i حيث** $i = 1,2,3$ ، **والحدث** B_j **يشير إلى عدد الأطفال j في الأسرة الواحدة حيث** $j = 1,2,3,4$ **وبالتالي فإن احتمال وجود طفلين في الأسرة وفي نفس الوقت يكون مستوي تعليم الأم في هذه الأسرة أقل من المتوسط هو** $P_r(A_1 \cap B_2)$ **حيث:**

$$P_r(A_1 \cap B_2) = \frac{7}{1000} = 0.007 \qquad \textbf{(A.58)}$$

٢- **احتمال وجود أكثر من طفلين في الأسرة وفي نفس الوقت يكون مستوي تعليم الأم عالي هو** $P_r\left[(A_3 \cap B_3) \cup (A_3 \cap B_4)\right]$ **وبما ان الحدثين** $(A_3 \cap B_3),(A_3 \cap B_4)$ **حدثين متنافيين فإن:**

$$P_r[(A_3 \text{ I } B_3) \text{ Y } (A_3 \text{ I } B_4)] = P_r(A_3 \text{ I } B_3) + P_r(A_3 \text{ I } B_4)$$

$$= \frac{174}{1000} + \frac{1}{1000} = \frac{175}{1000} = 0.175$$

٣- احتمال وجود ثلاثة أطفال في الأسرة ومستوي تعليم الأم متوسط هو $P_r(A_2 \text{ I } B_3)$ حيث:

$$P_r(A_2 \text{ I } B_3) = \frac{60}{1000} = 0.060$$

٤- احتمال أن يكون عدد الأطفال في الأسرة 3 أطفال بشرط أن يكون مستوي تعليم الأم متوسط هو $P_r(B_3 \mid A_2)$ حيث:

$$P_r(B_3 \mid A_2) = \frac{P_r(B_3 \text{ I } A_2)}{P_r(A_2)} = \frac{60/1000}{291/1000} = \frac{60}{291}$$

٥- احتمال أن يكون عدد الأطفال في الأسرة أكثر من طفلين بشرط أن يكون تعليم الأم عالي هو $P_r[(B_3 \mid A_3) \text{ Y } (B_4 \mid A_3)]$ وبما أن الحدثين $(B_3 \mid A_3), (B_4 \mid A_3)$ حدثين متنافيين فإن:

$$P_r[(B_3 \mid A_3) \text{ Y } (B_4 \mid A_3)] = P_r(B_3 \mid A_3) + P_r(B_4 \mid A_3)$$

$$= \frac{P_r(B_3 \text{ I } A_3)}{P_r(A_3)} + \frac{P_r(B_4 \text{ I } A_3)}{P_r(A_3)}$$

$$= \frac{174/1000}{330/1000} + \frac{1/1000}{330/1000} = \frac{175}{330}$$

٦- احتمال أن يكون عدد الأطفال في الأسرة أقل من 3 أطفال بشرط أن يكون مستوي تعليم الأم أقل من العالي هو $P_r[(B_3 \mid A_1) \text{ Y } (B_3 \mid A_2)]$ وبما أن الحدثين $(B_3 \mid A_1), (B_3 \mid A_2)$ حدثين متنافيين فإن:

$$P_r\left[(B_3 \mid A_1)\,Y\,(B_3 \mid A_2)\right] = P_r(B_3 \mid A_1) = P_r(B_3 \mid A_2)$$

$$= \frac{P_r(B_3\ I\ \ A_1)}{P_r(A_1)} + \frac{P_r(B_3\ I\ \ A_2)}{P_r(A_2)}$$

$$= \frac{70/1000}{379/1000} + \frac{60/1000}{291/100} = \frac{70}{379} + \frac{60}{291}$$

ملحق (٢)
ملخص لبعض التوزيعات الاحتمالية

التوزيع	نوعه	دالة الاحتمال أو دالة كثافة الاحتمال	التوقع (μ)	التباين (σ^2)
المنتظم	متقطع	$P_r(X) = \frac{1}{n}$, $X = 0,1,2,...,\ n$	$\mu = \frac{n+1}{2}$	$\sigma^2 = \frac{n^2+1}{12}$
برنولى	متقطع	$P_r(X) = P^X (1-P)^{1-X}$, $X = 0.1$	$\mu = P$	$\sigma^2 = P(1-P)$
ذات الحدين	متقطع	$P_r(X) = C_X^n P^X (1-P)^{n-X}$, $X = 0.1,..,\ n$	$\mu = nP$	$\sigma^2 = nP(1-P)$
بواسون	متقطع	$P_r(X) = \frac{e^{-\lambda}\lambda^X}{X!}$, $X = 0,1,2,..$	$\mu = \lambda$	$\sigma^2 = \lambda$
المنتظم	متصل	$f(X) = \frac{1}{b-a}$, $a \le X \le b$	$\mu = \frac{a+b}{2}$	$\sigma^2 = \frac{(b-a)^2}{12}$

التوزيع	نوعه	دالة الاحتمال أو دالة كثافة الاحتمال	التوقع (μ)	التباين (σ^2)
الاسى	متصل	$f(X)=\lambda e^{-\lambda X} \quad , X \geq 0 \quad , \lambda > 0$	$\mu = \frac{1}{\lambda}$	$\sigma^2 = \frac{1}{\lambda^2}$
المعتاد الطبيعي	متصل	$f(X)=\frac{1}{\sqrt{2\prod}\sigma} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \quad , -\infty < X < \infty$	$\mu = \mu$	$\sigma^2 = \sigma^2$
المعتاد القياسي	متصل	$f(X)=\frac{1}{\sqrt{2\prod}} e^{-\frac{1}{2}(X)^2} \quad , -\infty < X < \infty$	$\mu = 0$	$\sigma^2 = 1$
مربع كا (χ^2)	متصل	$f(X)=\frac{1}{\left(\frac{n}{2}-1\right)}\left(\frac{1}{2}\right)^{\frac{n}{2}} X^{\frac{n}{2}-1} e^{\frac{-1}{2}X}$ $X \geq 0 \quad , \quad n = 1,2,...$	$\mu = n$	$\sigma^2 = 2n$

التوزيع	نوعه	دالة الاحتمال أو دالة كثافة الاحتمال	التوقع (μ)	التباين (σ^2)
توزيع (T)	متصل	$f(X)=\dfrac{\Gamma[(k+1)/2]}{\Gamma(k/2)\sqrt{2\prod}(1+X^2/k)^{(k+1)/2}}$ $k>0$	$\mu=0$ $k>1$	$\sigma^2=\dfrac{k}{k-2}$ $k>2$
توزيع (F)	متصل	$f(X)=\dfrac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\dfrac{m}{n}\right)^{m/2}$ $m,n=1,2,\ldots$	$\mu=\dfrac{m}{n-2}$ $n>2$	$\sigma^2=$ $\dfrac{2n^2(n+m-2)}{m(n-2)^2(n-4)}$

ملحق (٣)
الأعداد العشوائية

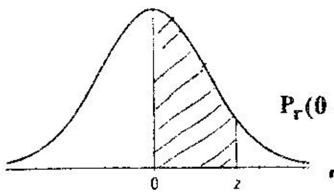
71274	84346	75444	85690	35384	87841	97411	78698	46796	33552
64017	01373	14665	31891	80997	14321	47741	59980	87739	38174
43747	17686	11045	15549	52779	65135	00275	95434	36337	24041
59688	48689	41591	47042	83615	93034	25077	64835	67798	30547
95016	73467	11447	59500	94921	15166	69217	26267	11316	22651
65207	30591	65947	58339	00952	32111	45459	14986	57395	34492
34510	78657	08883	49489	85619	52912	01662	[illegible]	78354	30631
56299	60624	91572	31734	18159	18927	31314	59682	41320	88602
02113	12579	86172	03819	69968	02616	72687	42699	04792	16510
00884	87979	45184	61572	20086	14498	29640	94263	90964	29278
89367	53577	97412	19603	57234	63055	49059	35761	72007	22751
99781	56740	42659	46617	21828	99831	45987	63450	66919	78252
92024	12100	76013	12587	86340	74880	79979	35906	38122	64917
82861	09215	87342	72789	76132	24468	93065	78968	03321	48081
11286	13011	67982	74101	44961	25468	14247	95934	50711	24492
68674	24686	14460	61242	92310	86810	87702	69811	53996	99517
24882	20749	94139	28785	74402	18561	79069	56838	30020	99707
21740	51134	39298	92203	66230	30636	58169	78982	50057	39908
46901	34825	28673	10404	97777	30782	04680	15319	84125	40937
92686	81702	74149	76326	01101	96278	90855	55145	89705	51199
98743	59366	94797	96803	69876	87533	19675	59246	65348	06606
11463	25619	38107	91053	58416	02720	86563	27443	99598	04074
76975	18636	54975	67422	57101	68857	35389	35641	34505	71552
67359	50379	81053	97357	00717	59504	34480	77127	23243	48682
34084	48031	27227	48912	10797	88917	93126	52945	79457	66528
94553	44441	83166	40056	73935	52103	13972	56781	31900	95037
20545	43211	34500	92233	53497	39401	78535	82360	57410	32060
05646	01152	13235	39168	69214	83852	00144	08105	94247	37189
54022	58295	96122	87620	74774	06884	81689	68392	25776	08748
32291	30700	32561	83579	94582	77930	06826	96855	97751	42664
66212	75061	65891	96896	15107	12985	97616	64241	08592	81036
98059	44951	23078	92793	80756	52799	20340	62969	81775	31065
23157	52179	24394	39833	89427	58771	26992	00649	25475	50888
28911	81929	91368	49372	43335	44465	43257	66893	34761	60423
85959	90369	02100	28727	83001	84166	20473	35305	38088	54795
34459	31400	58760	17157	73816	55527	54133	24605	56153	35354
04073	57781	36894	93000	57834	29343	98195	58425	97275	71392
22126	91330	95667	75737	36869	55209	41663	04943	59401	17039
10288	61685	25302	84097	13088	86840	04020	43046	01043	43157
75431	32853	72907	99432	65482	23011	70466	87386	67471	77629
90800	17425	28042	53770	98924	31863	84115	82488	23239	82185
19083	89475	05207	41284	83405	55825	31117	59821	96455	63796
10405	67911	77238	46262	42766	07215	02391	47316	78724	41170
34711	77325	99768	63455	44335	91028	27740	86163	81474	08159
73334	61941	16883	05012	63191	35763	60157	09617	25501	44989
79452	68381	71937	23274	60273	47091	82876	24641	03825	50894
13864	28746	32434	88325	99996	96130	39471	74020	56077	22133
73082	50271	83240	80065	09328	02940	41686	32758	89467	73553
43060	88221	35010	79829	71520	80453	95049	66352	77495	83256
15172	42061	33264	63832	48528	23258	13520	83222	45659	39074

Source: The Rand Corporation, *A Million Random Digits with 100,000 Normal Deviates* (Glencoe, Ill.: The Free Press, 1955), p. 377. Copyright © 1955 by the Rand Corporation. Reprinted by permission.

