الإحصاء الاجتماعي

د. عبدالله بن عمر النجار

المحاضرة الأولى

مراجعة لما تم تناوله في مقرر مبادئ الإحصاء

أساليب إجراء البحث الميداني

هناك سؤال مهم لابد من الإجابة عليه وهو:

هل تشمل الدراسة جميع مفردات المجتمع الإحصائي أم سيطبق على جزء منه؟

حالة اعتماد البحث على دراسة جميع مفردات المجتمع الإحصائي يسمى ذلك أسلوب الحصر الشامل

أما إذا أعتمد البحث على دراسة جزء فقط من مفردات المجتمع الإحصائي يسمى ذلك أسلوب العينة

اسلوب الحصر الشامل

يمكنا هذا الأسلوب من الحصول على كافة البيانات والمعلومات عن كافة مفردات المجتمع الإحصائي وبالتالي فإن النتائج التي نحصل عليها لا يوجد بها تحيز ولا تحتاج لتعديل لكنها تحتاج إلى وقت وجهد كبيرين

اسلوب العينات

يبدوا هذا الأسلوب على العكس من أسلوب الحصر الشامل حيث تقتصر الدراسة فيه على جزء من المجتمع الإحصائي، لذا فهذا الأسلوب يوفر الوقت و الجهد و التكاليف ويصلح للمجتمعات غير المحدودة. إلا ان أهم عيوب هذا النوع هو ما يسمى بخطأ التحيز **Sampling Bias** .

أقسام مجتمع البحث

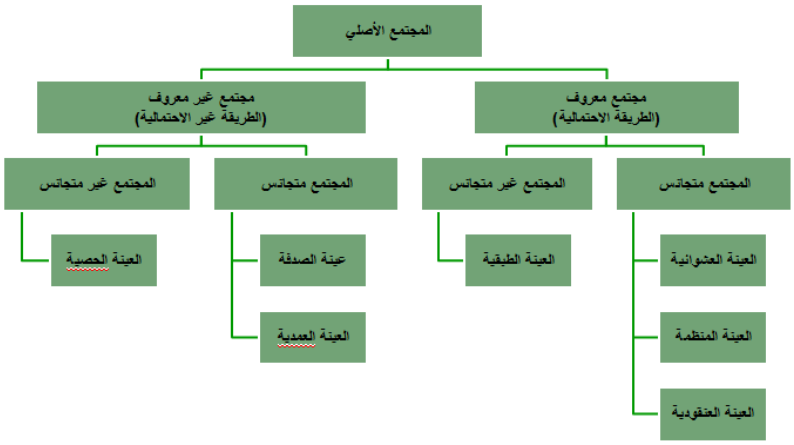
قسم بعض العلماء مجتمع البحث الى قسمين:

- المجتمع الكلي للبحث
- المجتمع الذي يمكن التعرف عليه

مجتمع البحث هو مصطلح علمي يراد به كل من يمكن أن تعمم عليه نتائج البحث

عينة البحث بأنها جزء من المجتمع اختير بطريقة علمية بشرط ان تمثل المجتمع ككل

طرق اختيار العينات

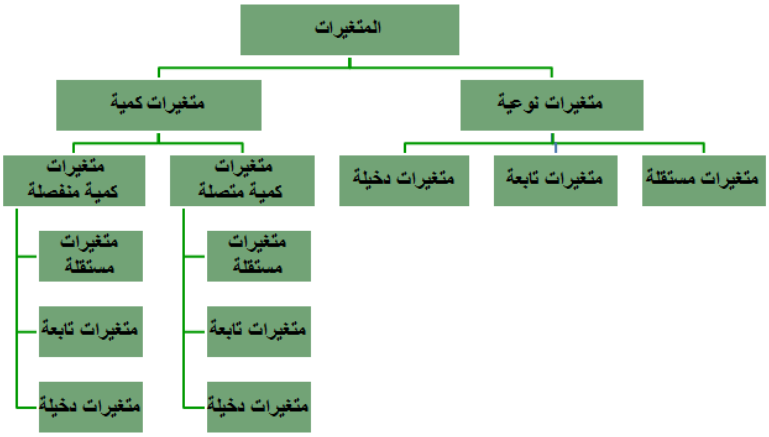


المتغير والثابت في البحث العلمي

المتغير: هو أي خاصية أو صفة سواء للأفراد أو الأشكال والتي تختلف من شخص لآخر ومن وقت لآخر مثل الطول، الذكاء ، التحصيل ويعمل الباحث على دراستها وقياسها.

الثابت: هي الصفات أو الظواهر التي لا تتغير، أو أي صفة أو خاصية تأخذ صفة واحدة ومن الممكن أخذ متغير وتحويله الى ثابت مثل درجة الحرارة في الغرفة. والباحث يسعى الى تثبيت عدد من المتغيرات في دراسته للتخلص من تأثيرها .

تصنيف المتغيرات



الخطوات الواجب مراعاتها بعد جمع البيانات

هناك عدد من الخطوات يجب على الباحث مراعاتها بعد جمع البيانات منها :

- تسجيل البيانات
- ترميز البيانات

1- الترميز الرقمي أو العددي

2- الترميز الأبجدي أو الحرفي

3- الترميز الأبجدي الرقمي

تصنيف البيانات

مراجعة وتنقية البيانات

طرق عرض البيانات

وهناك عدة طرق لعرض وتبويب البيانات الا أن من أبسط تلك الطرق للتعبير عن البيانات هى أن تدمج هذه البيانات في صيغة كتابية إلا أن هذه الطريقة يشوبها الكثير من العيوب

أما الطرق الفنية في عرض البيانات الاحصائية فهي:

- العرض الجدولى للبيانات
- العرض البيانى للبيانات

وسوف نتناول فى هذه المحاضرة العرض الجدولى للبيانات بينما نتعرض للعرض البيانى للبيانات فى المحاضرة التالية إن شاء الله تعالى.

أولا: العرض الجدولي

ويقصد بالعرض الجدولى للبيانات أن يتم تلخيص البيانات محل الدراسة وتصنيفها فى صورة جداول تعبر عن القيم التى أخذها المتغير من خلال البيانات التى جمعها و تكرار كل قيمة من تلك القيم.

أهمية الجداول الاحصائية:

- **تعبر عن الحقائق الكمية المعروضة بعدد كبير من الارقام في جداول بطريقة منظمة**
- **تلخيص المعلومات الرقمية الكثيرة العدد، المتغيرة القيم، مما يسهل التعرف عليها.**
- **الاستيعاب وبسهولة عدد كبير من الموضوعات**
- **اظهار البيانات بأكبر وضوح ممكن وأصغر حيز مستطاع**

تكوين الجداول:

تتكون اجزاء الجدول مما يلي:

رقم الجدول: يجب ان يرقم كل جدول حتى تسهل الاشارة اليه.

