الأكاديمية العربية الدولية

الأكاديمية العربية الدولية
Arab International Academy

الأكاديمية العربية الدولية

المقررات الجامعية

www.aiacademy.info | care@aiacademy.info

علم الإحصاء:

❖ تعاريف ومصطلحات:

1- علم الإحصاء:

وهو أحد أهم فروع الرياضيات وهو من أهم الفروع وهو يستخدم في كل مجالات الحياة كما ويدخل علم الإحصاء في كل العلوم الأخرى وكل العلوم مهما كانت تحتاج إلى علم الإحصاء **فمثلاً:**(العلوم العلمية – العلوم الإنسانية والأدبية – العلوم الرياضية – الفنون – العلوم العسكرية – العلوم الاجتماعية) لذلك يعتبر علم الإحصاء هو الأساس في كل العلوم.

- **ما هي وظيفة علم الإحصاء(بماذا يهتم علم الإحصاء):**

يهتم علم الإحصاء بشكل أساسي في جمع وتنظيم وتلخيص وعرض وتحليل وتفسير ونشر البيانات الإحصائية.

2- البيانات Data:

هي جزء من المعلومات الموثوق والمدعم بالأرقام أو معلومات تكون مرتبة ومنظمة ومخزنة بطريقة سهلة التداول و الفهم والتفسير ومدعمة بالأرقام ويمكن تحليلها وتفسيرها وليست كل المعلومات بيانات.

لماذا يعتبر علم الإحصاء ضروري لكل العلوم؟:

كل علم من العلوم فيه بيانات ويتعامل مع البيانات وبالتالي نحتاج إلى الإحصاء ليقوم بمعالجة وتحليل وتقييم وتفسير البيانات .

أو بتعبير أخر كل العلوم أساسها البيانات وهي تحتاج إلى دراسات وأبحاث تطبيقية واختبارات وكل بحث أو اختبار أو تجربة يحتاج إلى علم الإحصاء.

من أين نحصل على البيانات:؟

1– تأتي البيانات بشكل أساسي من مراقبة الظواهر ومشاهدة الظواهر الموجودة بالطبيعة.

2– من خلال إجراء التجارب والاختبارات.

3–من خلال الأبحاث والدراسات والمقابلات.

فوائد علم الإحصاء:

1–يمكن تحليل ومعالجة البيانات.

2–تفسير النتائج للأبحاث والتجارب والاختبارات.

3– الكشف عن صلاحية وجودة البيانات لأي بحث ولأي منتج ولأي تجربة جديدة.

4–تقييم أو تقويم الأبحاث والدراسات.

3- المجتمع Populatıon:

عدد لانهائي من الإفراد أو العناصر التي تتعايش مع بعضها البعض وتتميز بخصائص ومواصفات تميزها عن بقية المجتمعات حيث أنه لكل مجتمع خصائص ومميزات تميزه عن المجتمعات الأخرى.

المجتمع البشري يقسم إلى مجتمعات صغيرة (إحصائية) هذا المجتمع الإحصائي معروف عدده وتميزه مجموعة من الصفات الزمانية و المكانية وهذا النوع هو الذي يستخدم في دراسات الأبحاث.

أهم ميزة في المجتمع الإحصائي هو عدد الأفراد (**n**) والمتوسط (**X**).

4- العينة Sample:

هي جزء من المجتمع يجب ألا يقل عدد أفرادها عن (2 – 10 %) من عدد أفراد المجتمع وهي تؤخذ بطريقة عشوائية **وتوجد منها الأنواع التالية**:

أ– العينة العشوائية البسيطة:

هي أصغر وأبسط أنواع العينات تؤخذ بطريقة عشوائية ويجب أن لا يقل عدد أفرادها عن **(30)** فرد أو عنصر.

ب– العينة العشوائية الطبقية:

وهي تؤخذ من المجتمع إذا كان على شكل طبقات أو إذا كان الهدف هو دراسة طبقات المجتمع.

مثال: المجتمع العربي طبقات من ناحية الغنى (طبقة الأثرياء – طبقة الأغنياء – طبقة الوسط – طبقة الفقراء – طبقة المعدمين).

عند دراسة المجتمع : من كل طبقة نأخذ عينة حسب حجم الطبقة .

ج– العينة العشوائية المنتظمة:

تؤخذ بطريقة عشوائية ولكن في فترات منتظمة أو من أماكن منتظمة.

مثال: إذا أردنا أخذ عينات من مجتمع معين مثلاً شركة تنتج معجون أسنان وأردنا اختبار المنتجات فنقوم مثلا بأخذ العينات كل ساعة أو كل يوم أو كل أسبوع .

د- العينات العشوائية العنقودية:

تؤخذ إذا كان المجتمع مرتب بطريقة عنقودية أو كان على شكل تنظيمات هرمية .

مثال:التنظيم الإداري للمحافظات في سورية (محافظة– مدينة – منطقة – ناحية – قرية).

مثال:التنظيم في الجامعات (الجامعة – الكلية – السنوات الدراسية –الشعب– الفئات – المجموعات).

5- المتغيرات (المتحولات) Variable:

هي توابع ذات قيم متغيرة تتعلق دائماً بصفة معينة **وهي على نوعين**:

أ–المتغيرات الكمية. – **ب**–المتغيرات النوعية.

أ- المتغيرات الكمية:

هي المتغيرات التي قياسها بالقياسات المعروفة مثل الطول والوزن والحجم ويمكن التحكم بها.

ب- المتغيرات النوعية:

هي المتغيرات التي لا يمكن قياسها بالطرق المعروفة وهي عبارة عن تدريجات أو مستويات مثل اللون أو درجة الذكاء.

كما يمكن تقسيم المتغيرات من حيث الاحتمال إلى نوعين :

أ- المتغيرات العشوائية:

هي المتغيرات التي لا يمكن التحكم بها (مثل: درجة الحرارة – نسبة الرطوبة – سرعة الرياح).

ب– المتغيرات اللاعشوائية (النظامية):

يمكن التحكم بها وتحديدها مسبقاً(مثل كمية السماد المضافة ، جرعات من الدواء المأخوذة،عدد الريات).

6- الصفة المدروسة category:

هي الصفة التي يختارها الباحث ليدرس تغيرات هذه الصفة أثناء التجربة أو الاختبار.

مثال:من الصفات المدروسة (صفة الإنتاج أو الإنتاجية – طول النبات – لون الثمرة – حجم الثمرة).

7- القيمة الإحصائية Value:

هي القياس أو الرقم الذي نحصل عليه عند قياس الصفة المدروسة ومجموع القيم الإحصائية نسميها البيانات (Data).

8- التوزيع Distribution:

هو طريقة توزع القيم الإحصائية أو طريقة توزع البيانات للمجتمع أو العينة واحتمالاتها.

- **لدينا ثلاث أنواع من التوزيعات:**

1– التوزيع الثنائي:هو التوزيع الذي يمثل قيم من نوعين فقط أو احتمالين فقط مثل (صح،خطأ).

2– التوزيع الطبيعي:هذا التوزيع يمثل توزيع قيم المجتمعات الطبيعية مثل المجتمع البشري.

3– توزيع بواسون: يمثل هذا التوزيع توزيع الحوادث النادرة (الصقيع -الزلازل – البراكين).

=============   نهاية المحاضرة الأولى=============

د. مرهج دبيات

التحليل الإحصائيData analysıs

❖ تعاريف ومصطلحات:

التحليل الإحصائي: يعني تحليل البيانات وهو ترتيب وتبويب البيانات ومن ثم حساب المؤشرات الوصفية لها بالإضافة إلى تطبيق تحاليل واختبارات أخرى من أجل تفسير النتائج وعرضها بطريقة مفهومة بالاستعانة بوسائل العرض البياني وأخيرا تعميمها ونشرها.

المراحل الأساسية للبحث (أو التجربة) الإحصائي:

أولاً- مرحلة التخطيط: وضع الخطة لإجراء البحث.

ماذا تتضمن خطة البحث؟:

1- عرض المشكلات (مشكلات البحث):

وفيها يتم تقديم عرض مختصر ومركز حول مشكلة البحث بحيث يتم تسليط الضوء على المشكلة العلمية التي يتمحور حولها البحث.

2- تحديد أهداف البحث:

وهنا يتم صياغة الأهداف بدقة علمية إلى أقصى حد ممكن ويفضل أن تتم صياغة الأهداف على شكل أسئلة واضحة ومختصرة ودقيقة ولا يوجد عدد محدد للأهداف من الممكن أن يكون عدد الأهداف (1 – 2 – 3 – 4 – ....) وذلك حسب رغبة الباحث.

3- تحديد المادة التجريبية :

هي مادة التجربة التي تطبق عليها الأهداف الاختبارات والمعاملات المطلوبة.

4- الفرضيات الإحصائية:

يتم تحديد الفرضيات الإحصائية طبقاً لأهداف البحث وهي في الواقع عبارة عن تعبير رياضي أو تعبير بطريقة رياضية عن الأهداف ويتم هنا تحديد الفرضيات الأولية والبديلة لها في حال تم رفضها .

5- تحديد المتغيرات أو العوامل المطلوب دراستها من البحث.

تحدد العوامل في البحث من خلال تحديد المؤثرات على المادة التجريبية وتحديد نوعها (هل هي كمية أم نوعية) وهل هي عشوائية أم لا عشوائية.

6- المعاملات:

هي عبارة عن أشكال أو كميات أو تدرجات أو أنواع أو رموز للعوامل والمتغيرات في البحث.

7- المكررات:

هي عبارة عن عدد مرات تكرار المعاملات في البحث أي كم مرة نحتاج إلى إعادة المعاملات في التجربة.

8- الصفة المختارة (الصفة المدروسة):

وهي الصفة التي يختارها الباحث يراقب تغيرات قيم هذه الصفة خلال فترة البحث.وهي اما أن تكون صفة كمية أو نوعية وعادة يتم تحويل النوع الثاني الى نوع كمي حتى يصبح التحليل والتقويم لها ممكناً

9- المجال التجريبي :

وهو المجال الذي يتم تجربة المعاملات ضمنه وهذا المجال يمكن أ ن يكون كمي أو نوعي بحسب نوع المتغيرات أو العوامل المدروسة للبحث.

10- حجم العينة أو التجربة:

يرمز له بالرمز (**n**) وهو يساوي جداء عدد المعاملات (**t**)في عدد المكرارت(**r**).

$$n= r * t$$

11- الدقة :

وهي الدقة المطلوبة التي يجب أن تتوفر خلال البحث أو التجربة وهذه الدقة تختلف بحسب نوع البحث أو التجربة وهي عادة إما 1% أو 5% ولكن نادرا ما تكون 10 أو 25% .

12- التحاليل والاختبارات الإحصائية اللازمة.

يتم اختيار التحاليل والاختبارات الإحصائية اللازمة حسب أهداف التجربة وهذه تحتاج عادة الى خبرة أو بالاستعانة بالاختصاصيين في التحليل الإحصائي.

13- التصميم الرياضي:

لكل تجربة تصميم وسيتم دراسة التصاميم التجريبية في مقرر كامل أسمه تصميم التجارب.

14- الملاحظات Notıce: ويتم هنا كتابة أسم الباحث وفريق العمل بالإضافة إلى مواعيد تتعلق بالبحث وملاحظات أخرى.

ثانياً- تحليل البيانات :

يوجد للتحليل الإحصائي **نوعان:**

1- التحليل الإحصائي الأولي Primary:

نحتاج دائماً إلى إجراء تحليل إحصائي أولي لكل البيانات التجريبية أو بيانات بحث معين.

ماذا يشمل التحليل الإحصائي الأولي:

1-وصف البيانات بشكل نظري.

2- حساب المؤشرات الإحصائية الأساسية **Descriptive Statistics**

وتتضمن المؤشرات التالية:

- أ-المتوسط أو المعدل الحسابي Main أو Average.
- **ب**-الوسيط الحسابي Median.
- **ج**-المنوال أو القمة Mode.
- **د**-الخطأ القياسي Standard Error.

- **هـ**–التباين أو التشتت Variation.
- **و**–معامل الاختلاف Coefficient of Variation.

3– إنشاء جدول التوزيع التكراري.

4–العرض البياني للبيانات.

<u>2- التحليلى الإحصائى المتقدم Advanced : ويشمل:</u>

أ– البحث عن التوزيع الرياضي المناسب للبيانات واختبار مدى صلاحية البيانات باستخدام الاختبارات المناسبة ومعالجة القيم الشاذة والمفقودة.

ب–تطبيق التحاليل الإحصائية اللازمة مثل:(تحليل التباين المشترك –التحليل العنقودي – تحليل الصفوف).

ج–استخراج أو استنتاج النموذج الرياضي المناسب للبيانات.

د–تطبيق الاختبارات الإحصائية اللازمة مثل:(الاختبارات المعنوية – اختبارات مقارنة المتوسطات –اختبارات إضافية متعددة).

هـ–تحليل أو دراسة العلاقات باستخدام التحاليل المعروفة مثل:(تحليل الانحدار والارتباط).

<u>ملاحظة:</u>

دائماً يلزمنا إجراء التحليل الإحصائي الأولي أما التحليل الإحصائي المتقدم فيتم استخدامه عند الضرورة إذا كانت البيانات أو أهداف البحث تحتاج لذلك.

فيما يلي سيتم شرح **خطوات التحليل الإحصائي الأولي** بالتفصيل:

يشمل التحليل الإحصائي الأولي:

1. وصف البيانات بشكل نظري وهو يعني وصف نوع وطريقة جمع البيانات بالإضافة إلى الأساليب والطرق المستخدمة في ترتيب وجمع البيانات كما يشمل آليات الوصول إلى مصادر هذه البيانات إن كانت تقليدية أم أنها متطورة.
2. التوزيع التكراري : هو عملية ترتيب وتنظيم البيانات في جداول نطلق عليها جدول التوزيع التكراري بحيث يمكن إدارة ومعالجة البيانات بطريقة سهلة ودقيقة بدون فقدان أي تفاصيل مهمة في البيانات.
3. حساب المؤشرات الإحصائية للعينة أو العينات المتوفرة قيد التحليل ومن قم حساب مؤشرات وصفية تقديرية للمجتمع الذي أخذت منه العينات.
4. إنشاء العروض البيانية . وهنا يفضل استخدام الرسوم البيانية التي توضح الغاية من البحث أو التجربة بشكل واضح وذلك حسب المجال الذي ستستخدم البيانات فيه سواء المجال العلمي أم الإرشادي أم الدعاية والإعلان.

نهاية المحاضرة الثانية

المؤشرات الإحصائية

1- مقاييس النزعة المركزية:

النزعة المركزية طبيعة عامة لكل المجتمعات أو ميزة تميز كل المجتمعات الطبيعية وتعني النزعة المركزية رغبة أو نزوع الأفراد في المجتمع حول المركز أي في كل مجتمع من المجتمعات الطبيعية يحاول أفراد هذا المجتمع التجمع حول المركز.

مثال:

لوأخذنا أطوال المجتمع البشري في سوريا لو أخذانا أطوال (500) شخص نلاحظ أن الأطوال تتراوح حول مجال معين (150 – 180) سم بشكل عام .

المتوسط لهذه الأطوال أو المركز هو (165) سم .

إذا كان لدينا (500) شخص نلاحظ أن (400) شخص طوله قريب من (165) سم ونلاحظ عدد قليل طوله (180) سم.

مثال أخر:

لو أحذنا ضغط الدم مثلاً عند (500) شخص نلاحظ حوالي (400) شخص يكون ضغط الدم عندهم (12 / 8) ونلاحظ حوالي (30) شخص ضغط الدم عندهم (7/10) ونلاحظ حوالي (70) شخص ضغط الدم عندهم (9/14).

❖ مقاييس النزعة المركزية:

يوجد مقاييس خاصة لقياس النزعة المركزية وهي:

1- المتوسط أو المعدل الحسابي Avarage أوMain.

2- الوسيط Medıan.

3- المنوال أو القمة Mode.

1- المتوسط الحسابي أو المعدل الحسابي:

هو من أهم مقاييس النزعة المركزية وزهو الأكثر استخداماً ويتمتع بدلالة قوية أو هو مؤشر للنزعة المركزية.

❖ مميزات المتوسط الحسابي:

أ- سهل جداً للحساب حيث يمكن حسابه بسهولة كبيرة فهو يساوي المجموع / العدد.

ب- يعطي دلالة قوية جداً عن العينة التي تم الحساب منها.

ج- يمكن حسابه من أي نوع من العينات سواء كانت عينات بسيطة أو غير ذلك ولكن لا يمكن حسابه للمجتمع الطبيعي إلا إذا عرفنا العدد الكلي.

❖ **عيوب المتوسط الحسابي**:

أ- يتأثر المتوسط الحسابي بالقيم المتطرفة أي أنه يتأثر بمدى التباين أو التفاوت بين القيم.

مثال :

لدينا عينة مؤلفة من (3 – 4 – 5).

ولدينا عينة أخرى (1 – 2 – 9).

ولدينا عينة ثالثة مؤلفة من (1 – 1 – 10).

لو قمنا الآن بحساب المتوسط الحسابي لهذه العينات نلاحظ أن المتوسط هو نفسه بالرغم من اختلاف الأرقام المكونة لكل عينة من العينات والمتوسط يكون للعينات الثلاث السابقة (4).

فمشكلة المتوسط الحسابي يتأثر بالقيم المتطرفة ولايتأثر بمدى اختلاف أفراد العينة.

- **مقارنة بين العينة والمجتمع:**

العينة	المجتمع
عدد الأفراد معروف يرمز له ب(n) وبالتالي	عدد الأفراد غير معروف وغير محدد ويرمز

متوسط العينة معروف 'x ويمكن حسابه.	لعدد الأفراد بالرمز (N) لا يمكن حساب متوسط المجتمع (µ) لأن العدد الكلي غير معروف.
المتوسط في العينة هو حسابي أي يتم حسابه.	المتوسط في المجتمع تقديري أي يتم تخمينه ضمن مجال يسمى مجال الثقة.

كيف يمكن حساب متوسط أو المعدل العام للمجتمع:

بالنسبة للمجتمع يوجد لدينا متوسطين :

(µ) متوسط المجتمع وهو المتوسط الحقيقي.

(m)متوسط المجتمع التقديري أو التقريبي والذي يمكن حسابه تقديراً أو تقريباً ولا يمكن معرفة المتوسط الحقيقي للمجتمع.

– نقوم بحساب المتوسط التقريبي ثم نحسب مجال المتوسط الحقيقي.

المتوسط التقريبي نقدره بالطريقة التالية:

1–نأخذ أكبر عدد ممكن من العينات الممثلة للمجتمع.

2– نحسب متوسط كل عينة.

3– نحسب متوسط متوسطات للعينة.

4– نحصل بذلك على المتوسط الحسابي التقريبي.

❖ **درجات الثقة:**

مجال الثقة عند احتمال خطأ (1 % أو 5 % أو 10 % أو غير ذلك).

هناك مقولة لعالم إحصاء ألماني يقول فيها : "إن دراسة أي ظاهرة من ظواهر الطبيعة عرضة للأخطاء واحتمال كل حدث مرهون بالخطأ" .

كذلك أنشتاين يقول كل شي في الكون نسبي ولا يوجد شي مطلق.

=========== **نهاية المحاضرة الثالثة**=========

مقاييس التشتت والاختلاف

MEASURES OF DISPERSION AND VARIATION

مقدمة :

عرضنا في الفصل السابق النزعة المركزية ووجدنا هذه المقاييس مهمة من حيث أنها تعطينا فكرة جيدة ووافية عن عناصر العينة المدروسة ومدى نزعتها للتجمع حول المتوسط ؛فالمتوسط الحسابي قيمة قريبة من كل عناصر العينة تقريباً (باستثناء القيم الشاذة) ، لكن في حال تساوي قيمة المتوسط لعينتين بالرغم من اختلافهما من حيث قيم العناصر فكيف نفرق بين العينتين والحكم من حيث التجانس وغيره من المؤشرات الإحصائية.

من هنا كان لابد من وجود مؤشرات خاصة لاستخدامها في حال تساوي قيمة المتوسط لعينتين ،هذه المؤشرات الأكثر أهمية واستخداماً هي مقاييس التشتت والاختلاف وهي موضوع هذا الفصل التي سندرسها بالتفصيل مع الأمثلة اللازمة.

(3-1) المدى (RANGE) :

تعريف : المدى هو الفرق بين أكبر وأصغر قيمة في العينة المدروسة .في جدول التوزيع التكراري يتم حساب المدى على أنه الفرق بين الحد الأعلى للفئة الأخيرة والحد الأدنى للفئة الأولى ،أو هو الفرق بين مركز الفئة الأولى ومركز الفئة الأخيرة.

مثال : لدينا العينة التالية :

$$X=\{33,15,20,25,35\}$$

الحل : نلاحظ أن أصغر قيمة أي MIN (X)= 15 وأكبر قيمة هي MAX(X)= 35 وبالتالي فان : المدى هو التالي :

RANGE=MAX(X)-MIN(X)

= 35 – 15

= 20

(3-2) متوسط الانحراف المطلق ABSOLUTE DEFIATION MEAN

تعريف : يعرف متوسط الانحراف المطلق بأنه متوسط مجموع حاصل طرح كل عنصر من عناصر العينة من متوسط العينة بالقيمة المطلقة . و متوسط الانحراف المطلق يعطى بالعلاقة التالية :

$$S^2 = (\sum_1^n (X_i - \overline{X})^2) / (N-1)$$

ومن الجدير بالتنويه إلى أن متوسط الانحراف المطلق نادر الاستخدام في مجال الإحصاء التطبيقي في تحليل البيانات التجريبية للاختبارات والتجارب الزراعية.

لذا اكتفينا بهذا القدر بالحديث عن متوسط الانحراف المطلق

(3-3) التباين:VARIANCE

يعتبر التباين VARIANCE أهم مقياس من مقاييس التشتت والمستخدمة على نطاق واسع جداً خاصة في تحليل البيانات التجريبية للاختبارات والتجارب الزراعية وهو يعطى بالعلاقة التالية :

$$S^2 = (\sum_1^n (X_i - \overline{X})^2) / (N-1)$$

كما يمكن حساب التباين VARIANC بطريقة أخرى يطلق عليها طريقة تربيع القيم وهي كما يلي :

$$S^2 = (\sum_1^n X_i^2 - (\sum_i^n X_i^n)^2 / n) / (n-1)$$

ومن الجدير بالتنويه إلى أن الطريقة الاخيره أعلاه " طريقة تربيع القيم " تستخدم عندما تكون قيم العينة قيم كسرية.

مثال (2) :أوجد التشتت أو التباين والانحراف المعياري للعينة التالية باستخدام الطريقة المناسبة :

X = {5,6,7,1,3,2}

الحــل : يمكن الاستعانة بالعلاقة الخاصة بحساب التباين الاولى الأنفة الذكر وهي التالية :

$$S^2 = (\sum_{1}^{n}(X_i - \overline{X})^2)/(N-1)$$

كما يمكن الاستعانة بالعلاقة الخاصة بحساب التباين الثانية وهي التالية :

$$S^2 = (\sum_{1}^{n} X_i^2 - (\sum_{i}^{n} X_i^n)^2 / n)/(n-1)$$

من أجل ذلك علينا إنشاء الجدول المبسط التالي :

X	X-X`	(X-X`)^2	X^2
5	1	1	25
6	2	4	36
7	3	9	49
1	-3	9	1
3	-1	1	9
2	-2	4	4
∑ 24	0	28	124

الآن يمكن تطبيق إحدى العلاقتين السابقتين للحصول على التباين كما يلي:

S^2 = 28 / (6-1)

= 5.6

أما الانحراف المعياري للعينة فهو سهل الحساب باعتباره الجذر الموجب للتباين ،إذن هو يساوي:

$$S = \sqrt{S^2}$$

$$S = \sqrt{5.6}$$

$$= 2.37$$

◄ من الجدير بالذكر انه يمكن الملاحظة بسهولة أن الفرق بسيط بين قانون تباين العينة و قانون تباين المجتمع .

◄عند تساوي المتوسطات فانه يستخدم عادة الانحراف المعياري كأساس للمقارنة بين عينتين أو أكثر شرط أن تكون العينات مأخوذة من مجتمع واحد ،وبالتالي فالعينة ذات الانحراف المعياري الأقل هي التي تكون أكثر استقرارا من غيرها وعناصرها أقل تبعثرا وأكثر تجانسا.

◄ أما إذا لم يتحقق الشرط المذكور أنفا أي عندما لا تكون العينات مأخوذة من مجتمع واحد فان الانحراف المعياري لايمكن اتخاذه كأساس للمقارنة بين عينتين أو أكثر ويستخدم عادة مقياس آخر يطلق عليه عامل الاختلاف COFFICIENT OF VARIATION

عامل الاختلاف COFFICIENT OF VARIATION : يرمز عادة لعامل الاختلاف بـ C.V وهو يعرف بأنه أحد مقاييس التشتت أو التباين الهامة والذي يستخدم بشكل عام للمقارنة بين عينتين أو أكثر عندما لا تكون العينات مأخوذة من مجتمع واحد بل من مجتمعات مختلفة ، وهو أي عامل الاختلاف يعطى بالعلاقة التالية :

$$C.V = (S / X`) * 100$$

حيث أن :

C.V : عامل الاختلاف

S : الانحراف المعياري

X` : متوسط العينة

◄من الجدير بالذكر انه كلما كانت قيمة عامل الاختلاف C.V أصغر كلما دل ذلك على ثبات العينة وتجانس أفرادها وعناصرها أقل تبعثرا مقارنة بالعينة الأخرى الأكبر بقيمة عامل الاختلاف

مثـــال ( ) : قارن بين العينتين التاليتين علما أنهما من مجتمعين مختلفين :

X	3000	4500	4000	4100
Y	900	1200	1700	2200

الحــل : للمقارنة بين العينتين باعتبار أن العينات ليست مأخوذة من مجتمع واحد بل من مجتمعات مختلفة ، فان عامل الاختلاف يعطى بالعلاقة التالية :

$$C.V = (S / X`) * 100$$

وبتطبيق العلاقة السابقة الذكر نصل على قيمة عامل الاختلاف وهي :

$$C.V (X) = 637.7/3900 * 100$$

$$= 16 \%$$

$$C.V (Y) = 439.7/1500 * 100$$

$$= 29 \%$$

ومن الواضح أن عامل الاختلاف للعينة الاولى أصغر من عامل الاختلاف للعينة الثانية وبالتالي فان العينة الاولى أكثر تجانسا من العينة الثانية.

◄ من الجدير بالذكر انه بالرغم من أن الانحراف المعياري للعينة الثانية أصغر من الانحراف المعياري للعينة الاولى فإننا لا نقدر أن نحكم على مدى تجانس العين لان العينات ليست مأخوذة من مجتمع واحد بل من مجتمعات مختلفة ، وهذا ما يؤكد كل ما ذكرناه سابقاً.

الخطأ القياسي STANDARD ERROR : وهو يعرف بأنه أحد مقاييس التشتت أو التباين الهامة والذي يستخدم بشكل عام من أجل التعبير عن مدى التطابق ما بين متوسط الذي أخذت منه تلك العينة.وهذا يعني أنه كلما كان الخطأ القياسي أصغر

للعينة كلما كان الفرق أقل مابين متوسط العينة ومتوسط المجتمع . إن الخطأ القياسي يعطى بالعلاقة التالية :

$$S\bar{X} = S / \sqrt{N}$$

حيث أن : `SX : الخطأ القياسي

S الانحراف المعياري

N عدد عناصر العينة

◄ من الجدير بالذكر انه يمكن استخدام الخطأ القياسي لتحديد المجال الذي يقع ضمنه متوسط المجتمع حيث أن متوسط المجتمع يقع بشكل عام ضمن المجال :

$$\bar{X} \pm S\bar{X} * t$$

ويطلق على هذا المجال كما هو معروف في الإحصاء مجال الثقة.

========== **نهاية المحاضرة الثالثة الرابعة** ==========

نظرية الاحتمالات والتوزيعات

THIORY OF PROBABILITY AND DISTRIBUTIONS

5-1 **- مقدمة**: يرتبط الإحصاء والاحتمال ارتباطا وثيقا،فالاحتمال هو الأداة التي تمكن الإحصائي من استخدام المعلومات في عينة للقيام باستقراءات حول المجتمع الذي أخذت منه.

من اجل دراسة احتمال ظاهرة طبيعية أو حادثة معينة يتم من خلال مراقبة وقوع الحادثة وتسجيل تكرار وقوعها خلال فترة محددة.ثم نحسب احتمال الظاهرة من خلال تكرار التجربة أو المراقبة عددا كبيرا من المرات N ،مثلا، وكان عدد مرات وقوع الحادثة A هو n مرة فيكون بذلك احتمال الحادثة A هو كما يلي:

$$P(A) = n / N$$

ولوضع نموذج مخصص يعبر عن قيمة احتمال الحادثة A يجب الأخذ بعين الاعتبار مراعاة الشرطين التاليين :

$$0 <= P(A) <= 1$$

$$\Sigma P(A) = 1$$

تختلف الظواهر والحوادث من حيث درجة حدوثها أو وقوعها ومدى تكرار ذلك خلال فترات معينة ،وهذا ما يطلق عليه احتمال الظاهرة أو الحدث .بناء على ذلك يمكن أن نصنف الظواهر والحوادث بشكل عام إلى المجموعات التالية:

- حوادث مؤكدة : درجة احتمالها 100 %
- حوادث مستحيلة : درجة احتمالها 0.0 %
- حوادث محتملة أو متوقعة درجة احتمالها أكبر من 0.0 % وأصغر من 100 %

لنتذكر أن التجربة هي كل عملية تؤدي إلى قياس أو ملاحظة .وتنتج معظم التجارب ذات الأهمية قياسات عددية تتغير من عينة إلى أخرى وبالتالي يدعى هذا القياس بالمتحول العشوائي.وهذا يقودنا إلى الفقرة التالية.

5- 2- أنواع المتغيرات variables: يمكن تقسيم المتغيرات إلى نوعين وذلك حسب القيم التي يأخذها المتغير :

- متغير كمي QUALITIVE:هو المتغير الذي يأخذ قيم كمية أي قابلة للقياس كميا كوحدات القياس مثل وحدات الوزن،الطول ،الحجم ...إلى آخره. وكمثال على ذلك المبيد ،السماد،العمر،كمية الغذاء وهكذا...
- متغير نوعي QUANTITIVE: هو المتغير الذي يأخذ قيم نوعية غير قابلة للقياس كميا بوحدات القياس المذكورة آنفا بل يقاس بوحدات القياس النوعي كالدرجات والتدرجات والمستويات وهكذا ... إلى آخره.وكمثال على ذلك اللون ودرجة النجاح ودرجة المقاومة ...

- مصادر المتغيرات variables: يمكن تقسيم المتغيرات إلى نوعين وذلك حسب نوع التحكم أو المصدر إلى نوعين :

1. متغير عشوائي أو احتمالي RANDOM variables: هو المتغير الذي يأخذ قيم عشوائية محتملة لا يمكن التحكم بها مسبقا مثل المتغيرات الجوية كالحرارة والرطوبة وغيرها .
2. متغير غير احتمالي Controlled variable: المتغير الذي يأخذ قيم محددة يمكن التحكم بها مسبقا مثل المبيد ،السماد،العمر،كمية الغذاء وهكذا...

3-5-التوزيع الاحتمالي للمتحولات :ان الظواهر الطبيعية والتي تهمنا من الناحية الزراعية يمكن تمثيل قيمها كمجال لمتغير عشوائي.

بالتالي فإن المتغير العشوائي يمثل تابع عددي معرف على عينة القياسات أو القيم.وبما أن لكل قيمة أو قياس في العينة احتمالا معينا فانه يمكن القول أن التوزيع لمتحول عشوائي هو علاقة أو جدول أو خط بياني يقدم الاحتمال الموافق لكل قيمة من قيم المتحول العشوائي.

من هنا تستوضح لنا أهمية دراسة التوزيعات الاحتمالية المعروفة والتي لها تطبيقات عملية في مجال الزراعة.

إذن من خلال هذا الفصل سنقدم شرحا مفصلا لموضوع التوزيعات الاحتمالية كما يلي:

4-5- التوزيع الثنائي :يعتبر التوزيع الثنائي من التوزيعات الرياضية المعروفة والمهمة من الناحية التطبيقية .

1-4-5- تعريف : ليكن لدينا E تجربة ما وليكن A حدث مرتبط بهذه التجربة ولنرمز لـ n لعدد مرات تكرار التجربة و X عدد مرات ظهور أو وقوع الحدث A ،عندئذ فان احتمال ظهور الحدث A عدد من المرات مساو لـ r هو التالي:

$$P(x=r)=(r^n)*P^r*q^{n-r}$$

حيث أن : P هو احتمال ظهور أو وقوع الحدث A

q احتمال عدم ظهور أو وقوع الحدث A

run هي توافيق كل منهما وهي تحسب كما يلي:

$\binom{n}{r}=n$!/(n-r!)*r!

حيث أن : n عدد مرات تكرار

r عدد مرات احتمال ظهور الحدث A

ويحسب التوقع الرياضي للتوزيع الثنائي من العلاقة التالية :

$$M(x) = n * p$$

كما يحسب التباين

$$\sigma^2 (x) = n * p * q$$

كما يحسب الانحراف المعياري للتوزيع الثنائي من العلاقة التالية :

$$\delta(x)=\sqrt{n*p*q}$$

مثـــــال (1) :تم رش 100 شجرة برتقال في أحد البساتين لمعالجة أحد الأمراض علما أن فعالية المبيد النظرية هي 90% .والمطلوب هو حساب ما يلي :

1. اخترنا أحد الأشجار بعد الرش ،فما هو احتمال أن تكون هذه مصابة؟
2. ما هو احتمال أن تكون ثلاث أشجار مصابة؟
3. ما هو احتمال أن لا تكون ولا شجرة مصابة؟
4. ما هو احتمال أن يوجد شجرة مصابة على الأقل؟
5. ما هو أن يوجد شجرتين مصابتين على الأكثر؟

الحــــــل : لدينا مسبقا الافتراضات التالية:

$$P = 1 - q$$

$$n = 100$$

$$P(x = r) = (r^{n}) * P^{r} * q^{n-r}$$

$$(\overset{n}{r}) = n\,!/(n-r!)*r!$$

الآن نطبق العلاقات المذكورة مع مراعاة القيم المسبقة لكل متغير :

1. احتمال أن تكون هذه الشجرة مصابة:

$$P = 1 - q$$

$$P = 0.1$$

$$q = 0.9$$

$$r = 1$$

$$n = 100$$

$$P(x=1) = ({}_{1}{}^{100})*(0.1)^{1}*(0.90)^{99}$$

$$P(x=1) = 100*0.1*2.95*10^{-5}$$

$$P(x=1) = 0.000295$$

2. احتمال أن تكون ثلاث أشجار مصابة:

$$P = 1 - q$$

$$P = 0.1$$

$$q = 0.9$$

$$r = 3$$

$$P(x=3) = ({}_{3}{}^{100})*(0.1)^{3}*(0.90)^{79}$$

$$P(x=3) = 161700 * 0.001 * 3.63 * 10^{-5}$$

$$P(x=3) = 0.0059$$

3. احتمال أن لا تكون ولا شجرة مصابة:

$$P = 1- q$$

$$P = 0.1$$

$$q = 0.9$$

$$r = 0$$

$$P(x=0) = ({}_{0}^{100}) * (0.1)^{0} * (0.90)^{100}$$

$$P(x=0) = 1*1*2.65*10^{-5}$$

$$P(x=0) = 0.0000265$$

4. احتمال أن يوجد شجرة مصابة على الأقل:

$$P(x>=1) = 1- p(x<1) = 1-p(x=0)$$

$$P(x>=1) = 1- 0.0000265$$

$$P(x>=1) = 0.99997$$

5. أن يوجد شجرتين مصابتين على الأكثر :

$$P(x<=2) = p(x=2) + p(x=1) + p(x=0)$$

$$P(x=2) = ({}_{2}^{100}) * (0.1)^{2} * (0.90)^{89}$$

$$P(x=2) = 0.001485$$

$$P(x=1) = 0.000295$$

$$P(x=0) = 0.0000265$$

$$P(x<=2) = 0.001541$$

أما التوقع الرياضي لعدد الأشجار المصابة بعد الرش يمكن ان نحسبه كما يلي:

$$M(x) = n * p = 100 * 0.1 = 10$$

أي من المتوقع بقاء عشرة أشجار مصابة بعد الرش.

5-5- **توزيع بواسون** : نقول عن متحول عشوائي x أن توزيعه الاحتمالي يتبع توزيع بواسون إذا كان تابع التوزيع الاحتمالي له على الشكل التالي :

$$p(x = r) = (\lambda^r * e^{-\lambda}) / r\,!$$

حيث أن : λ الوسيط وهي تحسب من العلاقة التالية:

$$\lambda = n * p$$

ويحسب التوقع الرياضي لتوزيع بواسون من العلاقة التالية :

$$M(x) = \lambda$$

كما يحسب التباين لتوزيع بواسون من العلاقة التالية :

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

كما يحسب الانحراف المعياري لتوزيع بواسون من العلاقة التالية:

$$\delta^2 = \sqrt{\lambda}$$

تجدر الإشارة هنا إلى أنه يجب أن تتوفر الشروط التالية استخدام توزيع بواسون :

$$n > 50 \quad \text{و} \quad n * p < 5$$

مثـــــــــال :تم تجريب لقاح مضاد لأحد الأمراض على 2000 حيوان تجربة من حيوانات الأرانب مع العلم أن نسبة الضرر من اللقاح هي 0.2% والمطلوب :

1. ما هو التوقع الرياضي لعدد الأرانب المصابة بعد اللقاح؟
2. ،فما هو احتمال أن يكون أرنب واحد متضرر؟
3. ما هو احتمال أن لا يكون ولا أرنب واحد متضرر ؟
4. ما هو احتمال أن تكون ثلاث أرانب متضررة على الأقل؟

الحـــــــل : لدينا المعطيات الأولية التالية :

n = 2000   (>50)

n *p = 4   (<5)

إذا أمعنا النظر في المعطيات الأولية نلاحظ أنه تتوفر الشروط اللازمة لاستخدام توزيع بواسون .

1. التوقع الرياضي لعدد الأرانب المصابة بعد اللقاح هو :

M(x) = λ = n * p

M(x) = 2000 * 0.002

M(x) = 4

2. احتمال أن يكون أرنب واحد متضرر:

$$p(x = r) = (\lambda^r * e^{-\lambda}) / r!$$

λ = n * p = 4

r = 1

P(x=1) = 0.07

3. احتمال أن لا يكون ولا أرنب واحد متضرر :

$$p(x = r) = (\lambda^r * e^{-\lambda}) / r!$$

λ = n * p = 4

r = 0

P(x=1) = 0.07

4. احتمال أن تكون ثلاث أرانب متضررة على الأقل:

$$p(x = r) = (\lambda^r * e^{-\lambda}) / r!$$

λ = n * p = 4

$$r >= 3$$

$$P(x>=3) = P(x=3) + P(x=4) +\ldots P(x=n)$$

$$P(x>=3) = 1- P(x<3)$$

$$P(x>=3) = 1- [P(x=2) + P(x=2) + P(x=0)]$$

$$P(x>=3) = 1- (0.15+0.07+0.02)$$

$$P(x>=3) = 0.76$$

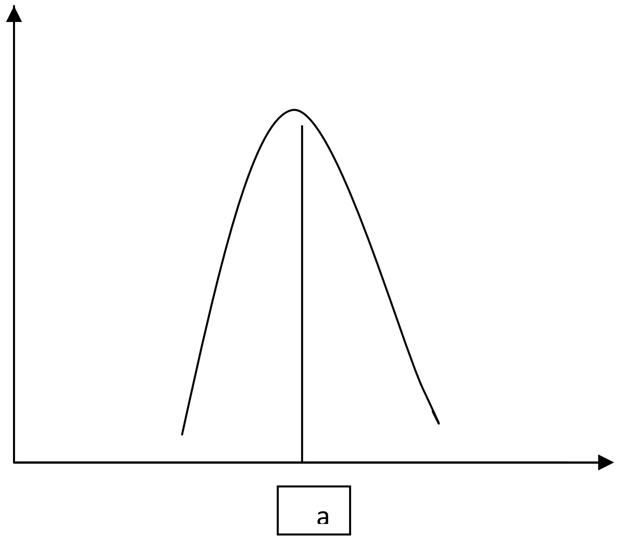
5-6- **التوزيع الطبيعي** :

5-6-1- تعريف : نقول عن متحول عشوائي x أن توزيعه الاحتمالي يتبع التوزيع الطبيعي إذا كان تابع التوزيع الاحتمالي له على الشكل التالي :

$$f(x) = (1 * e^{-(x-a)/2\delta^2}) / \delta * \sqrt{2\pi}$$

يعتبر التوزيع الطبيعي من التوزيعات الرياضية المعروفة والمهمة من الناحية التطبيقية .