ملحق (٣)
الأعداد العشوائية

71274	84346	75444	85690	35384	87841	97411	78698	46796	33552
64017	01373	14665	31891	80997	14321	47741	59980	87739	38174
43747	17686	11045	15549	52779	65135	00275	95434	36337	24041
59688	48689	41591	47042	83615	93034	25077	64835	67798	30547
95016	73467	11447	59500	94921	15166	69217	26267	11316	22651
65207	30591	65947	58339	00952	32111	45459	14986	57395	34492
34510	78657	08883	49489	85619	52912	01662	[illegible]	78354	30631
56299	60624	91572	31734	18159	18927	31314	59682	41320	88602
02113	12579	86172	03819	69968	02616	72687	42699	04792	16510
00884	87979	45184	61572	20086	14498	29640	94263	90964	29278
89367	53577	97412	19603	57234	63055	49059	35761	72007	22751
99781	56740	42659	46617	21828	99831	45987	63450	66919	78252
92024	12100	76013	12587	86340	74880	79979	35906	38122	64917
82861	09215	87342	72789	76132	24468	93065	78968	03321	48081
11286	13011	67982	74101	44961	25468	14247	95934	50711	24492
68674	24686	14460	61242	92310	86810	87702	69811	53996	99517
24882	20749	94139	28785	74402	18561	79069	56838	30020	99707
21740	51134	39298	92203	66230	30636	58169	78982	50057	39908
46901	34825	28673	10404	97777	30782	04680	15319	84125	40937
92686	81702	74149	76326	01101	96278	90855	55145	89705	51199
98743	59366	94797	96803	69876	87533	19675	59246	65348	06606
11463	25619	38107	91053	58416	02720	86563	27443	99598	04074
76975	18636	54975	67422	57101	68857	35389	35641	34505	71552
67359	50379	81053	97357	00717	59504	34480	77127	23243	48682
34084	48031	27227	48912	10797	88917	93126	52945	79457	66528
94553	44441	83166	40056	73935	52103	13972	56781	31900	95037
20545	43211	34500	92233	53497	39401	78535	82360	57410	32060
05646	01152	13235	39168	69214	83852	00144	08105	94247	37189
54022	58295	96122	87620	74774	06884	81689	68392	25776	08748
32291	30700	32561	83579	94582	77930	06826	96855	97751	42664
66212	75061	65891	96896	15107	12985	97616	64241	08592	81036
98059	44951	23078	92793	80756	52799	20340	62969	81775	31065
23157	52179	24394	39833	89427	58771	26992	00649	25475	50888
28911	81929	91368	49372	43335	44465	43257	66893	34761	60423
85959	90369	02100	28727	83001	84166	20473	35305	38088	54795
34459	31400	58760	17157	73816	55527	54133	24605	56153	35354
04073	57781	36894	93000	57834	29343	98195	58425	97275	71392
22126	91330	95667	75737	36869	55209	41663	04943	59401	17039
10288	61685	25302	84097	13088	86840	04020	43046	01043	43157
75431	32853	72907	99432	65482	23011	70466	87386	67471	77629
90800	17425	28042	53770	98924	31863	84115	82488	23239	82185
19083	89475	05207	41284	83405	55825	31117	59821	96455	63796
10405	67911	77238	46262	42766	07215	02391	47316	78724	41170
34711	77325	99768	63455	44335	91028	27740	86163	81474	08159
73334	61941	16883	05012	63191	35763	60157	09617	25501	44989
79452	68381	71937	23274	60273	47091	82876	24641	03825	50894
13864	28746	32434	88325	99996	96130	39471	74020	56077	22133
73082	50271	83240	80065	09328	02940	41686	32758	89467	73553
43060	88221	35010	79829	71520	80453	95049	66352	77495	83256
15172	42061	33264	63832	48528	23258	13520	83222	45659	39074

Source: The Rand Corporation, *A Million Random Digits with 100,000 Normal Deviates* (Glencoe, Ill.: The Free Press, 1955), p. 377. Copyright © 1955 by the Rand Corporation. Reprinted by permission.



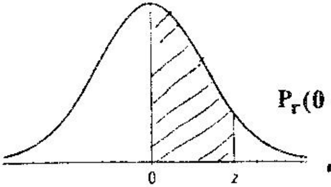
ملحق (٤)

التوزيع المعتاد القياسي

الجدول التالى يعطى المساحة تحت منحني دالة كثافة احتمال المتغير Z أي يعطي $P_r(0 < Z < z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4986	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

Source: The National Bureau of Standards. *Tables of Normal Probability Functions*. Applied Mathematics Series, no. 23 (Washington, D.C.: U.S. Government Printing Office, 1953). The original contains probabilities for values of z from 0 to 8.285, mostly in increments of .0001, and for areas from $\mu - z$ to $\mu + z$.



ملحق (٤)

التوزيع المعتاد القياسي

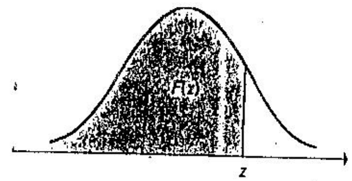
الجدول التالى يعطى المساحة تحت منحنى دالة كثافة احتمال المتغير Z أي يعطي $P_r(0 < Z < z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
.7	2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4986	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

Source: The National Bureau of Standards. *Tables of Normal Probability Functions*. Applied Mathematics Series, no. 23 (Washington, D.C.: U.S. Government Printing Office, 1953). The original contains probabilities for values of z from 0 to 8.285, mostly in increments of .0001, and for areas from $\mu - z$ to $\mu + z$.

تابع ملحق (٤)

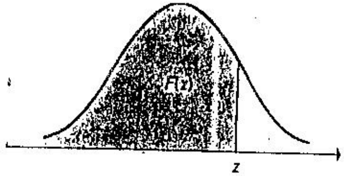
دالة التوزيع التراكمية للمتغير Z أي $P_r(Z < z)$



z	F(z)	z	F(z)	z	F(z)	z	F(z)	z	F(z)	z	F(z)
.00	.5000										
.01	.5040	.31	.6217	.61	.7291	.91	.8186	1.21	.8869	1.51	.9345
.02	.5080	.32	.6255	.62	.7324	.92	.8212	1.22	.8888	1.52	.9357
.03	.5120	.33	.6293	.63	.7357	.93	.8238	1.23	.8907	1.53	.9370
.04	.5160	.34	.6331	.64	.7389	.94	.8264	1.24	.8925	1.54	.9382
.05	.5199	.35	.6368	.65	.7422	.95	.8289	1.25	.8944	1.55	.9394
.06	.5239	.36	.6406	.66	.7454	.96	.8315	1.26	.8962	1.56	.9406
.07	.5279	.37	.6443	.67	.7486	.97	.8340	1.27	.8980	1.57	.9418
.08	.5319	.38	.6480	.68	.7517	.98	.8365	1.28	.8997	1.58	.9429
.09	.5359	.39	.6517	.69	.7549	.99	.8389	1.29	.9015	1.59	.9441
.10	.5398	.40	.6554	.70	.7580	1.00	.8413	1.30	.9032	1.60	.9452
.11	.5438	.41	.6591	.71	.7611	1.01	.8438	1.31	.9049	1.61	.9463
.12	.5478	.42	.6628	.72	.7642	1.02	.8461	1.32	.9066	1.62	.9474
.13	.5517	.43	.6664	.73	.7673	1.03	.8485	1.33	.9082	1.63	.9484
.14	.5557	.44	.6700	.74	.7704	1.04	.8508	1.34	.9099	1.64	.9495
.15	.5596	.45	.6736	.75	.7734	1.05	.8531	1.35	.9115	1.65	.9505
.16	.5636	.46	.6772	.76	.7764	1.06	.8554	1.36	.9131	1.66	.9515
.17	.5675	.47	.6803	.77	.7794	1.07	.8577	1.37	.9147	.167	.9525
.18	.5714	.48	.6844	.78	.7823	1.08	.8599	1.38	.9162	1.68	.9535
.19	.5753	.49	.6879	.79	.7852	1.09	.8621	1.39	.9177	1.69	.9545
.20	.5793	.50	.6915	.80	.7881	1.10	.8643	1.40	.9192	1.70	.9554
.21	.5832	.51	.6950	.81	.7910	1.11	.8665	1.41	.9207	1.71	.9564
.22	.5871	.52	.6985	.82	.7939	1.12	.8686	1.42	.9222	1.72	.9573
.23	.5910	.53	.7019	.83	.7967	1.13	.8708	1.43	.9236	1.73	.9582
.24	.5948	.54	.7054	.84	.7995	1.14	.8729	1.44	.9251	1.74	.9591
.25	.5987	.55	.7088	.85	.8023	1.15	.8749	1.45	.9265	1.75	.9599
.26	.6026	.56	.7123	.86	.8051	1.16	.8770	1.46	.9279	1.76	.9608
.27	.6064	.57	.7157	.87	.8078	1.17	.8790	1.47	.9292	1.77	.9616
.28	.6103	.58	.7190	.88	.8106	1.18	.8810	1.48	.9306	1.78	.9625
.29	.6141	.59	.7224	.89	.8133	1.19	.8830	1.49	.9319	1.79	.9633
.30	.6179	.60	.7257	.90	.8159	1.20	.8849	1.50	.9332	1.80	.9641