العنوان: يجب أن يعطي كل جدول عنوانا كاملا لتسهيل مهمة استخراج المعلومات منه، ويجب أن يكون هذا العنوان واضحا قصيرا بقدر الامكان، ويستخدم في بعض الاحيان عنوان توضيحي لبعض الجداول وذلك من أجل إعطاء معلومات إضافية عن بيانات الجدول.

الهيكل الرئيسي: ويتكون هيك الجدول من أعمدة وصفوف، ويعتبر ترتيب المعلومات في الاعمدة والصفوف أهم خطوة في تكوين الجدول.

العمود: إن كل جدول يتكون من عمود أو اكثر ويوجد تكل عمود عنوان يوضح محتوياته.

الحواشي: قد يحتوي الجدول على مفردات بيانات لا ينطبق عليها عنوان الجدول أو عنوان العمود، ففي هذه الحالة تستعمل الحواشي لتوضيح ذلك وذلك اما بترقيم الملاحظات او باستعمال علامة (*) .. الخ.

المصدر: قد تؤخذ بيانات الجدول من مصادر جاهزة لذلك يجب إظهار المصدر في أسفل الجدول حتى يمكن الرجوع اليه عند الحاجة.

أنواع الجداول الاحصائية:

تقسم الجداول تبعا لدرجة تعقيدها الى:

جداول بسيطة: وفيها يتكون كل من موضوع الجدول ومادته من بضع أسطر وخانات تتعلق بالتقسيمات الزمانية (أي الأمور التي يتناولها الجدول أمور تتسلسل حسب السنوات) أو المكنية (أي توزيع الظاهرة حسب المكان) أو مؤشرات وصفية بسيطة وبأرقام بسيطة أيضا.

جداول التوزيع التكرارى: وفيها تكون المعطيات مجمعة في فئات بمؤشر أو متغير واحد، ولكل فئة تكراراتها الخاصة عند ذلك المؤشر

جدول التوزيع التكراري المتجمع: وفيه تجمع التكرارات على التوالي من أحد طرفي الجدول الى طرفة الآخر فنحصل على التكرار الكلي (مجموعة التكرارات)، (فاذا بدأ من أعلى الى أسفل الجدول) سمي جدول تكراري متجمع صاعد، (واذا بدأ من أسفل الى أعلى الجدول) سمي جدول تكرار متجمع نازل أو هابط.

الجداول المزدوجة أو المركبة: وهي الجداول التي تتكون من متغيرين أو أكثر، وهذه المتغيرات قد توزع على أعمدة وحقول الجدول بصورة نظامية، تعبر عن الافكار العلمية التي يريد الباحث توضيحها توضيحا عدديا.

وقد أوضحنا في المحاضرة السابقة ما هي **البيانات** وعرفناها بأنها [**هي مجموعة المشاهدات أو القياسات التي تخص ظاهرة معينة تحت الدراسة**]

وعرفنا كذلك **المتغير** على أنه تلك الكمية التي نقوم بمشاهدتها أو قياسها ، كما ذكرنا أن البيانات إما أن تكون : نوعية أو كمية ، حيث : وتتوقف عملية تبويب وتصنيف البيانات على نوع البيانات الإحصائية المراد التعامل معها ودراستها والتى يمكن تقسيمها من حيث طريقة إعداد الجداول إلى التالي:

(ب – 1) **بيانات كمية متصلة :** وفيها يمكن أن يأخذ المتغير أي قيمة بين قيمتين (أي بيانات يمكن أن**تُقاس** ولا**تُعد** ، مثل :

- أطوال الطلاب في إحدى المدارس .
- أوزان العاملات بإحدى المصانع .
- الدخل السنوي لمنسوبي مؤسسة معينة .

وغيره من مثل هذه الأمثلة .

(ب – 2) **بيانات كمية متقطعة :** وفيها يمكن أن يأخذ المتغير قيمة رقم صحيح بدون كسور (مثلا إما **10** أو **11** وليس أي قيمة بينهما)، وبتعبير آخر هي بيانات يمكن أن**تُعد** ولا**تُقاس** ، مثل

عدد طلاب الفصول المختلفة في مدرسة ما

والبيانات المنفصلة إما أن تكون نوعية أو كمية متقطعة

ثانيا: البيانات الكمية المتصلة:

وفيها يتم توزيع البيانات في جدول تكراري ذوفئات، ويتم ذلك من خلال اتباع الخطوات التالية:

الخطوة الأولى: تحديد عدد الفئات

الخطوة الثانية: تحديد طول الفئة

الخطوة الثالثة: تعيين حدود الفئات

الخطوة الرابعة: توزيع التكرارات على الفئات

وهناك عدة ملاحظات يجب الإنتباه إليها عند عمل جدول التوزيع التكرارى لبيانات المتغير الكمى المتصل:

1– إن تحديد عدد الفئات يتوقف على أمور عدة منها:

- عدد المفردات محل الدراسة
- انتظام وتوزيع تلك البيانات
- طبيعة بيانات المشكلة محل الدراسة

2– طول الفئة لا بد أيضاً من تحديده بعناية حيث يمثل الوجه الآخر للعملة مع عدد الفئات، فمن الأفضل أن يكون تحديده بطريقة تجعل مركز الفئة قريباً من تركز البيانات بتلك الفئة بقدر الإمكان حيث يعبر مركز الفئة عن قيمة كل مفردة من المفردات التي تنتمي لتلك الفئة

3- أن تكون حدود الفئات واضحة بحيث لا يكون هناك أي تداخل فيما بينها.