5-6-2- منحنى التوزيع الطبيعي :يملك منحنى التوزيع الطبيعي شكلا ناقوسيا يشبه شكل الجرس على الشكل التالي :



ويمثل هذا الشكل السلوك الطبيعي للظواهر الطبيعية .حيث تبدأ من نقطة منخفضة ثم تزداد حتى تصل إلى الذروة ثم تعود وتنخفض إلى نفس المستوى السابق ،وهذا ما يطلق عليه قانون تناقص الغلة في الاقتصاد(كما ورد في كتاب الاقتصاد الزراعي في السنة الاولى).

ويحسب التوقع الرياضي للتوزيع الطبيعي من العلاقة التالية :

$$M(x) = a$$

كما يحسب التباين للتوزيع الطبيعي من العلاقة التالية :

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

كما يحسب الانحراف المعياري للتوزيع الطبيعي من العلاقة التالية:

$$\sigma (x) = \sigma$$

ملاحظات هامة :

- ان المساحة المحصورة تحت منحنى التوزيع الطبيعي والمحور الأفقي (x) يساوي الواحد كما هو متعارف عليه.
- ان إحداثيات التالية:

$(a , 1/\delta * \sqrt{2\pi})$

- ان مسقط الذروة لمنحنى التوزيع الطبيعي هي قيمة المتوسط ، وبالتالي فان المتوسط يقسم المساحة تحت منحنى التوزيع الطبيعي إلى قسمين متماثلين.
- ان ارتفاع الذروة يتعلق فقط بالانحراف المعياري لمنحنى التوزيع الطبيعي.

5-6-3-العلاقة بين منحنى التوزيع الطبيعي والاحتمالات:

تحدد العلاقة بين منحنى التوزيع الطبيعي والاحتمالات فيما يلي :

- ان المساحة المحصورة تحت منحنى التوزيع الطبيعي والمحور الأفقي (x) يساوي الواحد كما هو متعارف عليه.
- ان أي قيمة من العينة تقع على يمين أو يسار المتوسط أي أكبر منه أو أصغر منه .
- ان احتمال أي قيمة من العينة تقع على يمين المتوسط هو مابين 0 و 0.5 و احتمال أي قيمة من العينة تقع على يسار المتوسط هو مابين 0 و 0.5 .
- ان احتمال أي قيمة من العينة محصورة ضمن مجال المتوسط زائد أو ناقص انحراف معياري واحد هو $p = 0.68$.
- ان احتمال أي قيمة من العينة محصورة ضمن مجال المتوسط زائد أو ناقص انحرافين معياريين هو $p = 0.95$.
- ان احتمال أي قيمة من العينة محصورة ضمن مجال المتوسط زائد أو ناقص ثلاث انحرافات معيارية هو $p = 0.99$.

5-6-4- التوزيع الطبيعي المعياري : ان التوزيع الطبيعي المعياري هو منحنى معدل من التوزيع الطبيعي بحيث أصبحت ميزاته هبي التالية :

1. المتوسط الحساب مساو للصفر أي $M(z) = 0$
2. الانحراف المعياري $\sigma (x) = 1$

لقد قام المتخصصين بالإحصاء الرياضي باستنتاج منحنى التوزيع الطبيعي المعياري من خلال استبدال كل قيمة x بالقيمة التالية :

$$(x-a / \sigma)$$

استنتاج العلاقة الجديدة التي تمثل معادلة منحنى التوزيع الطبيعي المعياري وهي التالية :

$$z =(x-a / \sigma)$$

ومن ثم عمل هؤلاء المتخصصين أيضاً على وضع جدول لمنحنى الكثافة للتوزيع الطبيعي المعياري سمي جدول احتمالات z ؛ حيث أن قيمة z لا يمكن أن تتجاوز 5 .

في جداول z نجد أن لكل قيمة z ما يقابلها في قيمة التابع f(z) أو ϕ(z) .

مثـــــال (1) :لدى معاينة 100 سنبلة تبين أن القيمة الوسط هي 60 بانحراف معياري قدره 10 ،تم سحب سنبلة بشكل عشوائي ،ما هو احتمال أن يكون عدد حباتها هو 60 حبة؟

الحـــل : لدين هنا المعطيات الأولية : a= 50 و σ= 10 ، إذن نطبق القانون اللازم :

$$z =(x-a / \sigma)$$

$$z = (60-50)/10$$

$$z = 1$$

الآن نعود إلى جدول توزيع z ونستخرج أن :

$$\phi(z) = \phi(1) = 0.34$$

أي أن احتمال أن يكون عدد حباتها هو 60 حبة هو p= 0.34 .

5-6-5- العلاقة بين التوزيع الطبيعي والتوزيع الثنائي :عندما يكون عدد الأفراد في العينة n كبيرا جدا وعندما يكون كلا من قيمتي p ,q قريبتين من بعضهما البعض فان التوزيع الثنائي يتحول إلى التوزيع الطبيعي بالقيم التالية :

القيمة الوسط $a = n * p$

الانحراف المعياري $\delta = \sqrt{n * p * q}$

مثـــــال (2) : تم رش 1000 شجرة في أحد البساتين لمعالجة أحد الأمراض علما أن فعالية المبيد النظرية هي 98% .والمطلوب هو حساب ما يلي :

1. ما هو احتمال أن تكون ثلاث أشجار على الأقل مصابة؟
2. ما هو احتمال أن يوجد شجرتين مصابتين على الأكثر؟
3. ما هو ؟

الحـــــــل : لدينا مسبقا الافتراضات التالية:

P = 1- q

n = 1000

$$P(x=r) = (r^n) * P^r * q^{n-r}$$

$\binom{n}{r} = n$!/(n-r!)*r!

a = n * p = 20

$\delta = \sqrt{n * p * q}$ = 4.43

الآن نطبق العلاقات المذكورة مع مراعاة القيم المسبقة لكل متغير :

1. احتمال أن تكون ثلاث أشجار على الأقل مصابة:

P(x>=3) = 0.5 + p(3<x<20)

p(3<x<20) = ϕ(20-20/4.43)- ϕ(3-20/4.43)

p(3<x<20) = 0.4999

P(x>=3) = 0.5 +4.999 = 0.9999

2. احتمال أن يوجد شجرتين مصابتين على الأكثر:

P(x<=2) = 0.5 – p(2<x<20)

p(2<x<20) = ϕ(20-20/4.43)- ϕ(2-20/4.43)

p(2<x<20) = 0.49997

$$P(x<=2) = 0.5 - 0.49997= 3*10^{-6}$$

3. احتمال أن يوجد اقل من 70 شجرة مصابة:

$$P(x<70) = 0.5 + p(20<x<70)$$

$$p(20<x<70) = \phi(70-20/4.43)- \phi(0)$$

$$p(20<x<70) = 0.5$$

$$P(x<70) = 0.5 + 0.5 =1$$

5-6-6- العلاقة بين التوزيع الطبيعي والانحدار والارتباط : تتجلى العلاقة بين التوزيع الطبيعي والانحدار والارتباط في الحالة الخاصة التالية :عندما تكون العوامل المستقلة المدروسة مقاسة بواحدات قياس مختلفة فانه من الأفضل تحويل المقادير إلى التوزيع الطبيعي المعياري ؛وذلك يتم نقوم بالخطوات التالية :

- نطرح من كل قيمة من القيم المتوسط الحسابي للعينة.
- ناتج الطرح نقسمه على الانحراف المعياري للعينة

بعد إجراء الخطوتين السابقتين نكون قد حولنا المقادير إلى التوزيع الطبيعي المعياري وأصبحت تتميز بميزات التوزيع الطبيعي المعياري وهي التالية:

- المتوسط الحسابي للعينة يساوي الصفر
- الانحراف المعياري للعينة يساوي الواحد.

 من الجدير بالتنويه هنا إلى أنه عند تحويل المقادير إلى التوزيع الطبيعي المعياري فإن كلا من عاملي الانحدار والارتباط يصبحان متطابقين تماما.

*****

خلاصة الفصل الخامس

- يرتبط الإحصاء والاحتمال ارتباطا وثيقا،فالاحتمال هو الأداة التي تمكن الإحصائي من استخدام المعلومات في عينة للقيام باستقراءات حول المجتمع الذي أخذت منه.
- يمكن أن نصنف الظواهر والحوادث بشكل عام إلى المجموعات التالية:
  1. حوادث مؤكدة : درجة احتمالها % 100
  2. حوادث مستحيلة : درجة احتمالها % 0.0
  3. حوادث محتملة أو متوقعة درجة احتمالها أكبر من % 0.0 وأصغر من % 100
- التوزيع الاحتمالي للمتحولات :ان الظواهر الطبيعية والتي تهمنا من الناحية الزراعية يمكن تمثيل قيمها كمجال لمتغير عشوائي.
- التوزيع الثنائي :يعتبر التوزيع الثنائي من التوزيعات الرياضية المعروفة والمهمة من الناحية التطبيقية .
- توزيع بواسون : نقول عن متحول عشوائي x أن توزيعه الاحتمالي يتبع توزيع بواسون إذا كان تابع التوزيع الاحتمالي له على الشكل التالي :

$$p(x=r) = (\lambda^{r} * e^{-\lambda}) / r \; !$$

- يجب أن تتوفر الشروط التالية استخدام توزيع بواسون :

$n > 50$ و $n * p < 5$

- نقول عن متحول عشوائي x أن توزيعه الاحتمالي يتبع التوزيع الطبيعي إذا كان تابع التوزيع الاحتمالي له على الشكل التالي :

$$f(x) = (1 * e^{-(x-a)/2\delta^2}) / \delta * \sqrt{2\pi}$$

- ان المساحة المحصورة تحت منحنى التوزيع الطبيعي والمحور الأفقي (x) يساوي الواحد كما هو متعارف عليه.
- ان مسقط الذروة لمنحنى التوزيع الطبيعي هي قيمة المتوسط ، وبالتالي فان المتوسط يقسم المساحة تحت منحنى التوزيع الطبيعي إلى قسمين متماثلين.
- ان ارتفاع الذروة يتعلق فقط بالانحراف المعياري لمنحنى التوزيع الطبيعي.

تحدد العلاقة بين منحنى التوزيع الطبيعي والاحتمالات فيما يلي :

- ان المساحة المحصورة تحت منحنى التوزيع الطبيعي والمحور الأفقي (x) يساوي الواحد كما هو متعارف عليه.
- ان أي قيمة من العينة تقع على يمين أو يسار المتوسط أي أكبر منه أو أصغر منه .
- ان احتمال أي قيمة من العينة تقع على يمين المتوسط هو مابين 0 و 0.5 و احتمال أي قيمة من العينة تقع على يسار المتوسط هو مابين 0 و 0.5 .
- ان احتمال أي قيمة من العينة محصورة ضمن مجال المتوسط زائد أو ناقص انحراف معياري واحد هو $p = 0.68$.
- ان احتمال أي قيمة من العينة محصورة ضمن مجال المتوسط زائد أو ناقص انحرافين معياريين هو $p = 0.95$.
- ان احتمال أي قيمة من العينة محصورة ضمن مجال المتوسط زائد أو ناقص ثلاث انحرافات معيارية هو $p = 0.99$.
- ان التوزيع الطبيعي المعياري هو منحنى معدل من التوزيع الطبيعي بحيث أصبحت ميزاته هبي التالية :
  1. المتوسط الحساب مساو للصفر أي $M(z) = 0$
  2. الانحراف المعياري $\sigma(x) = 1$
- عندما يكون عدد الأفراد في العينة n كبيرا جدا وعندما يكون كلا من قيمتي p ,q قريبتين من بعضهما البعض فان التوزيع الثنائي يتحول إلى التوزيع الطبيعي بالقيم التالية :

القيمة الوسط $a = n * p$

الانحراف المعياري $\delta = \sqrt{n * p * q}$

- عندما تكون العوامل المستقلة المدروسة مقاسة بواحدات قياس مختلفة فانه من الأفضل تحويل المقادير إلى التوزيع الطبيعي المعياري
- عند تحويل المقادير إلى التوزيع الطبيعي المعياري فإن كلا من عاملي الانحدار والارتباط يصبحان متطابقين تماما.

*****

مسائل وحلول متقدمة

التوزيع الثنائيBINOMDIST

إرجاع الحدود الجبرية الفردية لاحتمالية التوزيع ذات الحدين. استخدم BINOMDIST في المسائل التي تتعلق بعدد ثابت من الاختبارات أو التجارب، وعندما تكون نتائج أية تجربة عبارة

عن نجاح أو فشل فقط، وعندما تكون التجارب مستقلة، وعندما يكون احتمال النجاح ثابت في كافة مراحل التجربة. فعلي سبيل المثال، يمكن لـ BINOMDIST حساب احتمالية أن يكون طفلان من ثلاثة أطفال سيتم ولادتهما ذكوراً.

بناء الجملة لاستخدامها في الحاسب الالكتروني:

BINOMDIST(number_s,trials,probability_s,cumulative)

Number_s عدد مرات النجاح في التجارب.

Trials عدد التجارب المستقلة.

Probability_s احتمال النجاح في كل تجربة.

Cumulative القيمة المنطقية التي تحدد نموذج الدالة. إذا كانت cumulative TRUE، تقوم BINOMDIST بإرجاع دالة التوزيع التراكمي، وهي لاحتمال أن هناك عدد مرات (number_s) من النجاح في الغالب؛ وإذا كانت FALSE ، تقوم بإرجاع دالة مجموع الاحتمالات، وهي لاحتمال أن هناك عدد مرات (number_s) من النجاح.

تنويهات

- يتم اقتصاص Number_s و trials إلى أعداد صحيحة.
- إذا كانت number_s، أو trials، أو probability_s غير رقمية، تقوم BINOMDIST بإرجاع قيمة الخطأ !VALUE#.
- إذا كانت number_s < 0 أو إذا كانت number_s > trials، تقوم BINOMDIST بإرجاع قيمة الخطأ !NUM#.
- إذا كانت probability_s < 0 أو إذا كانت probability_s > 1، تقوم

BINOMDIST بإرجاع قيمة الخطأ !NUM#.

.COMBIN(n,x)

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

A **B**

1	**بيانات**	**الوصف**
2	6	عدد مرات النجاح في التجارب
3	10	عدد التجارب المستقلة
4	0.5	احتمالية نجاح كل تجربة

الصيغة	**وصف (الناتج)**
=BINOMDIST(A2,A3,A4,FALSE)	احتمالية نجاح ست تجارب من عشر تجارب (0.205078)

التوزيع الطبيعي NORMSDIST

لإرجاع دالة التوزيع التراكمي المعياري العادي. يكون للتوزيع وسيطاً يساوي 0 (صفر) وانحراف معياري يساوي 1. استخدم هذه الدالة في جدول يتضمن نواحي المنحنى المعياري العادي.

بناء الجملة لاستخدامها في الحاسب الالكتروني:

NORMSDIST(**z**)

Z هي القيمة التي تريد توزيعها.

تنويهات

- إذا كانت z غير رقمية، تقوم NORMSDIST بإرجاع قيمة الخطأ !VALUE#.
- معادلة دالة الكثافة المعيارية العادية هي:

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
1	**صيغة**	**وصف (الناتج(**
2	=NORMSDIST(1.333333)	دالة توزيع التراكمي العادي عند 1.333333 (0.908789)

التوزيع الطبيعي المعياريSTANDARDIZE

إرجاع قيمة قياسية من توزيع ممثل بمتوسط وانحراف معياري.

بناء الجملة لاستخدامها في الحاسب الالكتروني:

STANDARDIZE(x,mean,standard_dev)

X هي القيمة التي تريد أن تجعلها قياسية.

Mean (الوسط) هو الوسط الحسابي للتوزيع.

Standard_dev (الانحراف المعياري) هو الانحراف المعياري للتوزيع.

تنويهات

- إذا كان standard_dev ≤ 0، تُرجع STANDARDIZE القيمة الخطأ !NUM#.
- معادلة القيمة القياسية هي:

مثـــــــــــال:

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

A	B

	بيانات	وصف
1	بيانات	وصف
2	42	القيمة المراد جعلها قياسية
3	40	الوسط الحسابي للتوزيع
4	1.5	الانحراف المعياري للتوزيع

الصيغة	الوصف (الناتج)
=STANDARDIZE(A2,A3,A4)	القيمة القياسية للعدد (42) للبيانات أعلاه (1.333333).

اختبار Z - TEST

إرجاع القيمة ثنائية الطرف P-value لـ z-test. يقوم z-test بإنشاء الدرجة القياسية لقيمة x بالنسبة لمجموعة البيانات (المصفوفة) وإرجاع الاحتمال ثنائي الطرف للتوزيع العادي. يمكنك استخدام هذه الدالة لتقدير احتمال استنتاج ملاحظة معينة من مجموعة بيانات معينة.

بناء الجملة لاستخدامها في الحاسب الالكتروني:

ZTEST(**array**,**x**,sigma)

Array (مصفوفة) مصفوفة أو نطاق البيانات الذي سيتم اختبار x معه.

X القيمة التي يتم اختبارها.

Sigma (سيجما) الانحراف المعياري لمجموعة البيانات (معطى). عند حذفها، يتم استخدام الانحراف المعياري للعينة.

تنويهات

- إذا كان المصفوفة خالياً، ترجع ZTEST قيمة الخطأ #N/A.

مثـــــــــال:

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.

3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات،** أشر إلى **تدقيق الصيغة،** ثم انقر فوق **وضع تدقيق الصيغة**.

	A
	بيانات
1	3
2	6
3	7
4	8
5	6
6	5
7	4
8	2
9	1
10	9

الصيغة	**الوصف (الناتج)**
=ZTEST(A2:A11,4)	قيمة P-value ثنائية الطرف لـ z-test لمجموعة البيانات الموجودة أعلاه، عند القيمة 4 (0.090574)

11

توزيع POISSON:

لإرجاع توزيع Poisson. يعتبر التنبؤ بعدد الأحداث خلال زمن محدد من التطبيقات الشائعة لتوزيع Poisson ، كالتنبؤ بعدد السيارات التي تصل إلى ساحة الرسوم في دقيقة واحدة مثلاً.

بناء الجملة لاستخدامها في الحاسب الالكتروني:

POISSON(x,mean,cumulative)

X عدد الأحداث.

Mean القيمة الرقمية المتوقعة.

Cumulative القيمة المنطقية التي تحدد النموذج لتوزيع الاحتمال التي يتم إرجاعها. إذا كان cumulative يساوي TRUE، تقوم POISSON بإرجاع احتمال Poisson التراكمي الذي فيه تتراوح عدد الأحداث العشوائية بين صفر وبما فيه x؛ وإذا كان FALSE، تقوم بإرجاع دالة احتمال Poisson غير التراكمية التي تقع فيها عدد الأحداث تساوي x تماماً.

تنويهات

- إذا لم يكن x عدداً صحيحاً، يتم اقتصاصها .
- إذا كانت x أو المتوسط قيمة غير رقمية، تُرجع POISSON قيمة الخطأ !VALUE#.
- إذا كان x ≥ 0، تُرجع POISSON القيمة الخطأ !NUM#.
- إذا كان المتوسط ≥ 0، تُرجع POISSON قيمة الخطأ !NUM#.

مثـــــــال :

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
1	**بيانات**	**الوصف**
2	2	عدد الإحداث
3	5	المتوسط المتوقع

الصيغة	وصف (الناتج)
=POISSON(A2,A3,TRUE)	احتمال Poisson التراكمي بالشروط الموضحة أعلاه (0.124652)
=POISSON(A2,A3,FALSE)	دالة احتمال Poisson غير التراكمية بالشروط الموضحة أعلاه(0.084224)

تمارين الفصل الخامس

1. لدينا البيانات التالية التي تبين تطور الإنتاج من أحد المحاصيل الزراعية وهو الشعير خلال الفترة 1998-2001

السنة	1998	1999	2000	2001
الإنتاج	222	456	340	875

والمطلوب :

- ارسم المنحنى البياني للبيانات التالية باستخدام النوع المناسب من الرسوم البيانية.
- احسب معدل الإنتاج من محصول الشعير خلال الفترة المحددة من 1998وحتى 2001 .

- احسب قيم الإنتاج المتوقعة باستخدام معادلة منحنى التوزيع الطبيعي.
- ما هو التوزيع المناسب لبيانات الإنتاج؟

2. لدينا البيانات التالية التي تبين تطور عدد الطلاب في الجامعة:

السنة	1998	1999	2000	2001
عدد الطلاب	60000	70000	40000	20000
عدد المدرسين	200	170	220	190

والمطلوب:

- مثل البيانات السابقة في الجدول أعلاه بالطريقة البيانية المناسبة.
- ما هو التوزيع المناسب للبيانات الواردة في الجدول السابق ؟
- احسب أعداد الطلاب المتوقعة باستخدام معادلة منحنى التوزيع الطبيعي، ثم باستخدام معادلة منحنى التوزيع الثنائي وقارن النتيجة.

ج- بعض توزيعات الاحتمال الخاصة

سنقدّم في هذا القسم توزيعين من أكثر توزيعات الاحتمال فائدة لحلّ مسائل الاستدلال الإحصائي ، هذان التوزيعان ، أحدهما توزيع لمتغير متقطّع ، والآخر توزيع لمتغيّر متّصل ، هما . بلا شكّ . أكثر ما جرى استخدامه من توزيعات في مزاولة الإحصاء .

17-7- التوزيع ذو الحدين Binomial Distribution :

اعتبر تجربة يمكن فيها تبويب كل النتائج الممكنة كنتائج ينتج عنها حدوث حدث ، فإذا نتج عنه حدوث A اعتبرناه نجاحاً وإلاّ ، فهو فشل . واستخدام كلمة "نجاح" هنا ما هو إلاّ طريقة مناسبة لوصف وقوع حدث ما ، ولا يعني بالضرورة أن وقوع الحدث مرغوب فيه . نتوقع أن التجربة تكررت عدداً من المرات نرمز له بالحرف n ، ولندخل متغيّراً عشوائياً x يمثّل العدد الكلي لمرات النجاح التي حصلنا عليها ، أي عدد مرات وقوع A عند تكرار التجربة n من المرات . يسمى المتغيّر العشوائي الذي من هذا النوع متغيراً ذي حدين .

دعنا نعدل تلك التجربة بحيث تصبح عبارة عن إلقاء قطعة النقود مرة واحدة بدلاً من ثلاث مرات . لنعرف النجاح بأنه الحصول على رأس ، ولتكرار التجربة ، ثلاث مرات : إذن فلدينا هنا n= 3 . عندئذٍ يمثّل المتغيّر العشوائي x عدد الرؤوس التي نحصل عليها في ثلاث رميات كما كانت الحال في بند 2 ، الفقرة (ب). وإذن فسيعطي الشكل (7-19) الفقرة (ب)، توزيع المتغيّر العشوائي ذي الحدين . هذا ، وقد أعدنا تصويره في الشكل (7-21) .

أما تجربة دحرجة زهرتي طاولة مع تعريف x كمجموع النقاط على الزهرتين فلا تعطينا متغيّراً من متغيّرات ذي الحدين ، إذ لا يمكن للمرء اعتبار كل دحرجة زهرة كناتج يعطينا نجاحاً أو فشلاً مع جعل x تمثّل مجموع عدد مرات النجاح . فهناك ست إمكانيات لكل من الزهرتين لا إمكانيتان فقط ، كما

تتطلّب مسألة ذي الحدين .

وتجربة سحب كرة من صندوق يحتوي على ثلاث كرات حمراء ، وكرتين سوداوتين وكرة خضراء ، ليست هي الأخرى تجربة من التجارب التي تؤدي إلى متغيّر من متغيّرات ذي الحدين في الصورة الحالية له ، إذ لدينا هنا ثلاث نتائج ممكنة بدلاً من نتيجتين . ومع هذا ، فإذا كان اهتمامنا محصوراً فقط في معرفة ما إذا كنا سنحصل على كرة حمراء ، فستصبح المسألة مسألة توزيع ذي الحدين . وسيمثّل المتغيّر x عدد الكرات الحمراء التي نحصل عليها عند إجراء التجربة n من المرات . هذا ، وفي مسائل متغيرات ذي الحدين ، يفهم دائماً أنه ، إذا كانت التجربة تتضمن سحب أشياء من أوعية ، فإن إرجاع الأشياء المسحوبة يتم قبل البدء في إجراء التجربة الآتية ، إذ إن تكرار التجربة يجب أن يكون تكراراً للتجربة الأصلية من كافة جوانبها .

الشكل (7–21): توزيع متغيّر عشوائي في تجربة رمي قطعة نقود

ولننظر الآن في مثال أعقد نوعاً لمتغيّر من متغيّرات ذي الحدين ، لنر كيف نحصل على توزيعه . لتكن التجربة الأساسية هي دحرجة زهرة طاولة مرة أخرى ، وليكن تعريف النجاح هو الحصول على آس (أي نقطة واحدة) . سنجري التجربة ثلاث مرات ، أي أن $n = 3$.

نحصل على توزيع المتغيّر العشوائي هذا ، يمكننا اتّباع الطريقة نفسها التي اتّبعناها في تجربة رمي قطعة النقود الممثلة في الشكل (7-21) . هذه الطريقة تتكوّن من النظر إلى مجال العيّنة الأصلي للتجربة الكاملة ثم اختزاله إلى مجال عيّنة جديدة للمتغيّر العشوائي x ، وذلك بتطبيق تعريف احتمال الأحداث على مجال العيّنة الأصلي . والآن ، عند كل دحرجة لزهرة طاولة ، ما علينا إلاّ أن نسجّل إذا كان ما حدث هو نجاح أو فشل ، حيث كلمة نجاح تعني الحصول على آس . فإذا استخدمنا الحرفين S , F ليمثّلا النجاح والفشل فسيكون لدينا ثمان نقط في مجال العيّنة ، كما كانت الحال عند رمي قطعة النقود ثلاث رميات ، وكان لابدّ من ظهور H أو T عند كل رمية . ولقد مثلنا هذه النقاط في جدول (7-5) باستخدام الحرفين S , F ليبيّنا النتائج المختلفة الممكنة . كما بيّنّا أيضاً في جدول (7-5) القيم المناظرة للمتغيّر العشوائي x .

جدول (7-5)

النتائج	SSS	SSF	SFS	FSS	SFF	FSF	FFS	FFF
قيم x	3	2	2	2	1	1	1	0

سنلاحظ أن هذا الجدول هو بالضبط نفس جدول مسألة رمي قطعة النقود ثلاث مرات نفسه والذي سبق أن بيّنّاه في الشكل (7-15) ، الفقرة (ب) . بيد أن حساب الاحتمالات قيم x

المختلفة تختلف اختلافاً بيّناً ، فاحتمالات هذه النتائج الثمانية الممكنة ليست متساوية كما كانت الحال في مسألة قطعة النقود . لنحسب هنا الاحتمالات باستخدام قاعدة الضرب للاحتمالات . فحيث أن دحرجات الزهرات الثلاثة مستقلة ، وإن احتمال النجاح (أي الحصول على آس) هو الآن 1/6 ينتج . على سبيل المثال . أن :

$$P\{SFS\} = \frac{1}{6}.\frac{5}{6}.\frac{1}{6} = \left(\frac{1}{6}\right)^2.\left(\frac{5}{6}\right)$$

وبحسابات مماثلة لهذه الخطوات نحصل على احتمالات كل من النتائج الثمان الممكنة ، وقد بيّنا نتائج مثل هذه الحسابات في جدول (6-7) .

جدول (6-7)

النتائج	SSS	SSF	SFS	FSS	SFF	FSF	FFS	FFF
الاحتمال	$\left(\frac{1}{6}\right)^3$	$\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$	$\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$	$\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)$	$\left(\frac{1}{6}\right).\left(\frac{5}{6}\right)^2$	$\left(\frac{1}{6}\right).\left(\frac{5}{6}\right)^2$	$\left(\frac{1}{6}\right).\left(\frac{5}{6}\right)^2$	$\left(\frac{5}{6}\right)^3$

وإذن ، نستطيع أن نقارن مجال العينة ذي الثماني نقاط مع احتمالاتها المخصصة إلى مجال العينة للمتغيّر العشوائي x ، وذلك بحساب احتمالات الأحداث المركبة المناظرة للقيم المختلفة للمتغيّر العشوائي . ولقد أجرينا ذلك بمقارنة الجدولين (5-7) , (6-7) . فمثلاً ، يتكوّن حدث x = 2 من الأحداث البسيطة SSF , SFS , FSS وعليه ، فإننا نحصل على احتمال هذا الحدث بجمع الاحتمالات المقترنة بتلك الثلاث . ويبيّن جدول (7-7) مثل هذه الحسابات . كما يعطينا هذا الجدول توزيع المتغيّر العشوائي المطلوب . أما شكله البياني فيعطيه الشكل (22-7) ، وفيه عبّرنا عن احتمالات بصورة عشرية لرقمين عشريين .

جدول (7-7)

x	0	1	2	3
P{x}	$\left(\frac{5}{6}\right)^3$	$3\left(\frac{1}{6}\right).\left(\frac{5}{6}\right)^2$	$3\left(\frac{1}{6}\right)^2.\left(\frac{5}{6}\right)$	$\left(\frac{1}{6}\right)^3$

وفي ضوء هذه النتائج ، يتّضح لنا أنه لا ينبغي للمرء أن يتوقّع الحصول على 3 آسات عند دحرجة ثلاث زهور طاولة . فمثل هذه النتيجة تحدث . على المدى البعيد مرة واحدة تقريباً كل $216 = 6^3$ مرة كما يتّضح لنا أنه لا ينبغي للمرء أن يقبل رهاناً . حتى لو كان مبنياً على المال . أنه سيحصل على الأقل على آس واحد عند دحرجة ثلاثة زهور طاولة ، إذ إن مثل هذه النتيجة تحدث في حوالي 42% ، وعليه فإذا كنت تودّ الفوز على أصدقائك البسطاء ، فراهنهم على أنهم لن يحصلوا على آس واحد على الأقل عند دحرجة ثلاثة زهور طاولة .

ويمكن استخدام الطريقة التي استخدمناها في المثالين التوضيحيين السابقين في أية مسألة خاصة قد تنشأ من مسائل ذي الحدين . فمثلاً إذا كانت هناك 5 إعادات لتجربة النجاح أو الفشل الأساسية ، فسيكون هناك $32 = 2^5$ نقطة في مجال العينة . عندئذٍ نحسب الاحتمال المقترن بكل نقطة ، ثم نربط بكل نقطة من نقط مجال العينة القيمة المناسبة للمتغيّر العشوائي x ، ثم بجمع الاحتمالات الخاصة بتلك النقاط التي تناظر قيمة معينة للمتغير x ، نحصل على الاحتمالات للقيم المختلفة للمتغيّر x ، وهذه الاحتمالات تعطي توزيع الاحتمال المطلوب لمتغيّر ذي الحدين للمتغير العشوائي x.

الشكل (7-22) : توزيع عدد من الآسات عند دحرجة زهرة طاولة ثلاث مرات

هذا ، وبسبب صعوبة السير في الحسابات السابقة في كل مرة تنشأ فيها مسألة من مسائل ذي الحدين ، فمن المناسب أن تكون لدينا صيغة يمكن تطبيقها في كل مسألة من تلك المسائل . ولهذا الغرض ، لندرس مسألة عامة من مسائل ذي الحدين ، لنجرِ تجربة يمكننا فيها على الدوام تبويب النتائج إما كنجاح أو كفشل ، لنفرض أن احتمال الحصول على نجاح هو قيمة معلومة ، ولنرمز إليها بالرمز P كذلك ، لنرمز للاحتمال المناظر للفشل بالرمز q . إذن $P + q = 1$ نجري التجربة n من المرات . فإذا رمزنا لعدد مرات النجاح التي سنحصل عليها في الإعادات التي عددها n للتجربة بالحرف x فإن المسألة تكون مسألة حساب الاحتمالات للقيم الممكنة المتنوّعة للمتغيّر العشوائي x . وكما سبق أن قلنا، يمكننا إجراء هذه الحسابات بطريقة منظّمة في أية مسألة معطاة ، لن نعطي هنا سوى نتائج تلك الحسابات وسنستخدم الرمز $P\{x\}$ ليرمز إلى الاحتمال الخاص بقيمة نموذجية لهذا المتغيّر العام من متغيرات ذي الحدين . هذه الاحتمالات التي تعرف بتوزيع ذي الحدين وتعطى بالصيغة العامة الآتية :

(1)– توزيع ذي الحدين :

$$P\{x\} = \frac{n!}{x!(n-x)!}.P^x.q^{n-x} \qquad (1)$$

وكما بيّنا سابقاً نقرأ الرمز n! "مضروب n " . أنه يرمز إلى حاصل ضرب كل الأعداد الصحيحة الموجبة من 1 إلى n ، فمثلاً : $5! = 1 \;\; 2 \;.\; 3 \;.\; 4 \;.\; 5$، وكما لاحظنا هناك تعرف القيمة 0! على أنها 1 .هذا ولقيم x , n الكبيرة تصبح الحسابات المتضمّنة في عملية تقييم الكمية $n! \,/\, x!\,(n-x)!$ حسابات صعبة . هذا الجدول يتضمّن قيم n من 2 إلى 20 ، وقيم x من 2 إلى n/2 أو أكبر . فليس من الضروري جدولة كل قيم x التي تزيد عن هذا ، ورأينا أن هذه الفكرة عند معالجة قيم x التي تكبر n/2 ، وبالرغم من أن هذا الجدول يمكّننا من الكتابة السريعة لاحتمالات ذي الحدين ، فإننا لن نستخدمه في المسائل التوضيحيةالآتية ، إذ يبدو أن من المرغوب فيه بالنسبة للطالب أن يألف الصيغة (1) ، وذلك بإجراء الحسابات المطلوبة عندما تكون n صغيرة نوعاً .

ومن المعتاد أن نتحدّث عن التجارب التي عددها n . كمحاولات مستقلة . لتجربة ما ، فيها *P* هو احتمال النجاح في محاولة واحدة . وبناءً على ذلك تكون P{x} هي احتمال الحصول على x من مرات النجاح في n من المحاولات المستقلة لتجربة ما فيها *P* احتمال النجاح في محاولة واحدة .

والتوزيعان التكراريان المبيّنان في شكلي(7-21),(7-22) هما حالتان خاصتان لتوزيع ذي الحدين العام المعطى بالصيغة (1) . ففي مسألة رمي قطعة النقود ، تكون $n=3$ $P=q=½$ وفي مسألة دحرجة زهرة الطاولة تكون $n = 3$, $P = 1/6$, and $q = 5/6$ ، وينبغي للقارئ أن يراجع قيم P{x} المعطاة في الشكل (7-21) ، وفي الشكل (7-22) ، وذلك باستخدام الصيغة (1) بهدف أن يألف استخدامها . وكمثال توضيحي لمثل هذه المراجعة ، نفترض أننا نريد حساب قيمة P{3} في مسألة زهرة الطاولة . إن تعويض القيم المناسبة في (1) يعطينا :

$$P\{3\} = \frac{3!}{3!.0!}.\left(\frac{1}{6}\right)^3.\left(\frac{5}{6}\right)^0$$

وحيث أن 1 = !0 ، وأن أي عدد . في الجبر . مرفوع للقوة الصفرية يساوي الوحدة فإن :

$(\frac{5}{6})^0 = 1$ إذن :

$$P\{3\} = \left(\frac{1}{6}\right)^3$$

هذا ، وبالرغم من أن المسائل التي استخدمناها في تقديم توزيع ذي الحدين كانت مرتبطة بألعاب الحظ ، إلاّ أن توزيع ذي الحدين مفيد للغاية في حلّ أنواع معينة من المسائل العملية ، وهنا ، فسنناقش بضعة مسائل بسيطة لا تتطلّب سوى حسابات سهلة عن طريق الصيغة (1) .

مثال :

احتمال أن يكون لوالدين . أعينهما من نوع معين من العيون الزرقاء . طفل ذو عينين زرقاوين هو ¼ . فإذا كان في الأسرة ستة أطفال ، فما احتمال أن يكون لنصفهم على الأقل عيون زرقاء ؟

لكي نحلّ هذه المسألة نعامل الأطفال الستة في الأسرة كمحاولات مستقلّة . عددها 6 . لتجربة ما ، احتمال النجاح في محاولة واحدة ¼ ، وعلى هذا لدينا هنا $P=\frac{1}{4}$ $n=6$ ومن الضروري حساب P{6} و P{5} , P{4} , P{3} ثم جمعها ، إذ إن هذه الاحتمالات تناظر الطرق التي درسناها .

$$P\{3\}=\frac{6!}{3!.3!}.\left(\frac{1}{4}\right)^3.\left(\frac{3}{4}\right)^3=\frac{540}{4096},$$

$$P\{4\}=\frac{6!}{4!.2!}.\left(\frac{1}{4}\right)^4.\left(\frac{3}{4}\right)^2=\frac{135}{4096},$$

$$P\{5\}=\frac{6!}{5!.1!}.\left(\frac{1}{4}\right)^5.\left(\frac{3}{4}\right)^1=\frac{18}{4096},$$

$$P\{6\}=\frac{6!}{6!.0!}.\left(\frac{1}{4}\right)^6.\left(\frac{3}{4}\right)^0=\frac{1}{4096}$$

ونحصل على احتمال الحصول على عدد من مرات النجاح يساوي ثلاثة على الأقل بجمع هذه الاحتمالات ، وبالتالي ، يحصل المرء . بكتابة $x \geq 3$ لتمثيل ثلاث مرات نجاح على الأقل . :

$$P\{x\geq 3\}=\frac{694}{4096}=.169.$$

وتبيّن هذه النتيجة أن هناك فرصة ضئيلة جداً أن يكون لأسرة مثل هذه الأسرة عدد كهذا من الأطفال زرق العيون ، ففي حوالي 17 من كل 100 من هذه الأسر ، يكون على الأقل نصف الأطفال ذوي عيون زرقاء .

مثال :

يضمن أحد أصحاب مصانع قطع معينة للسيارات ألاّ يحتوي أي صندوق من صناديق قطعة على أكثر من قطعتين معيبتين ، فإذا كان الصندوق يحتوي على عشرين قطعة ، ودلّت التجربة على أن طريقة التصنيع التي يتّبعها تنتج 2% من القطع المعيبة ، فما احتمال أن يحقق صندوق ما من صناديق قطع المصنع الضمان .