تابع ملحق (٤)

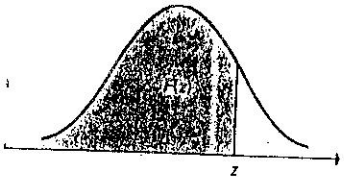
دالة التوزيع التراكمية للمتغير Z أي $P_r(Z < z)$



z	F(z)	z	F(z)	z	F(z)	z	F(z)	z	F(z)	z	F(z)
.00	.5000										
.01	.5040	.31	.6217	.61	.7291	.91	.8186	1.21	.8869	1.51	.9345
.02	.5080	.32	.6255	.62	.7324	.92	.8212	1.22	.8888	1.52	.9357
.03	.5120	.33	.6293	.63	.7357	.93	.8238	1.23	.8907	1.53	.9370
.04	.5160	.34	.6331	.64	.7389	.94	.8264	1.24	.8925	1.54	.9382
.05	.5199	.35	.6368	.65	.7422	.95	.8289	1.25	.8944	1.55	.9394
.06	.5239	.36	.6406	.66	.7454	.96	.8315	1.26	.8962	1.56	.9406
.07	.5279	.37	.6443	.67	.7486	.97	.8340	1.27	.8980	1.57	.9418
.08	.5319	.38	.6480	.68	.7517	.98	.8365	1.28	.8997	1.58	.9429
.09	.5359	.39	.6517	.69	.7549	.99	.8389	1.29	.9015	1.59	.9441
.10	.5398	.40	.6554	.70	.7580	1.00	.8413	1.30	.9032	1.60	.9452
.11	.5438	.41	.6591	.71	.7611	1.01	.8438	1.31	.9049	1.61	.9463
.12	.5478	.42	.6628	.72	.7642	1.02	.8461	1.32	.9066	1.62	.9474
.13	.5517	.43	.6664	.73	.7673	1.03	.8485	1.33	.9082	1.63	.9484
.14	.5557	.44	.6700	.74	.7704	1.04	.8508	1.34	.9099	1.64	.9495
.15	.5596	.45	.6736	.75	.7734	1.05	.8531	1.35	.9115	1.65	.9505
.16	.5636	.46	.6772	.76	.7764	1.06	.8554	1.36	.9131	1.66	.9515
.17	.5675	.47	.6803	.77	.7794	1.07	.8577	1.37	.9147	.167	.9525
.18	.5714	.48	.6844	.78	.7823	1.08	.8599	1.38	.9162	1.68	.9535
.19	.5753	.49	.6879	.79	.7852	1.09	.8621	1.39	.9177	1.69	.9545
.20	.5793	.50	.6915	.80	.7881	1.10	.8643	1.40	.9192	1.70	.9554
.21	.5832	.51	.6950	.81	.7910	1.11	.8665	1.41	.9207	1.71	.9564
.22	.5871	.52	.6985	.82	.7939	1.12	.8686	1.42	.9222	1.72	.9573
.23	.5910	.53	.7019	.83	.7967	1.13	.8708	1.43	.9236	1.73	.9582
.24	.5948	.54	.7054	.84	.7995	1.14	.8729	1.44	.9251	1.74	.9591
.25	.5987	.55	.7088	.85	.8023	1.15	.8749	1.45	.9265	1.75	.9599
.26	.6026	.56	.7123	.86	.8051	1.16	.8770	1.46	.9279	1.76	.9608
.27	.6064	.57	.7157	.87	.8078	1.17	.8790	1.47	.9292	1.77	.9616
.28	.6103	.58	.7190	.88	.8106	1.18	.8810	1.48	.9306	1.78	.9625
.29	.6141	.59	.7224	.89	.8133	1.19	.8830	1.49	.9319	1.79	.9633
.30	.6179	.60	.7257	.90	.8159	1.20	.8849	1.50	.9332	1.80	.9641

تابع ملحق (٤)

دالة التوزيع التراكمية للمتغير Z أي $P_r(Z < z)$

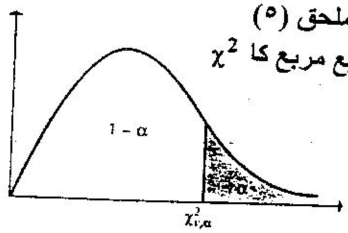


z	$F(z)$	z	$F(z)$	z	$F(z)$	z	$F(z)$	z	$F(z)$	z	$F(z)$
.00	.5000										
.01	.5040	.31	.6217	.61	.7291	.91	.8186	1.21	.8869	1.51	.9345
.02	.5080	.32	.6255	.62	.7324	.92	.8212	1.22	.8888	1.52	.9357
.03	.5120	.33	.6293	.63	.7357	.93	.8238	1.23	.8907	1.53	.9370
.04	.5160	.34	.6331	.64	.7389	.94	.8264	1.24	.8925	1.54	.9382
.05	.5199	.35	.6368	.65	.7422	.95	.8289	1.25	.8944	1.55	.9394
.06	.5239	.36	.6406	.66	.7454	.96	.8315	1.26	.8962	1.56	.9406
.07	.5279	.37	.6443	.67	.7486	.97	.8340	1.27	.8980	1.57	.9418
.08	.5319	.38	.6480	.68	.7517	.98	.8365	1.28	.8997	1.58	.9429
.09	.5359	.39	.6517	.69	.7549	.99	.8389	1.29	.9015	1.59	.9441
.10	.5398	.40	.6554	.70	.7580	1.00	.8413	1.30	.9032	1.60	.9452
.11	.5438	.41	.6591	.71	.7611	1.01	.8438	1.31	.9049	1.61	.9463
.12	.5478	.42	.6628	.72	.7642	1.02	.8461	1.32	.9066	1.62	.9474
.13	.5517	.43	.6664	.73	.7673	1.03	.8485	1.33	.9082	1.63	.9484
.14	.5557	.44	.6700	.74	.7704	1.04	.8508	1.34	.9099	1.64	.9495
.15	.5596	.45	.6736	.75	.7734	1.05	.8531	1.35	.9115	1.65	.9505
.16	.5636	.46	.6772	.76	.7764	1.06	.8554	1.36	.9131	1.66	.9515
.17	.5675	.47	.6803	.77	.7794	1.07	.8577	1.37	.9147	.167	.9525
.18	.5714	.48	.6844	.78	.7823	1.08	.8599	1.38	.9162	1.68	.9535
.19	.5753	.49	.6879	.79	.7852	1.09	.8621	1.39	.9177	1.69	.9545
.20	.5793	.50	.6915	.80	.7881	1.10	.8643	1.40	.9192	1.70	.9554
.21	.5832	.51	.6950	.81	.7910	1.11	.8665	1.41	.9207	1.71	.9564
.22	.5871	.52	.6985	.82	.7939	1.12	.8686	1.42	.9222	1.72	.9573
.23	.5910	.53	.7019	.83	.7967	1.13	.8708	1.43	.9236	1.73	.9582
.24	.5948	.54	.7054	.84	.7995	1.14	.8729	1.44	.9251	1.74	.9591
.25	.5987	.55	.7088	.85	.8023	1.15	.8749	1.45	.9265	1.75	.9599
.26	.6026	.56	.7123	.86	.8051	1.16	.8770	1.46	.9279	1.76	.9608
.27	.6064	.57	.7157	.87	.8078	1.17	.8790	1.47	.9292	1.77	.9616
.28	.6103	.58	.7190	.88	.8106	1.18	.8810	1.48	.9306	1.78	.9625
.29	.6141	.59	.7224	.89	.8133	1.19	.8830	1.49	.9319	1.79	.9633
.30	.6179	.60	.7257	.90	.8159	1.20	.8849	1.50	.9332	1.80	.9641