ومن هنا يمكن إعداد جداول التوزيعات التكرارية للمتغيرات المتصلة بثلاث صور هى:

- **الجداول التكرارية المنتظمة**
- **الجداول التكرارية غير المنتظمة**
- **الجداول التكرارية المفتوحة**

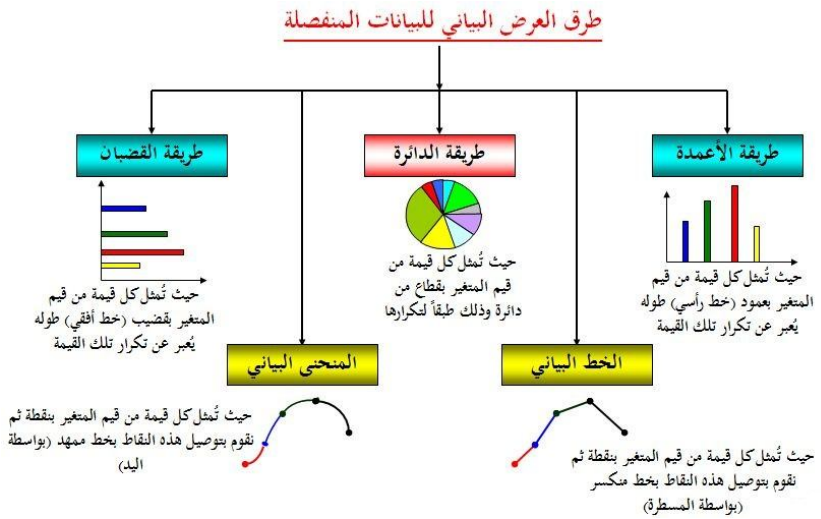
ثانيا: العرض البياني

تعريف الرسوم البيانية:

هي وسيلة مفيدة وفعالة لتوضيح وشرح الحقائق الرقمية وابراز العلاقة بين المتغيرات، واستقراء اتجاهاتها العامة بأسلوب يسهل فهمه وتذكره بمجرد النظر.

وتنطبق القواعد التي ذكرناها في العرض الجدولي على الرسوم البيانية، اذ يجب أن يرقم كل رسم ، ويعنون، ويمكن أن يستعمل الحواشي والمصدر وغيرها..

تختلف الرسوم البيانية حسب طبيعة ونوع البيانات المراد عرضها فاذا كانت البيانات اسمية أو رتبية (أي منفصلة) فإننا نستخدم أحد الأشكال البيانية التالية



1- طريقة الأعمدة:

ويتم عرض البيانات من خلال هذا الأسلوب من خلال عدة أنواع من الأعمدة البيانية وهي:

أ- الأعمدة البيانية البسيطة :

وهي عبارة عن مجموعة من الأعمدة الرأسية أو المستطيلات المتساوية القاعة والتي تتناسب ارتفاعاتها مع البيانات التي تمثلها

تمارين

أرادت باحث معرفة العلاقة بين حب الاستطلاع لدى الطلاب في السنوات الابتدائية وحل المسائل الرياضية، فاختار عشوائيا طلاب السنة الثالثة ثم اختار منهم عشوائيا 200 طالب، ثم قام بصياغة الفرضية التالية:

"لا توجد علاقة ذات دلالة إحصائية بين حب الاستطلاع وحل المسائل الرياضية"

ثم قام بتطبيق اختبار عليهم وذلك للحصول على البيانات اللازمة لاستنتاج العلاقة واتخاذ قرارات في ضوء ذلك المطلوب :

- **ما نوع الإحصاء الذي استخدمه الباحث في هذه الدراسة؟ علل ذلك ؟**
- **حدد مجتمع البحث في هذه الدراسة ، وما نوعه ؟**
- **حدد عينة الدراسة في هذه الدراسة ، وما نوعها؟**
- **حدد المتغير المستقل في هذه الدراسة ، وما نوعه ؟**
- **حدد المتغير التابع في هذه الدراسة ، وما نوعه ؟**
- **حدد في تصورتك المتغيرات الدخيلة التي من الممكن أن تؤثر على هذه الدراسة ؟**
- **حدد الفرضية التي يحاول الباحث اختبارها في هذه الدراسة ، وما نوعها ؟**
- **ما الوسيلة التي استخدمها الباحثة لجمع البيانات في هذه الدراسة ؟**

س ١ : المدى لمجموعة من البيانات المنفصلة هو :

○ أكثر القيم تكراراً في البيانات ○ أكبر قيمة في البيانات

○ أصغر قيمة في البيانات ● الفرق بين أكبر وأصغر قيمتين في البيانات

س ٢ : الجدول المرافق يبين درجات ٢٠ طالباً في أحد المقررات الدراسية :

الدرجة	92	93	94	95	96	97	98	99	100
التكرار	2	2	3	6	1	1	1	3	1

(أ) عدد الطلاب الحاصلين على 94 فأقل هو :

○ 3 ○ 0.15 ○ 4 ● 7

(ب) عدد الطلاب الحاصلين على درجة أقل من 94 هو :

○ 3 ○ 0.15 ● 4 ○ 7

(ج) نسبة الطلاب الحاصلين على 94 فأقل هي :

● 0.35 ○ 35% ○ 4 ○ 7

(د) النسبة المئوية للطلاب الحاصلين على 94 فأقل هي :

○ 0.35 ● 35% ○ 4 ○ 7

هامش للإجابة

(٢-أ) $2 + 2 + 3 = 7$

(٢-ب) $2 + 2 = 4$

(٢-ج) $\frac{7}{20} = 0.35$ خذ بالك : المطلوب نسبة (وليس نسبة مئوية)

(٢-د) $0.35 \times 100 = 35\%$ أيوه .. ، ده بقى نسبة مئوية

س ١ : الجدول المقابل يبين الجدول التكراري لأعمار عدد من الممرضات (لأقرب سنة) اللاتي تعملن في أحد أقسام إحدى المستشفيات ، من هذا الجدول أجب على الأسئلة التالية :

المتغير (العمر) x	التكرار (العدد) f	الزاوية المركزية
20	20	72^o
25	?	36^o
30	30	?
35	?	?
	$\sum f$	

(أ) عدد الممرضات ذات العمر 25 سنة هو :

● 10 ○ 20 ○ 30 ○ 40

(ب) الزاوية المركزية المناظرة للعمر 30 سنة هي :

○ 36^o ○ 72^o ● 108^o ○ 144^o

(ج) الزاوية المركزية المناظرة للعمر 35 سنة هي :

○ 36^o ○ 72^o ○ 108^o ● 144^o

(د) عدد الممرضات الكلي [أي مجموع التكرارات $\sum f$] هو :

○ 95 ● 100 ○ 105 ○ 110

هامش للإجابة

(١-أ) هناك تناسب بين التكرار والزاوية المركزية ، إذن : $72 \times ? = 36 \times 20$ ، $\therefore ? = 10$

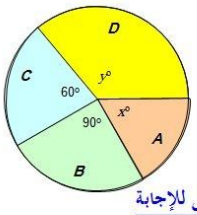
(١-ب) بنفس الأسلوب السابق $72 \times 30 = ? \times 20$ ، $\therefore ? = 108^o$

(١-ج) مجموع الزوايا المركزية يجب أن يكون 360^o

$\therefore 72 + 36 + 108 + ? = 360$ ، $\therefore ? = 144^o$

(١-د) هناك أكثر من طريقة أميزها الأسلوب المتبع في الجزئين (أ) ، (ب) :