يمكن عد هذه المسألة مسألة من مسائل توزيع ذي الحدين ، فيها $n = 20 \; , \; P = .02$ وسيحقق صندوق ما الضمان إذا كان عدد القطع المعيبة هو 0, 1, or 2 وتعطي احتمالات هذه الأحداث الثلاثة . عن طريق الصيغة (1) . بما يأتي :

$$P\{0\} = \frac{20!}{0!.20!}.(.02)^0.(.98)^{20} = (.98)^{20} = .668,$$
$$P\{1\} = \frac{20!}{1!.19!}.(.02)^1(.98)^{19} = 20.(.02)^2.(.98)^{18} = .273,$$
$$P\{2\} = \frac{20!}{2!.18!}.(.02)^2.(.98)^{18} = 190(.02)^2.(.98)^{18} = .053$$

وحيث أن هذه الأحداث متنافية ، يكون احتمال وجود قطعتين معيبتين على الأكثر أي $x \le 2$ هو مجموع هذه الاحتمالات ، وإذن فإن الجواب المطلوب هو :

$$P\{x \le 2\} = .994 .$$

مثال :

كمثال توضيحي أخير ، ندرس المسألة الآتية المتعلّقة باستخدام الحدس في الامتحان ، نفترض أن امتحاناً ما يتكوّن من عدد كبير من الأسئلة من النوع ذي الاختيار المتعدد . لكل سؤال خمس إجابات ممكنة ، واحدة منها فقط هي الإجابة الصحيحة ، فإذا كان الطالب يحصل على 3 نقاط لكل إجابة صحيحة ، في حين يحصل على (1-) نقطة لكل إجابة غير صحيحة ، وإذا كان احتمال حدسه للإجابة الصحيحة في كل سؤال من الأسئلة العشرة هي 1/3 فقط ، فما احتمال حصوله على مجموع كلي موجب عن تلك الأسئلة العشرة ؟

وإذا كانت x ترمز إلى عدد الأسئلة التي أجاب الطالب نفسها إجابة صحيحة فإن مجموعاً كلياً موجباً ينتج إذا كان $3x > 10 - x$ ، فإذا كان الطرف الأيسر لهذه المتباينة يعطي عدد النقاط الموجبة التي يحصل عليها ، في حين يعطي الطرف الأيمن عدد نقاط الخطأ ، وستتحقق هذه المتباينة إذا كانت $x > 10/4$ ، وهذا يعني أنه من الضروري الحصول على ثلاث إجابات صحيحة على الأقل لتحقيق مجموع كلي موجب ، وعلى هذا لتحقيق مجموع كلي موجب ، يعطي الاحتمال الآتي :

$$P\{x \geq 3\} = 1 - \sum_{x=0}^{2} \frac{10!}{x!.(10-x)!} \cdot \left(\frac{1}{3}\right)^{x} \cdot \left(\frac{2}{3}\right)^{10-x}$$
$$= 1 - \left[\left(\frac{2}{3}\right)^{10} + 10\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{9} + 45\left(\frac{1}{3}\right)^{2} \cdot \left(\frac{2}{3}\right)^{8}\right]$$
$$= .70.$$

إذن ، لدى الطالب فرصة ممتازة لكسب النقاط إذا كان احتمال حدسه لإجابة صحيحة في كبر القيمة 1/3 ، أما لو كان لا يعرف شيئاً عن موضوع الامتحان ، واختار واحداً من البدائل كيفما اتفق ، لكان بالطبع احتمال اختياره للإجابة الصحيحة عن ل سؤال هو 1/5 فقط ، بيد أننا قد افترضنا هنا أن الطالب ملمّ بموضوعه إلماماً كافياً يجعله يستبعد اثنين من الممكنات الخمسة معتبراً إياها إجابتين بعيدتين عن الصحة بعداً واضحاً ، فيحصر حدسه في الممكنات الثلاثة الباقية، أما لو لم يكن له مثل هذا الإلمام بالموضوع ، وبذا كان احتماله 1/5 ، فإن حسابات شبيهة بتلك الحسابات السابقة ترينا أن ليس هناك من جدوى للحدس .

غالباً ما تصير عملية حساب الاحتمالات ذي الحدين عملية ثقيلة ، حتى باستخدام جدول III ، وذلك عندما يتطلّب الأمر جمع عدد من هذه الاحتمالات ، شريطة ألاّ تكون في متناول أيدينا حاسبات إلكترونية . هذا ، وسنجد أن جدول III في الملحق مفيد في هذه الحالات ، إذا كانت قيمة n لا تتجاوز 10 ، فإن ذلك الجدول يعطي مجموع احتمالات الذيل الأيمن على الصورة :

$$P\{x \geq 10\} = \sum_{x=x0}^{n} \frac{n!}{x!.(n-x)!} . P^{x} . q^{n-x}$$

وحيث أن القيم التي لدينا للاحتمال P تبدأ بالقيمة 05. وتزداد إلى 50. في خطوات قيمة كل منها 05. ، فإنه من الواجب إجراء عمليات توليد القيم P الأخرى التي تقلّ عن 50. ، أما إذا كانت قيمة P تزيد عن 5. ، فإن المرء يستخدم في الجدول $q= 1-P$ بدلاً من $n-x_o+1$, P بدلاً من x_o ، ثم يطرح الناتج من الوحدة ليحصل $P\{x \geq x_o\}$ ، أما في حالة $n > 10$ فمن

الضروري الرجوع إلى تقريب تلك الاحتمالات ، وسنعرض في بند 4 طريقة من طرق إجراء هذا التقريب .

ولتوضيح استخدام جدول III ، ندرس تطبيقه على آخر مثال من الأمثلة الثلاثة السابقة . إن قيم جدول III التي تحتاج إليها $n= 10$, $P= 1/3 = .33$, $x_o= 3$ وباستخدام التوليد للقيمة $P= .33$ نحصل على $P\{x \geq 3\} = .70$ صحيحاً لرقمين، وهو الجواب نفسه الذي سبق أن حصلنا عليه .

18-7– خواص توزيع ذي الحدين Binomial Distribution Properties:

حيث أن لدينا من الباب الرابع صيغتين عامتين لحساب وسط وتباين متغيّر عشوائي متقطّع فإنه من الممكن استخدام هاتين الصيغتين للحصول على وسط وتباين أي متغير من متغيرات ذي الحدين . فالحسابات التي أنتجت لنا الجدولين (7-2) , (7-3) في الفقرة (ب) هل تلك التي نحتاج إليها للمتغير الخاص من متغيرات ذي الحدين الذي له $n = 3$ and $P = ½$.

ولكي نألف أكثر مثل هذه الحسابات ، دعنا ندرس مسألة حساب الوسط والتباين لمتغير ذي الحدين الذي يعطيه جدول (7-7) والذي يبيّن رسمه البياني في الشكل (7-22) .

هنا : $n = 3$, $P = 1/6$.

والقيم الممكنة للمتغيّر x هي $x_1= 0$, $x_2= 1$, $x_3= 2$, $x_4= 3$ ، أما القيم المناظرة للاحتمالات $P\{x_i\}$ فيعطيها جدول 3 ، وحساب قيمة μ التي بنيت على تلك القيم مبينة في جدول (7-8) .

من هذه الحسابات ، ينتج أن : $\mu = \frac{108}{216} = \frac{1}{2}$

جدول (7-8)

X	P{X}	XP{X}
0	$\frac{125}{216}$	0
1	$\frac{75}{216}$	$\frac{75}{216}$

2	$\frac{15}{216}$	$\frac{30}{216}$
3	$\frac{1}{216}$	$\frac{3}{216}$
		$\frac{108}{216}$

أما حسابات σ فيبيّنها جدول (7-9) .

جدول (7-9)

x	$P\{x\}$	$x - \mu$	$(x - \mu)^2$	$(x-\mu)^2\, P\{x\}$
0	$\frac{125}{216}$	$-\frac{1}{2}$	$\frac{1}{4}$	$\frac{125}{864}$
1	$\frac{75}{216}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{75}{864}$
2	$\frac{15}{216}$	$\frac{3}{2}$	$\frac{9}{4}$	$\frac{135}{864}$
3	$\frac{1}{216}$	$\frac{5}{2}$	$\frac{25}{4}$	$\frac{25}{864}$
				$\frac{360}{864}$

ومن هذه الحسابات ينتج ، أن :

$$\sigma = \sqrt{\frac{360}{864}} = \sqrt{\frac{5}{12}}$$

ومن الممكن أن نستخدم بعض الحيل الجبرية لإجراء حسابات مماثلة للتوزيع العام، الخاص بذي الحدين ، الذي تعطيه الصيغة (1) . إن مثل هذه الحسابات تمدّنا بصيغتين : واحدة للوسط والأخرى للانحراف المعياري ، يمكن استخدامها في كل مسائل ذي الحدين ، وحيث أن مثل هذه الحسابات معقّدة نوعاً ما ، فلن نجرّبها هنا ، بل سنعطي الصيغتين الناتجتين ، وسوف نستخدمها في حل مسائل ذي الحدين ، هاتان الصيغتان وهما في غاية البساطة ، هما :

$$\mu = nP \qquad (2)$$

$$\sigma = \sqrt{nPq}$$

وستتضح ميزة أن يكون لدينا مثل هاتين الصيغتين البسيطتين عندما نطبقهما في المسألة الحسابية التي أتممناها آنفاً ، فللتوزيع الخاص بجدول 3 ، لدينا : $P = 1/6$ ، و $n= 3$ إذن :

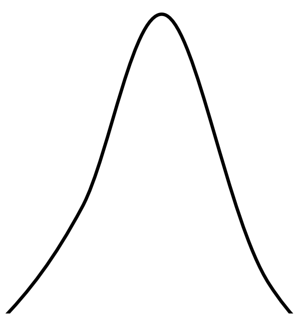
$$\mu = 3.\frac{1}{6} = \frac{1}{2}$$
$$\sigma = \sqrt{3.\frac{1}{6}.\frac{5}{6}} = \sqrt{\frac{5}{12}}$$

وهذا توفير كبير في الوقت متى قورن بالوقت الذي يحتاج إليه المرء لحساب هاتين الكميتين عن طريق الصيغتين العامتين ، كما في الجدولين (7-8),(7-9) .

19-7– التوزيع الطبيعي أو المعتدل Normal Distribution:

إن توزيع ذي الحدين الذي أوردناه في بند 1 هو أكثر توزيعات الاحتمال فائدة للمتغيرات المتقطّعة ، أما التوزيع الذي سندرسه في هذا البند فهو أكثرها فائدةً للمتغيّرات المتّصلة .

إن المدرّج التكراري في الشكل (7-21) ، الفقرة (ب) ، نموذج لكثير من التوزيعات التي نجدها في الطبيعة وفي الصناعة . فمثل هذه التوزيعات متماثلة ، وتتضاءل بسرعة عند الذيلين ، كما أن لها شكلاً يشابه شكل القوس . وسنرى حالاً أن هناك توزيعاً نظرياً يسمى التوزيع الطبيعي ، قد ثبتت فائدته الكبيرة ، لمثل هذه التوزيعات ، كما أنه مهم للغاية من نواحٍ أخرى والمنحني المرسوم في الشكل (7-21) الفقرة (ب) ، هو رسم بياني لتوزيع طبيعي خاص . أما الرسم البياني لتوزيع طبيعي عام فيعطيه الشكل (7-24) . هذا ، وبالرغم من كون التوزيع الطبيعي يعرف بمعادلة منحني التوزيع ، إلاّ أن هذه المعادلة لن تستخدم ، ولذا فلن نكتبها ، بل سننظر إلى المنحني نفسه كتعريف للتوزيع .



x

μ-3σ μ-2σ μ-σ μ μ+σ μ+2σ μ+3σ

الشكل **(7-24)** : توزيع طبيعي نموذجي

لعلّك قد سمعت عن المنحني الطبيعي فيما يتعلّق بتوزيع التقديرات لمناهج معيّنة في فصول كثيرة ، إن تقدير أداء الطلبة على أساس المنحني الطبيعي يفترض أن توزيع الناتج الذهني للطلبة ينبغي أن يماثل توزيع كثير من الخواص الجسدية لهم ، التي ندرك أنها تتوزّع طبيعياً على وجه التقريب . كذلك . في العادة . يفترض المعلّم الذي يستخدم مثل هذا النظام في وضع تقديراته لأداء الطلبة أنهم . وقد حضروا منهجه . كانت لديهم الخلفيات المناسبة التي يتطلّبها هذا المنهج ، وأنهم قد درسوا العدد المتوقّع من ساعات الدراسة . يا له من معلّم متفائل للغاية ! بيد أن إحدى صعوبات هذا النظام هي أنه يخصص التقديرات على أساس نسبي لا على أساس مطلق . فمثلاً ، إذا كان هناك فصل طلابي موهوب بدرجة غير عادية ، طلبته يشتغلون بجد ، فإنهم سيعطون نفس توزيع التقديرات التي تعطى لفصل من الكسالى الأغبياء . وبالطبع ، يمكننا الردّ على هذا النقد ، إذ إن الطلبة لا يختلفون من عام إلى آخر سوى باختلاف ضئيل جداً ، وعليه ، فليس من المتوقّع أن يتكوّن فصل كبير إما من طلبة مجدّين أذكياء أو من طلبة كسالى أغبياء . هذا ، وإذا كنت قد كرهت المنحني الطبيعي بسبب التقديرات التي ربما قد أعطيتها ، فعليك أن توقن أن المنحني الطبيعي في الواقع ليس هو الملوم ، وأنه ينبغي لك أن تكون كريماً بما فيه الكفاية كي تقترب من دراسته هنا دون تحيّز ، فهو منحن بالغ النفع في مجالات كثيرة من مجالات التطبيق بعيدة تماماً عن مسائل الدراسة .

لعلّنا نذكر من أن وسط توزيع ما يمثّل النقطة الواقعة على محور x التي عندها تتّزن رقيقة معدنية على شكل مدرّج تكراري التوزيع إذا وضعت على مرتكز . هذه الخاصة الهندسية للوسط توضح لنا أنه متى كان مدرّج تكراري ما متماثلاً حول محور رأسي ، فلابدّ أن يقع الوسط عند نقطة التماثل على محور x. وهذا صحيح أيضاً للقيمة النهائية للوسط عندما تجعل حجم العينة يكبر كبراً

بلا حدود ، ونجعل فترة الفئة تصغر بلا حدود . وهذا يفسّر لنا كذلك سبب وضع الرمز μ عند نقطة التماثل على محور x في الشكل (7-24) .

والآن نفترض أن الشكل النهائي للمدرّج التكراري الخاص بتوزيع تكراري ما ، بالمفهوم الذي سبق شرحه ، هو منحن طبيعي . إذن ، يمكننا أن نبيّن بطرق الرياضيات المتقدمة أن σ ، قيمة S النهائية ، لها التفسير الهندسي الآتي بالنسبة للمنحني الطبيعي النهائي :

(أ): المساحة الواقعة تحت المنحني الطبيعي بين $\mu+\sigma$ و $\mu - \sigma$ هي 68% من المساحات الكلية لأقرب 1% .

(ب): المساحة الواقعة تحت المنحني الطبيعي بين $\mu+ 2\sigma$ و $\mu - 2\sigma$ هي 95% من المساحات الكلية لأقرب 1% .

(ج): المساحة الواقعة تحت المنحني الطبيعي بين $\mu+3\sigma$ و $\mu - 3\sigma$ هي 99.7% من المساحة الكلية لأقرب 1% .

لقد وضعنا علامات على المحور في الشكل (7-24) نقسّمه إلى وحدات من σ ، مبتدئين من الوسط μ ، ويتّضح من هذا الشكل أنه ليس هناك تقريباً أية مساحة تحت المنحني فيما يلي: μ بمقدار 3σ من الوحدات ، بيد أن معادلة المنحني ترينا أن المنحني يمتدّ فعلاً من $- \infty$ to $+ \infty$.

هذا ، وقد سبق أن استخدمنا الخاصتين الأولى والثانية من الخواص السابقة لتفسير معنى *S* في الباب الثاني . ولقد وجدنا هناك أن النسبتين المئويتين هاتين صحيحتان على وجه التقريب للمدرجات التكرارية التي لها شكل يشابه المنحني الطبيعي .

الشكل (7-25) : توزيع طبيعي فيه $\mu= 0 , \sigma= 1$

الشكل (7-26) : توزيع طبيعي فيه $\mu= 0 , \sigma= 3$

وثم خاصية مثيرة للاهتمام للمنحني الطبيعي ، ألا وهي أن موقعه وشكله يتحددان كليةً بقيمتي μ , σ الخاصتين به . فقيمة μ تقع طبعاً عند مركز المنحني ، في حين تحدد قيمة σ مدى انتشاره، وحيث أن كل المنحنيات الطبيعية التي تمثّل توزيعات تكرارية نظرية لها مساحة كلية تساوي الوحدة ، فكلّما زادت σ قلّ بالضرورة ارتفاع المنحني وزاد انتشاره . ولقد أوضحنا هذا في الشكل (7-25) و (7-26) حيث رسمنا منحنيين طبيعيين لهما نفس الوسط 0 ، لكن لهما الانحرافين المعياريين 3 و 1 على الترتيب . وحقيقة كون شكل المنحني الطبيعي يتحدد كليةً بانحرافه المعياري تمكن المرء من اختزال المنحنيات الطبيعية جميعها إلى منحن معياري واحد بإجراء تغيير بسيط للمتغير . فمثلاً ، يمكن جعل منحني الشكل (7-25) يشابه منحني الشكل (7-26) بتغيير مقياس الرسم على محور x بحيث تمثّل الوحدة على محور الشكل -25) (7 ثلاث وحدات على محور الشكل (7-26) . إن هذا ت على وجه التقريب . يُناظر أخذ منحني الشكل (7-25) ومعالجته هو والمساحة التي تحته كما لو كان مصنوعاً من المطاط ، فنمدّه . مع الاحتفاظ بمساحته . إلى ما يساوي ثلاث مرات طوله الطبيعي . وبالعكس ، يمكن جعل الشكل (7-26) بأخذ الشكل (7-25) بضغط الشكل (7-26) إلى ثلث طوله الطبيعي . وفي هذا نفترض أن الذيلين الطويلين لهذين المنحنيين قد قطعا وقد أهملا . ولما كان أبسط منحن طبيعي يمكننا العمل به هو ذلك المنحني الذي يقع وسطه عند الصفر الذي يكون انحرافه المعياري

الوحدة ، فقد جرت العادة على اختزال المنحنيات الطبيعية إلى هذا المنحني المعياري كلّما احتجنا إلى إجراء الاختزال . والآن ، لكل نقطة على محور x لمنحن طبيعي ما نقطة مناظرة على محور x للمنحني الطبيعي المعياري يمكن تعيين قيمتها بذكر مدى بعدها عن نقطة الوسط للمنحني بدلالة الانحرافات المعيارية فمثلاً ، النقطة x = 6 في الشكل (7-26) تناظر النقطة x= 2 على المنحني الطبيعي المعياري في الشكل (7-25)، إذن يمكننا الحصول على القيمة x = 6 من الشكل (7-25) بأن نذكر أنها تقع إلى يمين الوسط صفر بمقدار انحرافين معياريين .

وبوجه عام ، إذا ناظرت نقطة x على محور منحن طبيعي ما، وسطه μ وانحرافه المعياري σ ، نقطة z على المنحني الطبيعي المعياري ، وقعت x إلى يمين μ بمقدار z من الانحرافات المعيارية . وعلى هذا ، تعطي العلاقة بين هاتين النقطتين المتناظرتين بالصيغة x = μ + zσ أو إذا عبّرنا عن z بدلالة x ، فإن :

$$z = \frac{x - \mu}{\sigma} \qquad (4)$$

هذه الصيغة تمكن المرء من إيجاد النقطة z على المنحني الطبيعي المعياري التي تناظر أية نقطة x على منحن طبيعي غير معياري . فمثلاً ، النقطة x = 4 في الشكل (7-26) تناظر النقطة :

$$x = (4 – 0) / 3 = 11/3$$

في الشكل (7-25) . وبهذه الطريقة ، التي بها نعبّر عن كل قيم x على منحني طبيعي ما بدلالة القيم المناظرة على المنحني الطبيعي المعياري ، يمكننا اختزال المنحنيات الطبيعية جميعها إلى منحن معياري واحد .

وفي المناقشات السابقة لاحظنا أن الانحراف المعياري s لا يتأثّر بإضافة ثابت إلى قيم فئة من القياسات ، لكنه يضرب في c إذا ضربت كل القياسات في c هذه الخاصية نفسها تظل صحيحة للمتغيّر العشوائي x وانحرافه المعياري النظري σ ، بمعنى أن المتغير c من المرات سيكون له الانحراف المعياري نفسه الذي للمتغيّر x ، لكن انحراف cx المعياري سيكون c من المرات قدر انحراف x المعياري .

وفي ضوء هذه الخواص يكون انحراف σ / $(x - \mu)$ المعياري $1/\sigma$مرة من انحراف $x - \mu$ المعياري أو من انحراف x المعياري، وعليه، إذا كانت σ هي الانحراف المعياري للمتغير x ، فإن الانحراف المعياري للمتغير $z = (x - \mu)/\sigma$ لابدّ أن يكون الوحدة . وحيث أن وسطه هو μ ، فبطرح μ من x نحصل على متغير $x - \mu$ وسطه صفر . وعلى هذا يكون وسط المتغير $z = (x - \mu)/\sigma$ هو أيضاً الصفر . إذ إن ضرب متغير ما في ثابت يضرب وسطه في ذلك الثابت ، وضرب الصفر في $1/\sigma$ لا يزال يعطينا الصفر ، وعلى هذا سيكون للمتغير $z = (x - \mu)/\sigma$ الوسط صفر ، والانحراف المعياري الوحدة . وبتغيير المتغير الذي تعطيه الصيغة (4) سيغير أي متغير x إلى متغير آخر وسطه صفر وانحرافه المعياري الوحدة وسيكون هذا صحيحاً سواء أكان المتغير x متغيراً طبيعياً أم لم يكن . ويقال عن متغير ما x قد تغير إلى المتغير z عن طريق الصيغة (4) . إنه مقيس بوحدات معيارية . وذلك بعد إتمام عملية التغيير .

وجدول IX في الملحق هو جدول لإيجاد المساحة الواقعة تحت أي جزء من المنحني الطبيعي للمتغير z أي للمنحني الطبيعي الذي وسطه الصفر وانحرافه المعياري الوحدة. وقيم z في هذا الجدول معطاة لرقمين عشريين ، الرقم العشري الثاني منهما يحدد العمود الواجب استخدامه فمثلاً نفترض أننا نريد حساب المساحة الواقعة تحت المنحني الطبيعي المعياري هذا من $z = 0$ إلى $z = 1000$– وقد بيّنا هذه المساحة المطلوبة ، هندسياً في الشكل (7-27) . نقرأ في جدول IV من أعلى إلى أسفل حتى نصل إلى القيمة 1.0 للمتغير z ، ثم نقرأ بعرض الجدول إلى الكمية المدرجة في العمود المعنون 00 . فنجد القيمة 3413 ، فتكون هي المساحة المطلوبة .

هذا ، ولا يعطينا جدول IV سوى المساحات من $z = 0$ إلى أية قيمة موجبة معينة للمتغير z . أما إذا كنا نريد مساحات واقعة إلى يسار $z= 0$ فلابد لنا من استخدام التماثل والاشتغال على ما يناظر تلك المساحات من النصف الأيمن للمنحني ، فمثلاً التماثل $P\{-1 < z < 1\}$ يعطيها مجموع : $P\{0 < z < 1/2\}$, $P\{0 < z < 2\}$.

والآن ، افرض أننا نريد حساب المساحة الواقعة تحت جزء من منحن طبيعي ما ، وسطه μ وانحرافه المعياري σ . وعلى سبيل المثال نفرض أننا نريد إيجاد المساحة من $x= \mu - \sigma$ إلى $x = \mu+\sigma$ في الشكل (7-22) . من التماثل ، ينتج أن تلك المساحة هي ضعف المساحة

من $x = \mu$ إلى $x = \mu+\sigma$ ومن الصيغة (4) ، نرى أن القيمتين $x = \mu+\sigma$ و $x = \mu$ تتناظران القيمتين $z=1$ و $z= 0$ ، وعليه فإن المساحة من $x = \mu$ إلى $x = \mu+\sigma$ لها القيمة نفسها التي للمساحة الواقعة تحت المنحني الطبيعي المعياري $z= 0$ إلى $z= 1$.

الشكل (7-27) : توزيع طبيعي معياري

وهذه هي نفس قيمة المساحة المناظرة تحت المنحني الطبيعي المعياري المحسوبة في جدول IX والمبينة بالمساحة المظللة في الشكل (7-27) . وحيث أننا قد وجدنا أن هذه المساحة تساوي 3413 . فإن المساحة من $x = \mu+\sigma$ إلى $x = \mu - \sigma$ هي ضعف هذا العدد ، أي 6826 ، وهذا العدد الذي نشأت عنه الخاصية (1) من خواص التوزيع الطبيعي والواردة في (3) .

مثال :

لتوضيح استخدام جدول IX ، ندرس المسألة الآتية : نفرض أن طول الطلبة الجامعيين الذكور هو متغير طبيعي، سنرمز له بالرمز x ، وأن وسطه 69 بوصة وانحرافه المعياري 3 بوصة . ما النسبة المئوية من مثل هؤلاء الطلبة لمن يزيد طولهم عن 72 بوصة ؟ .

هذه النسبة نحصل عليها بحساب الكمية $P\{x > 72\}$ خذ القيمة $x = 72$. إذن ، حيث $\sigma = 3$ و $\mu = 69$ فإن قيمة z المناظرة تعطيها :

$$z = \frac{72 - 69}{3} = 1$$

ونجد من جدول IX ، إن المساحة الواقعة تحت المنحني الطبيعي المعياري بين $z= 0$ و $z= 1$ هي 3413 ، وحيث أن المساحة الواقعة تحت النصف الأيمن من هذا المنحني هي 5 ، فينتج أن المساحة الواقعة إلى يمين $z= 1$ تساوي : $5000 - 3413 = 1587$ وعليه فإن ما يقرب من 16% من الطلبة ستكون قيمة z لهم تزيد عن الوحدة ، وبالتالي يكون طولهم أكبر من 6 قدم 72 بوصة .

مثال :

كمثال توضيحي آخر لاستعمال جدول IX ، نفترض أننا نريد إيجاد المساحة بين $x= 280$ ، $x= 220$ للمتغير x الذي له توزيع طبيعي له $\sigma= 20$ و $\mu= 230$ ، وقد بيّنا المساحة المطلوبة في الشكل (7-28) .

الشكل (7-28) : توزيع طبيعي خاص

أولاً، من الضروري أن نحسب قيمتي z المناظرتين ، وذلك عن طريق (4) . هاتان القيمتان هما :

$$z_1 = \frac{220-230}{20} = -50$$
$$z_2 = \frac{280-230}{20} = 2.50$$

قيمتـا z هاتـان قد بيّنتـا فـي شـكل (7-27) بواسـطة سـهمين رأسـيين . والآن المسـاحة المطلوبة هي المعطاة بالمساحة الواقعة تحت المنحنـي الطبيعي المعياري من (-.50) إلى 2.50 ونجد من جدول IX أن السماحة من z= 0 إلى z= 2.50 هي 4938 . نتيجة للتماثل ، تكون المساحة من z= -.50 إلى z= 0 هي المساحة نفسها من z= 0 إلى z= .50 ، وهذه المساحة الأخيرة نجدها من جدول IX ، وهي 1915 ، وعلى هذا ، تكون المساحة المطلوبة هي مجموع هاتين المسـاحتين أي 6853 . هذا ، ولـو أن المنحنـي الطبيعي فـي شكلي (7-27) و (7-28) كان سيبدو مختلفاً تماماً لو كنا استخدمنا مقياس الرسم نفسه على كل من المحورين ، إلاّ أننا قد رسمناه بحيث يبدو متشابهاً حتى يتّضح تكافؤ المساحات .

7-20- تقريب طبيعي لذي الحدين Normal Aproximation to Binomial:

من السهل حلّ المسائل المتعلقة بتوزيع ذي الحدين ، شريطة ألاّ يكون عدد المحاولات n كبيراً ، أما إذا كان العدد n كبيراً ، فسيصبح الحسابات التي يتضمنها استخدام الصيغة (1) طويلـة للغايـة ، وبالتـالي ، سيكون من المفيد جداً الحصـول علـى تقريب بسيط طيب للتوزيـع . ومثل هذا التقريب موجود فعلاً في صورة التوزيع الطبيعي المناسب . وبهدف فحص هذا التقريب ، لندرس بعض الأمثلة العددية .

نفترض أن n = 12 و P= 1/3 كون الشكل البياني لتوزيع ذي الحدين المناظر ، إن قيم P{x} باستخدام الصيغة (1) ، محسوبة لثلاثة أرقام عشرية ، هي :

(5)

P{8}=.015	P{4}=.238	P{0}=.008
P{9}=.003	P{5}=.191	P{1}=.046
P{10}=.000	P{6}=.111	P{2}=.127

P{11}=.000	**P{7}=.048**	**P{3}=.212**
P{12}=.000		

هذا وبالرغم من أن الشكل البياني المستخدم سابقاً لتوزيع ذي الحدين كان شكلاً خطياً بسبب طبيعة المتغير x المتقطع ، فإن هذا التوزيع سيرسم كمدرّج تكراري كي تسهل مقارنته بالمدرّجات التكرارية الخاصة بالتوزيع الطبيعي ، ولقد بيّنا المدرج التكراري الخاص بهذا التوزيع في الشكل (7-29) .

الشكل (7-29) : توزيع ذي الحدين ، فيه P= 1/3 و n= 12

إن ارتفاع أي مستطيل يساوي الاحتمال المعطى في (5) لعلامة الفئة المناظرة ، وحيث أن طول قاعدة أي مستطيل هو الوحدة ، فإن مساحة أي مستطيل تساوي . عددياً . ارتفاعه ، وعلى ذلك فإن هذه الاحتمالات تعطيها كذلك مساحات المستطيلات المناظرة ، ومن شكل هذا المدرج التكراري يبدو أن من الممكن توفيقه بالتقريب مع المنحني الطبيعي المناسب .

وحيث أن أي توزيع طبيعي يتحدد كلية بوسطه وانحرافه المعياري ، فمن الطبيعي هنا استخدام المنحني الطبيعي الذي له الوسط نفسه و الانحراف المعياري نفسه الخاصيان بتوزيع ذي الحدين ومن الصيغتين (2) ، ينتج أن $\mu = 12.\frac{1}{3} = 4 \ and \ \sigma = \sqrt{12.\frac{1}{3}.\frac{2}{3}} = 1.63$ ولقد وضعنا منحنياً طبيعياً بهذا الوسط ، وبهذا الانحراف المعياري فوق ، الشكل (7-29) ليعطي

شكل (7-30) ، ومنه يبدو أن التوفيق مقبول ، بالرغم من كون n= 12 قيمة صغيرة للعدد n ، ولا تعطينا النظريات المتقدّمة بتوفيق طيب سوى للقيم الكبيرة للعدد n .

وكاختبار لدقّة تقري المنحني الطبيعي ، وكتوضيح لكيفية استخدام طرق المنحني الطبيعي لتقريب احتمالات ذي الحدين ، ندرس بضعة مسائل مرتبطة بشكل (7-30) .

الشكل (7-30) : توزيع ذي الحدين ، فيه n= 12 و P= 1/3 مع منحني طبيعي موفق

مثال :

إذا كان احتمال أن يصيب رجلاً هدفاً هو 1/3 ، وأطلق الرجل 12 طلقة ، فما احتمال أن يصيب الهدف على الأقل 6 مرات ؟

إن الجواب المضبوط ، صحيح لثلاثة أرقام عشرية ، يمكن الحصول عليه بجمع القيم الواردة في (5) من x= 6 إلى x=12 ، وهذا المجموع يعطينا الجواب 177.

وهذا الجواب يمثّل هندسياً مساحة ذلك الجزء من المدرّج التكراري المرسوم في الشكل (7-30) الواقع إلى يمين x= 5.5 ، وعلى ذلك فلكي نجد تقريباً لذلك الاحتمال بطرق المنحني الطبيعي ،

ما علينا إلاّ أن نوجد المساحة تحت ذلك الجزء من المنحني الطبيعي الذي وفقناه والواقع إلى يمين x= 5.5 ، وحيث أن المنحني الطبيعي الموفق له σ= 1.63 و μ= 4 ، فينتج أن :

$$z = \frac{x-\mu}{\sigma} = \frac{5.5-4}{1.63} = 0.92$$

والآن ، من جدول IX ، نجد أن المساحة بين z= 0.92 و z= 0 هي 321 . إذن، المساحة الواقعة إلى يمين z= 0.92 هي 179 ، ويكون هذا هو التقريب المطلوب لاحتمال الحصول على الأقل على ست إصابات للهدف . وحيث أن الجواب المضبوط قد سبق أن حسبناه ووجدناه 177 ، فإننا نجد أن تقريب المنحني الطبيعي هنا تقريب جيد بالتأكيد .

مثال :

لاختبار دقة طرق المنحني الطبيعي لفترات قصيرة لنحسب احتمال إصابة الرجل للهدف 6 مرات بالضبط في 12 طلقة .

إن الجواب المحسوب عن طرق (5) صحيح إلى ثلاثة أرقام عشرية هو 111 ، وحيث أن هذا الجواب يساوي مساحة المستطيل الذي تجري قاعدته من 5.5 إلى 6.5 ، فلكي نقرّب هذا الجواب علينا حساب المساحة تحت المنحني الطبيعي الموفق بين x= 6.5 و x= 5.5 وعليه فبحساب قيمتي z وباستخدام جدول IX نجد أن :

$$z_2 = \frac{6.5-4}{1.63} = 1.53 \qquad , \quad A_2 = 4370$$
$$z_1 = \frac{5.5-4}{1.63} = 0.92 \qquad , \quad A_1 = 3212$$

وبطرح هاتين المساحتين ، نحصل على 116 ، وهو تقريب جيد أيضاً بالمقارنة مع القيمة المضبوطة 111 ، للاحتمال من هذين المثالين يتّضح أن طرق المنحني الطبيعي تعطي تقريبات جيدة حتى في بعض الحالات ، مثل تلك التي درسناها هنا، التي لا تكون فيها n كبيرة جداً .

والآن نفترض أننا لم نغيّر قيمة *P*= 1/3 ، لكن سمحنا للعدد n بالازدياد، إن المدرّج التكراري الناتج ، مثل ذلك المرسوم في الشكل (7-30) ، سيتحرّك إلى اليمين منتشراً ، وسيقلّ

ارتفاعه . ومن الصعب فحص مثل هذا المدرّج التكراري وملاحظة ما إذا كان من الممكن توفيق منحن طبيعي به . بيد أنه يمكننا تجنّب هذه التغيرات غير المرغوب فيها في المدرّج التكراري وذلك بأن ننتقل إلى المتغيّر المناظر بالوحدات المعيارية ، هذا يعني رسم المدرّج التكراري للمتغير :

$$z = \frac{x-\mu}{\sigma}$$

وذلك من (4) ، وعند استخدام (2) ، يأخذ المتغير المعياري z الصورة :

$$z = \frac{x-nP}{\sqrt{nPq}} \qquad (6)$$

وحيث أن المتغيّر z له توزيع وسطه الصفر وانحرافه المعياري الوحدة ، فإن مسلك مدرّج تكراري z سيكون مسلكاً مرضياً دون ما تجول ، ودون ما تنتشر ، بزيادة n كما كانت الحال بالنسبة لمدرّج تكراري .

ويبيّن الشكلان (7-31) و (7-32) مدرّجاً تكرارياً للمتغيّر z عندما تكون $P= 1/3$ في الحالتين $n= 48$ و $n= 24$ على الترتيب ، ومنهما يتّضح كيف يقترب توزيع z بسرعة من توزيع متغيّر طبيعي ، وسطه الصفر وانحرافه المعياري الوحدة . ويمكننا بطرق متقدمة إثبات أنه إذا احتفظنا بقيمة P ثابتة وسمحنا للعدد n بالازدياد ، فإن توزيع z سيقترب أكثر وأكثر من توزيع متغيّر طبيعي ، وسطه الصفر وانحرافه المعياري الوحدة . هذا ، ومن جهة نظر عملية، قد دلّتنا التجربة ، أن التقريب يكون طيباً متى كان $nP > 5$ في حالة $P \leq 1/2$ ومتى كان $nq > 5$ في حالة $P > 1/2$.

وحقيقة أن يكون للشكل المعياري لمتغيّر ذي الحدين توزيع يقترب من توزيع المتغيّر الطبيعي المعياري، إنما تعني أن المتغير ذي الحدين x مدرّج تكراري من الممكن توفيقه توفيقاً جيداً مع المنحني الطبيعي المناسب عندما تكون n كبيرة ، والمنحني الطبيعي المناسب هو بالطبع ذلك المنحني الذي تعطي الصيغتان (2) وسطه وانحرافه المعياري .

الشكل (7-31) : توزيع ذي الحدين للمتغير

***P* = 1/3 , *n*= 24 عندما (x – *nP*) /** $\sqrt{nPq}$

الشكل (7-32) : توزيع ي الحدين للمتغير

***P* = 1/3 , *n*= 48 عندما (x – *nP*) /** $\sqrt{nPq}$

وهناك حالات عديدة يكون فيها من المناسب أكثر الاشتغال بنسبة مرات النجاح x/n في n من المحاولات ، بدلاً من العمل بالعدد الفعلي x لمرات النجاح، فقسمة البسط والمقام في (6) على n يأخذ z الصورة :

$$z = \frac{\hat{P} - P}{\sqrt{\frac{Pq}{n}}} \qquad (7)$$

حيث رمزنا بالرمز $\hat{P}$ إلى نسبة مرات النجاح x/n . لاحظ أن قيمة z لم تتغير، لكن الذي تغير هو الصورة التي كتبت بها z . وبالتالي فلا يزال للمتغير z ، عندما تكون n كبيرة ، توزيع طبيعي تقريبي ، وسطه الصفر وانحرافه المعياري الوحدة . هذا يعني أنه عندما تكون n كبيرة ، فيكون لنسبة مرات النجاح $\hat{P}$ مدرّج تكراري يمكن توفيقه توفيقاً جيداً بالمنحني الطبيعي المناسب ، أما المنحني الطبيعي المناسب هنا فهو ذلك المنحني الذي يعطي وسطه وانحرافه المعياري بالصيغتين :

$$\mu_{\hat{P}} = P$$

$$\sigma_{\hat{P}} = \sqrt{\frac{Pq}{n}} \qquad (8)$$

مثال :

لتوضيح استخدام تقريب المنحني الطبيعي لتوزيع ذي الحدين في الصيغة (7).

لندرس المسألة التالية ، نفترض أن سياسياً ادّعى أن مسحاً في دائرته قد بيّن أن 60% من مرشحيه يوافقون على تأييده في تشريع مهم . فإذا فرضنا وقتياً أن هذه النسبة المئوية صحيحة ، وإذا أخذت عينة من دائرته تتكون من 400 ناخب ، فما احتمال أن تنتج هذه العينة تأييداً أقل من 50% .

إذا فرضنا أن أخذ عينة من 400 ناخب تشبه ممارسة لعبة من ألعاب الحظ 400 مرة ، يكون احتمال النجاح في كل مرة واحدة منها 6 ، فإن هذه المسألة ممكن أن تعالج كمسألة من مسائل توزيع ذي الحدين فيه n= 400 و P_6 .وباستخدام الصيغة (7) ، تكون :

$$z = \frac{5-6}{\sqrt{\frac{(6).(4)}{400}}} = -4.08$$

والآن ستكون نسبة العينة x=n أقل من 5 ، بشرط أن تكون x أقل من 4.08– وبالتماثل ، يكون احتمال أن تكون 4.08- >z مساوياً لاحتمال أن تكون 4.08 >z، هذا الاحتمال يعد صغيراً لدرجة أنه لا يستحق إدراجه في جدول IX . وعلى ذلك إذا حدث أن أيد السياسي أقل من 50% من العينة ، لكان لنا بالتأكيد أن نرفض ادّعاءه بحصوله على 60% من التأييد .

بيد أن هناك اعتراضات . قد تقوم وبحق . تعارض في اعتبار أخذ عيّنة من 400 ناخب مكافئاً لممارسة لعبة من ألعاب الحظ 400 مرة . كما أن هناك أسئلة تتعلّق باستقلال المحاولات وثبوت الاحتمال يجب أن تجيب عنها قبل أن يكون المرء سعيداً سعادةً كاملة بنموذج توزيع ذي الحدين لهذه المسألة .

مثال :

وكمثال ثانٍ لاستخدامه الصيغة (7) ، نحل المسألة التالية ، دلّت التجربة السابقة بامتحان اللغة الإنكليزية للمستجدين في الجامعة أن 50% فقط من الطلبة ينجحون فيه ، فإذا حضر الامتحان فصل جديد مكوّن من 200 مستجدّ ، فما احتمال أن ينجح على الأقل 55% منهم .

$$z = \frac{55-50}{\sqrt{\frac{(5).(5)}{200}}} = 1.41$$

لكن $x \geq 55$ إذا كانت $z \geq 1.41$ وبالعكس، إذن من جدول IX يكون احتمال $x \geq 55$ هو 08.