z	F(z)	z	F(z)	z	F(z)	z	F(z)	z	F(z)	z	F(z)
1.81	.9649	2.21	.9864	2.61	.9955	3.01	.9987	3.41	.9997	3.81	.9999
1.82	.9656	2.22	.9868	2.62	.9956	3.02	.9987	3.42	.9997	3.82	.9999
1.83	.9664	2.23	.9871	2.63	.9957	3.03	.9988	3.43	.9997	3.83	.9999
1.84	.9671	2.24	.9875	2.64	.9959	3.04	.9988	3.44	.9997	3.84	.9999
1.85	.9678	2.25	.9878	2.65	.9960	3.05	.9989	3.45	.9997	3.85	.9999
1.86	.9686	2.26	.9881	2.66	.9961	3.06	.9989	3.46	.9997	3.86	.9999
1.87	.9693	2.27	.9884	2.67	.9962	3.07	.9989	3.47	.9997	3.87	.9999
1.88	.9699	2.28	.9887	2.68	.9963	3.08	.9990	3.48	.9997	3.88	.9999
1.89	.9706	2.29	.9890	2.69	.9964	3.09	.9990	3.49	.9998	3.89	1.0000
1.90	.9713	2.30	.9893	2.70	.9965	3.10	.9990	3.50	.9998	3.90	1.0000
1.91	.9719	2.31	.9896	2.71	.9966	3.11	.9991	3.51	.9998	3.91	1.0000
1.92	.9726	2.32	.9898	2.72	.9967	3.12	.9991	3.52	.9998	3.92	1.0000
1.93	.9732	2.33	.9901	2.73	.9968	3.13	.9991	3.53	.9998	3.93	1.0000
1.94	.9738	2.34	.9904	2.74	.9969	3.14	.9992	3.54	.9998	3.94	1.0000
1.95	.9744	2.35	.9906	2.75	.9970	3.15	.9992	3.55	.9998	3.95	1.0000
1.96	.9750	2.36	.9909	2.76	.9971	3.16	.9992	3.56	.9998	3.96	1.0000
1.97	.9756	2.37	.9911	2.77	.9972	3.17	.9992	3.57	.9998	3.97	1.0000
1.98	.9761	2.38	.9913	2.78	.9973	3.18	.9993	3.58	.9998	3.98	1.0000
1.99	.9767	2.39	.9916	2.79	.9974	3.19	.9993	3.59	.9998	3.99	1.0000
2.00	.9772	2.40	.9918	2.80	.9974	3.20	.9993	3.60	.9998		
2.01	.9778	2.41	.9920	2.81	.9975	3.21	.9993	3.61	.9998		
2.02	.9783	2.42	.9922	2.82	.9976	3.22	.9994	3.62	.9999		
2.03	.9788	2.43	.9925	2.83	.9977	3.23	.9994	3.63	.9999		
2.04	.9793	2.44	.9927	2.84	.9977	3.24	.9994	3.64	.9999		
2.05	.9798	2.45	.9929	2.85	.9978	3.25	.9994	3.65	.9999		
2.06	.9803	2.46	.9931	2.86	.9979	3.26	.9994	3.66	.9999		
2.07	.9808	2.47	.9932	2.87	.9979	3.27	.9995	3.67	.9999		
2.08	.9812	2.48	.9934	2.88	.9980	3.28	.9995	3.68	.9999		
2.09	.9817	2.49	.9936	2.89	.9981	3.29	.9995	3.69	.9999		
2.10	.9821	2.50	.9938	2.90	.9981	3.30	.9995	3.70	.9999		
2.11	.9826	2.51	.9940	2.91	.9982	3.31	.9995	3.71	.9999		
2.12	.0830	2.52	.9941	2.92	.9982	3.32	.9996	3.72	.9999		
2.13	.9834	2.53	.9943	2.93	.9983	3.33	.9996	3.73	.9999		
2.14	.9838	2.54	.9945	2.94	.9984	3.34	.9996	3.74	.9999		
2.15	.9842	2.55	.9946	2.95	.9984	3.35	.9996	3.75	.9999		
2.16	.9846	2.56	.9948	2.96	.9985	3.36	.9996	3.76	.9999		
2.17	.9850	2.57	.9949	2.97	.9985	3.37	.9996	3.77	.9999		
2.18	.9854	2.58	.9951	2.98	.9986	3.38	.9996	3.78	.9999		
2.19	.9857	2.59	.9952	2.99	.9986	3.39	.9997	3.79	.9999		
2.20	.9861	2.60	.9953	3.00	.9986	3.40	.9997	3.80	.9999		

ملحق (٥)

توزيع مربع كا χ^2

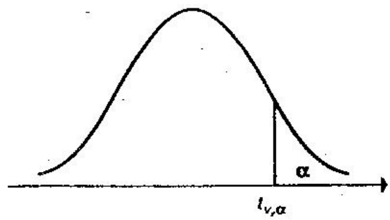


الجدول التالي يعطي مساحة الطرف الأيمن تحت المنحني أي $P_r(\chi^2 > \chi^2_{v,\alpha})$

v	α .995	.990	.975	.950	.900	.100	.050	.025	.010	.005
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

Reproduced with permission from C. M. Thompson, "Tables of percentage points of the chi-square distribution," *Biometrika*, 32 (1941).

ملحق (٦)



توزيع استيودنت (T)

الجدول التالي يعطي مساحة الطرف الأيمن تحت المنحني أي $P_r(T > T_{V,\alpha})$

v	α				
	0.100	0.050	0.025	0.010	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
∞	1.282	1.645	1.960	2.326	2.576

ملحق (٧)

التوزيع الاحتمالي لمتغير ذات الحدين

الجدول التالي يعطى الإحتمال عند $X = x$ حيث $p_r(X=x) = C_x^n \Pi^x (1-\Pi)^{n-x}$

n	*x*	π .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
1	0	.9500	.9000	.8500	.8000	.7500	.7000	.6500	.6000	.5500	.5000
	1	.0500	.1000	.1500	.2000	.2500	.3000	.3500	.4000	.4500	.5000
2	0	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
	2	.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500
3	0	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
	2	.0071	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
	3	.0001	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250
4	0	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625
	1	.1715	.2916	.3685	.4096	.4219	.4116	.3845	.3456	.2995	.2500
	2	.0135	.0486	.0975	.1536	.2109	.2646	.3105	.3456	.3675	.3750
	3	.0005	.0036	.0115	.0256	.0469	.0756	.1115	.1536	.2005	.2500
	4	.0000	.0001	.0005	.0016	.0039	.0081	.0150	.0256	.0410	.0625
5	0	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0312
	1	.2036	.3280	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1562
	2	.0214	.0729	.1382	.2048	.2637	.3087	.3364	.3456	.3369	.3125
	3	.0011	.0081	.0244	.0512	.0879	.1323	.1811	.2304	.2757	.3125
	4	.0000	.0004	.0022	.0064	.0146	.0284	.0488	.0768	.1128	.1562
	5	.0000	.0000	.0001	.0003	.0010	.0024	.0053	.0102	.0185	.0312
6	0	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156
	1	.2321	.3543	.3993	.3932	.3560	.3025	.2437	.1866	.1359	.0938
	2	.0305	.0984	.1762	.2458	.2966	.3241	.3280	.3110	.2780	.2344
	3	.0021	.0146	.0415	.0819	.1318	.1852	.2355	.2765	.3032	.3125
	4	.0001	.0012	.0055	.0154	.0330	.0595	.0951	.1382	.1861	.2344
	5	.0000	.0001	.0004	.0015	.0044	.0102	.0205	.0369	.0609	.0938
	6	.0000	.0000	.0000	.0001	.0002	.0007	.0018	.0041	.0083	.0156
7	0	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078
	1	.2573	.3720	.3960	.3670	.3115	.2471	.1848	.1306	.0872	.0547
	2	.0406	.1240	.2097	.2753	.3115	.3177	.2985	.2613	.2140	.1641
	3	.0036	.0230	.0617	.1147	.1730	.2269	.2679	.2903	.2918	.2734
	4	.0002	.0026	.0109	.0287	.0577	.0972	.1442	.1935	.2388	.2734
	5	.0000	.0002	.0012	.0043	.0115	.0250	.0466	.0774	.1172	.1641
	6	.0000	.0000	.0001	.0004	.0013	.0036	.0084	.0172	.0320	.0547
	7	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0016	.0037	.0078
8	0	.6634	.4305	.2725	.1678	.1001	.0576	.0319	.0168	.0084	.0039
	1	.2793	.3826	.3847	.3355	.2670	.1977	.1373	.0896	.0548	.0312
	2	.0515	.1488	.2376	.2936	.3115	.2965	.2587	.2090	.1569	.1094
	3	.0054	.0331	.0839	.1468	.2076	.2541	.2786	.2787	.2568	.2188
	4	.0004	.0046	.0185	.0459	.0865	.1361	.1875	.2322	.2627	.2734
	5	.0000	.0004	.0026	.0092	.0231	.0467	.0808	.1239	.1719	.2188
	6	.0000	.0000	.0002	.0011	.0038	.0100	.0217	.0413	.0703	.1094
	7	.0000	.0000	.0000	.0001	.0004	.0012	.0033	.0079	.0164	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0017	.0039
9	0	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176