$360 \times 20 = 72 \times \sum f$ ، $\therefore \sum f = 100$



س ٢ : الشكل المقابل يبين مبيعات أربع شركات ***A , B , C , D*** (لبيع لعب الأطفال) وذلك خلال عيد الفطر المبارك ، فإذا كان عدد اللعب الكلي التي تم بيعها بواسطة هذه الشركات هو **5400** لعبة ، أجب على الأسئلة التالية :

(أ) النسبة المئوية لمبيعات الشركة ***B*** هي :

○ 60% ○ 40% ○ 30% ● 25%

(ب) عدد اللعب التي باعتها الشركة ***B*** هو :

○ 2700 ○ 2250 ○ 900 ● 1350

(ج) عدد اللعب التي باعتها الشركتان ***A , D*** معاً هو :

○ 900 ○ 2250 ● 3150 ○ 1350

هامش للإجابة

(٢-أ)

$360 \times ? = 90 \times 100$

$? = 25\%$

100%	360°
?	90°

(٢-ب)

$\frac{25}{100} \times 5400 = 1350 \rightarrow \sum f$

(٢-ج)

الزاوية المركزية المناظرة لمبيعات الشركتين معاً تساوي

$360 - (90 + 60) = 210°$

5400	360°
?	210°

$360 \times ? = 210 \times 5400$

$? = 3150$

المحاضرة الثانية

أدوات جمع البيانات

Data Collection Instruments

يقصد بأداة جمع البيانات الوسيلة التي تتم بواسطتها عملية جمع البيانات بهدف اختبار فرضيات البحث أو الإجابة عن تساؤلاته. ويتوقف اختيار الأداة المناسبة لجمع البيانات اللازمة والتي ستستخدم في إجراء بحث معين على نوعية البحث نفسه وطبيعته ، وعلى الهدف من تطبيقه ، وعلى نوعية المفحوصين وخصائصهم ...الخ ، وقد يستخدم الباحث أداة واحدة فقط لجمع البيانات التي يحتاج إليها في بحثه ، وقد يستخدم أكثر من أداة إذا وجد مبررا لذلك. وتجدر الإشارة إلى أن خطوة جمع البيانات في البحث تعتبر من الخطوات الأساسية التي يبدأ منها عمل الباحث ، لذا فالهدف النهائي من إعداد وسائل وأدوات جمع البيانات هو الحصول على تلك المعلومات التي تخدم في تحقيق أغراض البحث ودراسة مشكلته ، وإيجاد الحلول المناسبة له .

تعني أدوات جمع البيانات تلك الوسائل التي تجمع بها المعلومات الازمة لإجابة أسئلة البحث أو اختبار فروضه . وتجمع المعلومات بواسطة واحدة أو أكثر من الأدوات التالية :

1- الاستبيان : وهو أداة لجمع المعلومات المتعلقة بموضوع بحث محدد عن طريق استمارة يجري تعبئتها من قبل المستجيب .

2- المقابلة : وهي تفاعل لفظي يتم بين شخصين في موقف مواجهة حيث يحاول أحدهما وهو القائم بالمقابلة أن يستثير بعض المعلومات أو التعبيرات لدى المبحوث والتي تدور حول آرائه وأفكاره لاستغلالها في بحث علمي أو للاستعانة بها في التوجيه والتشخيص والعلاج .

3- الملاحظة : تعني الانتباه المقصود والموجه نحو سلوك فردي أو جماعي معين بقصد متابعته ورصد تغيراته ليتمكن الباحث بذلك من وصف السلوك وتحليله وتقويمه .

4- الاختبارات المقننة : هي تلك الاختبارات التي تحافظ على صدقها (أي قياس ما أعدت لقياسه) وثباتها (أي الوصول الى النتائج نفسها لو تكرر تطبيقها) خاصة اذا اتبعت التعليمات المصاحبة له .

وسنركز في هذه المحاضرة على الإستبانة كأحد وسائل جمع البيانات المهمة في البحوث والدراسات الإجتماعية

تعتبر الاستبانة من الأدوات البحثية شائعة الاستخدام في أغلب البحوث والدراسات النفسية والاجتماعية ، وخاصة تلك التي تركز على جمع معلومات وبيانات متعلقة بمعتقدات ورغبات المستجيبين ، وكذلك الحقائق التي هم على علم بها . ولهذا نوه (كنث **Kenneth 1987**) إلى أن الاستبانة تستخدم بشكل رئيس في مجال الدراسات التي تهدف إلى استكشاف حقائق عن الممارسات الحالية واستطلاعات الرأي وميول الأفراد

وإذا كان الأفراد الذين يرغب الباحث في الحصول على بيانات بشأنهم متواجدين في أماكن متفرقة ، فإن وسيلة الاستبانة تمكنه من الوصول إليهم جميعا في وقت محدد ، وبتكاليف معقولة ، كذلك تعتبر الاستبانة وسيلة ناجحة لدراسة الحياة الشخصية للأفراد وخاصة تلك الجوانب من الحياة الخاصة التي لا يمارسها الأفراد إلا عندما ينفردون بأنفسهم بعيدا عن أعين المراقبين .

لذلك فإن تصميم استمارة الاستبانة يحتاج إلى عناية فائقة ، إذ تعتمد عليها مدى صحة ودقة البيانات المجموعة بواسطتها ، فيتطلب ذلك دراسة واسعة ، وإلماما تاما بأوضاع جمهور البحث من حيث نوعيتهم وخصائصهم ومدى استجابتهم .

وفي هذا الصدد اشار (ميلر **1991Miller**) إلى إنه يجب على الباحث مراعاة بعض القواعد عند بناء استمارة الاستبانة منها ما يتصل بشكلها وتنسيقها وملاءمتها ، ومنها ما يتعلق بصياغة بنودها وحساب صدقها وثباتها .

وإن ما يجب التنويه إليه هنا أن خطوة جمع البيانات في البحث تعتبر من الخطوات الأساسية والتي يبدأ منها عمل الباحث ، لذا فالهدف النهائي من إعداد الاستبانة هو الحصول على تلك البيانات التي تخدم في تحقيق أغراض البحث ودراسة مشكلته ، وإيجاد الحلول المناسبة لها.

خطوات بناء الاستبانة

تعتبر الإستبانة من أفضل و أكثر الطرق استخداما لجمع البيانات و تحليلها في العلوم الإجتماعية (هنيرسون ، موريس وجيبون **Henerson, Morris & Gibbon ، 1987**) ، لكن المتتبع لهذه الأداة يجد أن هناك أخطاء ونقص في إعدادها وضعف في بنائها مما قد يؤدي إلى عدم الدقة في جمع المعلومات بواسطتها ، وبناء عليه فإن النتائج المتحصل عليها تكون أقرب إلى الخطأ منها إلى الصواب (رودن ، **Roden 1998**) .