مثال :

وكمثال أخير ، اعتبر المسألة التالية . افرض أن %5 من الأفراد الذي يطعمون ضدّ الأنفلونزا يعانون ردود فعل لهذا التطعيم غير مستحبة وخطيرة ، باستخدام التقريب الطبيعي ، احسب احتمال أن يعاني مثل ردود فعل هذه أكثر من %8 من أفراد مطعمين عددهم 200 .

إذا طبقنا الصيغة (7) ، فإننا أولاً نحسب :

$$z=\frac{08-05}{\sqrt{\frac{(05).(95)}{200}}}=\frac{03}{\sqrt{00\ 02\ 37}}=\frac{03}{0154}=1.95$$

ثم نجد من جدول IX أن $P\{z>1.95\}=0256$ وهذا هو الاحتمال المطلوب لأن يعاني أكثر من %8 ردود فعل .

هذا ، ولا ينبغي أن نفترض من الأمثلة السابقة أنه من الممكن معالجة مسائل توزيع ذي الحدين جميعها معالجةً مرضية عن طريق التقريب الطبيعي ، حتى لو كانت n كبيرة نوعاً ، فمثلاً إذا كانت $P=1/20$ و $n=80$ ، فإن حساب قيم $P\{x\}$ ورسم هذه القيم بيانياً سيبيّنان أن التوزيع غير متماثل بالطريقة التي تكفي للسماح بتوفيق جيد بواسطة المنحني الطبيعي ، ذلك لأن الوسط هنا 4 ، والمتغيّر x لا يمكن أن يأخذ قيماً سالبة ، وبالتالي يتركّز الكثير من التوزيع قرب الصفر للتماثل ، ولقد نشأت من تلك الاعتبارات القاعدة التجريبية للتقريب الجيد ، وقد ذكرناها في الفقرة التي تلي الصيغة (6) .

21-7– استنتاج توزيع ذي الحدين Binomial Distribution Derivation :

سنبيّن فيما يلي صحّة الصيغة (1)، اعتبر تجربة ما من n من المحاولات ، ومتتابعة معينة من النواتج التي تنتج بالضبط عدداً x من مرات النجاح ، وعدداً x-n من مرات الفشل . إن واحدة من مثل تلك المتتابعات هي المتتابعة الآتية ، وفيها وقعت مرات النجاح جميعها أولاً ، ثم تلتها مرات الفشل .

$$\overbrace{SS.......S}^{x}\ \overbrace{FF........F}^{n-x}$$

وفيما يلي متتابعة أخرى فيها وقعت مرة فشل أولاً تلاها x من مرات النجاح المتتالية ثم تلي ذلك مرات الفشل الباقية :

$$F\overbrace{SS.......S}^{x}\ \overbrace{FF.........F}^{n-x-1}$$

وبسبب استقلال المحاولات يكون احتمال الحصول على أولى هاتين المتتابعتين هو:

$$\overbrace{P.P.......P}^{x}\,.\,\overbrace{q.q..........q}^{n-x} = P^{x}.q^{n-x}$$

واحتمال الحصول على المتتابعة الثانية هو :

$$q.\overbrace{P.P........P}^{x}\,.\,\overbrace{q.q.........q}^{n-x-1} = P^{x}.q^{n-x}$$

وعلى هذا ، يكون احتمال الحصول على المتتابعة الأولى هو نفسه احتمال الحصول على كل متتابعة تحقق شرط حدوث x من مرات النجاح ، n-x من مرات الفشل .

إن عدد الطرق التي يمكن بها حدوث الحدث المرغوب فيه يساوي عدد المتتابعات المختلفة التي يمكن كتابتها من النوع الذي قد شرحناه ، أي المحتوية على x من حروف S و n-x من حروف F . بيد أن هذا العدد يساوي عدد طرق اختيار x من المواقع من بين n من المواقع التي لدينا لكي نضع فيها الحرف S . أما المواقع الباقية وعددها n-x فستخصص تلقائياً للحرف F . ومن الممكن ترقيم المواقع التي عددها n ومعالجتها كبطاقات مرقمة عددها n ؛ عندئذٍ تكون المسألة مسألة اختيار x من تلك البطاقات . وحيث أن الأرقام المكتوبة على البطاقات هي التي تهمّنا لا ترتيب سحبها ، تصبح المسألة مسألة توافيق . ومن الاستنتاج الوارد في الفقرة (أ) ، ينتج أن عدد طرق اختيار x من البطاقات من بين n من البطاقات المتمايزة تعطيه صيغة التوافيق (13) من الفقرة (أ)، أي :

$$\binom{n}{x} = \frac{n!}{x!.(n-x)!}$$

إذن : هذا هو عدد المتتابعات التي تنتج بالضبط x من مرات النجاح .

وحيث أن كل متتابعة من هذه المتتابعات تمثّل طريقةً من الطرق المتنافية مثنى . التي يمكن أن يقع بها الحدث المرغوب فيه ، وحيث أن كل متتابعة من هذه المتتابعات لها احتمال الحدوث نفسه، وهو $P^{x}.q^{n-x}$ فينتج أن الاحتمال المطلوب ينتج من جمع هذا الاحتمال عدداً من المرات بعدد المتتابعات التي لدينا ، لكن عدد مثل تلك المتتابعات قد وجدناه $\binom{n}{x}$ ، إذن نحصل على $P\{x\}$ بضرب $P^{x}.q^{n-x}$ في $\binom{n}{x}$ ، وهذا يحقق الصيغة (1) .

- **أمثلة توضيحية إضافية :**

مثال (1) :

نفترض أن 60% من الطلبة في جامعة كبيرة يعملون لبعض الوقت ، نفترض أن x ترمز لعدد مثل هؤلاء الطلبة في عينة من 8 طلبة اختيروا كيفما اتّفق من ملفات التسجيل .

فبفرض أن حجم الجامعة كبير كبر يسمح بفرض أن تكون $P= 6$ ثابتة عندما تؤخذ عينة من ثمانية طلبة فقط .

(أ)– أوجد تعبيراً للاحتمال $P\{x\}$ يعطي توزيع x (بند 1) .

(ب)– استخدم الناتج في (1) لحساب $P\{x \geq 6\}$ (بند 2) .

(ج)– استعمل الصيغتين (2) لحساب وسط x وانحرافها المعياري (بند 3).

(د)– استخدم التقريب الطبيعي لتقريب القيمة التي حصلت عليها في (ب) (بند 4) وهنك الحلول :

(أ): يمكننا معاملة العينة المكونة من 8 طلبة كمحاولات مستقلة لتجربة ما فيها $P= 6$ ، وعليه تكون x متغيراً من متغيرات ذي الحدين يعطي توزيعه كما يلي :

$$P\{x\} = \frac{8!}{x!(8-x)!}.(.6)^x(.4)^{8-x}$$

$$(b) \quad P\{x \geq 6\} = P\{6\} + P\{7\} + P\{8\}$$

(ب):

$$= \frac{8!}{6!2!}.(.6)^6(.4)^2 + \frac{8!}{7!1!}(.6)^7(.4)^1 + \frac{8!}{8!0!}(.6)^8.(.4)^0$$
$$= (.6)^6[28(.4)^2 + 8(.6).(.4) + (.6)^2]$$
$$= 315$$

(ج) (c) $\mu = 8(6) = 4.8 \ , \ \sigma = \sqrt{8(.6).(.4)} = 1.39$

(د) (d) $x = \dfrac{5.5-4.8}{1.39} = .50 \ , \ P\{x > .50\} = .31$

مثال (2) :

يمكن أن يسع أحد المطاعم 50 عميلاً . وقد دلّت التجربة على أن 10% ممن يحجزون موائدهم لا يأتون ، نفترض أن المطعم قبل 55 حجزاً ، نفترض أن x ترمز إلى عدد العملاء الذين سيأتون .

(أ)- أوجد تعبيراً للتوزيع $P\{x\}$ الخاص بالمتغيّر x (بند 1) ز

(ب)- احسب وسط x وانحرافها المعياري عن طريق الصيغتين (2) (بند 2).

(ج)- استخدام التقريب الطبيعي لحساب احتمال استطاعة المطعم أن يسع جميع العملاء الذين يأتون . (بند 4) .

(أ): يمكن معاملة 55 حجزاً كخمسة وخمسين محاولة مستقلة لتجربة ما فيها $P= 9/10$ ، إذن x متغير من متغيرات ذي الحدين يعطي توزيعه بما يأتي :

$$P\{x\} = \frac{55!}{x!(55-x)!}(.9)^x.(.1)^{55-x}$$

(ب):

$$\mu = 55(.9) = 49.5 \ , \ \sigma = \sqrt{55(.9).(.1)} = 2.2$$

(ج): يمكن أن يجد كل العملاء أماكن لهم إذا حضر على الأكثر 55 منهم ، إذن فالمسألة هي أن تحسب $P\{x \le 50\}$ ، وهذا الاحتمال يمكن أن نحصل عليه بإيجاد المساحة الواقعة تحت مدرّج تكراري $P\{x\}$ إلى يسار 50.5 ، إن القيمة المناظرة للمتغيّر هي :

$$z = \frac{50.5 - 49.5}{2.2} = .45$$

إذن : $P\{x \le 50\} = P\{z \le .45\} = 6.7$

مثال (3) :

نفترض أن x تمثّل الوزن بالرطل لسمكة من نوع السالمون الكبير التي تصطاد عند مصب نهر معين ، ونفترض أن للمتغير x توزيعاً طبيعياً وسطه 30 وانحرافه المعياري 6 .

احسب احتمال أنه . إذا اصطاد سمكة سالمون . فإن وزنها :

(أ)– سيكون على الأقل 41 رطلاً (بند 3) .

(ب)– سيكون بين 40 , 20 رطلاً بما في ذلك 40 , 20 رطلاً (بند 3).

(أ): $$z = \frac{41-30}{6} = 1.83 \ , \ P\{x \ge 41\} = P\{z \ge 1.83\} = .03$$

(ب): $$z_1 = \frac{20-30}{6} = -1.67 \ , \ z_2 = \frac{40-30}{6} = 1.67$$

$$P\{20 \le x \le 40\} = P\{-1.67 \le z \le 1.67\} = 2P\{0 \le z \le 1.67\} = .90$$

مثال (4) :

ضبطت آلة مصممة لثقب الصفائح الحديدية كي تثقب ثقوباً قطرها 1.02 بوصة حتى نضمن بذلك ألاّ يقل الثقب عن 1 بوصة .

(أ)– فإذا كان الانحراف المعياري لقطر الثقوب التي تقبتها هذه الآلة هو 01. بوصة فما النسبة المئوية للثقوب التي تقل عن 1 بوصة ؟ (بند 3) .

(ب)– إذا كانت الصفائح التي يزيد قطر ثقوبها عن 1.05 بوصة غير صالحة للاستعمال ، فما النسبة المئوية للصفائح التي تحتوي ثقبين والتي سترفض أو تستبعد لأن أحد الثقبين على الأقل أصغر مما ينبغي أو أكبر مما ينبغي ؟ (بند 3) .

(أ): نفترض أن الأقطار لها توزيع طبيعي ، تصبح المسألة مسألة حساب $P\{x < 1.00\}$ حيث x هو هذا المتغيّر الطبيعي ، هنا :

$$z = \frac{1.00 - 1.02}{.01} = -2.00$$

إذن : $P\{x < 1.00\} = P\{x < -2.00\} = .023$

(ب): (b) $$z = \frac{1.05 - 1.02}{.01} = 3.00$$

$P\{x > 1.05\} = P\{z > 3.00\} = .001$

وعليه ، فإن احتمال أن يكون ثقب ما غير مرضي هو 024. ، أما احتمال أن يكون الثقبان كلاهما مرضياً فهو 95. = (976.) . (976.) ، إذن ستستبعد 5% من الصفائح .

- **تمرينات :**

1– باستخدام الصيغة (1) حقق القيم التي أعطيناها في جدول 3 من هذا الكتاب .

2– استخدم طريقة حصر النواتج الممكنة مع احتمالاتها المناظرة لاشتقاق توزيع لذي حدين الخاص بعدد الرؤوس التي تحصل عليها عند رمي قطعة نقود 4 مرات. راجع نتائجك باستخدام الصيغة (1).

3– حلّ المسألة 2 بالنسبة لعدد الآسات التي تحصل عليها عند دحرجة زهرة طاولة 4 مرات .

4- رميت قطعة نقود 5 مرات ، استخدم الصيغة (1) لتحصل على قيم P{x} ، حيث x ترمز إلى عدد الرؤوس ، ثم ارسم P{x} بيانياً كخطوط .

5- إذا كان احتمال أن تفوز بيد للبريدج ¼ ، ولعبت 5 أيادٍ ، احسب قيم P{x} حيث x ترمز لعدد مرات الفوز ، وذلك باستخدام الصيغة (1) .

6- إذا كان 20% من المنصهرات معيبة ، واشترينا صندوقاً به 8 منصهرات ، فما احتمال أن تكون على الأقل 6 من تلك المنصهرات صالحة ؟

7- نفترض أن نوعين من الأسبرين A , B لهما الكفاية نفسها في تخفيف الآلام ، إذا جرّب 10 أشخاص نوعي الأسبرين لتخفيف آلامهم ، فما هو احتمال أن يختار 8 منهم على الأقل النوع A .

8- وضع 12 زوجاً من حيوانات التجارب على نظامين مختلفين من أنظمة التغذية وكان تخصيص نظام التغذية لكل زوج عشوائياً . وبعد إتمام إجراء التجربة ، وجد فرق بين زيادة وزن الحيوان الذي اتبع النظام A وزيادة وزن الحيوان الذي اتّبع النظام B ، فإذا كان هذا الفرق موجباً ، سمينا النتيجة نجاحاً فما احتمال أن يقع على الأقل 9 مرات من مرات النجاح إذا لم يكن هناك فرق حقيقي في زيادة الوزن بالنسبة لنظامي التغذية ؟

9- في مسألة 6 ، ما عدد المنصهرات التي تحتاج إلى شرائها حتى يكون احتمال حصولك على الأقل على 6 منصهرات سليمة على الأقل 9 ؟

10- هل تتوقّع أن يكون توزيع ذي الحدين قابلاً للتطبيق عند حساب احتمال أن تزداد سوق الأوراق المالية على الأقل 20 يوماً خلال الشهر القادم إذا كانت لديك سجلاّت عن الأعوام الخمسة الأخيرة للنسبة المئوية التي زادت فيها ؟ اشرح .

11- اشرح لماذا لن يكون صحيحاً تماماً تطبيق توزيع ذي الحدين عند حساب احتمال أن تمطر على الأقل 10 أيام في تشرين الأول إذا عاملنا كل يوم من أيام تشرين الأول كمحاولة حدث وكان لدى المرء سجل للنسبة المئوية للأيام المطيرة في شهر تشرين الأول .

12– لمتغير ذي الحدين الذي لـه $P= ¼$ و $n= 8$ احسب الوسط والانحراف المعياري ، وحقق نتائجك باستخدام الصيغتين (2) .

13– احسب المسألة 5 وسط المتغير x وانحرافه المعياري وحقق نتائجك باستخدام الصيغتين (2) .

14– استنتج الصيغة $\mu= nP$ لتوزيع ذي الحدين العام $P\{x\}$ الذي تعطيه (1) بأن تكتب حدود $\sum_x^n = 0\, xP\{x\}$ ثم تخرج العامل المشترك nP ، وبعد ذلك نحسب $\sum_{x=0}^{n-1} Q(x)$ حيث $Q(x)$ هي نفس $P\{x\}$ لعدد من المحاولات تساوي الـ n-1 ، وأخيراً تتحقق من أن هذا المجموع الأخير ، الذي قيمته الوحدة ، هو العامل الآخر في $\Sigma xP\{x\}$. إن الصيغة تبدو واضحة لكن استخلاصها جبرياً ليس واضحاً .

15– ارجع إلى أحد المراجع المتقدمة في الإحصاء لمشاهدة الطريقة الجبرية المستخدمة في اشتقاق الصيغة $\sigma = \sqrt{nPq}$.

16– نفترض أن ما حصلت عليه في امتحان ما بالوحدات المعيارية (z) هو 8 . وأننا نفترض توزيع التقديرات توزيعاً طبيعياً، ما النسبة المئوية للطلبة الذين تتوقع أن يحصلوا على درجات أعلى من درجاتك؟

17– أعلن مدرّس ألعاب الرياضة في مدرسة ثانوية أن تقديراته للأحداث الرياضية الفردية تتبع الإنجازات النسبية لجميع فصوله ، فإذا كان يعطي تقدير A لعشرين في المائة وإذا دلّتنا الخبرة على أنه . بالنسبة للقفز العالي . يكون الوسط 10 بوصة والانحراف المعياري 4 بوصة ، ما العلو الذي على الطالب أن يقصد القفز إليه إذا كان يتوقّع الحصول على تقدير A ؟

18– إذا فرضنا أن تقديرات اختبار ذكاء لطلبة إحدى الكليات لها توزيع طبيعي ، وسطه 115 وانحرافه المعياري 8، احسب النسبة المئوية للطلبة الذين سيحصلون في اختبار ذكاء (أ) على أكثر من 130، (ب) على أقل من 100 ، (ج) ما بين 125 و 105 بما في ذلك هاتين الدرجتين.

19– دلّت التجربة بالنسبة لامتحان ما للغة الإنجليزية الأساسية أن توزيع الدرجات يقارب التوزيع الطبيعي الذي له الوسط 130 والانحراف المعياري 20 فإذا كانت درجة النجاح هي 100 فما النسبة المئوية من الطلبة الذين تتوقّع رسوبهم في الامتحان ؟

20– متوسّط عمر موتور كهربي 6 سنوات بانحراف معياري قدره سنتان ، فإذا كان لنا أن نعامل مدى حياة مثل هذا الموتور كمتغيّر طبيعي ، وإذا كان هذا الموتور مضموناً ، فما المدى الواجب أن يطبّق فيه الضمان حتى لا يتعطّل سوى 15% من الموتورات على الأكثر قبل نهاية مدة الضمان ؟

21– رميت قطعة نقود 8 مرات ، أوجد بالضبط ، مستخدماً الصيغة (1) وبالتقريب عن طريق تقريب المنحني الطبيعي ، احتمال الحصول على (أ) 6 مرات على وجه العملة ، (ب) 6 مرات على الأقل على وجه العملة .

22– إذا كان احتمال فوزك في لعبة ما 6 ، أوجد بالضبط ، باستخدام الصيغة (1) وبالتقريب ، عن طريق تقريب المنحني الطبيعي ، احتمال فوزك في 4 لعبات أو أكثر من 7 لعبات تلعبها .

23– إذا كان 30% من الطلبة أبصارهم غير سليمة ، فما احتمال أن يكون أبصار نصف الطلبة على الأقل في فصل من 18 طالباً غير سليم ، استخدم تقريب المنحني الطبيعي .

24– إذا كان 10% من صمامات الصورة في أجهزة التلفزيون تحترق قبل نهاية فترة ضمانها (أ) فما احتمال أن يضطرّ تاجر قد باع 100 من مثل هذه الصمامات أن يعطي على الأقل 14 صماماً منها بدلاً مما احترق ؟ (ب) ما احتمال أن يعطي على الأقل 5 صمامات وعلى الأكثر 14 صماماً ؟ استخدم تقريب المنحني الطبيعي .

25– إذا كان 20% من السائقين في مدينة معينة على الأقل حادث واحد خلال قيادتهم في عام ، ما احتمال أن تتجاوز النسبة المئوية 200 عميل من عملاء إحدى شركات التأمين في تلك المدينة 25% خلال السنة التالية ؟ استخدم تقريب المنحني الطبيعي .

26– إذا كان 20% من الدبابيس المصنّعة معيبة ، فما الاحتمال التقريبي لتجاوز نسبة الدبابيس المعيبة في صندوق به 200 دبوس 15% ؟

27– في أحد امتحانات الاختبارات المتعددة 20 سؤالاً ، مدرّجاً لكل منها أربع إجابات ممكنة ، واحدة منها فقط هي الإجابة الصحيحة ، فإذا استخدم أحد الطلبة حدسه في كل سؤال ، ما احتمال حصوله على الأقل على 8 إجابات صحيحة ؟

28– نفترض $P= 1/10$, $n= 20$ ، احسب $P\{0\}$ ، وعلى أساس قيمتها دافع عن عدم توقّع توفيق مقبول لمنحني طبيعي للتوزيع بأكمله هنا .

29– احسب لمسألة 31 قيمة $P\{x= 0\}$ بوساطة تقريب المنحني الطبيعي وقارن نتيجة مسألة 31 .

30– أعطِ مثالاً لتوزيع من توزيعات ذي الحدين يكون فيه $n> 100$ وقيمته P غير صغيرة ، ومع ذلك فلن يكون له تقريب جيد .

31– يحتوي صندوق على أوراق اللعب الآتية وعددها تسع :

(ثلاث) ، (أربع) ، (خمس) إسباتي .

(ثلاث) ، (أربع) بستوني .

(ثلاث) ، (أربع) قلوب .

(أربع) ، (خمس) ديناري .

فإذا كان لنا أن نسحب ورقة لعب من الصندوق ونكرر التجربة 10 مرات، مع إرجاع الورقة المسحوبة في كل مرة ، وإذا كانت X ترمز لعدد الورقات السوداء التي نحصل عليها .

(أ)– أوجد تعبيراً للاحتمال $p\{X\}$ الذي يعطي توزيع X .

(ب)– استخدم نتيجة (1) لحساب .

(ج) استخدم التقريب الطبيعي لتقريب القيمة التي حصلت عليها في (ب) .

(د) احسب الاحتمال التقريبي ، مستخدماً التقريب الطبيعي كي تكون نسبة الورقات السوداء في مائة محاولة للتجربة ، أقل من 6.

32- لفات لطعام الإفطار كتب عليها أنها تحوي 12 أوقية من الحبوب ، لكن آلة التعبئة عرضة للأخطاء ، بانحراف معياري قدره 1 أوقية .

(أ)- فإذا ضبطت الآلة لتعبئ اللفّة بما يساوي 12.1 أوقية من الحبوب ، ما النسبة المئوية للفات التي يقل وزنها عن المقرر بمقدار 1 أوقية على الأقل ؟

(ب)- إذا رغب المنتج ألاّ يكون هناك أكثر من 5% من لفاته وزنها أقل بمقدار 1 أوقية على الأقل ، فما هو وسط وزن التعبئة الذي ينبغي له استخدامه ؟

33- يمكن إحدى الطائرات أن تتّسع 300 مسافر ، 30 بالدرجة الأولى ، 270 بالدرجة السياحية ، فإذا قبلت شركة الطيران حجزاً لثلاثين من ركاب الدرجة الأولى و 290 من ركاب الدرجة السياحية ، وكان احتمال ألاّ يحضر شخص ممن يقومون بالحجز هو 1 ، فما احتمال أن تسع الطائرة كل المسافرين إذا كان من الممكن استخدام مقاعد الدرجة الأولى لمسافري الدرجة السياحية ؟

الفصل الرابع

الارتباط والانحدار

CORRELTION AND REGRESSION

مقدمة :لاشك أن الطبيعة زاخرة بالظواهر الطبيعية ،المعروفة وغير المعروفة،المألوفة والغريبة،وعلى الباحث في أي مجال كان قبل كل شيء ؛ان يتعرف على العوامل المؤثرة على الظاهرة المدروسة وما هي طريقة وشكل وقوة

التأثير.هذا ما يطلق عليه بشكل عام دراسة العلاقة بين المتغيرات relationship between variables .ان معرفة نوع وشكل وقوة العلاقة وبالتالي التأثير هي أساس النجاح في التعرف على الظاهرة المدروسة .

ولحسن الحظ انه تم تطوير وسائل إحصائية كثيرة بناء على أسس رياضية من أجل دراسة العلاقة بين المتغيرات relationship between variables.ان من أهم هذه التحاليل هي التالية :

1. تحليل الانحدار REGRESSION ANALYSIS
2. تحليل الارتباط CORRELATION ANALYSIS
3. تحليل التباين VARIANCE ANALYSIS

السؤال الآن الذي تطرح نفسه هو أي من التحاليل يتوجب استخدامه من أجل دراسة العلاقة بين المتغيرات،أي العوامل المؤثرة والظاهرة المدروسة ؟

لكن قبل الإجابة على السؤال المطروح علينا أن نقدم بعض الشروحات الضرورية لتوضيح المسألة.

- أنواع المتغيرات variables: يمكن تقسيم المتغيرات إلى نوعين وذلك حسب القيم التي يأخذها المتغير :
  - متغير كمي QUALITIVE:هو المتغير الذي يأخذ قيم كمية أي قابلة للقياس كميا كوحدات القياس مثل وحدات الوزن،الطول ،الحجم ...إلى آخره. وكمثال على ذلك المبيد ،السماد،العمر،كمية الغذاء وهكذا...
  - متغير نوعي QUANTITIVE: هو المتغير الذي يأخذ قيم نوعية غير قابلة للقياس كميا بوحدات القياس المذكورة آنفا بل يقاس بوحدات القياس النوعي كالدرجات والتدرجات والمستويات وهكذا ... إلى آخره.وكمثال على ذلك اللون ودرجة النجاح ودرجة المقاومة ...
- مصادر المتغيرات variables: يمكن تقسيم المتغيرات إلى نوعين وذلك حسب نوع التحكم أو المصدر إلى نوعين :

3. متغير احتمالي RANDOM variables: هو المتغير الذي يأخذ قيم عشوائية محتملة لا يمكن التحكم بها مسبقا مثل المتغيرات الجوية كالحرارة والرطوبة وغيرها .
4. متغير غير احتمالي Controlled variable: المتغير الذي يأخذ قيم محددة يمكن التحكم بها مسبقا مثل المبيد ،السماد،العمر،كمية الغذاء وهكذا...

فيما يلي سنتعرض بالتفصيل لدراسة هذه الوسائل الإحصائية كل على حدة.

1-4 - تحليل الانحدار REGRESSION ANALYSIS :

1-1-4 تعريف : يعرف تحليل الانحدار REGRESSION ANALYSIS بأنه أحد التحاليل الإحصائية الهامة التي تستخدم بشكل عام دراسة العلاقة بين المتغيرات relationship between variables (مثل التحاليل السابقة الذكر كتحليل الارتباط CORRELATION ANALYSIS

وتحليل التباين VARIANCE ANALYSIS) وهو يستخدم في الحالات التالية:

- عندما يكون العامل المؤثر (إي المتغير) من نوع كمي.
- عندما يكون العامل المؤثر من النوع متغير غير احتمالي Controlled variable.
- عندما يكون التأثير من جهة واحدة أي أن العامل المؤثر (المتغير) يؤثر على الصفة المختارة للظاهرة المدروسة والعكس غير صحيح.
- عندما تكون الغاية هي معرفة النموذج الرياضي للعلاقة ما بين العامل المؤثر (إي المتغير) و الصفة المختارة للظاهرة المدروسة.
- عندما تكون الغاية هي التنبؤ الرياضي (أو التوقع الرياضي) لقيم الصفة المختارة للظاهرة المدروسة.

 يجدر التنويه إلى أنه في تحليل الانحدار REGRESSION ANALYSIS يمكن

أن نأخذ بعين الاعتبار ما يلي:

- يطلق على العامل المؤثر (إي المتغير) اسم آخر وهو العامل المستقل(X) بينما يطلق على الصفة المختارة للظاهرة المدروسة اسم العامل التابع (Y) .
- يمكن ان يوجد في التجربة الحيوية أكثر من عامل مؤثر ($X_1,X_{2,...}X_n$) بينما لا يمكن أن يوجد أكثر من صفة مختارة واحدة للظاهرة المدروسة.
- في حال وجود أكثر من عامل مؤثر في التجربة الحيوية فانه لإجراء تحليل الانحدار REGRESSION ANALYSIS لابد من تحقيق الشروط السابقة الذكر على كل عامل سيدخل في تحليل الانحدار REGRESSION ANALYSIS ولكن في هذه الحالة سيتم استخدام تحليل الانحدار المتعدد MULTI-REGRESSION ANALYSIS أما إذا كان هناك عامل مؤثر وحيد في التجربة الحيوية فانه في هذه

الحالة سيتم استخدام تحليل الانحدار البسيط -LINEAR- REGRESSION ANALYSIS

4-1-2- أولا : تحليل الانحدار البسيط LINEAR-REGRESSION ANALYSIS :

إذا كان لدينا في تجربة حيوية أو زراعية عامل مؤثر واحد (X) وندرس صفة مختارة (Y) فان معادلة الانحدار البسيط هي من الشكل التالي :

$$Y = a + b*X$$

حيث أن : Y القيمة النظرية للصفة المختارة للظاهرة المدروسة

a الثابت للانحدار

b عامل الانحدار

X العامل المؤثر (إي المتغير)

أما كيفية حساب كل من a الثابت للانحدار و b عامل الانحدار فهي على الشكل التالي:

- حساب a الثابت للانحدار : يمكن حساب a ثابت الانحدار باستخدام العلاقة الرياضية التالية:

$$a = \bar{y} - b * \bar{x}$$

- حساب b عامل الانحدار : : يمكن حساب b عامل الانحدار باستخدام العلاقة الرياضية التالية:

$$b = \sum (x - \bar{x}) * (y - \bar{y} / \sum x - \bar{x})^2$$

- كما ويمكن حساب b عامل الانحدار باستخدام العلاقة الرياضية التالية:

$$b = (\sum x * y - \sum x * \sum x / n) / \sum x^2 - (\sum x)^2 / n$$

مثــال ( ) : عند دراسة العلاقة ما بين كمية السماد المضاف كغ/ها وصفة الإنتاج لأحد المحاصيل وجاء كان لدينا التحليل الإحصائي للبيانات الإحصائية DATA كان لدينا ما يلي :

السماد	الإنتاج
X	**Y**
1	**5**
2	**13**
3	**16**
4	**23**
5	**33**
6	**38**
7	**40**

SUMMARY OUTPUT

إحصائية　　مؤشرا ت

Regression Statistics	
Multiple R	0.989094
R Square	0.978307
Adjusted R Square	0.973968
Standard Error	2.164651
Observations	7

ANOVA تحليل التباين

	df	SS	MS	F	Significance F
Regression	1	1056.571	1056.571	225.4878	2.37E-05
Residual	5	23.42857	4.685714		
Total	6	1080			

تحليل الانحدار

	Coefficients	Standard Error	t Stat	P-value
Intercept a الثابت	0.57143	1.829464	-0.31235	0.767383
عامل الانحدار b	6.142857	0.409081	15.01625	2.37E-05

من الجدول السابق الذي يتضمن نتائج تحليل الانحدار REGRESSION ANALYSIS باستخدام أحد الطرق السابقة الذكر نلاحظ أن :

b عامل الانحدار = 6.1428

الثابت a = 0.57143

وبالتالي تكون معادلة الانحدار البسيط LINEAR-REGRESSION ANALYSIS هي من الشكل التالي :

$$y' = a + b * x$$

وبعد التعويض في المعادلة المذكورة نجد أن :

Y = 6.1428 +0.5714*X

أما تفسير ألأرقام الواردة في معادلة الانحدار المذكورة فهو على الشكل التالي:

- أولا b عامل الانحدار = 6.1428 +   بما أن b عامل الانحدار موجب فهذا يعني أن الانحدار ايجابي والعلاقة طر دية وبالتالي إذا زاد قيمة العامل X ازدادت قيمة العامل التابع وهو الإنتاج وإذا نقصت قيمة العامل X وهو السماد نقصت قيمة العامل التابع وهو الإنتاج والعكس صحيح.
- يزداد قيمة العامل التابع وهو الإنتاج بمعدل قيمة b عامل الانحدار = 6.1428 +   (طن /ها) عندما تزداد قيمة العامل X وهو السماد وحدة واحدة (1 كغ/ها).
- الثابت a   = 0.57143 يعني أن قيمة العامل التابع وهو الإنتاج تساوي الثابت a   = 0.57143 (طن /ها) عندما تكون قيمة العامل X وهو السماد مساوية للصفر أي عندما لا يوجد أي إضافة للعامل X وهو السماد.
- أما على سبيل المثال إذا كانت قيمة (b) عامل الانحدار قيمة سلبية فإننا نقول أن الانحدار سلبي والعلاقة عكسية وبالتالي إذا زادت قيمة العامل المستقل X نقصت قيمة العامل التابع وهو الإنتاج وإذا نقصت قيمة العامل X وهو السماد زادت قيمة العامل التابع وهو الإنتاج والعكس صحيح.

3-1-4 جودة معادلة الانحدار REGRESSION MODEL :

ان معادلة الانحدار REGRESSION MODEL المستخدمة عادة هي من الشكل الخطي البسيط ،لكن في معظم الحالات فان هذا الشكل الخطي البسيط قد يكون غير مناسبا لتمثيل البيانات الإحصائية DATA وبالتالي كيف يمكن ان نعرف هل هذا النموذج الرياضي أي معادلة الانحدار البسيط LINEAR-REGRESSION ANALYSIS هل هو مناسب أم لا .

للإجابة على السؤال المذكور يجب التأكد من أن الفرق بين البيانات التجريبية و البيانات الإحصائية DATA المحسوبة من المعادلة معادلة الانحدار REGRESSION MODEL

هو أقل ما يمكن أي بتعبير أخر يجب أن يكون المقدار التالي أقل ما يمكن :

$$\sum(y_i - \hat{y}_i)^2$$

حيث أن : y_i القيم الحقيقية أو التجريبية

y_i القيم المحسوبة من معادلة الانحدار البسيط.

كما أن خطأ التناسب والذي يعط بالعلاقة التالية :

$$ERROR = \sqrt{\sum (y_I - \hat{y})^2 / n - (k+1)}$$

حيث أن : k عدد العوامل المستقلة

n عدد عناصر العينة

y_i القيم الحقيقية أو التجريبية

y_i القيم المحسوبة من معادلة الانحدار البسيط.

من الطبيعي أنه كلما كان المقدار ERROR ،الذي يرمز له عادة S_d، صغيرا كلما كانت جودة معادلة الانحدار REGRESSION MODEL أكبر.هذا يعني علينا دائما البحث عن معادلة الانحدار REGRESSION MODEL الأنسب كما انه يوجد معايير أخرى لجودة معادلة الانحدار ستذكر لاحقا.

4-1-4- الانحدار المتعدد MULTI-REGRESSION ANALYSIS :

إذا كان لدينا في تجربة حيوية أو زراعية أكثر من عامل مؤثر واحد ($X_1, X_2, ...X_n$) وندرس صفة مختارة (Y) فان معادلة الانحدار المتعدد هي من الشكل التالي :

$$Y = a + b_1 * X_1 + b_2 * X_2 + ... b_n * X_n$$

حيث أن : Y القيمة النظرية للصفة المختارة للظاهرة المدروسة

a الثابت للانحدار

b عامل الانحدار

X_1 العامل المؤثر الأول

X_2 العامل المؤثر الثاني

X_n العامل المؤثر الأخير

وعادة نستخدم معادلة الانحدار المتعدد MULTI-REGRESSION ANALYSIS في حال وجود عاملين مستقلين اثنين وتصبح معادلة الانحدار المتعدد من الشكل التالي :

$$Y = a + b_1*X_1+b_2*X_2$$

أما بالنسبة لطريقة إيجاد عوامل الانحدار المذكورة في المعادلة أعلاه فان ذلك يتطلب الطريقة الرياضية المعروفة في حل جملة معادلات، ولكن نظرا لصعوبة هذه الطريقة الرياضية بالنسبة للطلاب غير المتخصصين بالإضافة لتوفر البرامج الإحصائية المتطورة (برامج الحاسوب SOFTWARE) فلا يوجد أي مبرر لتعلم هذه الطرق الرياضية في هذا الفصل.

مثـــال ( ) : لدينا البيانات التالية التي تبين العلاقة ما بين السماد (X_1) بالكغ/ها والري (X_2) وصفة الإنتاج (Y) طن/ها من أحد المحاصيل الزراعية حيث اجري تحليل الانحدار المتعدد MULTI-REGRESSION ANALYSIS كما يلي :

X1	**X2**	**Y**
1	2	5
2	5	13
3	3	16
4	8	23
5	6	33
6	5	38
7	9	40

SUMMARY OUTPUT

مؤشرات إحصائية	
Regression Statistics	
Multiple R	0.98962932
R Square	0.97936618
Adjusted R Square	0.96904927
Standard Error	2.36032433
Observations	7

تحليل التباين

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	1057.715476	528.8577	94.92825	0.000426
Residual	4	22.28452381	5.571131		
Total	6	1080			

تحليل الانحدار

العوامل	*Coefficients*	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.0547619	2.29767487	-0.02383	0.982127
X1	6.36428571	0.661612967	9.619349	0.000653
X2	-0.25833333	0.570072486	-0.45316	0.673928

من الجدول السابق الذي يتضمن نتائج تحليل الانحدار REGRESSION ANALYSIS باستخدام أحد برامج جاهزة نلاحظ أن :

b_1 عامل الانحدار = 6.36428571

b_2 عامل الانحدار =0.25833333-

الثابت a = -0.0547619

وبالتالي تكون معادلة الانحدار المتعدد MULTI-REGRESSION ANALYSIS هي من الشكل التالي :

$$y' = a + b_1 * x_1 + b_2 * x_2$$

وبعد التعويض في المعادلة المذكورة نجد أن :

$$Y = 6.1428 + 6.36428571 X_1 - 0.25833333 X_2$$

4-1-5- الانحدار اللاخطي NUNLINEAR-REGRESSION ANALYSIS :

لدى دراسة العلاقة فلا بد أن نستقرىء أولا فيما إذا كان يوجد علاقة أصلا بين المتغيرات (X و Y) وبمعنى آخر هل تقدم البيانات الإحصائية DATA المتوفرة دليلا كافيا على وجود علاقة خطية بين المتغيرات .

ان القول بأن العامل X والعامل Y لا يرتبطان ببعضهما خطيا يكافىء القول بأن b عامل الانحدار الخطي يساوي الصفر والعكس صحيح ؛أي أن نقول إذا أثبتنا ان b عامل الانحدار الخطي يساوي الصفر أو لا يتميز عن الصفر بشكل معنوي فهذا يعني أن العلاقة بين المتغيرات ليست خطية أو العامل X والعامل Y لا يرتبطان ببعضهما خطيا.

إذن إذا كنا نرغب في دراسة العلاقة بين المتغيرات لمعرفة هل العلاقة خطية أم لا علينا أن نختبر الفرضية الابتدائية (H0: b=0) مقابل الفرضية البديلة (H1: b ≠ 0) .

من أجل هذا الغرض فان الاختبار المناسب هو اختبار ت T-TEST عند مستوى المعنوية 5% ودرجة الحرية DF=N-2 والذي سبق شرحه في الفقرة السابقة من هذا الفصل.