n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
	2	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
10	0	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0105	.0574	.1298	.2013	.2503	.2668	.2522	.2150	.1665	.1172
	4	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0226	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
11	0	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014	.0005
	1	.3293	.3835	.3248	.2362	.1549	.0932	.0518	.0266	.0125	.0054
	2	.0867	.2131	.2866	.2953	.2581	.1998	.1395	.0887	.0513	.0269
	3	.0137	.0710	.1517	.2215	.2581	.2568	.2254	.1774	.1259	.0806
	4	.0014	.0158	.0536	.1107	.1721	.2201	.2428	.2365	.2060	.1611
	5	.0001	.0025	.0132	.0388	.0803	.1321	.1830	.2207	.2360	.2256
	6	.0000	.0003	.0023	.0097	.0268	.0566	.0985	.1471	.1931	.2256
	7	.0000	.0000	.0003	.0017	.0064	.0173	.0379	.0701	.1128	.1611
	8	.0000	.0000	.0000	.0002	.0011	.0037	.0102	.0234	.0462	.0806
	9	.0000	.0000	.0000	.0000	.0001	.0005	.0018	.0052	.0126	.0269
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0007	.0021	.0054
	11	.0000	.9000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0005
12	0	.5404	.2824	.1422	.0687	.0317	.0138	.0057	.0022	.0008	.0002
	1	.3413	.3766	.3012	.2062	.1267	.0712	.0368	.0174	.0075	.0029
	2	.0988	.2301	.2924	.2835	.2323	.1678	.1088	.0639	.0339	.0161
	3	.0173	.0852	.1720	.2362	.2581	.2397	.1954	.1419	.0923	.0537
	4	.0021	.0213	.0683	.1329	.1936	.2311	.2367	.2128	.1700	.1208
	5	.0002	.0038	.0193	.0532	.1032	.1585	.2039	.2270	.2225	.1934
	6	.0000	.0005	.0040	.0155	.0401	.0792	.1281	.1766	.2124	.2256
	7	.0000	.0000	.0006	.0033	.0015	.0291	.0591	.1009	.1489	.1934
	8	.0000	.0000	.0001	.0005	.0024	.0078	.0199	.0420	.0762	.1208
	9	.0000	.0000	.0000	.0001	.0004	.0015	.0048	.0125	.0277	.0537
	10	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0025	.0068	.0161
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0029
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
13	0	.5133	.2542	.1209	.0550	.0238	.0097	.0037	.0013	.0004	.0001
	1	.3512	.3672	.2774	.1787	.1029	.0540	.0259	.0113	.0045	.0016
	2	.1109	.2448	.2937	.2680	.2059	.1388	.0836	.0453	.0220	.0095
	3	.0214	.0997	.1900	.2457	.2517	.2181	.1651	.1107	.0660	.0349
	4	.0028	.0277	.0838	.1535	.2097	.2337	.2222	.1845	.1350	.0873
	5	.0003	.0055	.0266	.0691	.1258	.1803	.2154	.2214	.1989	.1571
	6	.0000	.0008	.0063	.0230	.0559	.1030	.1546	.1968	.2169	.2095
	7	.0000	.0001	.0011	.0058	.0186	.0442	.0833	.1312	.1775	.2095

n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
	8	.0000	.0000	.0001	.0011	.0047	.0142	.0336	.0656	.1089	.1571
	9	.0000	.0000	.0000	.0001	.0009	.0034	.0101	.0243	.0495	.0873
	10	.0000	.0000	.0000	.0000	.0001	.0006	.0022	.0065	.0162	.0349
	11	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0012	.0036	.0095
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
14	0	.4877	.2288	.1028	.0440	.0178	.0068	.0024	.0008	.0002	.0001
	1	.3593	.3559	.2539	.1539	.0832	.0407	.0181	.0073	.0027	.0009
	2	.1229	.2570	.2912	.2501	.1802	.1134	.0634	.0317	.0141	.0056
	3	.0259	.1142	.2056	.2501	.2402	.1943	.1366	.0845	.0462	.0222
	4	.0037	.0348	.0998	.1720	.2202	.2290	.2022	.1549	.1040	.0611
	5	.0004	.0078	.0352	.0860	.1468	.1963	.2178	.2066	.1701	.1222
	6	.0000	.0013	.0093	.0322	.0734	.1262	.1759	.2066	.2088	.1833
	7	.0000	.0002	.0019	.0092	.0280	.0618	.1082	.1574	.1952	.2095
	8	.0000	.0000	.0003	.0020	.0082	.0232	.0510	.0918	.1398	.1833
	9	.0000	.0000	.0000	.0003	.0018	.0066	.0183	.0408	.0762	.1222
	10	.0000	.0000	.0000	.0000	.0003	.0014	.0049	.0136	.0312	.0611
	11	.0000	.0000	.0000	.0000	.0000	.0002	.0010	.0033	.0093	.0222
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0019	.0056
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0009
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
15	0	.4633	.2059	.0874	.0352	.0134	.0047	.0016	.0005	.0001	.0000
	1	.3658	.3432	.2312	.1319	.0668	.0305	.0126	.0047	.0016	.0005
	2	.1348	.2669	.2856	.2309	.1559	.0916	.0476	.0219	.0090	.0032
	3	.0307	.1285	.2184	.2501	.2252	.1700	.1110	.0634	.0318	.0139
	4	.0049	.0428	.1156	.1876	.2252	.2186	.1792	.1268	.0780	.0417
	5	.0006	.0105	.0449	.1032	.1651	.2061	.2123	.1859	.1404	.0916
	6	.0000	.0019	.0132	.0430	.0917	.1472	.1906	.2066	.1914	.1527
	7	.0000	.0003	.0030	.0138	.0393	.0811	.1319	.1771	.2013	.1964
	8	.0000	.0000	.0005	.0035	.0131	.0348	.0710	.1181	.1647	.1964
	9	.0000	.0000	.0001	.0007	.0034	.0116	.0298	.0612	.1048	.1527
	10	.0000	.0000	.0000	.0001	.0007	.0030	.0096	.0245	.0515	.0916
	11	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0074	.0191	.0417
	12	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052	.0139
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0032
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
16	0	.4401	.1853	.0743	.0281	.0100	.0033	.0010	.0003	.0001	.0000
	1	.3706	.3294	.2097	.1126	.0535	.0228	.0087	.0030	.0009	.0002
	2	.1463	.2745	.2775	.2111	.1336	.0732	.0353	.0150	.0056	.0018
	3	.0359	.1423	.2285	.2463	.2079	.1465	.0888	.0468	.0215	.0085
	4	.0061	.0514	.1311	.2001	.2552	.2040	.1553	.1014	.0572	.0278
	5	.0008	.0137	.0555	.1201	.1802	.2099	.2008	.1623	.1123	.0667
	6	.0001	.0028	.0180	.0550	.1101	.1649	.1982	.1983	.1684	.1222
	7	.0000	.0004	.0045	.0197	.0524	.1010	.1524	.1889	.1969	.1746
	8	.0000	.0001	.0009	.0055	.0197	.0487	.0923	.1417	.1812	.1964
	9	.0000	.0000	.0001	.0012	.0058	.0185	.0442	.0840	.1318	.1746
	10	.0000	.0000	.0000	.0002	.0014	.0056	.0167	.0392	.0755	.1222
	11	.0000	.0000	.0000	.0000	.0002	.0013	.0049	.0142	.0337	.0667
	12	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0040	.0115	.0278
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0029	.0085
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002

n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
17	0	.4181	.1668	.0631	.0225	.0075	.0023	.0007	.0002	.0000	.0000
	1	.3741	.3150	.1893	.0957	.0426	.0169	.0060	.0019	.0005	.0001
	2	.1575	.2800	.2673	.1914	.1136	.0581	.0260	.0102	.0035	.0010
	3	.0415	.1556	.2359	.2393	.1893	.1245	.0701	.0341	.0144	.0052
	4	.0076	.0605	.1457	.2093	.2209	.1868	.1320	.0796	.0411	.0182
	5	.0010	.0175	.0068	.1361	.1914	.2081	.1849	.1379	.0875	.0472
	6	.0001	.0039	.0236	.0680	.1276	.1784	.1991	.1839	.1432	.0944
	7	.0000	.0007	.0065	.0267	.0668	.1201	.1685	.1927	.1841	.1484
	8	.0000	.0001	.0014	.0084	.0279	.0644	.1134	.1606	.1883	.1855
	9	.0000	.0000	.0003	.0021	.0093	.0276	.0611	.1070	.1540	.1855
	10	.0000	.0000	.0000	.0004	.0025	.0095	.0263	.0571	.1008	.1484
	11	.0000	.0000	.0000	.0001	.0005	.0026	.0090	.0242	.0525	.0944
	12	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0081	.0215	.0472
	13	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0021	.0068	.0182
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
18	0	.3972	.1501	.0536	.0180	.0056	.0016	.0004	.0001	.0000	.0000
	1	.3763	.3002	.1704	.0811	.0338	.0126	.0042	.0012	.0003	.0001
	2	.1683	.2835	.2556	.1723	.0958	.0458	.0190	.0069	.0022	.0006
	3	.0473	.1680	.2406	.2297	.1704	.1046	.0547	.0246	.0095	.0031
	4	.0093	.0700	.1592	.2153	.2130	.1681	.1104	.0614	.0291	.0117
	5	.0014	.0218	.0787	.1507	.1988	.2017	.1664	.1146	.0666	.0327
	6	.0002	.0052	.0301	.0816	.1436	.1873	.1941	.1655	.1181	.0708
	7	.0000	.0010	.0091	.0350	.0820	.1376	.1792	.1892	.1657	.1214
	8	.0000	.0002	.0022	.0120	.0376	.0811	.1327	.1734	.1864	.1669
	9	.0000	.0000	.0004	.0033	.0139	.0386	.0794	.1284	.1694	.1855
	10	.0000	.0000	.0001	.0008	.0042	.0149	.0385	.0771	.1248	.1669
	11	.0000	.0000	.0000	.0001	.0010	.0046	.0151	.0374	.0742	.1214
	12	.0000	.0000	.0000	.0000	.0002	.0012	.0047	.0145	.0354	.0708
	13	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0044	.0134	.0327
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0039	.0117
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0009	.0031
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
19	0	.3774	.1351	.0456	.0144	.0042	.0011	.0003	.0001	.0000	.0000
	1	.3774	.2852	.1529	.0685	.0268	.0093	.0029	.0008	.0002	.0000
	2	.1787	.2852	.2428	.1540	.0803	.0358	.0138	.0046	.0013	.0003
	3	.0533	.1796	.2428	.2182	.1517	.0869	.0422	.0175	.0062	.0018
	4	.0112	.0798	.1714	.2182	.2023	.1419	.0909	.0467	.0203	.0074
	5	.0018	.0266	.0907	.1636	.2023	.1916	.1468	.0933	.0497	.0222
	6	.0002	.0069	.0374	.0955	.1574	.1916	.1844	.1451	.0949	.0518
	7	.0000	.0014	.0122	.0443	.0974	.1525	.1844	.1797	.1443	.0961
	8	.0000	.0002	.0032	.0166	.0487	.0981	.1489	.1797	.1771	.1442
	9	.0000	.0000	.0007	.0051	.0198	.0514	.0980	.1464	.1771	.1762
	10	.0000	.0000	.0001	.0013	.0066	.0220	.0528	.0976	.1449	.1762
	11	.0000	.0000	.0000	.0003	.0018	.0077	.0233	.0532	.0970	.1442
	12	.0000	.0000	.0000	.0000	.0004	.0022	.0083	.0237	.0529	.0961
	13	.0000	.0000	.0000	.0000	.0001	.0005	.0024	.0085	.0233	.0518
	14	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0082	.0222