ومن هنا يمكن القول أن بناء استبانه على أسس علمية يؤدي إلى نتائج صادقة ودقيقة يحتاج إلى إتباع خطوات علمية محددة تتمثل فيما يلي :

أولا : الإطلاع على الدراسات والبحوث السابقة :

إن الهدف الأساسي من الرجوع إلى الدراسات السابقة هو تكوين فكرة عامة عن الظاهرة موضع الدراسة ، ومحاولة تحديد مشكلة البحث ، والتعرف على ما تم التوصل إليه في هذا الموضوع ، والعمل على حصر الموضوعات التي ستتضمنها الاستبانة ، والمساعدة على تحديد الكثير من فقرات الاستبانة بشكلها النهائي (توماس **Thomas 1999**) .

ثانيا : تحديد الأسئلة الرئيسة للبحث موضع الدراسة :

بعد الاطلاع على الدراسات السابقة وقبل الشروع في بناء الإستبانة لابد من تحديد الأسئلة الرئيسة التي يرغب الباحث الإجابة عليها (ديرشوسكي **Dereshiwsky 1993**) . ومن الخصائص الأساسية لهذه الأسئلة أن تكون محددة ، واضحة ، دقيقة ..الخ ، كذلك لابد أن تكون هذه الأسئلة محددة لنوع المعلومات التي من أجلها تبنى الإستبانة ، وقد تصاغ هذه الأسئلة بصورة عامة ، وقد تصاغ بصورة محددة . وإن ما يجب التنويه إليه هنا أن هذه الخطوة لا تتعلق ببناء فقرات الإستبانة ، بل هي تعتبر بمثابة الإطار العام للإستبانة ، والتركيز على هذه الأسئلة الأساسية يساعد الباحث على عدم إضافة فقرات ليس لها علاقة بالبحث .

وهناك عدد من الخطوات الأساسية التي تساعد الباحث على كتابة الأسئلة الرئيسة للبحث وتحديدها وهي :

- الرجوع إلى الدراسات السابقة من كتب وبحوث ورسائل علمية .
- مناقشة الموضوع مع المتخصصين .
- مناقشة الموضوع مع صناع القرار .
- النزول إلى الميدان للإطلاع على الواقع الفعلي للظاهرة موضع الدراسة .

ثالثا : تحديد الأسئلة الفرعية المبنية على الأسئلة الرئيسية

عادة الأسئلة الرئيسة للدراسة تحوي بعض الكلمات العامة والتي يمكن أن تحمل أكثر من معنى ، مثلا " التدريس ، التعليم الجماعي ، المهارات الإدارية ، تحقيق الذات ، إدارة الصف ، الاتجاهات الخ ، هذه الكلمات تحتاج إلى نوع من التحديد والتعريف ، وعادة يتم هذا من خلال التعريف الإجرائي لهذه الكلمات إضافة إلى وضع الأسئلة الفرعية التي تتناول بالتفصيل الموضوعات المندرجة تحت هذه الكلمات .

- **و قد أشار (كوكس Cox، 1997) أن الأسئلة الفرعية لابد أن تتصف بالآتي :**
  - **أن تكون قابلة للقياس**
  - **أن تكون دقيقة و تعالج موضوعا محددا .**
  - **أن تكون على مستوى واحد من الصياغة .**

- **و يمكن استنباط هذه الأسئلة الفرعية من خلال :**
  - **الرجوع إلى الكتب ، البحوث ، الدراسات العلمية ذات العلاقة بالموضوع .**
  - **الحوار و المناقشة مع المتخصصين .**

رابعا : الدراسة الاستطلاعية :

بعد الاطلاع على الدراسات السابقة وتحديد الاسئلة الرئيسة والفرعية ذات العلاقة بموضوع البحث يأتي دور الدراسة الاستطلاعية وذلك لوضع المشكلة المدروسة ضمن الاطار الحضاري لها ، لأن الدراسات السابقة كما هو معروف قد اجريت في مجتمع غير مجتمع البحث. فالدراسة الاستطلاعية تساعد على التأكد من أن التصميم الذي انتهى إليه الباحث واقعي ويمكن تنفيذه

زد على ذلك أن الدراسة الاستطلاعية تمكن الباحث من التالي :

- **التعرف على أفضل الأساليب لمخاطبة أفراد العينة .**
- **تحديد الموضوعات التي سوف تدور حولها اسئلة الاستبانة .**
- **تحديد شكل الأسئلة التي تدور حولها تلك الموضوعات .**
- **تحديد الصورة الاجمالية للاستبانة .**
- **تحديد أفضل ترتيب للأسئلة التي تدور حول موضوعات الاستبانة .**
- **تحديد الصياغة اللفظية للغة الاستبانة .**

خامسا : كتابة فقرات الإستبانة :

من أجل ضمان دقة المعلومات التي تحصل عليها من خلال وسيلة جمع البيانات ، فإنه لابد لفقرات الإستبانة أن تكون دقيقة ومحدده وغير قابلة لأكثر من تفسير (فنك Fink 1995) . إن الأسئلة الدقيقة تضمن لنا إجابة المستجيب على فقرات الإستبانة بطريقة تتوافق مع الهدف الذي وضعت من أجله . إن وضوح فقرات الإستبانة لايتيح للمستجيب قراءة الفقرة مرة واحدة لفهمها فقط بل يساعد ايضا على تقليص الوقت المتطلب لإكمال هذه الإستبانة وبالتالي ضمان إعادتها . يجب على الباحث اثناء كتابة فقرات الإستبانة أن يضع نفسه موضع المستجيب ، فالبعد عن استخدام الكلمات غير المفهومة مطلب اساسي لضمان دقة الإجابة.