مثـــال ( ):لدى دراسة العلاقة بين المتغيرين السماد X (كغ/ها) والإنتاج Y (طن/ها) حصلنا على البيانات الإحصائية DATA الواردة في الجدول التالي بالإضافة إلى تحليل الانحدار ومن ثم تطبيق اختبار ت T-TEST :

الإنتاج	السماد
Y	X
5	10
13	20
16	30
4	40
2	50

SUMMARY OUTPUT

تحليل الانحدار

Regression Statistics	
Multiple R	0.387298335
R Square	0.15
Adjusted R Square	-0.133333333
Standard Error	6.519202405
Observations	5

ANOVA

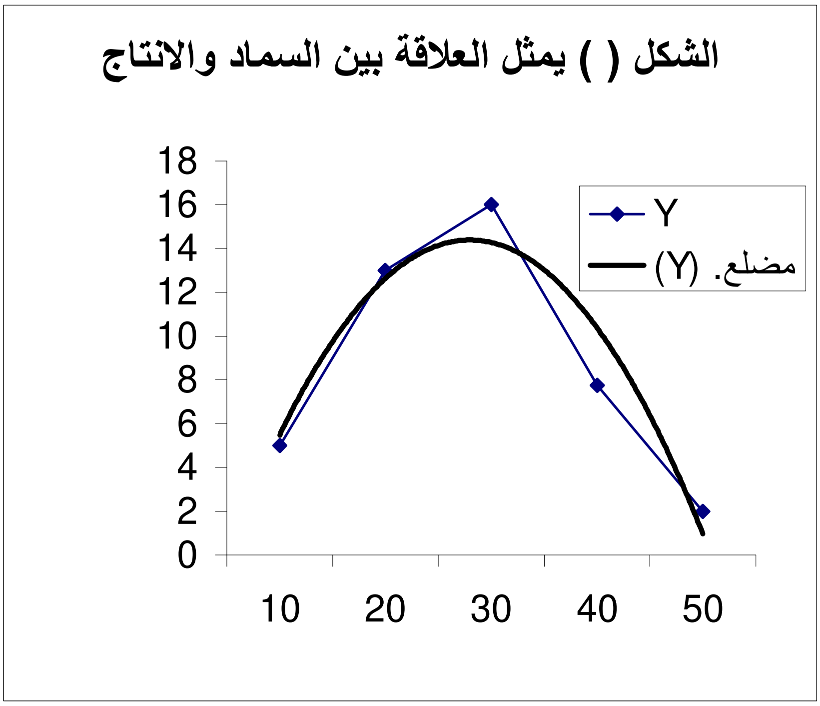
	df	*SS*	*MS*	*F*
Regression	1	22.5	22.5	0.529412
Residual	3	127.5	42.5	
Total	4	150		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	a = 12.5	6.837397166	1.828181	0.164969
X	b = -0.15	0.206155281	-0.72761	0.519498

من الجدول السابق أعلاه نلاحظ ان :   *t Stat* = -0.727

P-value = 0.519

نستنتج من قيم اختبار ت T-TEST المطبقة على عامل الانحدار b لمعرفة هل العلاقة خطية أم لا أي اختبار مقابل الفرضية البديلة (H1: b ≠ 0) نجد *ان* الفرضية الابتدائية (H0: b=0) مرفوضة وبالتالي نقبل الفرضية البديلة (H1: b ≠ 0) هذا يؤكد *ان عامل الانحدار الخطي غير معنوي ولا يتميز عن الصفر بشكل معنوي ،هذا يقود للقول انه توجد علاقة لاخطية كما يبدو من الشكل التالي:*



نلاحظ من الشكل الأخير أنه لا يمكن تمثيل العلاقة السابقة بخط مستقيم. إذن لدى دراسة العلاقة بين المتغيرين السماد X (كغ/ها) والإنتاج Y (طن/ها) وحصلنا

على البيانات الإحصائية DATA الواردة في الجدول *السابق* فان معادلة الانحدار اللاخطي التي توافق الشكل البياني السابق هي من الشكل التالي :

$$Y = a + b_1*X+b_2*X^2$$

حيث أن : Y القيمة النظرية للصفة المختارة للظاهرة المدروسة

a الثابت للانحدار

b_1 عامل الانحدار الخطي

b_2 عامل الانحدار التربيعي

وعادة نستخدم معادلة الانحدار اللاخطي MULTI-REGRESSION ANALYSIS في حال وجود عاملين مستقلين اثنين وتصبح معادلة الانحدار اللاخطي من الشكل التالي :

$$Y = a + b_1*X_1+b_2*X_2 +b_3*X^2+b_4*X^2+b_5*X1*X2$$

أما بالنسبة لطريقة إيجاد عوامل الانحدار المذكورة في المعادلة أعلاه فان ذلك يتطلب الطريقة الرياضية المعروفة في حل جملة معادلات، ولكن نظرا لصعوبة هذه الطريقة الرياضية بالنسبة للطلاب غير المتخصصين بالإضافة لتوفر البرامج الإحصائية المتطورة (برامج الحاسوب SOFTWARE) فلا يوجد أي مبرر لتعلم هذه الطرق الرياضية في هذا الفصل.

وهنا تبرز مسألة هامة وهي اختيار معادلة الانحدار اللاخطي المناسبة من بين مجموعة النماذج السابقة الذكر وهذه قضية تشكل تحد كبير خصوصا بالنسبة لغير المتخصصين وهو ما يطلق عليه PEST FIT ،لذا فإننا سنقدم وسيلة مساعدة سهلة لاختيار النموذج المناسب وهي التالية :

لدينا مقياسين أو مؤشرين إحصائيين اثنين هما :

- عامل التحديد R^2
- الخطأ القياسي للانحدار STADARD ERROR

ويمكن للباحث أن يعتمد على كلا المؤشرين الإحصائيين المذكورين لاختيار النموذج المناسب ،إذا أمكن ذلك ، ولكن يمكن الاعتماد على أحدهما إذا كان الآخر متوفر أو غير كاف أيضا.

فبالنسبة إلى المؤشر الأول عامل التحديد R^2 فهو عبارة عن النسبة المئوية للعامل المؤثر الذي أمكن كشفه من خلال النموذج المقترح ويمكن القول أن قيمته تتراوح بين 0%- 100% وكلما كانت أكبر كلما كان النموذج المقترح أكثر تناسبا مع البيانات الإحصائية DATA .

أما بالنسبة إلى المؤشر الثاني الخطأ القياسي للانحدار فيمكن القول أن قيمته ليس لها مجالا محدودا وكلما كانت أصغر كلما كان النموذج المقترح أكثر تناسبا مع البيانات الإحصائية DATA .لذا عند المقارنة بين نموذجين أو أكثر يمكن استخدام كلا المؤشرين الإحصائيين أو أحدهما كأساس لاختيار النموذج المناسب أو الأكثر تناسبا.

مثـــال (   ) :لنستعين بأحد الأمثلة السابقة وليكن المثال التالي الذي يقوم بدراسة العلاقة ما بين كمية السماد المضاف كغ/ها وصفة الإنتاج لأحد المحاصيل على الشكل التالي :

الإنتاج	السماد
Y	X2
5	10
13	20
16	30
7.73	40
5.51	50
10	60

6-1-4 اختبار معنوية الانحدار :ان المقصود من اختبار معنوية الانحدار هو التأكد إحصائيا من أن ؛ وبتعبير آخر ان اختبار معنوية الانحدار هو رياضيا اختبار العلاقة التالي :

$$b_i \neq 0$$

ولتحقيق ذلك فانه يتوجب استخدام اختبار ت (t-test) وهو أحد اختبارات المعنوية المعروفة وذلك حسب العلاقة التالية :

$$t = b / s_b$$

حيث أن : b عامل الانحدار

s_{ib} الخطأ القياسي للانحدار

مثـــال ( ) : سنستعين بالمثال السابق لتوضيح كيفية اختبار معنوية الانحدار من خلال إجراء اختبار معنوية عامل الانحدار باستخدام اختبار ت (t-test) كما يلي :

الإنتاج	السماد
Y	**X**
5	**1**
13	**2**
16	**3**
23	**4**
33	**5**
38	**6**
40	**7**

نطبق اختبار ت (t-test) على البيانات السابقة المذكورة في الجدول أعلاه كما يلي :

مؤشرات الانحدار	Coefficients	Standard Error	t Stat	P-value
Intercept a الثابت	-0.57143	1.829464	-0.31235	0.767383
b عامل الانحدار	6.142857	0.409081	15.01625	2.37E-05

من الجدول الأخير الوارد أعلاه نلاحظ ان قيمة المؤشر PARAMETER الناتجة عن استخدام اختبار ت (t-test) بالنسبة لعامل الانحدار b هي كما يلي :

T-STAT = 15.01625

P-value = 2.37E-05

نستنتج من ذلك أن الانحدار معنوي وبالتالي يكون انحدار العامل المستقل (X) على العامل التابع (Y) معنوي مؤكد إحصائيا حيث نتبع القاعدة التالية حين نحتاج للاستنتاج فيما إذا كان الانحدار معنوي أم لا كما يلي :

إذا كانت القيمة الاحتمالية لـ *P-value* اكبر أو تساوي 0.05 نقول أن الانحدار غير معنوي وبالتالي يكون انحدار العامل المستقل (X) على العامل التابع (Y)غير معنوي وغير مؤكد إحصائيا أما إذا كانت القيمة الاحتمالية لـ *P-value* أصغر من 0.05 نقول أن الانحدار معنوي وبالتالي يكون انحدار العامل المستقل (X) على العامل التابع (Y) معنوي مؤكد إحصائيا وناتج عن الصدفة.

 من الجدير بالذكر أنه يوجد طريقة أخرى عند الحاجة للاستنتاج فيما إذا كان الانحدار معنوي أم لا كما يلي :

إذا كانت القيمة الإحصائية T-STAT ،المحسوبة كما تسمى أحيانا، اكبر من قيمة T-TABLE ، الجدولية كما تسمى أحيانا، نقول أن الانحدار معنوي وبالتالي يكون

انحدار العامل المستقل (X) على العامل التابع (Y) معنوي مؤكد إحصائيا أما إذا كانت القيمة الإحصائية -T أصغر أو تساوي قيمة T-TABLE نقول أن الانحدار غير معنوي وبالتالي يكون انحدار العامل المستقل (X) على العامل التابع (Y) معنوي غير مؤكد إحصائيا وناتج عن الصدفة.

ويمكن للطالب أو القارئ مراجعة الكتاب العملي –الدرس العملي الرابع – للإطلاع على كيفية إجراء أو استخدام اختبار ت (t-test) لتوضيح كيفية اختبار معنوية الانحدار وتفسير النتائج حيث يتوفر المزيد من التفاصيل والأمثلة المناسبة.

2-4 - تحليل الارتباط CORRELATION ANALYSIS :

1-2-4 تعريف : يعرف تحليل الارتباط CORRELATION ANALYSIS بأنه أحد التحاليل الإحصائية الهامة التي تستخدم بشكل عام دراسة العلاقة بين المتغيرات relationship between variables (مثل التحاليل السابقة الذكر كتحليل الانحدار REGRESSION ANALYSIS وتحليل التباين VARIANCE ANALYSIS) وهو تلازم صفتين تلازما كميا.

2-2-4 - يستخدم تحليل الارتباط CORRELATION ANALYSIS في الحالات التالية:

- عندما يكون العامل المؤثر (إي المتغير) من نوع كمي.
- عندما يكون العامل المؤثر من النوع متغير احتمالي Random variable.
- عندما يكون التأثير من جهتين اثنتين أي أن العامل المؤثر (المتغير) يؤثر على الصفة المختارة للظاهرة المدروسة والعكس صحيح ،أي يوجد ما يسمى تأثير متبادل INTERACTION مابين العامل المؤثر (المتغير) و الصفة المختارة للظاهرة المدروسة.
- عندما يكون العامل المؤثر و العامل المتأثر (التابع) كلاهما صفة مدروسة ،أي عندما ندرس العلاقة بين المتغيرات الاحتمالية Random variables والتي هي في أغلب الأحيان صفات مدروسة .
- عندما نرغب في معرفة قوة العلاقة مابين العامل المؤثر و العامل المتأثر (الصفات المدروسة)، وذلك من خلال حساب عامل الارتباط كما سنرى بعد قليل.

- عندما نرغب في معرفة نسبة التأثير المتبادل INTERACTION مابين العامل المؤثر (المتغير) و الصفة المختارة للظاهرة المدروسة وذلك من خلال حساب عامل التحديد كما سنرى بعد قليل.

إذن ان تحليل الارتباط CORRELATION ANALYSIS هو تلازم صفتين تلازما أي ارتباطا كميا بين عاملين احتماليين أو صفتين بحيث إذا طرأ أي تغير بالزيادة أو النقصان على إحدى الصفتين يؤدي بالمقابل إلى تغير بالزيادة أو النقصان للصفة الأخرى.

بناء على ذلك يمكن أن نستنتج ما يلي :

1. إذا أدت الزيادة في إحدى الصفتين إلى زيادة في الصفة الأخرى نقول أن الارتباط ايجابي
2. إذا أدت الزيادة في إحدى الصفتين إلى نقص في الصفة الأخرى نقول أن الارتباط سلبي.

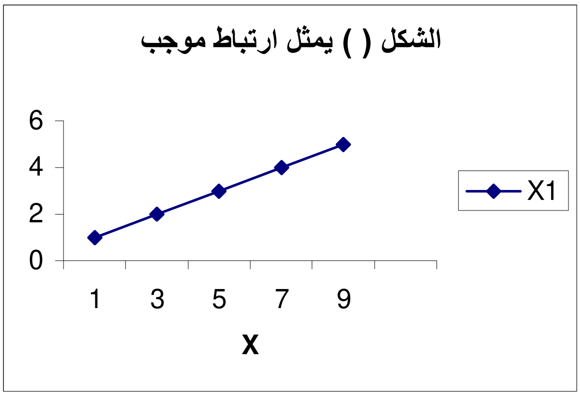
ان نتيجة تحليل الارتباط CORRELATION ANALYSIS هي قيمة وحيدة مهمة الا وهي عامل الارتباط ويرمز له بـ (r) وهنا نميز الحالات التالية :

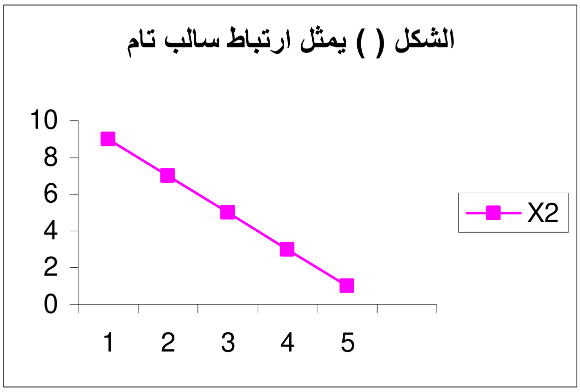
- ان قيمة عامل الارتباط (r) = 1 نقول ان الارتباط كامل وايجابي
- ان قيمة عامل الارتباط (r) = -1 نقول ان الارتباط كامل وسلبي
- معدوم
- ان قيمة عامل الارتباط 1-< (r) < 1 وبنفس الوقت قيمة عامل الارتباط (r) <> 0 نقول ان الارتباط حسب قيمته كما يلي:

1. إذا كانت قيمة عامل الارتباط (r) بين القيمتين 0 و 0.25 نقول ان الارتباط ضعيف (موجب أو سالب حسب إشارته)
2. إذا كانت قيمة عامل الارتباط (r) بين القيمتين 0.5و 0.25 نقول ان الارتباط متوسط (موجب أو سالب حسب إشارته)
3. إذا كانت قيمة عامل الارتباط (r) بين القيمتين 0.5و 0.75 نقول ان الارتباط قوي (موجب أو سالب حسب إشارته)

4. إذا كانت قيمة عامل الارتباط (r) بين القيمتين 1 و 0.75 نقول ان الارتباط قوي جدا (موجب أو سالب حسب إشارته)

والأشكال التالية (   ) تبين الحالات السابقة الذكر بشكل بياني.





4-2-3- أنواع الارتباط :في الحقيقة يوجد عدة أنواع من الارتباط بين المتغيرات الاحتمالية Random variables و الصفة المختارة للظاهرة المدروسة، لذا فإننا سنحاول دراسة الأشكال البسيطة منها وهي التالية :

1. الارتباط البسيط SIMPLE CORRELATION ANALYSIS :

وهو أبسط أشكال الارتباط المستخدمة في التحليل الإحصائي للبيانات الإحصائية DATA .في هذا الشكل المبسط يتم حساب عامل الارتباط (r) باستخدام إما العلاقة التالية:

$$r = \sum_{1}^{n}(x_i - \bar{x}) * (y_i - \bar{y}) / \sqrt{\sum_{i}^{n}(x_i - \bar{x})^2 * \sum_{i}^{n}(y_i - \bar{y})^2}$$

أو باستخدام العلاقة التالية التي تسمى عادة طريقة تربيع القيم :

$$r = (\sum x_i * y_i - \sum x_i * \sum y_i / n) / \sqrt{\left[\sum x_i^2 - (\sum x_i)^2 / n\right] * \left[\sum y_i^2 - (\sum y_i)^2 / n\right]}$$

حيث أن :

r عامل الارتباط

n عدد عناصر العينة

y_i القيم الحقيقية أو التجريبية للعامل الأول أو الصفة الاولى

x_i القيم الحقيقية أو التجريبية للعامل الثاني أو الصفة الثانية

x^- المتوسط الحسابي للعامل الأول أو الصفة الاولى

y^- المتوسط الحسابي للعامل الثاني أو الصفة الثانية

مثـــال ( ) : عند دراسة العلاقة ما بين صفة عدد الحبوب في السنبلة الواحدة وصفة الإنتاج لأحد المحاصيل وهو القمح عالي الإنتاج فكان لدينا ما يلي :

الإنتاج	عدد الحبوب/السنبلة

X	Y
9	5
7	13
5	16
3	23
1	33

ثم أجري التحليل الإحصائي للبيانات الإحصائية DATA فحصلنا على النتائج التالية:

تحليل الارتباط:المؤشرات	
Correlation Statistics	
عامل الارتباط	
r	0.986063603
R Square	0.972321429
Adjusted R Square	0.963095238
Standard Error	2.033060091
Observations	5

وكما يبدو من الجدول المذكور أعلاه نجد أن:

عامل الارتباط r = 0.986063603

وهنا يمكن ان نستنتج أن الارتباط ايجابي قوي جدا حسب الاعتبارات السابقة الذكر.

2. ارتباط الرتب : يستخدم ارتباط الرتب لدراسة العلاقة بين المتغيرات الاحتمالية Random variables عندما تكون بيانات هذه العوامل مرتبة أو يمكن ترتيبها في رتب معينة . في هذا الشكل يتم حساب عامل الارتباط (r) باستخدام العلاقة التالية:

$$r = 6 * \sum (x_i - y_i)^2 / n(n^2 - 1)$$

حيث أن :

y_i رتب العامل الأول أو الصفة الاولى

x_i رتب العامل الثاني أو الصفة الثانية

مثـــال( ) : لدينا ستة طلاب لديهم الدرجات التالية في الإحصاء والاقتصاد:

الاقتصاد	**الإحصاء**	**الطالب**
46	52	1
57	34	2
53	64	3
26	57	4
43	57	5
56	36	6

الحـــل:

- نقوم بفرز البيانات تصاعديا فنحصل على الجدول التالي :

الاقتصاد	رتب البرمجة	الإحصاء	رتب الإحصاء
26	6	34	6
43	5	36	5
46	4	52	4
53	3	57	2.5
56	2	57	2.5
57	1	64	1

بالنسبة لمادة الإحصاء نلاحظ أنها لها أعلى قيمة لذلك أعطيت المرتبة الاولى أما الدرجة 57 فهي مكررة مرتين لذلك تقع رتبتها في الوسط بين الرتبتين أي 2.5 = 2/(2+3) وهكذا تكتمل الرتب بالنسبة لبقية الدرجات .

- الآن نعيد كتابة البيانات كما يلي :

الاقتصاد	رتب الاقتصاد	الإحصاء	رتب الإحصاء	الطالب
46	6	52	6	1
57	5	34	5	2
53	4	64	4	3
26	3	57	2.5	4
43	2	57	2.5	5
56	1	36	1	6

- نطبق العلاقة السابقة الذكر لحساب عامل الارتباط (r) فنحصل على التالي :

$$r = 1- 6* \Sigma (x_i - y_i) / n(n^2 -1)$$

$$= 1 - 6*(56.5) / 6*(36 - 1)$$

$$= 1 - (56.5 / 35)$$

$$= 1- 1.61$$

$$= - 0.61$$

نلا حظ من النتيجة السابقة أن عامل الارتباط $r = -0.61$ وهذا يعني أن الارتباط سلبي ولكن قوي . بتعبير آخر ان ترتيب الطالب في مادة الإحصاء معاكس لترتيبه في مادة الاقتصاد ،هذا يقودنا للقول أنه إذا كانت علامة الطالب في مادة الإحصاء مرتفعة فان علامته ستكون في مادة الاقتصاد منخفضة والعكس صحيح.

 تجدر التنويه هنا إلى أنه يمكن للطالب أو القارئ العودة إلى كتاب الإحصاء وتصميم التجارب – القسم العملي لمزيد من المعلومات عن طريقة الحل ومزيد من الأمثلة المناسبة أيضا.

3. الارتباط الجزئي : يستخدم الارتباط الجزئي لدراسة العلاقة بين المتغيرات الاحتمالية Random variables عندما يوجد عدة متغيرات ونرغب بحساب الارتباط بين متغيرين أو صفتين محددتين علما أن الارتباط مع بقية المتغيرات معلومة .فإذا كان لدينا عدة متغيرات X1,X2,....Xn ورغبنا حساب الارتباط بين متغيرين أو صفتين محددتين X1 & X2 علما أن الارتباط مع بقية المتغيراتX3,....Xn معلومة؛عندئذ نسمي الارتباط المرغوب "الارتباط الجزئي" .ولحسا ب الارتباط الجزئي نستخدم العلاقة العامة التالية:

$$r_{12,3} = r_{12} - r_{13} * r_{23} / \sqrt{(1 - r_{13}^{\ 2}) * (1 - r_{23}^{\ 2})}$$

حيث أن : $r_{12,3}$ الارتباط الجزئي بين متغيرين أو صفتين محددتين X1 & X2

r_{12} عامل الارتباط البسيط بين متغيرين أو صفتين X1 & X2

r_{13} عامل الارتباط البسيط بين متغيرين أو صفتين & X1
X3

r_{23} عامل الارتباط البسيط بين متغيرين أو صفتين & X2
X3

 جدير بالقول ان علاقة الارتباط الجزئي علاقة رجعية أي أنه لحساب الارتباط الجزئي لعاملين آخرين معلومين يلزمنا حساب عوامل الارتباط البسيط المطلوبة .

مثال ( ) : لدى دراسة العلاقة ما بين صفة عدد الحبوب في السنبلة الواحدة ،عدد السنابل /النبات الواحد وصفة الإنتاج لأحد المحاصيل وهو القمح عالي الإنتاج حيث لدينا ما يلي :

الإنتاج	عدد الحبوب/السنبلة	عدد السنابل/النبات
X3	X1	X2
5	9	3
13	7	2
16	5	6
23	3	5
33	1	4

- نطبق العلاقة السابقة الذكر لحساب عوامل الارتباط الجزئية ($r_{12,3}, r_{13,2}, r_{23,1}$) فنحصل على التالي :

	X2	X1	X3
X2	1		
X1	-0.5	1	
X3	0.358568583	-0.986063603	1

من الجدول السابق أعلاه نجد أن :

X1 & X2 الارتباط الجزئي بين المتغيرين $r_{12,3}$ = -0.5

X1 & X3 الارتباط الجزئي بين المتغيرين $r_{13,2}$ =-0.986063603

X3 & X2 الارتباط الجزئي بين المتغيرين $r_{12,3}$ = 0.358568583

- جدير بالقول هنا أيضا ان الارتباط الجزئي (و الارتباط البسيط) بين أي متغير ونفسه يساوي الواحد كما هو وارد في الجدول السابق أعلاه.

4. الارتباط المتعدد : الارتباط المتعدد : يستخدم الارتباط المتعدد لدراسة العلاقة بين المتغيرات الاحتمالية Random variables عندما يوجد عدة متغيرات ونرغب بحساب الارتباط بين متغير أو صفة محددة X1 مع بقية لمتغيرات X2,X3,....Xn .فإذا كان لدينا عدة متغيرات X1,X2,....Xn ورغبنا حساب الارتباط بين متغير أو صفة محددة X1 مع بقية المتغيرات X3,....Xn ؛عندئذ نسمي الارتباط المرغوب "الارتباط المتعدد" .ولحسا ب الارتباط المتعدد نستخدم العلاقة العامة التالية:

$$r_{1,23} = \sqrt{r_{12}{}^{2} + r_{13}{}^{2}} - 2 * r_{12} * r_{13} * r_{23} / 1 - r^{2}{}_{23}$$

مثـــال ( ) :نستعين بالمثال السابق وهو دراسة العلاقة ما بين صفة عدد الحبوب في السنبلة الواحدة ،عدد السنابل /النبات الواحد وصفة الإنتاج لأحد المحاصيل وهو القمح عالي الإنتاج حيث لدينا ما يلي :

الإنتاج	عدد الحبوب/السنبلة	عدد السنابل/النبات
X3	X1	X2
5	9	3
13	7	2
16	5	6
23	3	5
33	1	4

نطبق العلاقة السابقة الذكر لحساب عامل الارتباط المتعدد ($r_{1,23}$) فنحصل على النتائج الواردة في الجدول التالي :

Correlation Statistics	
Multiple R عامل ا لارتباط ا لمتعدد	0.998213
R Square	0.996429
Adjusted R Square	0.992857
Standard Error	0.894427
Observations	5

كما يبدو من الجدول السابق أعلاه نجد أن :

$r_{1,23} = 0.998213$ الارتباط ا لمتعدد بين المتغير X1 والمتغيرين X2،X3

وهذا يدل على وجود انحدار متعدد ايجابي قوي جدا.

5. عامل التحديد :يعتبر عامل التحديد أحد المؤشرات الإحصائية الهامة جدا حيث يقدم لنا النسبة المئوية لتأثير أحد العوامل (متغير أو صفة) أو عدة عوامل على عامل واحد عند وجود ارتباط متعدد.و عامل التحديد هو رياضيا عبارة عن مربع عامل ا لارتباط أي أن :

$$B = r^2$$

حيث أن : B عامل التحديد

r عامل ا لارتباط

مثـــال ( ) : نستعين أيضا بالمثال السابق وهو دراسة العلاقة ما بين صفة عدد الحبوب في السنبلة الواحدة ،عدد السنابل /النبات الواحد وصفة الإنتاج لأحد المحاصيل وهو القمح عالي الإنتاج وبعد تحليل البيانات الإحصائية حصلنا على النتائج التالية:

Correlation Statistics

Multiple R	عامل ا لارتباط لمتعدد	0.998213
R Square	عامل التحديد	0.996429
Standard Error		0.894427
Observations		5

كما يبدو من الجدول السابق أعلاه نجد أن عامل التحديد B = 99.64% وهذا يعني أن كلا من عامل عدد الحبوب/السنبلة و عدد السنابل/النبات يؤثران على الإنتاج بنسبة 99.64% أما السبنسة الباقية وهي 0.36 % من التأثير فتعود إلى عوامل أخرى غير معروفة.

4-4-4- اختبار معنوية عامل ا لارتباط :

اختبار معنوية الارتباط :ان المقصود من اختبار معنوية الارتباط هو التأكد إحصائيا من أن ؛ وبتعبير آخر ان اختبار معنوية الارتباط هو رياضيا اختبار العلاقة التالي :

$$r_i \neq 0$$

ولتحقيق ذلك فانه يتوجب استخدام اختبار ت (t-test) وهو أحد اختبارات المعنوية المعروفة وذلك حسب العلاقة التالية :

$$t = r * \sqrt{n - k / (1 - r^2)}$$

حيث أن : k عدد العوامل الداخلة في الاختبار

r عامل ا لارتباط

n عدد المشاهدات في الاختبار

مثـــال ( ) : نعود ونستعين أيضا بالمثال السابق وهو دراسة العلاقة ما بين صفة عدد الحبوب في السنبلة الواحدة ،عدد السنابل /النبات الواحد وصفة الإنتاج لأحد المحاصيل وهو القمح عالي الإنتاج وبعد تحليل البيانات الإحصائية حصلنا على النتائج التالية:

Correlation Statistics

Multiple R	عامل ا لارتباط ا لمتعدد	0.998213
R Square	عامل التحديد	0.996429
Standard Error		0.894427
Observations		5

نطبق اختبار ت t-test السابق الذكر فنحصل على النتائج التالية :

	Coefficients	*Standard*	*t Stat*	*P-value*

	عامل الارتباط	Error		
r_2	-1.2	0.326599	- 3.67423	0.066743
r_1	-3.6	0.163299	- 22.0454	0.002051

من الجدول الأخير الوارد أعلاه نلاحظ ان قيمة المؤشر PARAMETER الناتجة عن استخدام اختبار ت (t-test) بالنسبة لعامل الارتباط r_2 هي كما يلي :

T-STAT = -22.0454

P-value = 0.066743

نستنتج من ذلك أن الارتباط غير معنوي وبالتالي يكون ارتباط العامل (X1) مع العامل التابع (Y) معنوي غير مؤكد إحصائيا حيث نتبع القاعدة التالية حين نحتاج للاستنتاج فيما إذا كان الارتباط معنوي أم لا كما يلي :

إذا كانت القيمة الاحتمالية لـ *P-value* أصغر أو تساوي من 0.05 نقول أن الارتباط معنوي وبالتالي يكون ارتباط العامل المستقل (X) مع العامل (Y) معنوي مؤكد إحصائيا أما إذا كانت القيمة الاحتمالية لـ *P-value* أكبر 0.05 نقول أن الارتباط غير معنوي وبالتالي يكون ارتباط العامل المستقل (X) مع العامل (Y) غير معنوي غير مؤكد إحصائيا وناتج عن الصدفة.

 من الجدير بالذكر أنه يوجد طريقة أخرى عند الحاجة للاستنتاج فيما إذا كان الارتباط معنوي أم لا كما يلي :

إذا كانت القيمة الإحصائية T-STAT ،المحسوبة كما تسمى أحيانا، اكبر من قيمة T-TABLE ، الجدولية كما تسمى أحيانا، عند درجة الحرية المناسبة نقول أن الارتباط معنوي وبالتالي يكون ارتباط العامل المستقل (X) على العامل (Y) معنوي مؤكد إحصائيا أما إذا كانت القيمة الإحصائية -T أصغر أو تساوي قيمة T-TABLE نقول أن الارتباط غير معنوي وبالتالي يكون ارتباط العامل المستقل (X) مع العامل (Y) معنوي غير مؤكد إحصائيا وناتج عن الصدفة.

ويمكن للطالب أو القارئ مراجعة الكتاب العملي –الدرس العملي الرابع – للإطلاع على كيفية إجراء أو استخدام اختبار ت (t-test) لتوضيح كيفية اختبار معنوية الارتباط وتفسير النتائج حيث يتوفر المزيد من التفاصيل والأمثلة المناسبة.

نعود ونذكر بانه يمكن للطالب أو القارئ مراجعة الكتاب العملي –الدرس العملي الرابع – للإطلاع على كيفية إجراء أو استخدام اختبار ت (t-test) لتوضيح كيفية اختبار معنوية الارتباط وتفسير النتائج حيث يتوفر المزيد من التفاصيل والأمثلة المناسبة.

3-4-السلاسل الزمنية TIME SERIES :

1-3-4- مقدمة: من المعروف ان العديد من الظواهر يتعلق تطورها بالزمن،وبالتالي فان دراسة تطور مثل هذه الظواهر ومراقبة تغيراتها وتسجيل المشاهدات لها يتم باستخدام ما يسمى أسلوب السلاسل الزمنية TIME SERIES .

تتألف السلاسل الزمنية عادة من بيانات على شكل جدول مؤلف من عمودين أو سطرين أحدهما يمثل قيم الزمن والثاني يمثل قيم الظاهرة المدروسة.

مثــال ( ) : لدينا بيانات السلسلة الزمنية التالية التي تشمل تطور صادرات إحدى المعامل مقدرة بآلاف الليرات السورية :

الصادرات	العام
200	1990
150	1991
350	1992
100	1993
400	1994

من الجدير بالتنويه إلى أنه يوجد بعض الشروط العامة للسلاسل الزمنية TIME SERIES نذكر منها ما يلي :

- وحدة المكان : أي يجب ان تجمع البيانات الخاصة بالسلسلة الزمنية من مكان واحد ،أي بتعبير آخر يجب أن يكون عامل المكان ثابت لا يتغير بتغير الزمان.
- أي يجب ان تجمع البيانات الخاصة بالسلسلة الزمنية بشرط أن تكون وحدة قياس الزمن واحدة ومتساوية.بناء على هذا فان السلاسل الزمنية يمكن أن تنقسم حسب دورية المشاهدات إلى الأنواع التالية :

1. سلاسل زمنية سنوية حيث تكون المعطيات أو البيانات الإحصائية DATA سنوية مثل إنتاج القطن السنوي .
2. سلاسل زمنية فصلية حيث تكون المعطيات أو البيانات الإحصائية DATA فصلية مثل إنتاج المعمل الربع سنوي .
3. سلاسل زمنية أسبوعية حيث تكون المعطيات أو البيانات الإحصائية DATA شهرية مثل إنتاج المعمل أسبوعي.

كما أنه يمكن تنقسم السلاسل الزمنية TIME SERIES حسب العلاقة بين المشاهدات إلى الأنواع التالية :

1. السلاسل الزمنية المترابطة (غير مستقلة) :في هذا النوع من السلاسل تكون كل قيمة من قيم السلسة الزمنية مرتبطة مع القيمة التي تسبقها .أي بتعبير آخر ان كل رقم من أرقام السلسة الزمنية ناتج عن إضافة أو طرح مقدار معين من الرقم السابق.كمثال على ذلك عدد المتخرجين من الطلاب كمهندسين في القطر العربي السوري حيث أن عدد المتخرجين هذا العام هو عبارة عن مجموع المتخرجين من السنوات الماضية مع عدد المتخرجين هذا العام وهكذا..
2. السلاسل الزمنية غير المترابطة (مستقلة) :في هذا النوع من السلاسل تكون كل قيمة من قيم السلسة الزمنية مستقلة تماما عن القيمة التي تسبقها .أي بتعبير آخر ان كل رقم من أرقام السلسة الزمنية مستقل تماما عن أي رقم من ألأرقام السابقة. كمثال على ذلك إنتاج الهكتار الواحد من القمح هذا العام.

4 -3-2- المتوسط الحسابي للسلسلة الزمنية : يختلف المتوسط الحسابي للسلسلة الزمنية حسب نوع السلسلة كما يلي :

- المتوسط الحسابي للسلاسل الزمنية المترابطة (غير مستقلة):يتم حساب المتوسط الحسابي هنا باستخدام العلاقة التالية :

$$\bar{X} = (X_1/2 + X_2 + ...X_n/2)/(n-1)$$

مثـــال ( ) : لدينا السلسلة الزمنية التالية :

عدد العمال	العام
300	1990
310	1991
320	1992
350	1993
400	1994

والمطلوب حساب المتوسط الحسابي :

الحـــــل : نطبق العلاقة السابقة الذكر أعلاه فينتج لدينا أن :

$$X` = (300/2 \ + 310 + \ldots\ldots + 400/2)/4$$

$$= 1330/4 = 332.5$$

- المتوسط الحسابي للسلاسل الزمنية غير المترابطة (مستقلة) أيضا يتم حساب المتوسط الحسابي هنا باستخدام العلاقة التالية :

$$\bar{X} = (\sum X)/N$$

وهنا نجد ان حساب المتوسط الحسابي لهذا النوع سهل جدا وكأنها عينة من الأرقام وتتم عملية الحساب بالتطبيق المباشر كما شرح سابقا في فصل مقاييس النزعة المركزية.

3-3-4-مؤشرات السلاسل الزمنية: يقصد بمؤشرات السلاسل الزمنية تلك العمليات الحسابية التي تطبق على السلاسل الزمنية من أجل بيان خصائصها ،وتقسم هذه مؤشرات السلاسل الزمنية إلى الأنواع التالية :

1. مؤشرات أساسية : وهي المؤشرات التالية :
   i. النمو المطلق : وهي تعطى من خلال :رقم المقارنة – رقم الأساس
   ii. وتيرة النمو: وهي تعطى من خلال: رقم المقاربة÷ رقم الأساس ×100

iii. زيادة النمو : وهي تعطى من خلال: وتيرة النمو -100 حيث أن رقم الأساس هو أي رقم من السلسلة الزمنية ويفضل أن يكون الرقم الأول.

مثـــال ( ) : : لدينا السلسلة الزمنية التالية :

العام	الصادرات	النمو المطلق	وتيرة النمو	زيادة النمو
1990	200	-	100	-
1991	150	-50	75	-25
1992	350	150	175	75
1993	100	-100	50	-50
1994	400	200	200	100

2. مؤشرات متتالية: وهي المؤشرات التالية حيث تتم مقارنة كل عنصر مع الرقم الذي يسبقه:
   i. النمو المطلق : وهي تعطى من خلال :رقم المقارنة – الرقم الذي يسبقه
   ii. وتيرة النمو: وهي تعطى من : رقم المقاربة÷ الرقم الذي يسبقه ×100
   iii. زيادة النمو : وهي تعطى من خلال: وتيرة النمو - 100

مثـــال ( ) :نستعين بالمثال السابق ونقوم الآن بحساب المؤشرات المتتالية كما يلي في الجدول التالي :

العام	الصادرات	النمو المطلق	وتيرة النمو	زيادة النمو
1990	200	-	100	-
1991	150	-50	75	-25

1992	350	200	231	256
1993	100	-250	28	-203
1994	400	300	400	372

نعود ونذكر بانه يمكن للطالب أو القارئ مراجعة الكتاب العملي –الدرس العملي الرابع – للإطلاع على كيفية إجراء أو استخدام اختبار ت (t-test) لتوضيح كيفية حساب مؤشرات السلاسل الزمنية وتفسير النتائج حيث يتوفر المزيد من التفاصيل والأمثلة المناسبة.

4-3-4- تعديل السلاسل الزمنية TIME SERIES :نظرا لان السلاسل الزمنية تتعرض لعدد من المؤثرات الخارجية لذا يلزم إجراء تعديل للسلاسل الزمنية .من هذه التعديلات نورد ما يلي :

- التعديل البسيط : في هذا النوع من التعديل يتم تقسيم السلسلة إلى عدة أقسام أو فترات متساوية ويحسب المتوسط الحسابي لكل فترة ومن ثم يتم دراسة تطور السلسلة وتحديد الاتجاه العام لها .

مثـــال ( ) : لدينا السلسلة الزمنية التالية :

العام	الصادرات	المجموع لثلاث سنوات	المتوسط لثلاث سنوات
1985	40		
1986	45	126	42
1987	41		
1988	50		
1989	45	150	50
1990	55		
1991	51		

1992	60	179	58.6
1993	65		
1994	70		
1995	75	219	73
1996	74		

نلاحظ من الجدول أعلاه أن المتوسط الحسابي يتزايد مع الزمن وهذا يدل على أن الاتجاه العام تصاعدي.

- طريقة المتوسط المنزلق :تستخدم هذه الطريقة عندما يكون تقسيم السلسلة إلى عدة أقسام أو فترات متساوية غير ممكن وبالتالي نلجأ إلى طريقة المتوسط المنزلق وهي تتم من خلال حساب المتوسط الحسابي لفترة ما ولتكن (3-6) سنوات ثم نحذف العنصر الأول من القسم الأول أو الفترة الاولى ونضيف عنصر جديد من السلسلة وهكذا ..أخيرا يتم دراسة تطور السلسلة وتحديد الاتجاه العام لها .

مثـــال ( ) :لدينا السلسلة الزمنية التالية :

العام	الصادرات	المجموع لثلاث سنوات	المتوسط لثلاث سنوات
1988	3500		
1989	4000	11600	3866.6
1990	4100	12300	4100
1991	4200	12200	4066.6
1992	3900	12400	4133.3
1993	4300	12200	4066.6

1994	4000	12800	4266.6
1995	4500		

- نلاحظ من الجدول أعلاه أن المتوسط الحسابي يتقلب بين الزيادة والنقصان مع الزمن وهذا يدل على أن الاتجاه العام يتخذ شكلا منحنيا ومتذبذبا .
- 4-3-5- التحليل الاقتصادي للسلاسل الزمنية TIME SERIES :
- يعتمد التحليل الاقتصادي للسلاسل الزمنية على معرفة القوى المؤثرة فيها ،فالسلاسل الزمنية تخضع لتغيرات دورية شبه منتظمة تعكس تأثير عوامل مختلفة .وتصنف القوى المؤثرة في السلاسل الزمنية بشكل عام ضمن الفئات التالية:
  1. قوى الاتجاه العام
  2. التغيرات الموسمية
  3. التغيرات الدورية
  4. القوى العشوائية أو العرضية.

  وفيما يلي سنشرح طرق تحديد هذه القوى بشيء من التفصيل.