n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
						π					
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0022	.0074
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
20	0	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.3774	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.1887	.2852	.2293	.1369	.0669	.0278	.0100	.0031	.0008	.0002
	3	.0596	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0133	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0022	.0319	.1028	.1746	.2023	.1789	.1272	.0746	.0365	.0148
	6	.0003	.0089	.0454	.1091	.1686	.1916	.1712	.1244	.0746	.0370
	7	.0000	.0020	.0160	.0545	.1124	.1643	.1844	.1659	.1221	.0739
	8	.0000	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.1201
	9	.0000	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602
	10	.0000	.0000	.0002	.0020	.0099	.0308	.0686	.1171	.1593	.1762
	11	.0000	.0000	.0000	.0005	.0030	.0120	.0336	.0710	.1185	.1602
	12	.0000	.0000	.0000	.0001	.0008	.0039	.0136	.0355	.0727	.1201
	13	.0000	.0000	.0000	.0000	.0002	.0010	.0045	.0146	.0366	.0739
	14	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0049	.0150	.0370
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0049	.0148
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0046
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

تابع ملحق (٧)
دالة التوزيع التراكمية لمتغير ذي الحدين

الجدول التالى يعطى $P_r(X \le x)$ أي : $P_r(X \le x\ ) = \sum_{i=1} C^{n}_{x_i} \Pi \cdot (1-\Pi)^{n-}$

n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.500
2	0	.902	.81	.722	.64	.562	.49	.422	.36	.302	.25
	1	.998	.99	.978	.96	.937	.91	.877	.84	.797	.75
	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	0	.857	.729	.614	.512	.422	.343	.275	.216	.166	.125
	1	.993	.972	.939	.896	.844	.784	.718	.648	.575	.500
	2	1.00	.999	.997	.992	.984	.973	.957	.936	.909	.875
	3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
4	0	.815	.656	.522	.41	.316	.24	.179	.13	.092	.062
	1	.986	.948	.89	.819	.738	.652	.563	.475	.391	.312
	2	1.00	.996	.988	.973	.949	.916	.874	.821	.759	.687
	3	1.00	1.00	.999	.998	.996	.992	.985	.974	.959	.937
	4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
5	0	.774	.59	.444	.328	.237	.168	.116	.078	.05	.031
	1	.977	.919	.835	.737	.633	.528	.428	.337	.256	.187
	2	.999	.991	.973	.942	.896	.837	.765	.683	.593	.500
	3	1.00	1.00	.998	.993	.984	.969	.946	.913	.869	.812
	4	1.00	1.00	1.00	1.00	.999	.998	.995	.99	.982	.969
	5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
6	0	.735	.531	.377	.262	.178	.118	.075	.047	.028	.016
	1	.967	.886	.776	.655	.534	.42	.319	.233	.164	.109
	2	.998	.984	.953	.901	.831	.744	.647	.544	.442	.344
	3	1.00	.999	.994	.983	.962	.93	.883	.821	.745	.656
	4	1.00	1.00	1.00	.998	.995	.989	.978	.959	.931	.891
	5	1.00	1.00	1.00	1.00	1.00	.999	.998	.996	.992	.984
	6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
7	0	.698	.478	.321	.21	.133	.082	.049	.028	.015	.008
	1	.956	.85	.717	.577	.445	.329	.234	.159	.102	.062
	2	.996	.974	.926	.852	.756	.647	.532	.42	.316	.227
	3	1.00	.997	.988	.967	.929	.874	.80	.71	.608	.500
	4	1.00	1.00	.999	.995	.987	.971	.944	.904	.847	.773
	5	1.00	1.00	1.00	1.00	.999	.996	.991	.981	.964	.937
	6	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.996	.992
	7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
8	0	.663	.43	.272	.168	.10	.058	.032	.017	.008	.004
	1	.943	.813	.657	.503	.367	.255	.169	.106	.063	.035
	2	.994	.962	.895	.797	.679	.552	.428	.315	.22	.145
	3	1.00	.995	.979	.944	.886	.806	.706	.594	.477	.363
	4	1.00	1.00	.997	.99	.973	.942	.894	.826	.74	.637
	5	1.00	1.00	1.00	.999	.996	.989	.975	.95	.912	.855
	6	1.00	1.00	1.00	1.00	1.00	.999	.996	.991	.982	.965
	7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.996
	8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
9	0	.63	.387	.232	.134	.075	.04	.021	.01	.005	.002
	1	.929	.775	.599	.436	.30	.196	.121	.071	.039	.020
	2	.992	.947	.859	.738	.601	.463	.337	.232	.15	.090
	3	.999	.992	.966	.914	.834	.73	.609	.483	.361	.254
	4	1.00	.999	.994	.98	.951	.901	.828	.733	.621	.500

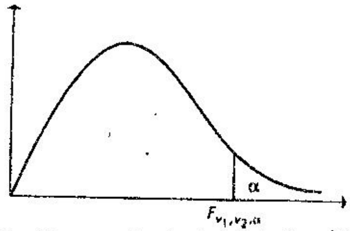
n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.500
	5	1.00	1.00	.999	.997	.99	.975	.946	.901	.834	.746
	6	1.00	1.00	1.00	1.00	.999	.996	.989	.975	.95	.910
	7	1.00	1.00	1.00	1.00	1.00	1.00	.999	.996	.991	.980
	8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998
	9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
10	0	.599	.349	.197	.107	.056	.028	.013	.006	.003	.001
	1	.914	.736	.544	.376	.244	.149	.086	.046	.023	.011
	2	.988	.93	.82	.678	.526	.383	.262	.167	.10	.055
	3	.999	.987	.95	.879	.776	.65	.514	.382	.266	.172
	4	1.00	.998	.99	.967	.922	.85	.751	.633	.504	.377
	5	1.00	1.00	.999	.994	.98	.953	.905	.834	.738	.623
	6	1.00	1.00	1.00	.999	.996	.989	.974	.945	.898	.828
	7	1.00	1.00	1.00	1.00	1.00	.998	.995	.988	.973	.945
	8	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.995	.989
	9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
	10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
11	0	.569	.314	.167	.086	.042	.02	.009	.004	.001	.000
	1	.898	.697	.492	.322	.197	.113	.061	.03	.014	.006
	2	.985	.91	.779	.617	.455	.313	.20	.119	.065	.033
	3	.998	.981	.931	.839	.713	.57	.426	.296	.191	.113
	4	1.00	.997	.984	.95	.885	.79	.668	.533	.397	.274
	5	1.00	1.00	.997	.988	.966	.922	.851	.753	.633	.500
	6	1.00	1.00	1.00	.998	.992	.978	.95	.901	.826	.726
	7	1.00	1.00	1.00	1.00	.999	.996	.988	.971	.939	.887
	8	1.00	1.00	1.00	1.00	1.00	.999	.998	.994	.985	.967
	9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.994
	10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
	11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
12	0	.54	.282	.142	.069	.032	.014	.006	.002	.001	.000
	1	.882	.659	.443	.275	.158	.085	.042	.02	.008	.003
	2	.98	.889	.736	.558	.391	.253	.151	.083	.042	.019
	3	.998	.974	.908	.795	.649	.493	.347	.225	.134	.073
	4	1.00	.996	.976	.927	.842	.724	.583	.438	.304	.194
	5	1.00	.999	.995	.981	.946	.882	.787	.665	.527	.387
	6	1.00	1.00	.999	.996	.986	.961	.915	.842	.739	.613
	7	1.00	1.00	1.00	.999	.997	.991	.974	.943	.888	.806
	8	1.00	1.00	1.00	1.00	1.00	.998	.994	.985	.964	.927
	9	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997	.992	.981
	10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997
	11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
	12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
13	0	.513	.254	.121	.055	.024	.01	.004	.001	.00	.000
	1	.865	.621	.398	.234	.127	.064	.03	.013	.005	.002
	2	.975	.866	.692	.502	.333	.202	.113	.058	.027	.011
	3	.997	.966	.882	.747	.584	.421	.278	.169	.093	.046
	4	1.00	.994	.966	.901	.794	.654	.501	.353	.228	.133
	5	1.00	.999	.992	.97	.92	.835	.716	.574	.427	.291
	6	1.00	1.00	.999	.993	.976	.938	.871	.771	.644	.50
	7	1.00	1.00	1.00	.999	.994	.982	.954	.902	.821	.709
	8	1.00	1.00	1.00	1.00	.999	.996	.987	.968	.93	.867
	9	1.00	1.00	1.00	1.00	1.00	.999	.997	.992	.98	.954
	10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.996	.989