ولقد اشار (فنك Fink 1995 ، وكوكس Cox 1996) إلى بعض الإرشادات التي تساعد الباحث على كتابة فقرات الإستبانة بطريقة جيدة:

- **أن تكون الفقرة واضحة وبسيطة .**
- **تجنب استخدام المصطلحات العامة ، والكلمات الغامضة (اكتب باللغة التي يفهمها المستجيب) .**
- **استخدام الاسئلة القصيرة المحددة المعنى .**
- **صياغة العبارات بصورة لاتوحي بالتحيز إلى أحد الاتجاهات .**
- **مراعاة عدم وضع اسئلة تمس شعور المفحوص أو عقائده .**
- **صياغة الاسئلة والعبارات بصورة تسمح بمعرفة شدة الاستجابة .**
- **تجنب صياغة الاسئلة بالنفي .**
- **تجنب الاسئلة التي تحوي على فكرتين .**

سادسا : الشكل العام للاستبانة :

تعرف الإستبانة بأنها عبارة عن استمارة تضم مجموعة من الأسئلة توجه للأفراد بغية الحصول على بيانات معينة ، وهذه الأسئلة التي تتضمنها الإستمارة يمكن تقسيمها إلى نوعين تبعاً لأسلوب الحصول على البيانات :

- **الأسئلة المباشرة: وهي التي تهدف إلى الحصول على المعلومات بطريقة واضحة وصريحة .**
- **الأسئلة غير المباشرة: وهي التي يمكن من خلال الإجابة عنها استنتاج البيانات المطلوبة .**

وكذلك يمكن تقسيمها إلى نوعين وفقاً لأسلوب تقنينها :

- **أسئلة مغلقة: وهي التي تحدد إجابة الفرد في إطار المتغيرات المحددة كأن تكون نعم ولا ، أو موافق ، وغير موافق .. الخ .**
- **أسئلة مفتوحة: وهي التي تسمح للمستجيب بالاجابة الحرة دون التقيد بإجابات معينة .**

ولا بد أن يراعي الباحث عند اعداده الإستبانة جملة من الأمور تتعلق بالشكل العام للإستبانة منها:

- **طول الاستبانة : يجب أن يكون طول الاستبانة معقول ، فعندما تكون الاستبانة طويلة فهذا يؤدي إلى عدم الإجابة الكاملة على بنودها. لذا ينصح بأن تكون المدة المحددة للإجابة على الاستبانة من 10 إلى 12 (كوكس cox 1996) .**
- **تصنيف الفقرات : العبارات ذات الإستجابة الموحدة من الأولى أن تكون مع بعض ولذا يجب مراعاة عدم تشتيت المستجيب في الانتقال من شكل استجابة إلى شكل آخر .**
- **استغلالية الصفحات : يجب أن لا توزع المعلومة المراد الإجابة عليها على أكثر من صفحة حتى لا يؤدي ذلك إلى ازعاج المستجيب في الرجوع إلى معلومات في صفحات سابقة .**
- **المسافات : يجب عدم ضغط المعلومات والفقرات في صفحات محددة مما يجعلها مزدحمة وغير واضحة للمستجيب .**
- **وضوح الخط المستخدم : ينبغي أن يكون الخط المستخدم في كتابة فقرات الإستبانة واضح ومقروء للجميع من حيث الخط ومقاسه .**
- **المراجعة اللغوية لمحتويات الاستبانة : ينبغي على الباحث المراجعة اللغوية لجميع محتويات الاستبانة لأن الخطأ الإملائي قد يؤدي إلى خطأ في الاستجابة مما قد يؤثر على النتائج المتحصلة .**

سابعاً : اختبار الإستبانة :

اختبار الاستبانة يعني التأكد من أنها أصبحت صالحة للإستخدام من حيث المدلول والمحتوى لجمع المعلومات حول المشكلة قيد البحث . وبهذا المفهوم لاختبار الاستبانة يمكن التفريق بين:

- **الاختبار الذي يهدف إلى تصحيح المدلول اللفظي لكل بند من بنود الاستبانة وإزالة ما يمكن أن يؤدي إلى غموض أو عدم معرفة المراد منه. ويتم ذلك من خلال عرض الاستبانة على من لهم خبرة علمية في مجال البحث .**
- **الاختبار الذي يهدف إلى التأكد من مدى صدق الاستبانة ومكانها ويتم ذلك باختيار عدة أشخاص من مجتمع البحث ثم يطلب منهم إجابة الاستبانة ويقاس في ضوء استجابتهم مدى صدق الاستبانة وثباتها .**
- **ا لاختبار الذي يهدف إلى التأكد من مدى جدية المجيب في إجابته للإستبانة ، وذلك من خلال تنويع صياغة سؤال أو أكثر ذي مدلول واحد ليتبين له من خلال مقارنة الاجابة مدى جديتها .**

ثامناً : كتابة تعليمات الإجابة :

بالاضافة إلى الإعتناء بمحتوى وشكل الاستبانة لا بد من تزويد المجيب بتعليمات واضحة للإجابة على بنود هذه الاستبانة ، تكون على شكل رسالة مصاحبة يوضح فيها المشكلة قيد الدراسة بإختصار ، والهدف من بحثها . ومدى أهمية مشاركة المجيب في تحقيق ذلك الهدف .

تاسعاً : توزيع الاستبانة ومتابعتها :

بعد أن يقوم الباحث ببناء الإستبانة واختبارها من حيث سلامة المدلول اللفظي لبنودها وصحة معناها ، وحساب معامل صدقها وثباتها، يتعين عليه اختيار الطريقة المناسبة التي سيستخدمها للتوزيع ومنها :

- **التوزيع المباشر: وهو أن يقوم الباحث بنفس أو من يمثله بتسليم الاستبانة لأفراد العينة .**
- **التوزيع غير المباشر: وفيها يقوم الباحث بإرسال الإستبانة عبر البريد .**

إن التوزيع بكلا الطريقتين يحتاج إلى متابعة حثيثة من قبل الباحث، وذلك لتدني نسبة المستجيبين والتي تعتبر من أهم العقبات التي تقف في طريق الباحثين عند استخدامهم للإستبانة كوسيلة لجمع البيانات

عاشراً : تبويب وترميز بيانات الإستبانة بالطريقة المناسبة :

بعد أن يتم جمع المعلومات من خلال الإستبانة يقوم بمراجعتها وذلك بهدف استبعاد الإستمارات التي لم يجب عليها أولاً ، ثم التأكد من مدى جدية المجيب في إجابته من خلال مراجعة البنود التي وضعت لقياس هذا الأمر، ومن ثم يبدأ عملية التبويب من خلال الآتي:

- **وضع رقم لكل إستبانة .**
- **وضع رقم لكل عبارة أو سؤال .**
- **وضع رقم لكل إجابة من إجابات العبارة أو السؤال .**

حادي عشر : تفريغ معلومات الاستبانة وادخالها بالطريقة المناسبة في الحاسب الآلي :

بعد أن يتم ترميز بيانات الإستبانة وإعطاء رقم لكل استمارة وبند، واجابة ، يتم بعد ذلك تفريغ هذه المعلومات وإدخالها بالطريقة المناسبة في الحاسب الآلي (باستخدام احد البرامج الإحصائية المناسبة) ويجب هنا مراعاة المتغيرات موضع الدراسة وطريقة تحليلها . لأن طريقة إدخال هذه البيانات في الحاسب تؤثر بطريقة مباشرة على النتائج المتحصلة

ثاني عشر : تحليل بيانات الاستبانة :

بعد أن يتم ادخال بيانات الإستبانة في الحاسب الآلي يأتي دور معالجة هذه البيانات معالجة رقمية وذلك من خلال تطبيق أساليب الإحصاء بنوعيه الوصفي والاستنتاجي ، وهنا يحتاج الباحث إلى التأني في الاختيار المناسب للأسلوب الإحصائي لأن ذلك قد يؤثر بطريقة مباشرة على النتائج المتحصلة.