1- الاتجاه العام : ان استخراج معادلة الاتجاه العام يماثل حساب معاملات معادلة الانحدار البسيط من الشكل التالي :

$$Y = a + b*X$$

حيث أن : Y القيمة النظرية للصفة المختارة للظاهرة المدروسة

a الثابت للمعادلة الممثلة للسلسلة الزمنية

b عامل للمعادلة الممثلة للسلسلة الزمنية

X العامل المؤثر (إي المتغير)

أما كيفية حساب كل من a الثابت للمعادلة الممثلة للسلسلة الزمنية و b عامل للمعادلة الممثلة للسلسلة الزمنية فهي على الشكل التالي:

- حساب a الثابت لمعادلة الممثلة للسلسلة الزمنية : يمكن حساب a ثابت المعادلة الممثلة للسلسلة الزمنية باستخدام العلاقة الرياضية التالية:

$$a = \bar{y} - b * \bar{x}$$

- حساب b عامل لمعادلة الممثلة للسلسلة الزمنية : : يمكن حساب b عامل لمعادلة الممثلة للسلسلة الزمنية باستخدام العلاقة الرياضية التالية:

$$b = \sum (x - \bar{x}) * (y - \bar{y} / \sum x - \bar{x})^2$$

- كما ويمكن حساب b عامل لمعادلة الممثلة للسلسلة الزمنية باستخدام العلاقة الرياضية التالية:

$$b = (\sum x * y - \sum x * \sum x / n) / \sum x^2 - (\sum x)^2 / n$$

ان المعادلة الممثلة للسلسلة الزمنية المستخدمة عادة هي من الشكل الخطي البسيط ،لكن في معظم الحالات فان هذا الشكل الخطي البسيط قد يكون غير مناسبا لتمثيل البيانات الإحصائية DATA للسلسلة الزمنية وبالتالي كيف يمكن ان نعرف هل هذا النموذج الرياضي للسلسلة الزمنية هل هو مناسب أم لا .ان هذا السؤال قد تمت الإجابة عليه في الفقرة السابقة (4-1-3) وهي تقدير جودة المعادلة الممثلة للسلسلة الزمنية وهذا الكلام ينطبق كليا على المعادلة الممثلة للسلسلة الزمنية ويمكن العودة إلى الفقرة المذكورة للاطاع على كيفية اختيار النموذج المناسب للمعادلة الممثلة للسلسلة الزمنية.

معلومات متقدمة

ADVANCED INFORMATION

معامل الارتباط CORREL

إرجاع معامل الارتباط لنطاقات خلايا مصفوفة 1 ومصفوفة 2. استخدم معامل الارتباط لتحديد العلاقة بين خاصيتين. على سبيل المثال، يمكنك فحص العلاقة بين متوسط درجة الحرارة في مكان واستخدام مكيفات الهواء.

بناء الجملة لاستخدامها في الحاسب:

CORREL(array1,array2)

مصفوفة 1 نطاق خلايا من القيم.

مصفوفة 2 نطاق خلايا ثاني من القيم.

تنويهات

- يجب أن تكون الوسائط أرقام، أو أسماء، أو مصفوفات، أو مراجع تحتوي على أرقام.
- إذا احتوت وسيطة صفيف أو مرجع على نص، أو قيم منطقية، أو خلايا فارغة، يتم تجاهل تلك القيم؛ وبالرغم من ذلك، يتم تضمين الخلايا التي تحتوي على قيمة الصفر (0).
- إذا كان مصفوفة 1 ومصفوفة 2 لهما رقمين مختلفين من نقاط البيانات، تقوم CORREL بإرجاع قيمة الخطأ #N/A.
- إذا كان أي من مصفوفة 1 أو مصفوفة 2 فارغاً، أو إذا كان ق (الانحراف المعياري) لقيمهما يساوي صفر، تقوم CORREL بإرجاع قيمة الخطأ #DIV/0!.

مثـــــــــــال :

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

B **A**

1	بيانات 1	بيانات 2
2	3	9
3	2	7
4	4	12
5	5	15
6	6	17

الصيغة	وصف (الناتج)
=CORREL(A2:A6,B2:B6)	معامل الارتباط الخاص بمجموعتي البيانات الموجودة أعلاه(0.997054)

التنبؤ **FORECAST**

لحساب أو التنبؤ بقيمة مستقبلية باستخدام قيم موجودة. تكون القيمة المتوقعة عبارة عن قيمة حل لقيمة س المعطاة. القيم المعطاة هي قيم س وقيم ص الموجودة، وقيم التنبؤ بالقيمة الجديدة باستخدام الانحدار الخطي. يمكنك استخدام هذه الدالة للتنبؤ بالمبيعات، ومتطلبات المخزون، واتجاهات السوق المستقبلية .

بناء الجملة لاستخدامها في الحاسب:

FORECAST(known_x's،known_y's،x)

X (س) نقطة البيانات التي تريد التنبؤ بقيمتها.

Known_y's (معطيات ص) صفيف أو نطاق البيانات التابع.

Known_x's(معطيات س) صفيف أو نطاق البيانات المستقل.

تنويهات

- إذا كانت x غير رقمية، تقوم FORECAST بإرجاع قيمة الخطأ !VALUE#.
- إذا كان known_y's و known_x's فارغين أو تحتويان على عدد مختلف من نقاط البيانات، تقوم FORECAST بإرجاع قيمة الخطأ N/A#.
- إذا كان تباين known_x's يساوي صفراً، تقوم FORECAST بإرجاع قيمة الخطأ !DIV/0 #.

مثـــــــــــال :

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

النمو (التزايد الأسي المتوقع) GROWTH

حساب باستخدام البيانات الموجودة. تقوم GROWTH بإرجاع قيم ص لسلسة من قيم س الجديدة التي تعينها باستخدام قيم س و قيم ص الموجودة. يمكنك أيضاً استخدام دالة ورقة العمل GROWTH لتناسب أحد المنحنيات الأسية لقيم س و قيم ص الموجودة.

بناء الجملة لاستخدامها في الحاسب :

GROWTH(**known_y's**,known_x's,new_x's,const)

Known_y's (معطيات ص) هي مجموعة من قيم y (ص) التي تعرفها بالفعل في العلاقة y = b*m^x.

- إذا كان مصفوفة Known_y's في عمود مفرد، يتم تفسير كل عمود لـ Known_x's كمتغير مفرد.
- إذا كان مصفوفة Known_y's في صف مفرد، يتم تفسير كل صف من Known_x's كمتغير مفرد.
- إذا كان أي من الأعداد في known_y's يساوي 0 (صفر) أو سالباً، تقوم GROWTH بإرجاع قيمة الخطأ !NUM#.

Known_x's (معطيات س) مجموعة اختيارية من قيم x (س) تَعرفها بالفعل في العلاقة y = b*m^x.

- يمكن للصفيف known_x's أن يتضمن مجموعة أو أكثر من المتغيرات. إذا تم استخدام متغير واحد فقط، يمكن أن تكون known_y's وknown_x's في أي شكل من النطاقات، طالما كان تتضمن أبعاد متساوية. إذا تم استخدام أكثر من متغير واحد، يجب أن تكون known_x's كمية موجهة (أي نطاق بارتفاع صف واحد أو بعرض عمود واحد).
- إذا تم حذف known_x's، يفترض أن تكون مصفوفة {3,2,1،...} الذي يكون بنفس حجم known_y's.

New_x's (قيم س الجديدة) هي قيم س الجديدة التي تريد من GROWTH إرجاعها لقيم ص المطابقة.

- يجب أن تتضمن new_x's عمود (أو صف) لكل متغير مستقل، تماماً مثل known_x's. إذا كانت known_y's في عمود مفرد، فيجب أن يكون عدد الأعمدة في known_x's وnew_x's متساوي. إذا كانت known_y's في صف مفرد، فيجب أن يكون عدد الصفوف في known_x's وnew_x's متساوي.
- إذا تم حذف new_x's، يفترض أن تكون نفس known_x's.
- إذا تم حذف كل من known_x's و new_x's، سيفترض أن يكونا مصفوفة {1،2،3،...} الذي يكون بنفس حجم known_y's.

Const (ثابت) قيمة منطقية تحدد ما إذا كان سيتم فرض الثابت b ليساوي 1.

- إذا كانت const تساوي TRUE أو محذوفة، يتم حساب b بالشكل المعتاد.
- إذا كانت const تساوي FALSE، يتم تعيين b تساوي 1 ويتم ضبط القيمة m بحيث y = m^x.

ملاحظات

- يجب إدخال الصيغ التي ترجع المصفوفات كصيغ مصفوفات بعد تحديد عدد الخلايا الصحيح.
- عند إدخال ثابت صفيف لوسيطة مثل known_x's، استخدم فواصل لفصل القيم في نفس الصف وفواصل منقوطة لفصل الصفوف.

مثـــــــــال :

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

يستخدم هذا المثال نفس البيانات المستخدمة في مثال LOGEST. توضح الصيغة الأولى القيم المناظرة للقيم المعطاة. تقوم الصيغة الثانية بتوقع قيم الأشهر التالية، إذا استمر الاتجاه الأسي.

	A	B	C
	الشهر	**الوحدات**	**الصيغة (الوحدات المناظرة(**
1	11	33,100	=GROWTH(B2:B7,A2:A7)
2	12	47,300	
3	13	69,000	
4	14	102,000	
5	15	150,000	
6	16	220,000	
7	**الشهر**	**الصيغة (الوحدات المتوقعة)**	
	17	=GROWTH(B2:B7,A2:A7, A9:A10)	
	18		

ملاحظة يجب إدخال الصيغة في المثال كصيغة صفيف. بعد نسخ المثال إلى ورقة عمل فارغة، حدد النطاق C2:C7 أو B9:B10 بادئاً بخلية الصيغة. اضغط على المفتاح F2، ثم اضغط CTRL+SHIFT+ENTER. إذا لم يتم إدخال الصيغة كصيغة صفيف، ستكون النتائج الوحيدة هي 32618.20377 و 320196.7184.

عامل الانحدار الخطي LINEST

حساب الإحصائيات لخط باستخدام طريقة "القيمة الصغرى لمجموع المربعات" لحساب خط مستقيم يناسب بياناتك بالشكل الأمثال، وإرجاع صفيف يصف الخط. نظراً لأن هذه الدالة تقوم بإرجاع صفيف من القيم، يجب إدخالها كصيغة صفيف.

معادلة الخط هي:

$y = mx + b$ أو

$y = m1x1 + m2x2 + ... + b$ (في حالة وجود نطاقات متعددة لقيم x)

حيث تكون قيم y التابعة هي دالة لقيم x المستقلة. وتعد قيم m مُعاملات مطابقة لكل قيمة من قيم x، وتكون b قيمة ثابتة. لاحظ أن y و x و m يمكن أن تكون كميات موجهة. ومصفوفة الذي تقوم LINEST بإرجاعه هو {mn,mn-1,...,m1,b}. يمكن أن تقومLINEST أيضاً بإرجاع إحصائيات انحدار إضافية.

بناء الجملة لاستخدامها في الحاسب :

LINEST(**known_y's**,known_x's,const,stats)

Known_y's (معطيات ص)   مجموعة من قيم y (ص) التي تعرفها مسبقاً في العلاقة $y = mx + b$.

- إذا كان مصفوفة Known_y's في عمود مفرد، يتم تفسير كل عمود لـ Known_x's كمتغير مفرد.
- إذا كان مصفوفة known_y's في صف مفرد، يتم تفسير كل صف لـ known_x's كمتغير منفصل.

Known_x's (معطيات س)   مجموعة اختيارية من قيم x (س) التي قد تعرفها مسبقاً في العلاقة $y = mx + b$.

- يمكن للصفيف known_x's أن يتضمن مجموعة أو أكثر من المتغيرات. إذا تم استخدام متغير واحد فقط يمكن أن تكون known_y's وknown_x's نطاقات من أي شكل، طالما كانت ذات أبعاد متساوية. إذا تم استخدام أكثر من متغير واحد، يجب أن تكون known_x's كمية موجهة (أي نطاق بارتفاع صف واحد أو بعرض عمود واحد).
- إذا تم حذف known_x's، يفترض أن تكون مصفوفة {1,2,3،...} الذي يكون بنفس حجم known_y's.

Const   (ثابت) قيمة منطقية تحدد ما إذا كان سيتم فرض الثابت b ليساوي 0 (صفر).

- إذا كانت const تساوي TRUE أو محذوفة، يتم حساب b بالشكل المعتاد.
- إذا كانت const تساوي FALSE، يتم تعيين b لتساوي صفر ويتم ضبط القيم m لتلائم $y = mx$.

Stats   (إحصائيات) قيمة منطقية تحدد ما إذا كان سيتم إرجاع إحصائيات انحدار إضافية.

- إذا كانت stats تساوي TRUE، تقوم LINEST بإرجاع إحصائيات الانحدار الإضافية، بحيث يكون مصفوفة الذي يتم إرجاعه هو {mn,mn-1,...,m1,b;sen,sen-1,...,se1,seb;r2,sey;F,df;ssreg,ssresid}.
- إذا كانت stats تساوي FALSE أو تم حذفها، تقوم LINEST بإرجاع معاملات m والثابت b فقط.

تكون إحصائيات الانحدار الإضافية كما يلي.

الإحصائية	**الوصف**
se1,se2,...,sen	قيم الخطأ المعيارية للمُعاملات.m1,m2,...,mn
Seb	قيمة الخطأ المعياري للثابت b (seb = #N/A عندما تكون const تساوي FALSE).
r2	مُعامل التحديد. يقوم بمقارنة قيم y المقدرة وقيم y الفعلية، وتتراوح قيمته من صفر إلى 1. إذا كانت 1، يوجد ارتباط تام في العينة - لا يوجد فرق بين قيمة y المقدرة وقيمة y الفعلية. ومن ناحية أخرى، إذا كان معامل التحديد صفر، لا تفيد معادلة الانحدار في التنبؤ بقيمة .y للحصول على معلومات حول كيفية حسابr2 ، راجع "تنويهات" في هذا الموضوع لاحقاً.
sey	الخطأ المعياري للتقدير.Y
F	الإحصاءF ، أو قيمة F التي تمت ملاحظتها. استخدم إحصاء F لتحديد ما إذا كانت العلاقة التي تمت ملاحظتها بين المتغيرات التابعة والمتغيرات المستقلة تحدث عشوائياً.
df	درجات الحرية. استخدم درجات الحرية لتساعدك في العثور على قيم F الحرجة في جدول إحصائي. قارن القيم التي تجدها في الجدول بالإحصاء Fالذي يتم إرجاعه بواسطة LINEST لتحديد مستوى الثقة للنموذج.
ssreg	انحدار مجموع المربعات.
ssresid	باقي مجموع المربعات.

تنويهات

- يمكنك وصف أي خط مستقيم بواسطة الميل وتقاطع Y:

الميل (m):
لإيجاد ميل خطٍ ما، يُكتب عادة m، خذ نقطتين على الخط، (x1,y1) و(x2,y2)؛ فيكون الميل مساوياً (y2 - y1)/(x2 - x1).

تقاطع Y ‏(b):
التقاطع y لخطٍ ما، يُكتب عادة b، هو قيمة y عند تقاطع الخط مع محور y.

معادلة الخط المستقيم هي y = mx + b. بمجرد معرفة قيم m وb، يمكنك حساب أية نقطة على الخط بواسطة تضمين قيمة y أو x في تلك المعادلة. يمكنك أيضاً استخدام الدالة TREND.

- إذا كان لديك متغير x مستقل واحد فقط، يمكنك الحصول على قيم الميل وتقاطع y مباشرةً باستخدام الصيغتين التاليتين:

الميل:
=INDEX(LINEST(known_y's,known_x's),1)

تقاطع Y:
=INDEX(LINEST(known_y's,known_x's),2)

- تعتمد دقة الخط الذي المحسوب بواسطة LINEST على درجة التبعثر في بياناتك. كلما كانت البيانات أكثر خطية، زادت دقة نموذج LINEST. تستخدم LINEST طريقة القيمة الصغرى لمجموع المربعات لتحديد الشكل الأمثل للبيانات.
- يمكن للدالتين LINEST وLOGEST أن تقوم بحساب الخط المستقيم أو المنحنى الأسي الذي يلائم بياناتك بشكل أفضل. على أي حال، يجب عليك أن تقرر أي النتيجتين أمثل للبيانات. يمكنك حساب TREND(known_y's,known_x's) للخط المستقيم، أو GROWTH (known_y's)، known_x's للمنحنى الأسي. تقوم هذه الدالات، دون وسيطة new_x's، بإرجاع صفيف للقيم y التي تم تكهنها على هذا الخط أو المنحنى في نقاط البيانات الحقيقية. يمكنك بعد ذلك مقارنة القيم التي تم التنبؤ بها بالقيم الحقيقية. قد تريد تخطيطها معاً لمقارنتها مرئياً.
- في تحليل الانحدار، يحسب Microsoft Excel الفرق التربيعي لكل نقطة بين قيمة y المقدرة لهذه النقطة وقيمة y الفعلية. يُسمى مجموع الفروق التربيعية هذه بباقي مجموع المربعات. ثم يحسب Microsoft Excel مجموع فرق المربعات بين قيم y الفعلية ومعدل قيم Y، الذي يُسمى إجمالي مجموع المربعات (انحدار مجموع المربعات + باقي مجموع المربعات). كلما صغر مجموع باقي المربعات، مقارنةً بإجمالي مجموع المربعات، كبرت قيمة معامل التحديد r2، وهو مؤشر على كيفية شرح المعادلة الناتجة من تحليل الانحدار للعلاقة بين المتغيرات.
- يجب إدخال الصيغ التي تقوم بإرجاع مصفوفات كصيغ مصفوفات.

- عند إدخال ثابت صفيف مثل known_x's كوسيطة، استخدم الفواصل لفصل القيم في نفس الصف والفواصل المنقوطة لفصل الصفوف. قد تختلف الأحرف الفاصلة وفقاً للإعدادات المحلية لديك في **الإعدادات الإقليمية** أو **الخيارات الإقليمية** في **لوحة التحكم**.
- لاحظ أن قيم y التي توقعتها بواسطة معادلة الانحدار قد لا تكون صالحة إذا كانت خارج نطاق قيم y المستخدمة لتحديد المعادلة.

المثال 1: الميل وتقاطع Y

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
	معطيات ص	**معطيات س**
1	1	0
2	9	4
3	5	2
4	7	3
5	**الصيغة**	**الصيغة**

=LINEST(A2:A5,B2:B5,,FALSE)

ملاحظة يجب إدخال الصيغة في المثال كصيغة صفيف. بعد نسخ المثال إلى ورقة عمل فارغة، حدد النطاق A7:B7 بدءاً بخلية الصيغة. اضغط على المفتاح F2، ثم اضغط على CTRL+SHIFT+ENTER. إذا لم يتم إدخال الصيغة كصيغة صفيف، تكون النتيجة الوحيدة 2.

عندما يتم الإدخال كمصفوفة، يتم إرجاع الميل (2) والتقاطع y (1).

مثال 2: انحدار خطي بسيط :

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
	الشهر	**المبيعات**
1	1	3100
2	2	4500
3	3	4400
4	4	5400
5	5	7500
6	6	8100
7	**الصيغة**	**وصف (الناتج)**
	=SUM(LINEST(B2:B7, A2:A7)*{9,1})	تقدير مبيعات الشهر التاسع(11000)

بشكل عام، SUM({m,b}*{x,1}) تساوي mx+b، وهي قيمة y المقدرة لقيمة x المعطاة. ويمكنك أيضاً استخدام الدالة TREND.

مثال 3: انحدار خطي متعدد

بفرض أن أحد المطورين التجاريين يفكر في شراء مجموعة من المباني الإدارية الصغيرة في منطقة تجارية منشأة.

يمكن للمطور استخدام تحليل انحدار خطي متعدد لتقدير قيمة مبنى إداري في منطقة معطاة استناداً إلى المتغيرات التالية.

المتغير	يشير إلى
y	القيمة المقدرة للمبنى الإداري
x1	مساحة الطابق بالقدم المربع
x2	عدد المكاتب
x3	عدد المداخل
x4	عمر المبنى الإداري بالسنوات

يفترض هذا المثال وجود علاقة ثابتة بين كل متغير مستقل (x1 وx2 و x3و x4) والمتغير التابع (y)، وهي قيمة المباني الإدارية في المنطقة.

يختار المطور عينة عشوائية مكونة من 11 مبنى إداري من بين 1.500 مبنى إداري ويحصل على البيانات التالية. "مدخل مفرد" تعني مدخل للاستقبال فقط.

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B	C	D	E
1 2	**مساحة الطابق(x1)**	**المكاتب (x2)**	**المداخل (x3)**	**العمر (x4)**	**القيمة المقدرة (y)**
3	2310	2	2	20	142,000
4	2333	2	2	12	144,000
5	2356	3	1.5	33	151,000

6	2379	3	2	43	150,000
7	2402	2	3	53	139,000
8	2425	4	2	23	169,000
9	2448	2	1.5	99	126,000
10	2471	2	2	34	142,900
11	2494	3	3	23	163,000
12	2517	4	4	55	169,000
	2540	2	3	22	149,000

صيغة

=LINEST(E2:E12,A2:D12,TRUE,TRUE)

ملاحظة يجب إدخال الصيغة في المثال كصيغة صفيف. بعد نسخ المثال إلى ورقة عمل فارغة، حدد النطاق A14:E18 بدءاً بخلية الصيغة. اضغط على المفتاح F2، ثم اضغط على CTRL+SHIFT+ENTER. إذا لم يتم إدخال الصيغة كصيغة صفيف، تكون النتيجة الوحيدة -234.2371645.

عند الإدخال كمصفوفة، يتم إرجاع إحصائيات الانحدار التالية. استخدم هذا الأساس لتعريف الإحصائية التي تريدها.

يمكن الحصول الآن على معادلة الانحدار المتعدد، y=m1*x1 + m2*x3 + m4*x4 +b، باستخدام القيم من الصف 14:

y = 27.64*x1 + 12,530*x2 + 2,553*x3+ 234.24*x4 + 52,318

يمكن للمطور الآن أن يقوم بتقدير قيمة مبنى إداري في نفس المنطقة مساحته 2.500 قدم مربع، ويتكون من ثلاثة مكاتب ومدخلين وعمره 25 سنة، باستخدام المعادلة التالية:

y = 27.64*2500 + 12530*3 + 2553*2 - 234.24*25 + 52318 = $158,261

أو يمكنك نسخ الجدول التالي إلى الخلية A21 من مصنف المثال.

مساحة الطابق(x1)	المكاتب (x2)	المداخل (x3)	العمر (x4)	القيمة المقدرة(y)
2500	3	2	25	=D14*A22 + C14*B22 + B14*C22 + A14*D22 + E14

يمكنك أيضاً استخدام دالة TREND لحساب هذه القيمة.

مثال 4: استخدام إحصائيات F وR^2

في المثال السابق، يكون معامل التحديد، أو r2، هو 0.099675 (انظر الخلية A17 في ناتج LINEST)، الذي يشير إلى وجود علاقة قوية بين المتغيرات المستقلة وسعر البيع. يمكنك استخدام إحصائية F لتحديد ما إذا كانت تلك النتائج، مع قيمة r2 المرتفعة هذه، قد حدثت عشوائياً.

افترض الآن أنه لا يوجد بالفعل علاقة بين المتغيرات، بل أنك قد رسمت بمجرد عينة نادرة من 11 مبنى إداري تؤدي إلى عرض علاقة قوية للتحليل الإحصائي. يُستخدم المصطلح "ألفا" لاحتمال خطأ استنتاج وجود علاقة.

توجد علاقة بين المتغيرات إذا كانت الإحصائية F الملاحظة أكبر من قيمة F الحرجة. يمكن الحصول على القيمة الحرجة F بالرجوع إلى جدول القيم الحرجة F في العديد من كتب الإحصاء. لقراءة الجدول، افترض اختبار منفرد، استخدم قيمة ألفا قدرها 0.05، وبالنسبة لدرجات الحرية (مختصرة في معظم الجداول كـ v1 وv2)، استخدم v1=k = 4 و v2 = n - (k + 1) = 11 - (4 + 1) = 6، حيث k هي عدد المتغيرات في تحليل الانحدار وn هي عدد نقاط البيانات. القيمة الحرجة F هي 4.53.

قيمة f الملاحظة هي 459.753674 (الخلية A18)، وهي فعلياً أكبر من القيمة الحرجة F التي تكون قدرها 4.53. لذلك، تكون معادلة الانحدار مفيدة في توقع القيمة المقدرة للمباني الإدارية في هذه المنطقة.

مثال 5: حساب الإحصائيات T

يحدد اختبار افتراضي آخر ما إذا كان كل معامل ميل مفيداً في تحديد القيمة المقدرة لمبنى إداري في المثال 3. مثلاً، لاختبار معامل العمر للأهمية الإحصائية، قم بقسمة 234.24-

(معامل ميل العمر) على 13.268 (الخطأ المعياري المقدر لمعاملات العمر في الخلية A15).
فيما يلي هو قيمة الملاحظة:

$$t = m4 \div se4 = -234.24 \div 13.268 = -17.7$$

إذا قمت بمراجعة جدول في دليل إحصاء، ستجد أن قيمة t الحرجة وحيدة الطرف، مع 6 درجات للحرية وألفا=0.05 تكون 1.94. ولأن القيمة المطلقة لـ t، وهي 17.7، أكبر من 1.94، يصبح العمر متغير مهم عند تقدير القيمة المقدرة لمبنى إداري. يمكن اختبار كل متغير من المتغيرات المستقلة الأخرى للأهمية الإحصائية بطريقة مماثلة. فيما يلي قيم t الملحوظة لكل من المتغيرات المستقلة.

المتغير	**قيمة t الملاحظة**
مساحة الطابق	5.1
عدد المكاتب	31.3
عدد المداخل	4.8
العمر	17.7

تحتوي كافة تلك القيم على قيم مطلقة أكبر من 1.94؛ لذلك فإن كافة المتغيرات المستخدمة في معادلة الانحدار مفيدة في توقع القيمة المقدرة للمباني الإدارية في هذه المنطقة.

ميل خط الانحدار الخطي TREND:

إرجاع قيم بطول اتجاه خطي. وملائمة خط مستقيم (باستخدام طريقة القيمة الصغرى لمجموع المربعات) للصفيفين known_y's (معطيات ص) وknown_x's (معطيات س). إرجاع القيم y على طول هذا الخط لصفيف new_x الذي حددته.

بناء الجملة لاستخدامها في الحاسب :

TREND(known_y's,known_x's,new_x's,const)

Known_y's (معطيات ص)   مجموعة من قيم y (ص) التي تعرفها مسبقاً في العلاقة

$$y = mx + b$$

- إذا كان مصفوفة Known_y's في عمود مفرد، يتم تفسير كل عمود في Known_x's كمتغير منفصل.
- إذا كان مصفوفة known_y's في صف مفرد، يتم تفسير كل صف في known_x's كمتغير منفصل.

Known_x's (معطيات س) مجموعة اختيارية من قيم x (س) التي قد تعرفها مسبقاً في العلاقة y = mx + b.

- يمكن للصفيف known_x's أن يتضمن مجموعة أو أكثر من المتغيرات. إذا تم استخدام متغير واحد فقط يمكن أن تكون known_y's وknown_x's نطاقات من أي شكل، طالما كان لديها أبعاد متساوية. إذا تم استخدام أكثر من متغير واحد، يجب أن تكون known_x's كمية موجهة (أي نطاق بارتفاع صف واحد أو بعرض عمود واحد).
- إذا تم حذف known_x's، يفترض أن تكون مصفوفة {1,2,3،...} الذي يكون بنفس حجم known_y's.

New_x's (معطيات س الجديدة) قيم x (س) الجديدة التي ترغب في أن تقوم TREND بإرجاع قيم y (ص) المناظرة لها.

- يجب أن تتضمن new_x's عمود (أو صف) لكل متغير مستقل، تماماً مثلما تفعل known_x's. إذا كانت known_y's في عمود مفرد، فيجب أن يكون عدد الأعمدة في known_x's وnew_x's متساوي. إذا كانت known_y's في صف مفرد، فيجب أن يكون عدد الصفوف في known_x's وnew_x's متساوي.
- إذا تم حذف new_x's، يفترض أن تكون مثل known_x's.
- إذا تم حذف كلاً من known_x's وnew_x's، يفترض أن تكونا مصفوفة {1,2,3،...} الذي يكون له نفس حجم known_y's.

Const (ثابت) قيمة منطقية تحدد ما إذا كان سيتم فرض الثابت b ليساوي 0 (صفر).

- إذا كانت const تساوي TRUE أو مهملة، يتم حساب b بالشكل المعتاد.
- إذا كانت const تساوي FALSE، يتم تعيين b لتساوي 0 (صفر)، ويتم ضبط القيم m لكي تكون y = mx.

تنويهات

- للحصول على معلومات حول كيفية قيام Microsoft Excel بملائمة خط ببيانات، انظر LINEST.
- يمكنك استخدام TREND لملائمة منحنى متعدد الحدود بإعادة حساب نفس المتغير مرفوعاً إلى قوى أسية مختلفة. على سبيل المثال، افترض أن العمود A يحتوي على قيم y والعمود B يحتوي على قيم x. يمكنك إدخال x^2 في العمود C، وx^3 في العمود D، وهكذا، ثم إعادة حساب العمود B خلال D بمقارنته بالعمود A.

- يجب إدخال الصيغ التي تقوم بإرجاع مصفوفات كصيغ مصفوفات.
- عند إدخال ثابت صفيف لوسيطة مثل known_x's، استخدم فواصل لفصل القيم في نفس الصف وفواصل منقوطة لفصل الصفوف.

مثال

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

تعرض الصيغة الأولى القيم المطابقة للقيم المعطاة. وتتنبأ الصيغة الثانية بقيم الشهر التالي، إذا استمر الاتجاه الخطي.

	A	B	C
1	**الشهر**	**التكلفة**	**الصيغة (التكلفة المطابقة)**
2	1	$133,890	=TREND(B2:B13, A2:A13)
3	2	$135,000	
4	3	$135,790	
5	4	$137,300	
6	5	$138,130	
7	6	$139,100	
8	7	$139,900	
9	8	$141,120	
10	9	$141,890	

11	10	$143,230
12	11	$144,000
13	12	$145,290

الشهر	**الصيغة (التكلفة المتوقعة)**
13	=TREND(B2:B13, A2:A13,A15:A19)
14	
15	
16	
17	

ملاحظة يجب إدخال الصيغة في المثال كصيغة صفيف. بعد نسخ المثال إلى ورقة عمل فارغة، حدد النطاق C2:C13 أو B15:B19 بدءاً بخلية الصيغة. اضغط على المفتاح F2، ثم اضغط على CTRL+SHIFT+ENTER. إذا لم يتم إدخال الصيغة كصيغة صفيف، تكون النتائج الوحيدة هي 133953.3333 و 146171.5152.

المنحنى الأسي LOGEST :

في تحليل الانحدار، تقوم بحساب أحد المنحنى الأسي الذي يلائم بياناتك وإرجاع صفيف قيم يصف المنحنى. ولأن هذه الدالة تقوم بإرجاع صفيف من القيم، يجب إدخالها بصيغة مصفوفة.

تكون المعادلة للمنحنى هي:

y = b*m^x

y = (b*(m1^x1)*(m2^x2)*_) (if there are multiple x-values)

حيث تكون قيمةy (ص) التابعة عبارة عن دالة من قيم x (س) المستقلة. وتكون قيم m عبارة عن أساسات تناظر كل أس لقيمة x. وتكون b قيمة ثابتة. لاحظ أنه يمكن أن تكون x و y و m كميات متجهة. ويكون مصفوفة الذي تقوم LOGEST بإرجاعه هو {mn,mn-1,...,m1,b}.

بناء الجملة لاستخدامها في الحاسب :

LOGEST(known_y's,known_x's,const,stats)

Known_y's (معطيات ص) مجموعة قيم y (ص) التي تعرفها بالفعل في العلاقة y = b*m^x.

- إذا كان مصفوفة Known_y's في عمود مفرد، يتم تفسير كل عمود لـ Known_x's كمتغير مفرد.
- إذا كان مصفوفة known_y's في صف مفرد، يتم تفسير كل صف لـ known_x's كمتغير منفصل.

Known_x's (معطيات س) مجموعة اختيارية من قيم x (س) تَعرفها بالفعل في العلاقة y = b*m^x.

- من الممكن أن يتضمن مصفوفة known_x's مجموعة أو أكثر من المتغيرات. في حالة استخدام متغير واحد فقط، فمن الممكن أن يكون known_y's وknown_x's عبارة عن نطاقات لأي شكل طالما أن أبعادهما متساوية. وفي حالة استخدام أكثر من متغير واحد، يجب أن تكون known_y's نطاقاً من الخلايا ارتفاعه صف واحد أو عرضه عمود واحد (والذي يعرف أيضاً بكمية موجهة).
- إذا تم حذف known_x's، يفترض أن تكون مصفوفة {1,2,3،...} الذي يكون بنفس حجم known_y's.

Const (ثابت) قيمة منطقية تحدد ما إذا كان سيتم فرض الثابت b ليساوي 1 (صفر).

- إذا كانت const تساوي TRUE أو مهملة، يتم حساب b بالشكل المعتاد.
- إذا كانت const تساوي FALSE، يتم تعيين b لتساوي 1، وتتلاءم قيم m مع y = m^x.

Stats (إحصائيات) قيمة منطقية تحدد ما إذا كان سيتم إرجاع إحصائيات انحدار إضافية.

- إذا كانت stats تساوي TRUE، تقوم LOGEST بإرجاع إحصائيات الانحدار الإضافية، وبذلك يكون مصفوفة الذي تم إرجاعه هو {mn,mn-1,...,m1,b;sen,sen-1,...,se1,seb;r 2,sey; F,df;ssreg,ssresid}.
- إذا كانت stats تساوي FALSE أو محذوفة، تقوم LOGEST بإرجاع معاملات m فقط والثابت b.

للحصول على مزيد من المعلومات حول إحصائيات الانحدار الإضافية، انظر LINEST.

تنويهات

- كلما كان شكل رسم بياناتك أكثر تمثيلاً للمنحنى الأسي، كلما زادت ملائمة الخط المحسوب لبياناتك. ومثل LINEST، تقوم LOGEST بإرجاع صفيف قيم يصف العلاقة بين القيم، لكن تقوم LINEST بملائمة بياناتك على خط مستقيم؛ بينما تقوم LOGEST بملائمة بياناتك على منحنى أسي. لمزيد من المعلومات، انظر LINEST.
- عندما يكون لديك متغير x مستقل واحد فقط، يمكنك الحصول على قيم الميل (m) والتقاطع y مباشرةً مع قيم b باستخدام الصيغ التالية:

الميل (m):
INDEX(LOGEST(known_y's,known_x's),1)

تقاطع y مع (b) :
INDEX(LOGEST(known_y's,known_x's),2)

يمكنك استخدام المعادلة y = b*m^x لتوقع القيم المستقبلية لـ y، لكن يوفر Microsoft Excel الدالة GROWTH للقيام بذلك نيابةً عنك. لمزيد من المعلومات، انظر GROWTH.

- يجب إدخال الصيغ التي تقوم بإرجاع مصفوفات كصيغ مصفوفات.
- عند إدخال ثابت صفيف مثل known_x's كوسيطة، استخدم الفواصل لفصل القيم في نفس الصف والفواصل المنقوطة لفصل الصفوف. قد تختلف الأحرف الفاصلة وفقاً للإعدادات المحلية لديك في **الإعدادات الإقليمية** أو **الخيارات الإقليمية** في **لوحة التحكم**.
- يجب ملاحظة أن قيم ص التي تم توقعها بواسطة معادلة الانحدار قد تكون غير صالحة إذا كانت خارج نطاق قيم ص الذي استخدمته لتحديد المعادلة.

مثال 1 معاملات m والثابت b

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
1	**الشهر**	**الوحدات**
2	11	33,100

3	12	47,300
4	13	69,000
5	14	102,000
6	15	150,000
7	16	220,000

صيغة **الصيغة**

=LOGEST(B2:B7,A2:A7, TRUE, FALSE)

ملاحظة يجب إدخال الصيغة في المثال كصيغة صفيف. بعد نسخ المثال إلى ورقة عمل فارغة، حدد النطاق A9:B9 بدءاً بخلية الصيغة. اضغط على المفتاح F2، ثم اضغط على CTRL+SHIFT+ENTER. إذا لم يتم إدخال الصيغة كصيغة صفيف، تكون النتيجة الوحيدة 1.463275628.

عندما يتم الإدخال كمصفوفة، يتم إرجاع معاملات m والثابت b.

y = b*m1^x1 أو استخدام القيم من مصفوفة:

y = 495.3 * 1.4633x

يمكنك تقدير المبيعات للأشهر المستقبلية عن طريق تبديل رقم الشهر بـ x في هذه المعادلة، أو يمكنك استخدام الدالة GROWTH.

مثال 2 إحصائيات كاملة

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
	الشهر	**الوحدات**
1	11	33,100
2	12	47,300
3	13	69,000
4	14	102,000
5	15	150,000
6	16	220,000
7	**صيغة**	

=LOGEST(B2:B7,A2:A7, TRUE, TRUE)

ملاحظة يجب إدخال الصيغة في المثال كصيغة صفيف. بعد نسخ المثال إلى ورقة عمل فارغة، حدد النطاق A9:B13 بدءاً بخلية الصيغة. اضغط على المفتاح F2، ثم اضغط على CTRL+SHIFT+ENTER. إذا لم يتم إدخال الصيغة كصيغة صفيف، تكون النتيجة الوحيدة 1.463275628.

عند الإدخال كمصفوفة، يتم إرجاع إحصائيات الانحدار التالية. استخدم هذا الأساس لتعريف الإحصائية التي تريدها.

يمكنك استخدام إحصائيات الانحدار الإضافية (الخلايا A10:B13 في صفيف الإخراج أعلاه) لتحديد مدى فائدة المعادلة في توقع القيم المستقبلية.

هام تشبه الأساليب التي تستخدمها لاختبار معادلة ما باستخدام LOGEST تلك الأساليب المستخدمة في LINEST. وبالرغم من ذلك، تستند الإحصائيات الإضافية التي يتم إرجاعها بواسطة LOGEST إلى النموذج الخطي التالي:

$$\ln y = x1 \ln m1 + \ldots + xn \ln mn + \ln b$$

يجب أن تتذكر هذا دائماً عندما تقيّم الإحصائيات الإضافية، خاصةً القيمتين sei و seb، واللتين يجب مقارنتهما بـ ln mi وln b، وليس بـ mi وb. للحصول على مزيد من المعلومات، راجع دليل الإحصائيات المتقدمة.

ميل خط الانحدار الخطي: **SLOPE**

إرجاع ميل خط الانحدار الخطي المار خلال نقاط البيانات في known_y's (معطيات ص) و known_x's (معطيات س). الميل هو ناتج قسمة مقدار المسافة العمودية على المسافة الأفقية بين أي نقطتين على الخط، والذي يمثل معدل التغيير على الانحدار الخطي.

بناء الجملة لاستخدامها في الحاسب :

SLOPE(known_y's,known_x's)

Known_y's (معطيات ص) هى صفيف أو نطاق من الخلايا يحتوي على نقاط بيانات رقمية تابعة.

Known_x's (معطيات س) هى مجموعة من نقاط البيانات المستقلة.

تنويهات

- يجب أن تكون الوسائط إما أرقام، أو أسماء، أو مصفوفات، أو مراجع تحتوي على أرقام.
- إذا احتوت وسيطة صفيف أو مرجع على نص، أو قيم منطقية، أو خلايا فارغة، يتم تجاهل تلك القيم؛ رغم ذلك يتم تضمين الخلايا التي تحتوي على القيمة صفر.
- إذا كانت known_y's وknown_x's خالية أو بها عدد مختلف لنقاط البيانات، تُرجع SLOPE قيمة الخطأ #N/A.