n	x	π									
		.05	.10	.15	.20	.25	.30	.35	.40	.45	.500
	11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998
	12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
14	0	.488	.229	.103	.044	.018	.007	.002	.001	.00	.000
	1	.847	.585	.357	.198	.101	.047	.021	.008	.003	.001
	2	.97	.842	.648	.448	.281	.161	.084	.04	.017	.006
	3	.996	.956	.853	.698	.521	.355	.22	.124	.063	.029
	4	1.00	.991	.953	.87	.742	.584	.423	.279	.167	.090
	5	1.00	.999	.988	.956	.888	.781	.641	.486	.337	.212
	6	1.00	1.00	.998	.988	.962	.907	.816	.692	.546	.395
	7	1.00	1.00	1.00	.998	.99	.969	.925	.85	.741	.605
	8	1.00	1.00	1.00	1.00	.998	.992	.976	.942	.881	.788
	9	1.00	1.00	1.00	1.00	1.00	.998	.994	.982	.957	.910
	10	1.00	1.00	1.00	1.00	1.00	1.00	.999	.996	.989	.971
	11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998	.994
	12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
	13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
15	0	.463	.206	.087	.035	.013	.005	.002	.00	.00	.000
	1	.829	.549	.319	.167	.08	.035	.014	.005	.002	.000
	2	.964	.816	.604	.398	.236	.127	.062	.027	.011	.004
	3	.995	.944	.823	.648	.461	.297	.173	.091	.042	.018
	4	.999	.987	.938	.836	.686	.515	.352	.217	.12	.059
	5	1.00	.998	.983	.939	.852	.722	.564	.403	.261	.151
	6	1.00	1.00	.996	.982	.943	.869	.755	.61	.452	.304
	7	1.00	1.00	.999	.996	.983	.95	.887	.787	.654	.500
	8	1.00	1.00	1.00	.999	.996	.985	.958	.905	.818	.696
	9	1.00	1.00	1.00	1.00	.999	.996	.988	.966	.923	.849
	10	1.00	1.00	1.00	1.00	1.00	.999	.997	.991	.975	.941
	11	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994	.982
	12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.996
	13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
16	0	.44	.185	.074	.028	.01	.003	.001	.00	.00	.000
	1	.811	.515	.284	.141	.063	.026	.01	.003	.001	.000
	2	.957	.789	.561	.352	.197	.099	.045	.018	.007	.002
	3	.993	.932	.79	.598	.405	.246	.134	.065	.028	.011
	4	.999	.983	.921	.798	.63	.45	.289	.167	.085	.038
	5	1.00	.997	.976	.918	.81	.66	.49	.329	.198	.105
	6	1.00	.999	.994	.973	.92	.825	.688	.527	.366	.227
	7	1.00	1.00	.999	.993	.973	.926	.841	.716	.563	.402
	8	1.00	1.00	1.00	.999	.993	.974	.933	.858	.744	.598
	9	1.00	1.00	1.00	1.00	.998	.993	.977	.942	.876	.773
	10	1.00	1.00	1.00	1.00	1.00	.998	.994	.981	.951	.895
	11	1.00	1.00	1.00	1.00	1.00	1.00	.999	.995	.985	.962
	12	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997	.989
	13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998
	14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
17	0	.418	.167	.063	.023	.008	.002	.001	.00	.00	.000
	1	.792	.482	.252	.118	.05	.019	.007	.002	.001	.000
	2	.95	.762	.52	.31	.164	.077	.033	.012	.004	.001
	3	.991	.917	.756	.549	.353	.202	.103	.046	.018	.006
	4	.999	.978	.901	.758	.574	.389	.235	.126	.06	.025
	5	1.00	.995	.968	.894	.765	.597	.42	.264	.147	.072
	6	1.00	.999	.992	.962	.893	.775	.619	.448	.29	.166

n	x	.05	.10	.15	.20	.25	.30	.35	.40	.45	.500
						π					
	7	1.00	1.00	.998	.989	.96	.895	.787	.641	.474	.315
	8	1.00	1.00	1.00	.997	.988	.96	.901	.801	.663	.500
	9	1.00	1.00	1.00	1.00	.997	.987	.962	.908	.817	.685
	10	1.00	1.00	1.00	1.00	.999	.997	.988	.965	.917	.834
	11	1.00	1.00	1.00	1.00	1.00	.999	.997	.989	.97	.928
	12	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997	.991	.975
	13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994
	14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
	15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
18	0	.397	.15	.054	.018	.006	.002	.00	.00	.00	.000
	1	.774	.45	.224	.099	.039	.014	.005	.001	.00	.000
	2	.942	.734	.48	.271	.135	.06	.024	.008	.003	.001
	3	.989	.902	.72	.501	.306	.165	.078	.033	.012	.004
	4	.998	.972	.879	.716	.519	.333	.189	.094	.041	.015
	5	1.00	.994	.958	.867	.717	.534	.355	.209	.108	.048
	6	1.00	.999	.988	.949	.861	.722	.549	.374	.226	.119
	7	1.00	1.00	.997	.984	.943	.859	.728	.563	.391	.240
	8	1.00	1.00	.999	.996	.981	.94	.861	.737	.578	.407
	9	1.00	1.00	1.00	.999	.995	.979	.94	.865	.747	.593
	10	1.00	1.00	1.00	1.00	.999	.994	.979	.942	.872	.760
	11	1.00	1.00	1.00	1.00	1.00	.999	.994	.98	.946	.881
	12	1.00	1.00	1.00	1.00	1.00	1.00	.999	.994	.982	.952
	13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.995	.985
	14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.996
	15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
	16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
19	0	.377	.135	.046	.014	.004	.001	.00	.00	.00	.000
	1	.755	.42	.198	.083	.031	.01	.003	.001	.00	.000
	2	.933	.705	.441	.237	.111	.046	.017	.005	.002	.000
	3	.987	.885	.684	.455	.263	.133	.059	.023	.008	.002
	4	.998	.965	.856	.673	.465	.282	.15	.07	.028	.010
	5	1.00	.991	.946	.837	.668	.474	.297	.163	.078	.032
	6	1.00	.998	.984	.932	.825	.666	.481	.308	.173	.084
	7	1.00	1.00	.996	.977	.923	.818	.666	.488	.317	.180
	8	1.00	1.00	.999	.993	.971	.916	.815	.667	.494	.324
	9	1.00	1.00	1.00	.998	.991	.967	.913	.814	.671	.500
	10	1.00	1.00	1.00	1.00	.998	.989	.965	.912	.816	.676
	11	1.00	1.00	1.00	1.00	1.00	.997	.989	.965	.913	.820
	12	1.00	1.00	1.00	1.00	1.00	.999	.997	.988	.966	.916
	13	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997	.989	.968
	14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.997	.990
	15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999	.998
	16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000
20	0	.358	.122	.039	.012	.003	.001	.00	.00	.00	.000
	1	.736	.392	.176	.069	.024	.008	.002	.001	.00	.000
	2	.925	.677	.405	.206	.091	.035	.012	.004	.001	.000
	3	.984	.867	.648	.411	.225	.107	.044	.016	.005	.001
	4	.997	.957	.83	.63	.415	.238	.118	.051	.019	.006
	5	1.00	.989	.933	.804	.617	.416	.245	.126	.055	.021
	6	1.00	.998	.978	.913	.786	.608	.417	.25	.13	.058
	7	1.00	1.00	.994	.968	.898	.772	.601	.416	.252	.132
	8	1.00	1.00	.999	.99	.959	.887	.762	.596	.414	.252
	9	1.00	1.00	1.00	.997	.986	.952	.878	.755	.591	.412

n	x	π									
		.05	.10	.15	.20	.25	.30	.35	.40	.45	.500
	10	1.00	1.00	1.00	.999	.996	.983	.947	.872	.751	.588
	11	1.00	1.00	1.00	1.00	.999	.995	.98	.943	.869	.748
	12	1.00	1.00	1.00	1.00	1.00	.999	.994	.979	.942	.868
	13	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994	.979	.942
	14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994	.979
	15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.998	.994
	16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	.999
	17	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000

ملحق (٨)
التوزيع الاحتمالي لفشر (F)



الجدول التالي يعطي مساحة الطرف الأيمن تحت المنحني اي $P_r(X < F_{V_1,V_2,\alpha})$ عند

درجات الحرية V_1, V_2

$\alpha = .05$

DENOMINATOR v_2	NUMERATOR v_1																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.67	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64

$\alpha = .01$

DENOMINATOR v_2	NUMERATOR v_1: 1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.48	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.30	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.37	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.30	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	3.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	2.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.16	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.49	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00

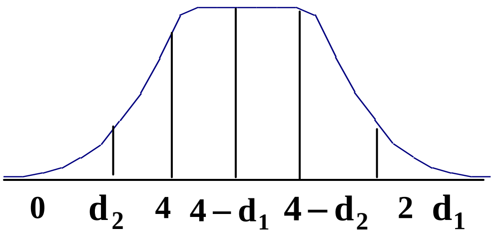
ملحق رقم (٩)

التوزيع الاحتمالي لكولومجروف – سيمزوف – القيم الحرجة لـ K_n^{α}

n \ α	0.10	0.05	0.025	0.01	0.005
1	0.90	0.950	0.975	0.990	0.995
2	0.684	0.776	0.842	0.900	0.929
3	0.565	0.636	0.708	0.785	0.829
4	0.493	0.565	0.624	0.689	0.734
5	0.447	0.509	0.563	0.627	0.669
6	0.410	0.468	0.519	0.577	0.616
7	0.381	0.436	0.473	0.538	0.576
8	0.358	0.410	0.454	0.507	0.542
9	0.339	0.387	0.430	0.480	0.513
10	0.323	0.369	0.409	0.457	0.489
11	0.308	0.352	0.391	0.437	0.468
12	0.296	0.338	0.375	0.419	0.449
13	0.285	0.326	0.361	0.404	0.433
14	0.275	0.314	0.349	0.390	0.418
15	0.266	0.304	0.338	0.377	0.404
16	0.258	0.295	0.327	0.366	0.392
17	0.250	0.286	0.318	0.355	0.381
18	0.244	0.279	0.309	0.346	0.370
19	0.237	0.271	0.301	0.337	0.361
20	0.232	0.265	0.294	0.329	0.352