أهم الصعوبات التي تواجه الباحثون في بناء الاستبانة

أولا : الصعوبات التي تواجه الباحثون قبل بناء الاستبانة واستخدامها كوسيلة لجمع البيانات :

1. **اختيار الاستبانة كوسيلة لجمع بيانات الدراسة :**

إن عملية اختيار الاستبانة أو بناءها تحتاج من الباحث أن ينتبه إلى أمرين هامين :

تحديد الهدف المراد من الاستبانة تحقيقه .، فثمة متخصصون يواجهون صعوبات في فهم الهدف الذي وضعت لأجله هذه الاداة أو تلك .

ضرورة تحديد المدرسة الفلسفية أو النفسية التي تكون إطارا للبحث وبناء أدواته وتصميم منهجيته .

لذا ينبغي على الباحث أن يكون دقيقا عند اختياره للاستبانة كأداة لجمع بيانات بحثه .

2. تحكيم الاستبانة من قبل مجموعة من المتخصصين :

بعد أن تأخذ الاستبانة صيغة نهائية صالحة من وجهة نظر الباحث الذي صممها ، يقوم الباحث نفسه بعرضها على نوعين من الخبراء هما :

- الخبراء المنهجيون والذين يدرسون الاستبانة من حيث شكلها العام وتكوينها ، ومن خلال هذه المراجعة ممكن اكتشاف العيوب الفنية في الاستبانة ومحاولة تداركها في الوقت المناسب .
- خبراء في المادة العلمية والذين يقومون بمراجعة المادة العلمية ذات العلاقة بموضوع البحث ، ومن خلال هذه المراجعة يمكن اكتشاف مواطن الضعف أو النقص في الموضوعات الواردة في الاستبانة .

ويجب على الباحث أن يكون موضوعيا عند استشارة الخبراء ، وأن لايتردد في تعديل الاستبانة تبعا لملاحظاتهم عليها .

3. عدم القيام بالدراسة الاستطلاعية :

إن من المشاكل الأساسية التي تواجه الباحثين عند بناؤهم للاستبانة هو عدم القيام بالدراسة الاستطلاعية والتي يهدف من خلالها الباحث إلى تحديد الاتجاه أو الجانب الذي يريد قياسه تحديدا واضحا ، لأن إعداد المقياس كما هو معروف يتطلب الاعتماد على خصائص الجماعة ونوعية المواقف التي تتصل بالاتجاه موضع القياس . فهذا الأمر يتطلب الاتصال ببعض من سوف يطبق عليهم المقياس وذلك عن طريق المقابلات الشخصية لمعرفة أبعاد الاتجاه ومحدداته والمتغيرات التي ترتبط به ، بل وما هو أهم من ذلك جميعا وهو معرفة ما ذا نريد أن نقيس. وإنه من خلال هذه الدراسة الأولية (والتي تفتقدها الكثير من دراسات الباحثين في محيطنا، وهذا يعتبر بلا شك قصور) يتمكن الباحث من جمع عدد كبير من التعبيرات والجمل والتعليقات والصيغ اللفظية التي قد تصلح تماما لتكوين وحدات وبنود الاستبانة .

ثانيا : الصعوبات التي تواجه الباحثون والتي تتعلق بكتابة أسئلة أو عبارات الاستبانة :

1. صياغة الأسئلة أو العبارات بطريقة تؤثر على المستجيب :

إن صياغة الأسئلة بطريقة تؤثر على المستجيب تؤدي إلى رفض الإجابة من قبل المستجيب أو تعمد الإجابة الخاطئة ، لذا يجب أن لا تكون صياغة السؤال من النوع الذي يمس شعور المستجيب أو عقائده ...بصورة تجعله يعترض على السؤال مما قد يدفعه إلى رفض الإجابة على الاستبانة ككل .

2. طول الأسئلة أو العبارات :

يجب على الباحث أن يراعي عند كتابته لأسئلة أو عبارات الاستبانة أن لا تكون طويلة ، لأنه كلما كانت الاسئلة أو العبارات طويلة أدى ذلك إلى إحجام ونفور المستجيب عن الإجابة عليها . فالمستجيب عندما ينظر لسؤال يتكون من عدة أسطر قد يحجم عن إجابته ، وذلك لما يتطلبه من وقت طويل لقراءته واستيعابه ، بينما عندما يكون قصير ، يكون ذلك دافعا قويا للإجابة عليه .

3. عدم الوضوح والدقة في صياغة الأسئلة أو العبارات :

وتظهر هذه المشكلة عندما يستخدم الباحث بعض الكلمات التي قد لا يتفق على مدلولها الباحث والمستجيب مثل (غالبا ، كثيرا ، نادرا ..الخ) ، فما هو غالب أو كثير بالنسبة للباحث قد يراه المستجيب نادرا أو قليلا . أو صياغة العبارات أو الأسئلة بصورة غير محددة ، فهذا لا يمكن المستجيب من معرفة المطلوب تماما . أو صياغة العبارات أو الأسئلة بصورة قابلة للتأويل لأكثر من مفهوم . أو صياغة العبارات أو الأسئلة بطريقة تكون محتاجة إلى عمق في التفكير مما قد يؤدي إلى إهمال المستجيب الإجابة على الاستبانة ، لأن بعض المستجيبين ليس عنده الدافع الذي يدفعه للتفكير العميق . أو استخدام الكلمات التي لا يعرف المستجيب معناها وذلك عند صياغة السؤال أو العبارة ، وهذا بطبيعة الحال يؤدي إلى نتائج مضللة بسبب عدم اختيار الكلمات المناسبة للمستجيب .

وإن ما يعين الباحث على اختيار الكلمات المناسبة مراعاة المستوى التعليمي والثقافي للمستجيب ، فكلمات العبارة أو السؤال الذي يجيب عليه طلبة الجامعة يجب أن تختلف عن كلمات السؤال الذي يجيب عليه طلبة المرحلة المتوسطة مثلا .

4. صياغة الاسئلة أو العبارة بصورة افتراضية :

إن صياغة الاسئلة أو العبارة بصورة افتراضية لا تعكس الواقع للجماعة المفحوصة ، وهذا بطبيعة الحال يعطي نتائج مضللة ، لذا من المفروض أن تكون وحدات الاستبانة حقيقية وليست افتراضية وخاصة في مقاييس الاتجاهات ، فالمطلوب هو أن يعبر المفحوص عما يشعر به فعلا وما يقوم به حقيقة ، وليس عما يجب أن يكون أو من المحتمل أن يحدث .