مثـــــــــــــال :

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
1	**معطيات ص**	**معطيات س**

2	2	6
3	3	5
4	9	11
5	1	7
6	8	5
7	7	4
8	5	4

الصيغة	**الوصف (الناتج)**
=SLOPE(A2:A8,B2:B8)	ميل خط الانحدار الخطي المار خلال نقاط البيانات أعلاه (0.305556)

الخطأ المعياري: STEYX

إرجاع الخطأ المعياري لقيمة ص المتوقعة لكل قيمة س في خط الانحدار. الخطأ المعياري هو مقياس لمقدار الخطأ عند توقع قيمة ص لقيمة س الفردية.

بناء الجملة لاستخدامها في الحاسب :

STEYX(**known_x's**،**known_y's**)

Known_y's (معطيات ص) هى صفيف أو نطاق من نقاط البيانات التابعة.

Known_x's (معطيات س) هى صفيف أو نطاق من نقاط البيانات المستقلة.

تنويهات :

- يجب أن تكون الوسائط إما أرقاماً أو أسماء، أو مصفوفات، أو مراجع تحتوي على أرقام.

- إذا احتوت وسيطة صفيف أو مرجع على قيم نصية أو منطقية أو خلايا فارغة، يتم تجاهل تلك القيم؛ رغم ذلك يتم تضمين الخلايا التي تحتوي على القيمة صفر.
- إذا كانت Known_y's وKnown_x's فارغة أو بها عدد مختلف من نقاط البيانات، تقوم STEYX بإرجاع قيمة الخطأ N/A#.
- معادلة الخطأ المعياري للقيمة ص المتوقعة هي:

مثـــــــــال :

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
	معطيات ص	**معطيات س**
1	2	6
2	3	5
3	9	11
4	1	7
5	8	5
6	7	4
7	5	4
8	**الصيغة**	**الوصف (الناتج)**
	=STEYX(A2:A8,B2:B8)	الخطأ المعياري لقيمة ص المتوقعة لكل قيمة س في خط الانحدار (3.305719)

تمارين غير محلولة

3. لدينا البيانات التالية التي تبين تطور الإنتاج من أحد المحاصيل الزراعية وهو القمح خلال الفترة 1998-2001:

السنة	1998	1999	2000	2001
الإنتاج	400	456	500	680

والمطلوب :

- ارسم المنحنى البياني للبيانات السابقة باستخدام الرسوم البيانية الخطية.
- احسب معدل الإنتاج من محصول الشعير خلال الفترة المحددة من 1998وحتى 2001 .
- احسب المؤشرات الإحصائية اللازمة التي تم شرحها في الفصل السابق الثاني والثالث.

- طبق تحليل الانحدار على البيانات الواردة في الجدول السابق الخطي واختبر معنوية عامل الانحدار .
- استنتج النموذج الرياضي المناسب للبيانات السابقة.
- هل البيانات المذكورة تمثل سلسلة زمنية،وما هي مؤشراتها؟

4. لدينا البيانات التالية التي تبين تطور عدد الطلاب في الجامعة و عدد المدرسين خلال الفترة 1998 - 2001:

السنة	1999	2000	2001	2002
عدد الطلاب	50000	60000	50000	30000
عدد المدرسين	200	270	220	210

والمطلوب:

- مثل البيانات السابقة في الجدول أعلاه بالطريقة البيانية المناسبة.
- احسب المؤشرات الإحصائية اللازمة لعدد الطلاب التي تم شرحها في الفصل السابق لكلا العينتين.
- الانحدار المتعدد على البيانات الواردة في الجدول السابق واختبر معنوية عامل الانحدار .
- طبق تحليل الارتباط لدراسة مدى الارتباط بين عدد الطلاب و عدد المدرسين
- احسب عامل التحديد وفسره.
- احسب المؤشرات الخاصة بالسلسلة الزمنية لكل من عدد الطلاب و عدد المدرسين.
- اجر التعديلات المناسبة على السلسلة الزمنية لكل من عدد الطلاب و عدد المدرسين مع الشرح.
- ارسم المنحنى البياني للبيانات السابقة باستخدام الرسوم البيانية الخطية واستنتج معادلة الانحدار المناسبة من الرسم.
- استنتج مؤشرات جودة النموذج الرياضي للبيانات السابقة لكل من عدد الطلاب و عدد المدرسين .

العينات Samples

أولاً : مقدمة :

رأينا عند دراستنا للأبحاث السابقة من مختلف نواحيها ، بأننا ندرس جزءاً من الظاهرة أو الظاهرة ككل دون أن نعلل وجود تطابق أو اختلاف بين الكل والجزء ، وفي بعض الأحيان لا يستطيع الباحث دراسة مفردات الظاهرة الداخلة في البحث جميعها ، لذلك نلجأ إلى بحث العينات التي تعدّ أسلوباً من الأساليب الرئيسة في تحليل معطيات المراقبة الإحصائية ، وفي السنوات الأخيرة تقدّمت البحوث الإحصائية تقدّماً كبيراً نتيجة تطبيق نظرية العينات على هذه البحوث ، وأصبح لها فروع مستقلّة وكبيرة في علم الإحصاء النظري والتطبيقي، فعند دراسة المجتمع يتحتّم علينا أن نكتفي بفحص جزء منه ، فالذي يريد أن يأخذ كمية من الدواء لا يتفحّص كل قطرة منه ، ولكنه يفحص كمية صغيرة من هذا الدواء ، واختبار جزء من المجتمع لدراسة يعرف بالمعاينة والجزء المختار يعرف بالعينة ، وتعد طريقة المعاينة العشوائية أبسطّ وأدقها طريقة للمعاينة ، حيث أن اختيار أفراد العينة يبنى على محض الصدفة ، ويشترط في العينة المختارة من المجتمع أن تتّصف بصفتين :

1- أن تكون متّسقة : ويعني ذلك أن الوسط الحسابي للعينة ($\overline{X}$) يقترب من الوسط الحسابي للمجتمع (M) ويتحقق ذلك كلّما كبر حجم العينة .

2- أن تكون غير متميّزة : فإذا أخذنا من المجتمع عدّة عيّنات فإن الوسط الحسابي لأوساط هذه العينات يكون مساوياً للوسط الحسابي للمجتمع (M) .

- **اختبارات المعنوية :**

لكي نتمكّن من تعميم نتائج البحث يجب أن نتأكّد من صحّتها ، ويتم ذلك بوضع النظرية الفرضية لاختبار مدى الاعتماد على الإحصاءات المقدّرة من التجربة أو العيّنة ثم تختبر إحصائياً وبنتيجة هذه الاختبارات تقبل النظرية الفرضية أو ترفض .

وللحكم على صحّة النظرية الفرضية من عدمها يجري اختبار المعنوية لمحاولة معرفة هل الفروق بين العيّنات أو المعاملات هي فروق عشوائية بين عيّنات المجتمع الواحد ، أم أن الفروق بين العيّنات أو المعاملات هي في الحقيقة فروق

أساسية ناتجة عن تأثير المعاملات المختلفة ، وأن تلك العينات مأخوذة من مجتمعين مختلفين ، وفي الحالة الأخيرة ترفض النظرية الفرضية .

فعند فحص النظرية الفرضية (H_1) ، نقول إن هذه الفرضية خطأ إذا كانت قيمة فحص الاحتمال المحسوبة أقل من مستوى الدلالة المفروض ، نقول إن النتيجة ذات دلالة إحصائية ، أما إذا كانت قيمة الاحتمال المحسوبة أكبر من مستوى الدلالة المفروض فإننا نقول بصحّة النظرية الفرضية ، و وإن النتيجة ليست بذات دلالة .

والمتّبع في الإحصاء غالباً مستويان للدلالة 5% و 1% .

- **ثانياً اختبار مربّع كاي (X^2) Chi Square :**

تقسم البيانات التي يحصل عليها الباحث من التجارب إلى قسمين :

أ- القياسات Measurement Date:

وتعبّر القياسات عن البيانات التي يحصل عليها الباحث عن طريق قياس أفراد المتغيّر العشوائي لصفة ما كالطول أو الوزن أو كمية المحصول ، وتكلّمنا عن طريقة جمع البيانات وكيفية تبويبها ثم تحليلها في الفصول السابقة .

ب- التعدادات Enumeration date :

يعبّر التعداد عن البيانات التي يحصل عليها الباحث عن طريق تسجيل عدد الأفراد أو عدد القياسات أو التكرارات التي تقع في قسم أو فئة معيّنة ، وذلك كما في حالة عمل جداول التوزيع التكراري ، حيث يقوم الباحث بتقسيم الصفة المدروسة إلى فئات أو أقسام أو مجموعات يقع كل منها داخل مدى معيّن ، ثم حصر عدد الأفراد أو تكرار الأفراد الذي يقع في كل فئة أو قسم منها ، وسنتناول في هذا الفصل طريقة اختبارات البيانات العددية ، ويعد الإحصاء المسمى اختبار مربع كاي لحسن المطابقة أو اختبار التطابق النسبي من أهم الطرق التي تستعمل في مقارنة مجموعة من النتائج المشاهدة أو المستحصل عليها من تجربة حقيقية بمجموعة أخرى فرضية وضعت على أساس النظرية الفرضية التي يُراد اختبارها ، وتعتمد هذه الطريقة على افتراض وجود عيّنة عشوائية بها عدد N من الأفراد ، قسّمت إلى عدد من الفئات المتشابهة بحيث يقع كل فرد في العينة في إحدى هذه الفئات ، ثم مقارنة التكرارات المشاهدة بالتكرارات النظرية بقصد معرفة مدى انطباق التكرارات المشاهدة على تلك النظرية ، وذلك باستعمال اختبار مربع كاي .

- **اختبار المعنوية بوساطة مربع كاي :**

يستخدم اختبار مربع كاي أحياناً لاختبار حسن المطابقة والتوفيق بين البيانات التي تحصلنا عليها وبين منحني أو معادلة معينة ، فإذا كان لدينا بيانات والمطلوب اختبارها إذا كانت هذه البيانات تتّفق مع معادلة وراثية معروفة فإنه يمكن حساب قيمة مربع كاي من المعادلة الآتي :

$$v^2 = \sum \frac{(O-E)^2}{E}$$

حيث : O : التكرار المشاهد أو القيم النظرية .

E : التكرار المتوقّع .

وتدل هذه المعادلة على أن مربع كاي عبارة عن مجموع مربعات الفروق بين التكرارات المشاهدة مقسوماً على التكرارات المتوقعة .

أمثلة :

أ- مربع كاي لدرجة حرية واحدة :

تعتمد المقارنة هنا على مجموعتين أو فئتين فقط ، ولذلك تكون درجات الحرية في هذا النوع من المقارنة عبارة عن درجة حرية واحدة فقط .

مثال :

قام أحد الأشخاص باختبار قدرة أحد المبيدات الحشرية على قتل نوع معين من الحشرات، وكانت التعليمات المرفقة بهذا المبيد تفيد بأن المبيد يمكنه أن يقتل 80% من عدد الحشرات بعد رشّها به ، ولاختبار ما إذا كانت هذه التعليمات تنطبق على الظروف العملية أو المشاهدة بالحقل ، أجريت تجربة حقيقية رشّ فيها المبيد على النباتات التي تحوي عدداً معروفاً من الحشرات ، ثم حصرت عدد الحشرات الحية والميتة ، فكانت النتائج كالتالي ، حشرات حية عددها 100 ، حشرات ميتة وعددها 300 ، فهل تُشير هذه النتائج إلى صحة التعليمات المرفقة بالمبيد ؟

الحل :

لحل هذه المسألة نضع النظرية الفرضية التي تقول بصحة التعليمات المرفقة بالمبيد ، أي أنه يقتل 80% من الحشرات ثم تقوم بحساب التكرارات النظرية والمتوقعة كالتالي :

عدد الحشرات الحية بعد استعمال المبيد : $400.\frac{20}{100} = 80$

عدد الحشرات الميتة بعد استعمال المبيد : $400.\frac{80}{100} = 320$

المشاهدات	التكرارات المتوقعة O	التكرارات النظرية E	$\frac{(O-E)^2}{E}$
الحشرات الحية	100	80	5
الحشـــــــــرات الميتة	300	320	1.25
المجموع	400	400	6.25

بالنظر إلى جدول K^2 عند احتمال 5% ودرجة حرية 1 نجد أن تساوي 3.84 ، وبما أن قيمة معامل كاي مربع أكبر من الجدولية ، لذا ترفض النظرية الفرضية ، أي أن النسبة المشاهدة في الحقل لا تنطبق على النظرية الموجودة في التعليمات المرفقة مع المبيد (80%) .

مثال2 :

حصل باحث في مجال وراثة النباتات على البيانات التالية : – 101 – 315 32 – 108 وافترض أنها تتبع نسبة الأشكال المظهرية : 1 :3 :3 :9، فهل تتّفق هذه البيانات مع النظرية المذكورة ؟

نوجد القيم المتوقعة بضرب المجموع الكلي في النسبة الخاصة بكل نوع على الترتيب .

المشاهدة O	المتوقع E	O - E	$\frac{(O-E)^2}{E}$
315	312.8	2.2	0.02
101	104.3	-3.3	0.10

0.13	**3.7**	**104.3**	**108**
0.20	**-2.6**	**34**	**32**
0.45	**0**	**556**	**556**

إذاً $K^2 = 0.45$ ، درجة الحرية 3 = 1 – 4 .

من الجدول نجد أن عند 5% = 7.81 .

نجد أن K^2 المحسوبة أقل من K^2 الجدولية ، ولذلك فهي غير معنوية ، أي أن قيمة K^2 التي حصلنا عليها من هذه البيانات لا تسمح برفض النظرية الفرضية ، ونستنتج أن البيانات تتوافق مع النسبة 1 :3 :3 :9 .

- **اختبار العلاقة بين صفتين :**

إذا قمنا بفحص كل فرد في المجتمع لخاصتين معينتين ، وكل خاصة صنّفت إلى مجموعات فإننا قد نرغب في معرفة إذا كانت هذه الخصائص مستقلّة بعضهاعن بعضها الآخر ، فمثلاً إذا كان لدينا عدد من النباتات في خط طويل ، وتمّت دراستها من حيث لون الأوراق ومعدّل النمو ، يمكننا وضع النظرية الفرضية الدالّة على أن لون الأوراق ومعدّل النمو مستقلان بعضهما عن بعضهما الآخر ، ومن ثم نُخضع هذه النظرية للاختبار ، وبالمثل إذا كان لدينا مجموعة من الطلاب وقسّمناها حسب لون الشعر ولون العيون ، ونريد أن نعرف إن كانت هناك علاقة ما بين لوني الشعر والعيون ، ففي هذه الحالات جميعها يمكننا الاستفادة من اختبار مربع كاي لإجراء الاختبار .

مثال:

الجدول الآتي يبيّن توزيع 250 بادرة من بادرات الفول البلدي حسب معدّل النمو ولون الأوراق ، باستخدام هذه البيانات اختبر إمكانية وجود علاقة بين الصفتين .

لون الشعر	معدل النمو		ضعيف	المجموع
	جيد	مقبول		
أخضر	55	79	4	138
أخضر مصفر	11	60	15	86
أصفر	1	6	19	26
المجموع	67	145	38	250

والاتحادات المتوقعة من هذه البيانات إذا كانت الصفتان مستقلتين هي :

أخضر جيد : 37 = 250 / (167 . 138)

أخضر مقبول: 80 = 250 / (145 . 138)

أخضر /ضعيف: 21 = 250 / (38 . 138)

أخضر مصفر/جيد: 23 = 250/ (67 . 86)

أخضر مصفر/مقبول: 49.9 = 250/ (145 . 86).

أخضر مصفر/ضعيف: 13.1 = 250 / (38 . 86) .

أصفر جيد : 7 =250 / (67 . 26)

أصفر مقبول: 15.1 = 250/ (145 . 26)

أصفر ضعيف : 3.9 = 250 / (38 . 26) .

نضع البيانات الآتية في جدول :

المجموع	معدل النمو			لون الأوراق
	ضعيف	مقبول	جيد	
138	21	80	37	أخضر
86	13.1	49.9	23	أخضر ـ مصفر
26	3.9	15.1	7	أصفر
250	38	145	67	المجموع

عدد الصفوف في الجدول (r) = 3 ، وعدد الأعمدة في الجدول C = 3 .

ودرجات الحرية 4 = 2 . 2 = (C-1) (r-1) .

مربع كاي المحسوب :

$$K^2 = \frac{(55-37)^2}{37} + \frac{(79-80)^2}{80} + \frac{(4-21)^2}{21} + \frac{(11-23)^2}{23} + \frac{(60-49.9)^2}{49.9} + \frac{(15-13.1)^2}{13.1} + \frac{(1-7)^2}{7} + \frac{(6-15.1)^2}{15.1} + \frac{(19-3.9)^2}{3.9} = \frac{(18)^2}{37} + \frac{(1)^2}{80} + \frac{(17)^2}{21} + \frac{(12)^2}{23} + \frac{(10.1)^2}{49.9} + \frac{(1.9)^2}{13.1} + \frac{(6)^2}{7} + \frac{(9.1)^2}{15.1} + \frac{(15.1)^2}{3.9} = 100.2$$

ثم نستخرج قيمة K^2 من الجدول عند درجة حرية 4 ومعنوية 5% ، وهي 9.48 ، نجد أن K^2 المحسوبة أكبر من K^2 الجدولية ، فمعنى ذلك نرفض النظرية الفرضية القائلة باستقلال العينتين بمستوى معنوية 5% ، أي نستنتج أن هناك علاقة بين الصفتين .

- **الخطوات الأساسية لبحث العينة :**

عند إجراء البحوث الإحصائية بوساطة العيّنات ، على الباحث أن يكون ملمّاً بأهمية البيانات المطلوبة ، وكيفية استخدامها ، والخطوات المتّخذة عند تصميم بحث العينة :

1- تعريف المشكلة ، وضع الأسئلة التي يجب أن نجد لها أجوبةً من خلال البحث .

2- تحديد المجتمع الإحصائي المراد معاينته والمفردات الداخلة فيه .

3- نتائج الأبحاث السابقة لهذه المشكلة .

4- تحديد البيانات المطلوبة ، هل هي عن الحبوب ، أو الأبقار ، أو الأغنام .

5- تحديد طريقة جمع البيانات : الاتّصال المباشر أو غير المباشر ، أو كليهما أو الطريقة الميدانية ، أو السجلاّت الرسمية .

6- تحديد وحدة القياس : هل هي الهكتار ، أو الفدان ، أو الدونم ، والكغ... إلخ.

7- تكوين إطار العينة : مثلاً تحديد المساحة المزروعة بالقمح في القطر العربي السوري يكون إطار العينة هو أراضي القطر العربي السوري المزروعة بالقمح كافة .

8- اختيار وحدة المعاينة : الإنسان ، أو الأسرة ، ....

9- اختيار نوع العيّنة : سنتكلّم عنها لاحقاً .

10-تهيئة العناصر البشرية والمادية، مثل تدريب العدادين والباحثين، وغير ذلك.

11-تحليل نتائج البيانات وإعلان النتائج بعد التدقيق والمراجعة .

رابعاً : دراسة العينات الكبيرة Large Samples:

ذكرنا سابقاً أن يشترط في العيّنة المختارة من المجتمع أن تتّصف بصفتين : متّسقة ، وغير متحيّزة . فإذا كان عدد المتغيّرات المأخوذة لعيّنة من مجتمع يساوي أو أكبر من 30 متغيّر فإن الانحراف المعياري للعيّنة $S\bar{x}$ يساوي الانحراف المعياري للمجتمع σ، وتسمى هذه بنظرية العيّنات الكبيرة، أما إذا كان عد المتغيّرات في العينة المدروسة أقل من 30 متغيّراً فعند ذلك نتّبع نظرية العيّنات الصغيرة وتكون قيمة $S\bar{x}$ تختلف عن σ .

- **توزيع الأوساط الحسابية للعيّنات :**

نفرض أننا نأخذ من مجتمع ما موزّعاً طبيعياً عدداً من العيّنات ، عدد المتغيّرات كل منها N متغيّراً (N > 30) ، فإن الأوساط الحسابية لهذه العيّنات هي :

$(\overline{x_n},......,\overline{x_3},\overline{x_2},\overline{x_1})$ ، وانحرافاتها المعيارية هي : $(S\overline{x_{12}},......,S\overline{x_2},S\overline{x_1})$ ، وهي متساوية وتساوٍي ***S*** .

هذه الأوساط الحسابية للعيّنات تختلف بعضها عن بعضها الآخر ، ولذلك تتوزّع إحصائياً وفق نظام معيّن يُطلق عليه اسم توزيع الأوساط الحسابية ، ولهذا التوزيع الإحصائي وسط حسابي $\overline{x}_n$ وانحراف معياري $S\overline{x}$.

ولقد وجد أن قيمة الوسط الحسابي لهذه الأوساط الحسابية يساوي الوسط الحسابي للمجتمع ، أي أن :

$$\overline{x}_m = M$$

وأن الانحراف المعياري لها : $$S\overline{x} = \frac{\sigma}{\sqrt{n}}$$

حيث: $S\overline{x}$: الانحراف المعياري للعينة .

σ : الانحراف المعياري للمجتمع .

n : عدد أفراد كل عينة من العينات .

أي أن الوسط الحسابي للأوساط الحسابية للعينات المأخوذة من هذا المجتمع جميعهايساوي الوسط الحسابي للمجتمع . كما أن الانحراف المعياري لأوساط العيّنات المأخوذة من المجتمع أصغر بكثير من الانحراف المعياري للمجتمع .

فإذا كانت المتغيّرات لمجتمع ما موزّعة توزيعاً طبيعياً بوسط حسابي (*M*) ، وانحراف معياري σ ، وأخذت عيّنات عشوائية بكل منها *N* متغيّراً فإن الأوساط الحسابية لهذه العيّنات $\overline{x}_m$ تكون موزّعة توزيعاً طبيعياً بوسط حسابي قدره (*M*) وانحراف معياري قدره :

$$S\overline{x} = \frac{\sigma}{\sqrt{N}}$$

وبذلك يمكن تطبيق خصائص المنحني الطبيعي ، حيث تمثّل الأوساط الحسابية للعيّنات على المحور الأفقي وتكراراتها على المحور العمودي ، وبالتالي يمكن قياس المساحة ***Z*** وحساب الاحتمال ، حيث أن :

$$Z = \frac{\bar{x} - M}{S\bar{x}}$$

حيث : $\bar{x}$: هو الوسط الحسابي للعينة .

M : الوسط الحسابي للمجتمع .

$S\bar{x}$: الانحراف المعياري للعينة .

مثال :

وجد أن أطوال جنود أحد المعسكرات البالغ عددهم 1000 جندي موزعة توزيعاً طبيعياً بوسط حسابي قدره 168.2 سم بانحراف معياري قدره 6.07سم ، احسب الاحتمال لأن يكون الوسط الحسابي لأطوال 100 جندي منهم أكثر من 169.9سم .

الحل :

$$\bar{x}_m = M = 168.2$$

$$S\bar{x} = \frac{\sigma}{\sqrt{N}} = \frac{6.07}{\sqrt{100}} = 0.607$$

والاحتمال المطلوب هو المساحة تحت منحني الاحتمال الطبيعي الواقعة إلى يمين Z :

$$Z = \frac{\bar{x} - M}{S\bar{x}} \Rightarrow \frac{169.9 - 168.2}{0.607} = 2.81$$

ومن جدول Z نجد أن المساحة المقابلة لـ Z = 0.4975 .

إذاً الاحتمال المطلوب : 0.5 – 0.4975 = 0.0025 .

أي أن احتمال الوسط الحسابي لعينة مؤلفة من 100 جندي أن يتجاوز 169.9سم هو 0.0025 أي 0.25% .

- **توزيع الفروق للعينات :**

لنفترض أنه لدينا مجتمعان موزعة تغيراتها بتوزيعين طبيعيين مستقلين ، فيكون وسطاهما الحسابيين $\bar{x}_{m1}$ و $\bar{x}_{m2}$ وانحرافهما المعياريين $\sigma_{\bar{x1}}$ $\sigma_{\bar{x2}}$, على

الترتيب ، فإذا أخذنا من المجتمع الأول عدداً من العينات بحيث تحوي كل عينة على ***N_1*** متغيراً ، فإن الأوساط الحسابية لهذه العينات $(\bar{\bar{x}}_n, \ldots\ldots\ldots, \bar{\bar{x}}_3, \bar{\bar{x}}_2 . \bar{\bar{x}}_1)$ حيث تتوزّع طبيعياً بوسط حسابي $\bar{x}_m$ وانحراف معياري $S\bar{x}$ ، وإذا أخذنا من المجتمع الثاني عدداً من العينات بحيث تحوي كل عينة على ***N_2*** متغيراً ، فإن الأوساط الحسابية لهذه العينات $\bar{\bar{x}}_n, \ldots\ldots, \bar{\bar{x}}_2, \bar{\bar{x}}_1$ تتوزّع طبيعياً أيضاً بوسط حسابي $\overline{\overline{x_M}}$ وانحراف معياري $S\bar{\bar{x}}$.

فإذا أخذنا من المجتمعين عدداً كبيراً من أزواج العينات $\bar{x}_1 \bar{\bar{x}}_1$ و $\bar{x}_2 \bar{\bar{x}}_2$ ... الخ، وحسبنا الفرق بين الوسطين الحسابيين لكل زوج من العينات المأخوذة $(\bar{x}_2 - \bar{\bar{x}}_2), (\bar{x}_1 - \bar{\bar{x}}_1)$ ... فإن الفروق تتوزّع إحصائياً (طبيعياً) بوسط حسابي لهذه الفروق قدره :

$$m.(\bar{x} - \bar{\bar{x}}) = \bar{x}_m - \bar{\bar{x}}_m$$

وانحرافه المعياري :

$$S(\bar{x} - \bar{\bar{x}}) = \sqrt{S^2 \bar{x} + S^2 \bar{\bar{x}}}$$

حيث : $\bar{x}_m = x_m$ و $\bar{\bar{x}}_m = x_m$.

كذلك : $\bar{x}$ الوسط الحسابي للعينة الأولى .

$\bar{\bar{x}}$ الوسط الحسابي للعينة الثانية .

$$S\bar{x} = \frac{\sigma}{\sqrt{n_1}} \qquad\qquad S\bar{\bar{x}} = \frac{\sigma}{\sqrt{n_2}}$$

وذلك عندما يكون المجتمعان المدروسان مستقلين بعضهماعن بعضهما الآخر .

فإذا مثلنا الفروق بين أوسط العينات $(\bar{x} - \bar{\bar{x}})$ على المحور الأفقي وتكراراتها على المحور العمودي ينتج توزيعاً طبيعياً ، ويكون :

$$Z = \frac{(\bar{x} - \bar{\bar{x}}) - M.(\bar{x} - \bar{\bar{x}})}{S.(\bar{x} - \bar{\bar{x}})}$$

مثال :

في أحد مصانع المصابيح الكهربائية (A) قدّر الوسط الحسابي لحياة المصابيح فوجد أنه 1400 ساعة بانحراف معياري قدره 200 ساعة ، بينما في مصنع آخر (B) وجد أن الوسط الحسابي لحياة المصابيح هو 1200 ساعة بانحراف معياري قدره 100 ساعة ، واختيرت عيّنة من كل مصنع عدد مصابيحها 125 مصباحاً ، وتمّ اختبارها . ما هو احتمال أن تكون مدة حياة مصابيح المصنع A أكبر من بـ 160 ساعة على الأقل من مصابيح المصنع B.

الحل :

$$M.(\bar{x}-\bar{\bar{x}}) = \bar{x}_m - \bar{\bar{x}}_m = 1400 - 1200 = 200$$ ساعة

(باعتبار أن $M.\bar{x} = M$, $M.\bar{\bar{x}} = M$) :

$$S\bar{x} = \frac{\sigma}{\sqrt{N_1}} = \frac{200}{\sqrt{125}}$$

$$S\bar{\bar{x}} = \frac{\sigma}{\sqrt{N_2}} = \frac{100}{\sqrt{125}}$$

ومنه يكون :

$$S(\bar{x}-\bar{\bar{x}}) = \sqrt{S^2\bar{x} - S^2\bar{\bar{x}}}$$

$$S(\bar{x}-\bar{\bar{x}}) = \sqrt{\left(\frac{200}{\sqrt{125}}\right)^2 + \left(\frac{100}{\sqrt{125}}\right)^2} = 20$$

ومنه تكون قيمة ***Z*** :

$$Z = \frac{\bar{x}-\bar{\bar{x}} - M.(\bar{x}-\bar{\bar{x}})}{S(\bar{x}-\bar{\bar{x}})} =$$

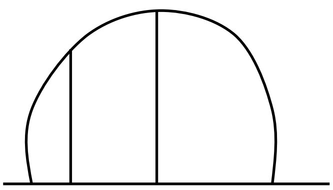
$$Z = \frac{160-200}{20} = -2$$

والاحتمال المطلوب هو المساحة الواقعة إلى يمين (Z= -2) .

من الجدول القيمة المقابلة هي : 0.4772 ،

والاحتمال المطلوب هو : 0.9772= 0.5+04772

إذاً الاحتمال المطلوب 97.72%



- **حدود الثقة Confidence Limits :**

ويسميها البعض حدود الثقة ، وأن لمجال الشك أو لفترة الثقة حدين أعلى وأدنى ، وفي التالي جدول يبين العلاقة بين مجال الثقة وقيمة Z المقابلة :

50% ، 68% ، 80% ، 95% ، 95.45% ، 0.96% ، 98% ، 99% ،99.73% : **مجال الثقة**

Z : 3 – 2.58 – 2.33 – 2.05 – 2 – 1.96 – 1.645 – 1.28 – 1 – 0.6745

الوسط الحسابي لعينة مؤلفة من 64 متغيراً هو 60 بانحراف حيادي قدره /3/ أوجد حدود الثقة 98% للوسط الحسابي للمجتمع .

$$S\bar{x} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{64}} = 0.375$$

ومن الجدول إن قيمة Z المناظرة للمساحة 0.49 = 2 / 0.98 على جانبي خط التناظر هي 2.33 =Z ، وبالتعويض بالمعادلة :

$$Z = \frac{|\bar{x} - M|}{S\bar{x}}$$

$$2.33 = \frac{|60 - M|}{0.375}$$

ومنه : $\boldsymbol{m = 6 \pm 0.87}$

أي أنه باحتمال 0.98 يقع الوسط الحسابي لهذا المجتمع بين (60.87 – 59.13).

- **نظرية العينات الصغيرة (توزيع ستودنت) :**

Small Sampling Theory, Student's Distribution

درس ستودنت هذا الموضوع وتوصّل إلى معرفة أنه لا يمكن تطبيق الجدول المبني على المنحني الطبيعي Z لتحليل البيانات التي نحصل عليها من العينات الصغيرة (أقل من 30) لأنه كلما قلّ عدد الأفراد في العينة ، كلّما كبر الفرق بين متوسط العينة ومتوسط المجموع الحقيقي ، مما ينتج عنه زيادة فرصة الوقوع في استنتاج خاطئ من النتائج نتيجة استعمال جداول الاحتمالات .

وعنـد عـدم معرفـة الانحـراف المعيـاري للمجتمـع (σ) يمكـن اسـتبداله بالانحراف المعياري للعينة $S\bar{x}$ وذلك عندما يكون عدد المتغيرات في العينة أكثر من 30 ، أما عندما يكون عدد المتغيرات N في العينة أقل من 30 فيلجأ إلى تقدير الانحراف المعياري للمجتمع من الانحراف المعياري للعينة من المعادلة :

$$\sigma = S\bar{x}.\sqrt{\frac{N}{N-1}}$$

حيث : σ : أحسن تقدير للانحراف المعياري للمجتمع .

$S\bar{x}$: الانحراف المعياري للعينة .

N : عدد المتغيرات في العينة .

أولاً : المقارنة متوسطة عينة بمتوسط المجتمع الذي أخذت منه في هذه تكون :

$$t = \frac{\bar{x} - M}{Sd'}$$

$$Sd' = \frac{\sigma}{\sqrt{N-1}}$$

حيث: $\bar{x}$: المتوسط الحسابي للعينة .

M : الوسط الحسابي للمجتمع .

Sd' : الخطأ المعياري .

تستخرج قيمة t من جدول المقابلة لدرجة الحرية N-1 الموجودة في التجربة عند احتمال 0.05 و 0.01 .

إذا كانت قيمة t المحسوبة أكبر من قيمة t الجدولية عند احتمال 5% و 1% دلّ أن الفرق مؤكد جداً .

أما إذا كانت t المحسوبة أكبر من t عند مستوى 5% وأصغر من t عند 1% دلّ على أن الفرق مؤكد وليس راجعاً للصدفة ، وفي كلتا الحالتين ترفض النظرية الفرضية القائلة : أنه لا يوجد فرق بين الأفراد والمجموعات . أما إذا كانت قيمة t المحسوبة أصغر من t الجدولية عند المستويين 1% و 5% فإن ذلك يدل على أن الفرق غير مؤكّد (معنوي) ، وتقبل النظرية الفرضية .

من المعادلة يمكن أن نحدد الحدين الأعلى والأدنى :

$$\bar{x} - M = \pm t S'd$$
$$M = \bar{x} \pm t S'd$$

نستخرج t عند المستوى 5% وعند المستوى 1% .

مثال:

أخذ من مجتمع طبيعي ذي وسط حسابي 30 عينة مؤلفة من 16 متغيراً ، وجد أن انحرافها المعياري هو 3 وكان انحراف وسطها عن الوسط الحسابي للمجتمع هو 2 فهل يعد هذا الانحراف ذا دلالة:

الحل :

توضع النظرية الفرضية وهي أن انحراف الوسط الحسابي للعينة عن الوسط الحسابي للمجتمع ليس بذات دلالة .

$$T = \frac{\bar{x} - M}{S'd}$$
$$S'd = \frac{\sigma}{\sqrt{N-1}} = \frac{3}{\sqrt{16-1}} = 0.78$$
$$T = \frac{2}{0.78} = 2.56$$

من الجدول t بدرجة حرية N-1 =15 والمستوي 0.5 هي 2.131 نجد أن t المحسوبة أكبر من t الجدولية عند مستوي 5% فالنتيجة ذات دلالة، أما قيمة t الجدولية عند مستوى 1% هي 2.947 ، نجد أن t المحسوبة أصغر من t الجدولية عند مستوى 1% فالنتيجة ليست بذات دلالة بل يرجع للصدفة .

- **سادساً : مقارنة أزواج الأفراد :**

تستعمل في الحالات التي يكون فيها من الواضح أن الفروق بين أزواج الأفراد

أكبر منها بين فردي الزوج الواحد نفسه بحيث قد تغطي الفروق بين الأزواج على

الفروق بين فردي الزوج .

ويستعمل هذا التصميم في تجارب التغذية ، فلو أريد مثلاً تجربة تأثير نوعين من الفئران على نسبة الدهن في الحليب أو معدّل النمو أو تجارب التسمين مثلاً ، وجرّبنا

نوعي الغذاء على مجموعتين متساويتين فإننا ندخل في نتائج التجربة الفروق التي تنتج عن الاختلافات بين أفراد المجموعتين من حيث التركيب الوراثي أو العمر أو تأثير البيئة ، ... الخ، كما يستعمل هذا التصميم في تجربة العقاقير الطبية وذلك بإعطائها لنفس الفرد على فترات مختلفة قبل تعاطي الدواء وبعده أو في أخذ قياسين مختلفين على الفرد نفسه على فترات مختلفة ، ونحسب قيمة t :

$$T = \frac{\overline{x_1} - \overline{x_2}}{S'd}$$

$$S'd = \frac{Sd}{\sqrt{n}}$$

$$Sd = \sqrt{\frac{\sum (D-\overline{D})^2}{n-1}} = \sqrt{\frac{\sum d^2}{n-1}}$$

حيث : $S'd$: الخطأ المعياري .

Sd : الانحراف المعياري .

مثال:

لاختبار تأثير دواء أو مصل معين على 8 أشخاص استعمل تركيزان وكان دليل التركيز هو البثور أو الحبوب في الجسم ودونت النتائج كالآتي :

$(D-d)^2$	$d= D-d$	الفرق التركيزين $D= x_1-x_2$	عدد البثور X_1	X_2	رقم الورقة
25	-5	-1	9	10	1
4	2	6	17	11	2
81	9	13	31	18	3
0	0	4	18	14	4
9	-3	1	7	6	5
9	-3	1	8	7	6

1	1	3	20	17	7
1	1	5	10	5	8
130	**.**	**32**	**120**	**88**	**.**

$$\overline{x_1} = \frac{\sum x_1}{N} = \frac{120}{8} = 15 \qquad \overline{x_2} = \frac{\sum x_2}{N} = \frac{88}{8} = 11$$

$$D' = \frac{\sum D}{N} = \frac{32}{8} = 4$$

$$T = \frac{\overline{x_1} - \overline{x_2}}{S'd}$$

$$S'd = \frac{Sd}{\sqrt{N}} \qquad Sd = \sqrt{\frac{\sum (D-d)^2}{n-1}} = \sqrt{\frac{130}{7}} = 4.31$$

$$S'd = \frac{4.31}{\sqrt{8}} = 1.52$$

$$T = \frac{15-11}{1.52} = 2.64$$

من الجدول عند درجة حرية ***N***-1 وهي 7 ومستوى 3.499 =1% ، 5%= %5= 2.365 .

نجد أن ***t*** المحسوبة أكبر من t الجدولية عند 5% وأصغر من *t* الجدولية عند 1% نستنتج أن هناك فرقاً بين التركيزين ذا دلالة عادية ، أي أن النظرية الفرضية الموضوعة خاطئة ، بمعنى آخر أن التركيز الثاني أفضل من التركيز الأول .

- **اختبار t لمقارنة المجموعات :**

أ- عندما يكون عدد الأفراد في المجموعتين واحداً :

إذا تساوى عدد الأفراد في المجموعتين المراد مقارنتهما أي تساوى عدد أفراد المجموعة الأولى N_1 مع أفراد المجموعة الثانية N_2 فإن طريقة تحليل مثل هذه التجارب تشابه الطريقة التالية :

$$t = \frac{\overline{x_1} - \overline{x_2}}{S'd}$$ بدرجات حرية $(n_1 + n_2 - 2)$

$\overline{x_1}$: الوسط الحسابي للعينة الأولى .

$\overline{x_2}$: الوسط الحسابي للعينة الثانية .

$S'd$: الخطأ المعياري .

$$S'd = \sqrt{\frac{S^2P}{n_1} + \frac{S^2P}{n_2}}$$

التباين المشترك للفرق بين المتوسطين :

$$S^2P = \frac{SS_1 + SS_2}{n_1 + n_2 - 2}$$

ثم نقارن بين t المحسوبة و t الجدولية عند درجات حرية (n_1+n_2-2) والمستويين 1% و 5% .

مثال :

في تجربة لمعرفة أثر استعمال مصل معين على مقاومة أحد الأمراض اختير 20 شخصاً بطريقة عشوائية ، ترك عشرة أشخاص منهم بدون مصل ، وأعطي العشرة الباقين المصل ، ورتبت النتائج في الجدول الآتي ، حيث تُشير x_1 إلى الأشخاص الذين أعطوا المصل و x_2 إلى الأشخاص الذين لم يأخذوا المصل ، فهل يوجد فرق حقيقي بين الأشخاص الذين أخذوا المصل والذين لم يأخذوا المصل ؟

الحل :

تنصّ الفرضية النظرية الإحصائية أنه لا يوجد فرق حقيقي بين الأشخاص الذين أعطوا المصل والذين لم يأخذوا المصل ، فننشئ الجدول الآتي :

x_1	x_2	x_1^2	x_2^2
5.1	**3.1**	**37.21**	**9.61**
5.6	**3.8**	**31.36**	**14.44**
6.4	**4.3**	**40.96**	**18.49**
5.8	**3.9**	**33.64**	**15.21**
6.0	**4.0**	**36.00**	**16.00**

5.7	**4.0**	**32.49**	**16.00**
5.9	**4.5**	**34.81**	**20.25**
7.5	**4.2**	**56.25**	**17.64**
7.0	**4.3**	**49.00**	**18.49**
6.0	**4.9**	**36.00**	**24.01**
62.0	**41.0**	**387.72**	**170.14**

$$\overline{x_1} = \frac{62}{10} = 6.2 \qquad \overline{x_2} = \frac{41}{10} = 4.1$$

$$T = \frac{\overline{x_1} - \overline{x_2}}{S'd} = \frac{6.2 - 4.1}{S'd}$$

$$S'd = \sqrt{\frac{S^2P}{N_1} + \frac{S^2P}{N_2}}$$

$$SS_1 = \sum x_1^2 - \frac{(\sum x_1)^2}{N} = 387.72 - \frac{(62)^2}{10} = 3.32$$

$$SS_2 = \sum x_2^2 - \frac{(\sum x_2)^2}{N} = 170.14 - \frac{(41)^2}{10} = 2.04$$

$$S^2P = \frac{SS_1 + SS_2}{n_1 + n_2 - 2} = \frac{3.32 + 2.04}{18} = 0.2978$$

$$S'd = \sqrt{\frac{0.2978}{10} + \frac{0.2978}{10}} = 0.244$$

$$T = \frac{6.2 - 4.1}{0.244} = 8.61$$

من الجدول وأمام درجة حرية 18 وعند مستوى 5% 2.101 ، وعند مستوى 1% 2.878 ، نجد أن t المحسوبة أكبر من t الجدولية عند المستويين ، فالفرق ذو دلالة معنوية عالية وترفض النظرية الفرضية أي أن الأشخاص الذين أعطوا المصل كانوا ذوي دلالة عالية .