ملحق رقم (١٠)

التوزيع الاحتمالي لديربن واتسون (توزيع D)



الجدول يعطي قيم d_1 ، d_2 عند درجات الحرية (1 , n)

n \ α	0.05		0.025		0.01	
	d_1	d_2	d_1	d_2	d_1	d_2
15	1.08	1.36	0.95	1.23	0.81	1.07
16	1.10	1.37	0.98	1.24	0.84	1.09
17	1.13	1.38	1.01	1.25	0.87	1.10
18	1.16	1.39	1.03	1.26	0.90	1.12
19	1.18	1.40	1.06	1.28	0.93	1.13
20	1.20	1.41	1.08	1.29	0.95	1.15
21	1.22	1.42	1.10	1.30	0.97	1.16
22	1.24	1.43	1.12	1.31	1.00	1.17
23	1.26	1.44	1.14	1.32	1.02	1.19
24	1.27	1.45	1.16	1.33	1.04	1.20
25	1.29	1.45	1.18	1.34	1.05	1.21
26	1.30	1.46	1.19	1.35	1.07	1.22
27	1.32	1.47	1.21	1.36	1.09	1.23
28	1.33	1.48	1.22	1.37	1.10	1.24

يعطى القيم الدالة $e^{-\lambda}$ عند القيمة المختلفة لـ λ $(0<\lambda<1)$

λ	0	1	2	3	4	5	6	7	8	9
0.0	1.00	.990	.9802	.9704	.9608	.9512	.9414	.9324	.9231	.9139
0.1	.9048	.8958	.8869	.8781	.8694	.8607	.8521	.8437	.8353	.8270
0.2	.8187	.8106	.8025	.7945	.7866	.7788	.7711	.7634	.7558	.7483
0.3	.7408	.7334	.7261	.7189	.7118	.7047	.6977	.6907	.6839	.6771
0.4	.6703	.6636	.6570	.6505	.6440	.6376	.6313	.6250	.6188	.6126
0.5	.6065	.6005	.5945	.5886	.5827	.5770	.5712	.5655	.5599	.5543
0.6	.5488	.5434	.5379	.5326	.5273	.5220	.5169	.5117	.5066	.5016
0.7	.4966	.4916	4868	.4819	.4771	.4724	.4677	.4630	.4584	.4538
0.8	.4493	.449	.4404	.4360	.4317	.427	.4232	.4190	.4148	.4107
0.9	.4066	.4025	.3985	.3946	.3906	.3867	.3829	.3791	.3753	.3716

قائمة المراجع

أولاً: المراجع العربية

١- أ.د. جلال مصطفى الصياد ، أ.د مصطفى جلال (٢٠٠٠): "المعاينة الإحصائية" مكتبة عين شمش – القاهرة.

٢- د. ربيع زكي عامر (١٩٨٩): " تحليل الانحدار – أساليبه وتطبيقاته باستخدام البرنامج الجاهز (SPSS/PCT) " . معهد الدراسات والبحوث الإحصائية – جامعة القاهرة.

٣- أ.د سمير عاشور ، أ.د سامية أبو الفتوح (١٩٩٤): "مقدمة لنظرية العينات" – معهد الدراسات والبحوث الإحصائية – جامعة القاهرة.

٤- أ.د سمير عاشور (١٩٨٦): "مقدمة في الإحصاء التحليلي" - معهد الدراسات والبحوث الإحصائية – جامعة القاهرة.

٥- أ.د سمير عاشور ، أ.د سامية أبو الفتوح (١٩٩٧): "التحليل الإحصائي باستخدام حزم SPSS" - معهد الدراسات والبحوث الإحصائية – جامعة القاهرة.

٦- أ.د عبد الله توفيق الهلباوي (١٩٨٧): "الإحصاء التطبيقي" – مكتبة عين شمس – القاهرة.

٧- أ.د عفاف على حسن الدش (١٩٩١): "نماذج الانحدار" مكتبة عين شمس- القاهرة.

٨- أ.د عفاف على حسن الدش (٢٠٠٥): "الإحصاء التطبيقي للتجاريين" – الجزء الأول – الطبعة السادسة – جهاز نشر وتوزيع الكتاب الجامعي – جامعة حلوان.

٩- أ.د عفاف على حسن الدش (١٩٨٧): "بحوث العمليات واتخاذ القرارات" – مكتبة عين شمس – القاهرة.

١٠- أ.د محمد عبد السميع عناني (١٩٩٣): "مبادئ الاقتصاد القياسي

النظـري والتطبيقـي" – الطبعـة الثانيـة – جامعـة الزقـازيق – جمهورية مصر العربية.

١١- موراى د. شبيجل (١٩٨٩): "الإحصاء" سلسلة ملخصات شوم – نظريات ومسائل الدار الدولية للنشر والتوزيع – القاهرة.

١٢- أ.د فتحي محمد على (١٩٨٤): "الإحصاء المتقدم" – مكتبة عين شمس – القاهرة.

١٣- أ.د نادية مكاري (١٩٧٢): "الإحصاء الاقتصادي" – كلية الاقتصاد والعلوم السياسية – القاهرة.

ثانياً: المراجع الأجنبية

14- Anderson, T.W. (1971): "An Introduction to Multivariate Statistical Analysis" , John Wiley & Sond, Inc. , New York.

15- David J.B , Andrew , F.F. and Sally I.M.C. (1991): "Statistical Technique for Manpower Planning" , Second Edition , Joho Wiley & Sons , New York.

16- Frank , S. (1993): "Applied Mathematics for Business , Economic and The Social Sciences" , McGraw-Hill International Edition , New York.

17- Forgiel , Director (1978): "The Statistics Problems Solver" , Research and Education Association , New York.

18- Gallant , A.R. (1987): "Nonlinear Statistics Models", Hohn Wiley & Sons , Inc. , New York.

19- Hamburg , M. (1970): "Statistical Analysis for

Decision Making" , Harsart , Brace & World , Inc. , New York.

20- Harry Frank and Steven Althoes (1994): "Statistics – Concepts and Applications" , Cambridge University Press.

21- Heinz Kohler (1994): "Statistics for Business and Economics" Third Edition , Harper Collins College Publishers , U.S.A.

22- Janes T. , George P. Lerry Tery S. (2001): "Statistics for Business and Economics" Eighth Edition , Prentice Hall , New Jersey.

23- Lawrence L. Lapin (1994): "Quantitative Methods For Business Decisions , With Cases" , Sixth Edition , The Dryden Press , New York.

24- Lawrence L. Lapin (1993): "Statistics for Modern Business Decisions" Sixth Edition , The Dryden Press , New York.

25- Kendal , M.G. and Studart , A. (1982): "The Advanced Theory of Statistics" , Vol. 11.2 , Charles Griggin & Company Limited , London.

26- Mcallister , H.E. (1995): "Elements of Business and Economic Statistics: Learning By Objectives" , Hohn Wiley & Sons Inc. , New York.

27- Mood , A.S. and Graybill , F.A. (1963): "Introduction to The Theory of Statistics" , McGraw Hill Book Company , New York.

28- Prem S. Mann (1995): "Statistics For Business and

Economics" , Hohn Wiley & Sons Inc. , New York.

29- Panl Willan , Betty M. (2003): "Statistics For Business and Economics" , Fifth Edition , Inc. , New Jersey.

30- Spiegel , M.R. (1972): "Schaums Outline of Theory and Problems of Statistics" , McGraw-Hill International Book Company , New York.

31- Shao & Shao (1990): "Mathematics For Management and Finance" , Sixth Edition , South-Western – Publishing Co. , West Chicago.

32- William A. Spurr and Charles P. Bonini (1973): "Statistical Analysis For Business Decisions" , Richard D. Irwin , Inc. , London.

كتب للمؤلفة

- الإحصـــاء التطبيقـــي – الجـــزء الأول (٢٠٠٠) ، (٢٠٠٥) جهـــاز نشـــر وتوزيع الكتاب الجامعي – جامعة حلوان.
- الإحصاء التطبيقي – الجزء الثاني (٢٠٠١) ، (٢٠٠٤) ، (٢٠٠٦) جهاز نشر وتوزيع الكتاب الجامعي – جامعة حلوان.
- رياضيات الأعمال جامعة حلوان (٢٠٠٢) ، (١٩٩٩) ، (١٩٩٧) جهاز نشر وتوزيع الكتاب الجامعي – جامعة حلوان.
- مقدمة في بحوث العمليات (٢٠٠٢) جهاز نشر وتوزيع الكتاب الجامعي – جامعة حلوان.
- رياضـيات الاسـتثمار (٢٠٠٦) ، (٢٠٠٣) ، (٢٠٠١) ، (٢٠٠٠) جهـاز نشر وتوزيع الكتاب الجامعي – جامعة حلوان.
- الإحصاء التطبيقي للتجاريين (١٩٩٧) ، (١٩٩٤) ، (١٩٩٢) مكتبة عين شمس – القاهرة.
- الإحصاء وصناعة القرارات (١٩٩٦) - مكتبة عين شمس – القاهرة.
- نماذج الانحدار (١٩٩٠) - مكتبة عين شمس – القاهرة.
- بحوث العمليات واتخاذ القرارات (١٩٨٧) - مكتبة عين شمس – القاهرة.