ب- مقارنة مجموعتين مختلفتين في عدد أفراد كل منهما :

يحدث في كثير من التجارب ألاّ يكون عدد الأفراد في المجموعتين المراد مقارنتهما بعضهما مع بعضهما الآخر متساوياً ، فقد تكون التجربة عند بدايتها متساوية في العدد ، ولكن لظروف خارجة عن إرادة الباحث ينقص عدد أفراد إحدى المجموعتين، كما في التجارب التي تجري على الحيوانات ، فقد يموت أحدها أثناء التجربة أو تصيب الآفات إحدى القطع المزروعة ، واختلاف عدد الأفراد في المجموعتين لا يغير في الأسس الإحصائية التي بنيت عليها طرق التحليل ، ولكن يحدث تعديلاً طفيفاً في طريقة الحساب لأن $N_2 \neq N_1$ ، فلحساب الخطأ المعياري

$$T = \frac{\overline{x_1} - \overline{x_2}}{Sd}$$

$$Sd = \sqrt{S^{2'} d_1 + S^{2'} d_2}$$

$$S'd_1 = \frac{SP}{\sqrt{n_1}} \qquad S'd_2 = \frac{SP}{\sqrt{n_2}}$$

$$Sd = \sqrt{\frac{S'd_1}{n_1} + \frac{S'd_2}{n_2}}$$

مثال :

في إحدى تجارب التغذية على فئران ، استعمل نوعان من الغذاء، أحدهما به نسبة منخفضة من البروتين ، والآخر به نسبة عالية من البروتين ، وقدر وزن الفئران وعمرها 28 يوماً، ثم قدّرت الزيادة في الوزن حينما بلغ عمرها84 يوماً، وكان عدد أفراد المجموعة الأولى 12 والثانية 7 ، وفيما يلي بيان الزيادة في الوزن الناتجة عن هذين النوعين من الفئران x_1 و x_2 ، والمطلوب تحليل نتائج التجربة.

الحل :

تنص النظرية الفرضية أنه لا يوجد فرق بين الغذائين في التأثير على معدّل الزيادة في وزن الفئران .

عالي البروتين **x_1**	***منخفض البروتين*** **x_2**	x_1^2	x_2^2
134	**70**	**17956**	**4900**
146	**118**	**11316**	**13924**

104	**101**	**10816**	**10201**
119	**85**	**14161**	**7225**
124	**107**	**15376**	**11449**
161	**132**	**25921**	**17424**
107	**94**	**11449**	**8836**
83		**6889**	
113		**12769**	
129		**16641**	
97		**9409**	
123		**15129**	
1440	**707**	**177832**	**73959**

$\overline{x_1} = 120 \qquad \overline{x_2} = 101$

$$T = \frac{\overline{x_1} - \overline{x_2}}{S'd} = \frac{120 - 101}{S'd}$$

$$SP = \sqrt{\frac{SS_1 + SS_2}{n_1 + n_2 - 2}}$$

$$SS_1 = \sum x_1^2 - \frac{(\sum x_1)^2}{N} = 177832 - \frac{(1440)^2}{12} = 5032$$

$$SS_2 = \sum x_2^2 - \frac{(\sum x_2)^2}{N} = 73959 - \frac{(707)^2}{7} = 2552$$

$$SP = \sqrt{\frac{5032 + 2552}{12 + 7 - 2}} = \sqrt{446.11}$$

$$S'd_1 = \frac{\sqrt{446.11}}{\sqrt{12}} \qquad S'd_2 = \frac{\sqrt{446.11}}{\sqrt{7}}$$

$$Sd = \sqrt{\frac{446.11}{12} + \frac{446.11}{7}} = 10.04$$

$$T = \frac{120 - 101}{10.04} = 1.89$$

من الجدول t الجدولية عند5% =2.11 ، وعند 1% = 2.898 درجة حرية (17) نجد أن t المحسوبة أصغر من t الجدولية ، فنقول ليس بذي دلالة والنظرية الفرضية مقبولة .

العينات والاختبارات

SAMPLES AND TESTS

6-1- مقدمة: إن هدف الاختبار الإحصائي هو اختبار فرضية تتعلق إما بمؤشرات إحصائية أساسية (أو ما يسمى أحيانا وسيط) أو أنها تتعلق بالتوزيع المفترض لمجتمع ما.

فكثيرا من الأحيان يتطلب الأمر معرفة قانون توزيع مجتمع ما غير معروف فعلا وبالتالي لابد من الافتراض بان هذا المجتمع يتبع أحد التوزيعات المعروفة ومن ثم يتم اختبار هذه الفرضية لمعرفة مدى صحتها أي لرفضها أو قبولها.

وقد يكون العكس أحيانا حيث يتطلب الأمر معرفة المؤشرات الإحصائية ألأساسية لمجتمع ما مثل المتوسط الحسابي أو الانحراف المعياري وما شابه ذلك، وبالتالي لابد من وضع الفرضيات التي تفترض قيما افتراضية للمؤشرات الإحصائية ومن ثم يتم اختبار هذه الفرضيات لمعرفة مدى صحتها أي لرفضها أو قبولها.

6-2- الاختبار الإحصائي : يتضمن الاختبار الإحصائي العناصر الرئيسية التالية:

1. الفرضية الابتدائية ويرمز لها بـ :H0
2. منطقة الرفض والقبول
3. الفرضية البديلة ويرمز لها بـ :H1
4. احصاء الاختبار

إن العناصر الرئيسية الأربع المذكورة هي ضرورية لكل اختبار إحصائي وأي تغير في أحد من هذه العناصر الأربع يؤدي إلى اختبار جديد.

وفيما يلي سندرس بعض هذه العناصر الأربع بشيء من التفصيل :

1. الفرضية الابتدائية (:H0) : وهي الفرضية التي سيجري اختبار مدى صحتها أي لرفضها أو قبولها .وهذه الفرضية تنص تتضمن بشكل عام قيم افتراضية لوسيط ما للمجتمع أو العينة المدروسة؛مثل الافتراض بان هذا المجتمع يتبع أحد التوزيعات المعروفة أو تفترض قيما افتراضية للمؤشرات الإحصائية كان تفترض أن قيمة المتوسط μ يساوي قيمة افتراضية μ_0 وهكذا...

إن اتخاذ القرار للتأكد من صحة الفرضية الابتدائية أي لرفضها أو قبولها يعتمد أساسا على المعلومات التي تحويها العينة أو العينات المسحوبة بطريقة عشوائية من المجتمع المدروس.

2. منطقة الرفض والقبول :تقسم مجموعة كل القيم المأخوذة التي يمكن أن يأخذها احصاء الاختبار إلى مجموعتين أو منطقتين ندعو أحداهما منطقة الرفض والأخرى منطقة القبول. وان الرفض أو القبول يتوقف على أين تقع القيمة التي يأخذها احصاء الاختبار ، فإذا وقعت القيمة التي يأخذها احصاء الاختبار ،محسوبا من العينة التي بين أيدينا ،في منطقة الرفض ترفض الفرضية الابتدائية ونقبل الفرضية البديلة.أما وقعت القيمة التي يأخذها احصاء الاختبار ،محسوبا من العينة التي بين أيدينا ،في منطقة القبول نقبل الفرضية الابتدائية و ترفض الفرضية البديلة.
يخضع اتخاذ القرار للتأكد من صحة الفرضية الابتدائية ، أي لرفضها أو قبولها، إلى نوعين من الخطأ وهما التالين:

1. الخطأ من النوع الأول ERROR TYPE I
2. الخطأ من النوع الثاني ERROR TYPE II

وينص الخطأ من النوع الأول ERROR TYPE I إلى أنه يمكن أن نرفض الفرضية الابتدائية بينما هي في الواقع صحيحة .أما الخطأ من النوع الثاني ERROR TYPE II فيحصل عندما نقبل الفرضية الابتدائية بينما هي في الواقع خاطئة.

والجدول التالي يبن الحالات السابقة عند اتخاذ القرار للتأكد من صحة الفرضية الابتدائية أو عدم صحتها :

القرار	الفرضية الابتدائية	
	صحيحة	خاطئة.
الرفض	الخطأ من النوع الأول	قرار صحيح
القبول	قرار صحيح	الخطأ من النوع الثاني

وعادة نقيس جودة الاختبار الإحصائي باحتمالي الخطأ من النوع الأول و الخطأ من النوع الثاني ،ونرمز لهما بـ α و β على الترتيب.ويمكن التعبير عن الاحتمال α بأنه أحتما أن يقع احصاء الاختبار في منطقة الرفض علما بأن الفرضية الابتدائية صحيحة .ويدعى α أيضا بمستوى الأهمية أو المعنوية .ومن الواضح أن زيادة حجم منطقة الرفض سيزيد من

قيمة α وفي نفس الوقت فانه يؤدي إلى تناقص قيمة β وذلك من أجل حجم ثابت للعينة n .أما تخفيض منطقة الرفض فسيزيد من قيمة β وفي نفس الوقت فانه يؤدي إلى تناقص قيمة α.وعند زيادة حجم العينة n فانه من الطبيعي أن توفر لنا العينة الأكبر قدرا لأكبر من المعلومات نتخذ على ضوئها قرارنا وهذه ستؤدي إلى تناقص كل من α و β .

يمكن أن نقسم الفرضية الابتدائية إلى نوعين :

- الفرضية الابتدائية أحادية الجانب (البسيطة):نقول عن فرضية ابتدائية أنها بسيطة إذا كانت تختبر قيمة واحدة فقط بدون احتمالات أخرى .كمثال على ذلك نقول : $H0:\ a > 50$ ، ففي هذه الحالة لا يوجد سوى قيمة واحدة فقط بدون احتمالات أخرى و الفرضية البديلة تنص كما يلي : $H1:\ a< 50$.
- الفرضية الابتدائية ثنائية الجانب(المركبة): نقول عن فرضية ابتدائية أنها مركبة إذا كانت تختبر قيمة واحدة فقط ولكن بعدة احتمالات.كمثال على ذلك نقول : $H0:\ a = 50$ وبالتالي فان الفرضية البديلة تنص كما يلي : $H1: a \neq 50$ ، ففي هذه الحالة يوجد احتمالين وهما $a>50$ والاحتمال الآخر $a< 50$.ويمكن أن تتوضع منطقة الرفض كما يلي :
  - إلى جانبي منطقة القبول عند ذلك نقول عن الاختبار الإحصائي بأنه ثنائي الجانب كما أشرنا قبل قليل.
  - إلى جانب واحد من منطقة القبول(إلى اليمين أو اليسار) عند ذلك نقول عن الاختبار الإحصائي بأنه أحادي الجانب كما أشرنا قبل قليل.والأشكال التالية توضح هذه الحالات.

3. الفرضية البديلة (H1:):توضع الفرضية البديلة كبديل في حال رفض الفرضية الابتدائية :H0 .
4. احصاء الاختبار: يتم استخدام قيم العينة المدروسة لحساب عدد واحد يأخذ دور صانع القرار و يدعى احصاء الاختبار.

3-6-الاختبارات المعنوية:تشمل الاختبارات المعنوية مجموعتين أساسيتين وذلك حسب مجال التطبيق وهما المجموعتان التاليتان:

- الاختبارات الخاصة بالعينات كبيرة الحجم :في حال العينات كبيرة الحجم يستخدم اختبار Z الذي يقوم على احصاء الاختبار الذي يتبع التوزيع الطبيعي .وهذا الاختبار يعتمد عليه بشكل خاص من أجل المقارنات التالية:
- مقارنة متوسطين لمجتمعين علما أن التشتت لهما معلوم.
- مقارنة متوسطين لمجتمعين علما أن التشتت لهما غير معلوم.

- مقارنة متوسط المجتمع مع متوسط العينة علما أن الانحراف المعياري للمجتمع معروف.

يكفي في هذا المجال أن نذكر أن العلاقة التي تعطينا اختبار Z-test التي سبق وأن شرحت في الفصل السابق كما تم شرح علاقة هذا الاختبار مع منحنى التوزيع الطبيعي .كذلك الأمر يمكن مراجعة الكتاب الثاني "الجزء العملي " للإطلاع على المزيد حول تطبيقات اختبار Z .

- الاختبارات الخاصة بالعينات صغيرة الحجم :في حال العينات صغيرة الحجم يستخدم الاختبارات التالية:

أ- اختبار t-test الذي يقوم على احصاء الاختبار الذي يتبع توزيع ستيودنت Student distribution .وهذا الاختبار يعتمد عليه بشكل خاص من أجل المقارنات التالية:

- مقارنة متوسط المجتمع مع متوسط العينة علما أن الانحراف المعياري للمجتمع غير معروف.
- اختبار الفرق بين نسبتين من مجتمع واحد.
- اختبار الفرق بين نسبتين من مجتمعين نفترض أن تباينهما متماثل
- اختبار الفرق بين متوسطين لعينتين نفترض أنهما متساويتين بعدد العناصر (العينات غير المستقلة)
- اختبار الفرق بين متوسطين لعينتين نفترض أنهما غير متساويتين بعدد العناصر (العينات المستقلة).
- اختبار القيم الشاذة وحذفها

وفيما يلي سنشرح بعضا من هذه الاختبارات والتي تلقى تطبيقا في مجال التجارب والاختبارات الزراعية والحيوية وهي التالية:

1. اختبارات النسبة المئوية بين نسبتين من مجتمع واحد: نستخدم مثل هذا الاختبار عندما تكون البيانات الإحصائية DATA عبارة عن نسبة مئوية،ومن أجل ذلك فإننا نستخدم احصاء الاختبار التالي :

$$T = p - p_0 / \sqrt{p_0 * q_0} / n$$

مثـــــــــال (1) :إذا كانت نسبة الأشجار المعتادة للإصابة بالمرض هي 15% ،تم رش الأشجار بأحد المبيدات فكانت نسبة الأشجار المريضة المتبقية هي 9% :فهل يعتبر المبيد المستخدم فعالا أم لا ؟

الحـــــــل : نضع الفرضية الابتدائية وهي التالية :

$$H0: p = p_0 = 0.15$$

مقابل ذلك تكون الفرضية البديلة هي التالية :

H1: $p \neq p_0$

نطبق العلاقة السابقة الذكر لحساب قيمة احصاء الاختبار كما يلي :

$$T = 0.09 - 0.15 / \sqrt{0.15 * 0.85} / 100$$

$t = -1.6$

نخرج قيمة t_{tab} الجد ولية عند مستوى المعنوية 10% فنجد أن :

$t_{tab} = 1.65$

الآن نقارن قيمة مع قيمة t_{cct} المحسوبة من العلاقة السابقة أعلاه

نجد أن t_{tab} الجد ولية أكبر من قيمة t_{cct} المحسوبة وبالتالي نقبل الفرضية الابتدائية أي أن أثر المبيد غير معنوي و نستنتج بالنهاية أن المبيد غير فعال.

2. اختبارات النسبة المئوية بين نسبتين من مجتمعين لهما تباين واحد: نستخدم مثل هذا الاختبار عندما تكون البيانات الإحصائية DATA عبارة عن نسبة مئوية من مجتمعين تساوى فيها الانحراف المعياري،ومن أجل ذلك فإننا نستخدم احصاء الاختبار التالي :

$$T = p - p_0 / s * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

حيث أن :

$$s = \sqrt{(n_1 * p_1 * q_1 + n_2 * p_2 * q_2)} / (n_1 + n_2 - 2)$$

P_1 : النسبة الاولى

P_2 : النسبة الثانية

q_1 : متمم النسبة الاولى

q_2 : متمم النسبة الثانية

مثــــــال (2) :تبين أن نسبة الأشجار المعتادة للإصابة بالمرض هي 20% ،تم رش الأشجار بأحد المبيدات فكانت نسبة الأشجار المريضة المتبقية هي 18 % :فهل تعتبر هاتان النسبتان مختلفتان علما أنه تم معاينة 20 شجرة في المرة الاولى و15 شجرة في المرة الثانية؟

الحـــــل : نضع الفرضية الابتدائية وهي التالية :

$H0:\ p_1 = p_2$

مقابل ذلك تكون الفرضية البديلة هي التالية :

$H1:\ p_1 \neq p_2$

$Df:\ n_1 + n_2 - 2 = 33$

نطبق العلاقة السابقة الذكر لحساب قيمة احصاء الاختبار كما يلي :

$$T = 0.20 - 0.18 / s * \sqrt{\frac{1}{20} + \frac{1}{18}}$$

حيث أن S تحسب من العلاقة:

$$s = \sqrt{(20 * 0.20 * 0.80 + 15 * 0.18 * 0.82) / (15 + 20 - 2)}$$

$$T_{cct} = 0.14$$

نخرج قيمة t_{tab} الجد ولية عند مستوى المعنوية 5% ودرجة حرية 33 فنجد أن :

$$t_{tab} = 2.02$$

الآن نقارن قيمة مع قيمة t_{cct} المحسوبة من العلاقة السابقة أعلاه

نجد أن t_{tab} الجد ولية أكبر من قيمة t_{cct} المحسوبة وبالتالي نقبل الفرضية الابتدائية أي أن تأثير المبيد غير معنوي و نستنتج أن المبيد لم يكن فعالا.

3. اختبار الفرق بين متوسطين لعينتين نفترض أنهما متساويتين بعدد العناصر (العينات غير المستقلة):في مثل هذه الحالة نستخدم مثل هذا الاختبار عندما تكون البيانات الإحصائية DATA عبارة عن عينتين متساويتين بعدد العناصر أي $n_1 = n_2$،ومن أجل ذلك فإننا نستخدم احصاء الاختبار التالي :

$$t = \bar{x}_1 - \bar{x}_2 / s_{\bar{d}}$$

$$s_{\bar{d}} = s_d / \sqrt{n}$$

$$s_d = \sqrt{\sum (d - \bar{d})^2 / (n-1)}$$

حيث أن : s_d الانحراف المعياري للفروق بين عناصر العينتين

$x`_1$ متوسط العينة الاولى

$x`_2$ متوسط العينة الثانية

مثـــــال (3): لدينا العينتين التاليتين :

والمطلوب اختبار الفرق بين متوسطين العينتين حيث أنهما متساويتين بعدد العناصر:

العينة الاولى	العينة الثانية
Y	X
5	10
13	20
16	30
7.73	40
5.51	50
10	60

نطبق العلاقة السابقة الذكر لحساب قيمة احصاء الاختبار t – test فنحصل على النتائج كما يلي :

t-Test: Two-Sample Assuming Equal Variances

النتائج	*X*	*Y*
Mean	35	9.54
Variance	350	18.80868
Observations	6	6
Pooled Variance	184.40434	
Hypothesized Mean Difference	0	
df	10	
t Stat	3.247382587	
P(T<=t) one-tail	0.004379611	
t Critical one-tail	1.812461505	
P(T<=t) two-tail	0.008759222	
t Critical two-tail	2.228139238	

الآن نضع الفرضية الابتدائية وهي التالية :

$$H0: \; x`_1 = x`_2$$

مقابل ذلك تكون الفرضية البديلة هي التالية :

$$H1: \; x`_1 \neq x`_2$$

نخرج قيمة t_{tab} الجد ولية عند مستوى المعنوية 5% ودرجة حرية 10 كما هو وارد في الجدول السابق فنجد أن :

$t_{tab} = 2.22$

الآن نقارن قيمة مع قيمة t_{cct} المحسوبة من العلاقة السابقة كما هو وارد في الجدول السابق أعلاه حيث أن $t_{clt} = 3.2473$

نجد أن t_{clt} المحسوبة أكبر من قيمة t_{tab} الجد ولية وبالتالي نرفض الفرضية الابتدائية أي أن متوسطي العينتين مختلفين والفرق بينهما معنوي مؤكد إحصائيا وليس عائدا للصدفة..

4. اختبار الفرق بين متوسطين لعينتين نفترض أنهما ليستا متساويتين بعدد العناصر (العينات المستقلة):في مثل هذه الحالة نستخدم مثل هذا الاختبار عندما تكون البيانات الإحصائية DATA عبارة عن عينتين غير متساويتين بعدد العناصر أي $n_1 \neq n_2$،ومن أجل ذلك فإننا نستخدم احصاء الاختبار التالي :

$$t = \bar{x}_1 - \bar{x}_2 / s_{\bar{d}}$$

$$s_{\bar{d}} = \sqrt{\frac{sp^2}{n_1} + \frac{sp^2}{n_2}}$$

$$sp^2 = s_1 + s_2 /(n_1 + n_2 - 2)$$

مثـــال (4) : لدينا العينتين التاليتين :

والمطلوب اختبار الفرق بين متوسطين العينتين حيث أنهما غير متساويتين بعدد العناصر:

العينة الاولى	العينة الثانية
Y	X

10	5
20	13
30	16
40	**10**
50	
60	

نطبق العلاقة السابقة الذكر لحساب قيمة احصاء الاختبار t – test فنحصل على النتائج كما يلي :

t-Test: Two-Sample Assuming Unequal Variances

	X	*Y*
Mean	35	11
Variance	350	22
Observations	6	4
Hypothesized Mean Difference	0	
df	6	
t Stat	3.003913896	
P(T<=t) one-tail	0.01194363	
t Critical one-tail	1.943180905	
P(T<=t) two-tail	0.023887261	
t Critical two-tail	2.446913641	

الآن نضع الفرضية الابتدائية وهي التالية :

$$H0:\ x`_1 = x`_2$$

مقابل ذلك تكون الفرضية البديلة هي التالية :

$$H1:\ x`_1 \neq x`_2$$

نخرج قيمة t_{tab} الجد ولية عند مستوى المعنوية 5% ودرجة حرية 6 كما هو وارد في الجدول السابق فنجد أن :

$$t_{tab} = 2.446913$$

الآن نقارن قيمة مع قيمة t_{cct} المحسوبة من العلاقة السابقة كما هو وارد في الجدول السابق أعلاه حيث أن $t_{clt} = 3.00391$

نجد أن t_{clt} المحسوبة أكبر من قيمة t_{tab} الجد ولية وبالتالي نرفض الفرضية الابتدائية أي أن متوسطي العينتين مختلفين والفرق بينهما معنوي مؤكد إحصائيا وليس عائدا للصدفة مثل النتيجة السابق.

5. اختبار القيم الشاذة وحذفها: كثيرا ما ينتج لدينا في تجربة ما بين قيم العينة قيم شاذة ومتطرفة تؤثر سلبا على التحليل الإحصائي ولابد من إزالتها.والمهم هنا هو كيفية التعرف على القيم الشاذة والمتطرفة .هذا ما سنشرحه فيما يلي :
يستخدم اختبار t-test لهذه الغاية حيث أ ننا نستخدم احصاء الاختبار التالي :

$$t = x_0 - \overline{x} / s$$

حيث أن : x_0 القيمة المراد اختبارها

s الانحراف المعياري للعينة

$x`$ المتوسط الحسابي للعينة

وتكون الفرضيات الإحصائية كما يلي:

الفرضية الابتدائية :

$$H0: \quad x_0 = x`$$

الفرضية البديلة :

$$H1: x_0 \neq x`$$

مثــــــــال (5) :لدينا العينة التالية :

$$X = \{ 200, 300,250,400,400,500,550\}$$

هل يمكن أن نعتبر القيمة 550 قيمة شاذة أم لا ؟

نطبق العلاقة السابقة الذكر لحساب قيمة احصاء الاختبار t – test فنحصل على النتائج كما يلي :

$$x` = 122.93$$

$$t = 550 – 383.75 / 122.93$$

$$t = 1.35$$

نخرج قيمة t_{tab} الجد ولية عند مستوى المعنوية 10% ودرجة حرية 7 فنجد أن :

$$t_{tab} = 1.89$$

الآن نقارن قيمة مع قيمة t_{cct} المحسوبة من العلاقة السابقة كما هو وارد في الجدول السابق أعلاه حيث أن $t_{clt} = 1.35$

نجد أن t_{clt} المحسوبة أصغر من قيمة t_{tab} الجد ولية وبالتالي نقبل الفرضية الابتدائية أي أن القيمة المعنية 550 ليست قيمة شاذة لأنها لا تختلف عن قيمة المتوسط بشكل معنوي والفرق غير حقيقي وغير مؤكد إحصائيا.

ب- اختبار مربع كاي χ^2 : يستخدم هذا الاختبار عند وجود نوعين من المؤشرات أو البيانات وهي التالية :

- مؤشرات أو بيانات حقيقية أو تجريبية observed يحصل عليها الباحث من خلال تجربة معينة.

- مؤشرات أو بيانات نظرية : يحصل عليها الباحث من خلال تطبيق معادلة أو نموذج رياضي arithmetical model أو معلومات مرجعية من خبرات سابقة في أماكن أخرى وظروف مختلفة.

في مثل هذه الحالة باستخدم اختبار مربع كاي χ^2 لقياس الفرق بين المؤشرات أو البيانات الحقيقية أو التجريبية وبين المؤشرات أو البيانات النظرية.

ولهذه الغاية يمكننا أن نستخدم احصاء الاختبار التالي :

$$\chi^2 = \sum((O_i - E_i)^2 / E_i)$$

حيث أن : O_i القيمة الحقيقية أو التجريبية

E_i القيمة النظرية.

بعد الآن نقارن قيمة χ^2 الجد ولية مع قيمة χ^2 المحسوبة عند درجة حرية $df = k-1$ حيث k عدد قيم أو المشاهدات للعينة أو المجموعات إذا كانت العينة مقسمة إلى مجموعات وعند مستوى المعنوية 5% لمعرفة هل الفروق بين البيانات الحقيقية (التجريبية) وبين المؤشرات أو البيانات النظرية هل هي معنوية حقيقية مؤكدة إحصائيا أم غير ذلك.

مثـــال (6) : لدينا العينتين التاليتين الاولى عينة من البيانات الحقيقية أو التجريبية وأخرى من البيانات النظرية:

البيانات النظرية	البيانات الحقيقية
y	x
12	10
22	20
25	30
44	40
66	50
48	60

والمطلوب اختبار هل الفروق بين البيانات الحقيقية أو التجريبية وبين المؤشرات أو البيانات النظرية هل هي معنوية حقيقية مؤكدة إحصائيا أم لا:

نطبق العلاقة التي تساعدنا في حساب قيمة مربع كاي المحسوبة وهي التالية :

$$\chi^2 = \sum((O_i - E_i)^2 / E_i)$$

بعد تبديل المتغيرات في العلاقة السابق نحصل على الجدول التالي :

	البيانات الحقيقية	البيانات النظرية	الجد ولية $\chi^2 = 11.07$	
	x	y	(x-y)^2	(x-y)^2/y
	10	12	4	0.333333
	20	22	4	0.181818
	30	25	25	1
	40	44	16	0.363636
	50	66	256	3.878788
	60	48	144	3
Total Σ	210	217	449	8.757576

الآن نقارن قيمة χ^2 الجد ولية 11.07 وهي تساوي مع قيمة χ^2 المحسوبة عند درجة حرية df = k-1= 5 (حيث k عدد قيم أو المشاهدات للعينة أو المجموعات إذا كانت العينة مقسمة إلى مجموعات) وعند مستوى المعنوية 5% فنجد أن χ^2 الجد ولية أكبر من قيمة χ^2المحسوبة والتي تساوي 8.757576 كما هو وارد في الجدول السابق وبالتالي نستنتج أن الفروق بين البيانات الحقيقية (التجريبية) وبين المؤشرات أو البيانات النظرية هي ليست معنوية وليست حقيقية وغير مؤكدة إحصائيا وتعود للصدفة أو لعوامل خارجية مجهولة.

ج- اختبار f-test : يعد اختبار f-test من الاختبارات المعنوية الهامة جدا وهو يتبع توزيع فيشر ، Fisher distribution كما أن هذا الاختبار يتعلق بتحليل التباين بشكل كامل لذلك سيتم شرحه بالتفصيل في الفصل القادم.

4-6- مجال أو حدود الثقة Confidence interval :إن المقصود من مجال أو حدود الثقة هو تحديد المجال (الحد الأدنى والحد الأعلى) لأحد المؤشرات الإحصائية Parameter مع حساب الاحتمال وقوع هذا المؤشر ضمن المجال المحدد.وعادة يتم استخراج مجال أو حدود الثقة من العلاقة التي تستخدم لحساب احصاء اختبار المعنوية لهذا المؤشر.

يقوم الاحصائيون عادة بحساب مجال أو حدود الثقة للحالات التالية :

1. حساب مجال أو حدود الثقة لمتوسط مجتمع طبيعي تباينه غير معروف
2. حساب مجال أو حدود الثقة لمتوسط مجتمع طبيعي تباينه معروف
3. حساب مجال أو حدود الثقة لمتوسط عينة كبيرة الحجم .
4. حساب مجال أو حدود الثقة لمعاملات الانحدار
5. حساب مجال أو حدود الثقة للقيم التابعية لمعادلة الانحدار .

تعتبر طريقة حساب مجال أو حدود الثقة للحالات الخمسة السابقة الذكر أمرا صعبا وخاصة بالنسبة لغير المتخصصين ،إلا أن توفر برامج التحليل الإحصائي في برمجيات الجداول الالكترونية مثل برنامج Excel وبرنامج SPSS وبرنامج

LOTUS 123 وغيرها جعل إنجاز مثل هذه الحسابات أمرا سهلا جدا لذلك سنكتفي بشرح مثال واحد عن تطبيق هذه البرمجيات لحساب مجال أو حدود الثقة لمعاملات الانحدار نظرا لأهميتها وكفايتها كما في المثال التالي:

مثـــال (7) : عند دراسة العلاقة ما بين كمية السماد المضاف كغ/ها وصفة الإنتاج لأحد المحاصيل وجاء كان لدينا التحليل الإحصائي للبيانات الإحصائية DATA كان لدينا ما يلي :

الإنتاج	السماد
Y	**X**
5	**1**

2	**13**
3	**16**
4	**23**
5	**33**
6	**38**
7	**40**

أجري التحليل الإحصائي للانحدار وحصلنا على النتائج التالية:

معاملات الانحدار	*Coefficients*	*الحد الأدنى Lower 95%*	*الحد الأعلى Upper 95%*
a	0.57142857	-5.274208	4.131351
b	6.14285714	5.0912837	7.194431

نلاحظ من ألجدول السابق أن عامل الانحدار $b = 6.14285714$ يقع ضمن مجال الثقة الوارد في الجدول السابق مابين 5.0912837 و 7.194431 باحتمال قدره : ($1-\alpha = 5\%$) .

*****

مسائل وحلول متقدمة

اختبار مربع كايCHITEST

إرجاع اختبار الاستقلالية. تقوم CHITEST بإرجاع القيمة الناتجة من التوزيع كاي تربيع (γ2) لإحصاء البيانات ودرجات الحرية المناسبة. يمكنك استخدام اختبارات γ2 في تحديد ما إذا كانت النتائج الفرضية تحققها تجربة ما.

بناء الجملة لاستخدامها في البرنامج في الحاسب:

CHITEST(**actual_range**,**expected_range**)

Actual range (النطاق الفعلي) نطاق البيانات الذي يحتوي على الملاحظة المراد اختبارها مقارنةً بالقيم المتوقعة.

Expected range (النطاق المتوقع) نطاق البيانات الذي يحتوي على نسبة حاصل ضرب مجاميع الصفوف ومجاميع الأعمدة إلى المجموع الكلي.

تنويهات

- إذا كان actual range و expected range لهما عدد مختلف من نقاط البيانات، تقوم CHITEST بإرجاع قيمة الخطأ #N/A.
- يحسب اختبار γ2 أولاً إحصائية γ2 ثم يلخص فرق القيم الفعلية عن القيم المتوقعة.

مثـــــــال:

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B	C
	رجال (فعلي)	**نساء (فعلي)**	**الوصف**
	58	35	موافقة
1	11	25	محايد
2	10	23	عدم موافقة
3			
4	**رجال (متوقع)**	**نساء (متوقع)**	**الوصف**
5	45.35	47.65	موافقة
6	17.56	18.44	محايد
7	16.09	16.91	عدم موافقة
8	**الصيغة**	**وصف (الناتج(**	
	=CHITEST(A2:B4,A6:B8)	إحصائية γ2 للبيانات أعلاه هي 16.16957 مع درجتين من درجات الحرية.	

مجال الثقة لوسط مجموعة بيانات**CONFIDENCE**

إرجاع فترة الثقة لوسط مجموعة بيانات. فترة الثقة هي النطاق الواقع على أي من جانبي وسط مجموعة البيانات. فعلي سبيل المثال، إذا طلبت أحد المنتجات عبر البريد، يمكنك تحديد، بمستوى معين من الثقة أقرب آخر موعد لوصول المنتج.

بناء الجملة

CONFIDENCE(alpha,standard_dev,size)

Alpha (ألفا) مستوى الأهمية المُستخدم في حساب مستوى الثقة. يساوي مستوى الثقة 100*(1 - ألفا)%؛ بمعنى أن ألفا 0.05 تشير إلى مستوى ثقة قدره 95 بالمائة.

Standard_dev (الانحراف المعياري) الانحراف المعياري لمحتوى نطاق البيانات ومفترض أنه معطى.

Size (الحجم) حجم العينة.

تنويهات

- إذا كانت أية وسيطة غير رقمية، تُرجع SERIESSUM القيمة الخطأ !#VALUE.
- إذا كانت alpha ≥ 0 أو إذا كانت alpha ≤ 1، تقوم CONFIDENCE بإرجاع قيمة الخطأ !#NUM.
- إذا كانت standard_dev ≥ 0، تقوم CONFIDENCE بإرجاع قيمة الخطأ !#NUM.
- إذا لم يكن size عدداً صحيحاً، يتم اقتصاصه.
- إذا كان size < 1، تقوم CONFIDENCE بإرجاع قيمة الخطأ !#NUM.
- بفرض أن ألفا تساوي 0.05، نحتاج إلى حساب الناحية الواقعة تحت المنحنى المعياري الطبيعي الذي يساوي (1 - ألفا)، أو 95 بالمائة. تساوي هذه القيمة ± 1.96. ولهذا تكون فترة الثقة:

مثـــــــــــال:

بفرض أننا قد لاحظنا في عينة تتكون من 50 راكباً أن متوسط طول الرحلة إلى العمل هو 30 دقيقة مع انحراف معياري للمحتوى 2.5. فيمكن الثقة بنسبة 95 بالمائة أن وسط المحتوى يقع في الفاصل الزمني:

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
1	**بيانات**	**الوصف**
2	0.05	مستوى الأهمية

3	2.5	الانحراف المعياري للمحتوى
4	50	حجم العينة

الصيغة	وصف (الناتج)
=CONFIDENCE(A2,A3,A4)	فترة الثقة لوسط مجموعة البيانات .بمعنى آخر، أن متوسط طول الرحلة إلى العمل يساوي 30 ± 0.692951 دقيقة، أو من 29.3 إلى 30.7 دقيقة (0.692951).

توزيع الاحتمال F : **FDIST**

لإرجاع توزيع الاحتمال F. يمكنك استخدام هذه الدالة لتحديد ما إذا كان هناك درجات للاختلاف بين مجموعتي البيانات. على سبيل المثال، يمكنك فحص نقاط الاختبار للفتيان والفتيات المتقدمين لمدرسة ثانوية جديدة وتحديد إذا كانت التباين بين الفتيات مختلف عن ذلك الموجود بين الفتيان .

بناء الجملة

FDIST(**x**,**degrees_freedom1**,**degrees_freedom2**)

X القيمة التي يتم تقييم الدالة عندها.

Degrees_freedom1 (درجات الحرية 1) هي بسط درجات الحرية.

Degrees_freedom2 (درجات الحرية 2) هي مقام درجات الحرية.

تنويهات

- إذا كانت أي وسيطة منها غير رقمية، تقوم FDIST بإرجاع قيمة الخطأ !VALUE#.
- إذا كانت x سالبة، تقوم FDIST بإرجاع قيمة الخطأ !NUM#.
- إذا لم تكن (درجات_الحرية 1) degrees_freedom1 أو (درجات_الحرية 2) degrees_freedom2 أعداداً صحيحة، يتم اقتصاصها.
- إذا كانتdegrees_freedom1 < 1 أو degrees_freedom1 ≥ 10^10، تقوم FDIST بإرجاع قيمة الخطأ !NUM#.

- إذا كانت degrees_freedom2 < 1 أو degrees_freedom2 ≥ 10^10، تقوم FDIST بإرجاع قيمة الخطأ #NUM!.
- يتم حساب FDIST كـ FDIST=P(F >x)، حيث F متغير عشوائي له التوزيع F.

مثـــــــــال:

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
	بيانات	**الوصف**
1	15.20675	القيمة التي يتم تقييم الدالة عندها
2	6	بسط درجات الحرية
3	4	مقام درجات الحرية
4	**الصيغة**	**وصف (الناتج)**
	=FDIST(A2,A3,A4)	Fاحتمالية التوزيع حسب الشروط الموضحة أعلاه(0.01)

اختبار_F **FTEST**

لإرجاع ناتج اختبار_F. يقوم اختبار F بإرجاع الاحتمال وحيد الطرف الذي تتباين فيه Array1 و Array2 بشكل غير مختلف اختلافاً كبيراً. استخدم هذه الدالة لتحديد ما إذا كانت عينتان بهما تباينات مختلفة. على سبيل المثال، عند معرفة درجات اختبارات من المدارس الحكومية والمدارس الخاصة، يمكنك اختبار ما إذا كانت هذه المدارس بها مستويات مختلفة من تباين درجات الاختبارات.

بناء الجملة

FTEST (array2،array1)

Array1 (مصفوفة 1) مصفوفة أو نطاق البيانات الأول.

Array2 (مصفوفة 2) مصفوفة أو نطاق البيانات الثاني.

تنويهات

- يجب أن تكون الوسائط إما أرقاماً أو أسماء، أو مصفوفات، أو مراجع تحتوي على أرقام.
- إذا احتوت وسيطة مصفوفة أو مرجع على نص، أو قيم منطقية، أو خلايا فارغة، يتم تجاهل تلك القيم؛ رغم ذلك يتم تضمين الخلايا التي تحتوي على قيمة الصفر (0).
- إذا كان عدد نقاط البيانات في array1 أو array2 أقل من 2، أو إذا كان تباين array1 أو array2 صفراً، تقوم FTEST بإرجاع قيمة الخطأ !DIV/0 #.

مثـــــــال :

1. قم بإنشاء مصنف أو ورقة عمل فارغة.
2. حدد المثال في موضوع التعليمات. لا تحدد رؤوس الصفوف أو الأعمدة.
3. اضغط CTRL+C
4. في ورقة العمل، حدد خلية A1، واضغط على CTRL+V.
5. للتبديل بين عرض النتائج وعرض الصيغ التي تقوم بإرجاعها، اضغط `+CTRL، أو في القائمة **أدوات**، أشر إلى **تدقيق الصيغة**، ثم انقر فوق **وضع تدقيق الصيغة**.

	A	B
1	**مجموعة بيانات 1**	**مجموعة بيانات 2**
2	6	20
3	7	28
4	9	31
5	15	38
6	21	40