5. احتواء الأسئلة أو العبارات على فكرتين أو أكثر :

إن صياغة العبارة أو السؤال الذي يحوي على أكثر من فكرة لا تمكن المستجيب من أن يجيب بما يتفق مع رأيه بصورة صحيحة ومناسبة ، مثال : الاختبار الجيد قصير وذو عبارات واضحة .

فالمستجيب هنا قد يرى أن أحد هاتين الصفتين للاختبار له أثر والآخر ليس له أثر ، أو أن أحدهما له أثر قوي جدا والآخر أثره قليل ، لكن صياغة العبارة أو السؤال لا تمكن المستجيب من أن يجيب بما يتفق مع رأيه ، ومن هنا لابد من تعديل صياغة هذه العبارة إلى عبارتين منفصلتين كالآتي :

- الاختبار الجيد قصير .
- الاختبار الجيد عباراته واضحة .

ثالثا : الصعوبات التي تواجه الباحثون والتي تتعلق بكتابة الاستجابات :

1. عدم اشتمال الخيارات على جميع الإجابات المحتملة على السؤال أو العبارة ، أو كون بعض الاستجابات غالبة والبعض شاذة :

فمن المشاكل التي تتعرض لها كتابة الإجابات هو عدم اشتمالها على جميع الاجابات المحتملة على السؤال أو العبارة ، أو كون بعض هذه الإجابات غالبة والبعض الآخر شاذة ، فهنا لا يستطيع المستجيب المفاضلة بين الإجابات ، لأنه سيختار قطعا الإجابة الغالبة ، مثلا :

"مدى قيام المدرسة الإبتدائية بدورها في تربية النشء"

( ) لا تقوم بدورها إطلاقا .

( ) تقوم ببعض ما يجب أن تقوم به .

هنا نلاحظ إجابات متنافرة ومتباعدة جدا ، مما يحتم التركيز على واحدة منها ، فبالتالي سوف يختار معظم أو كل المستجيبين الإجابة رقم (ب) مثلا. لذا يجب أن لا يكون إحتمال إختيار إجابة لسؤال أكبر في خيار واحد دون غيره من الإجابات الموفرة لذلك السؤال ، وذلك حتى يتضح الفرق بين إجابات المستجيبين .

2. كون الخيارات غير مستقلة في مدلولها بعضها عن بعض :

هذه المشكلة تدفع بالمستجيب إلى التردد بين إجابتين ، كأن يجد أنه يختار خيارين مثلا ، على أنه من الأولى أن يكون هناك نوع من التمييز والاستقلالية بين الاستجابات المتاحة

3. عدم مراعات اتجاه العبارات (الإيجابي والسلبي) :

يقع كثير من الباحثين عند صياغته لعبارات الاستبانة لمشكلة كون هذه العبارات مصاغة بإتجاه إيجابي نحو الظاهرة مرضع الدراسة أو باتجاه سلبي ، إن الاسراف في هذا الامر يؤدي إلى نتائج عكسية على الاستجابات ، لذا ينبغي على الباحث أن يراعي التوازن عند صياغة العبارات.

رابعا : الصعوبات التي تواجه الباحثون والتي تتعلق بالاستبانة بعد اكتمالها :

1. ترتيب عبارات الاستبانة :

بحيث لا تكون عبارات كل بعد من أبعاد الاستبانة في شكل متتالي مما يجعلها منفصلة عن الأبعاد الأخرى ، مما يؤدي بالتالي إلى إرهاق المستجيب من خلال الإجابة على اسئلة متتالية في بعد واحد ، أو يؤدي ذلك إلى نسق معين في الإجابة ، لذا يجب أن ترتب عبارات المقياس بطريقة عشوائية .

2. ترميز البيانات وإدخالها في الحاسب :

يعاني كثير من الباحثين من عملية ترميز البيانات التي تم جمعها من خلال الاستبانة ، مما يؤدي إلى ترجمتها بشكل خاطئ ، فتؤدي بالتالي إلى نتائج عكسية ، لأن عملية الترميز تعتمد عليها عملية إدخال البيانات وتحليلها .

3. اختيار الاسلوب الاحصائي المناسب لتحليل بيانات الاستبانة :

إن المتتبع لواقع الأبحاث في مجال العلوم السلوكية والاجتماعية يلاحظ تناقضا في النتائج بين الدراسات التي تبحث الموضوع الواحد، وهذا التناقض يعود بالدرجة الأولى إلى سوء استخدام الإحصاء ، وعدم تحرى الدقة في تحليل البيانات واختيار الأداة الإحصائية المناسبة ، وذلك لان الإحصاء أداة من أدوات البحث ، وانه إذا احسن استخدام هذه الأداة فإنها تعطي نتائج يعتمد عليها، وأما إذا أساء الباحث استخدام الإحصاء ، فإنه بالتالي يصل إلى نتائج مضللة .

ومع أن جميع الأساليب الإحصائية تستعمل في تلخيص البيانات وتحليلها ، إلا أن لكل منها حدودها ، ومما يساعد على معرفة حدود الأساليب الإحصائية هو معرفة الأساس النظري الذي يقوم عليه كل أسلوب إحصائي ، لان كل أسلوب إحصائي له افتراضات معينة ،فإذا كانت هذه الافتراضات صحيحة بالنسبة لبيانات معينة ، فإن العملية الإحصائية تكون مناسبة للتطبيق في معالجة تلك البيانات.

4. حساب صدق الاستبانة :

إن مشكلة حساب صدق الاستبانة من المشاكل المهمة التي تواجه الباحثين عند بناء أي استبانة ، فالمقصود بصدق الاستبانة هو مدى الكفاءة التي تتصف بها الاداة في قياس ما وضعت لقياسه . ويتبع الباحثون طرقا عديدة لحساب الصدق ، ومع ذلك فإن طبيعة البحث هي التي تقرر الطريقة المناسبة لذلك . لذا ينبغي على الباحث تحري الدقة في هذا الامر واختيار الاسلوب المناسب لحساب الصدق

5. حساب ثبات الاستبانة :

المقصود بثبات الاستبانة هو مدى اتساق البيانات التي تم جمعها بواسطة هذه الاداة ، والباحث عادة يلجأ إلى معرفة هذا الاتساق لأن أداة البحث معرضة دائما للخطأ ، ولذا فالمشكلة الرئيسية التي تواجه الباحث في بداية بحثه أن يحدد حجم الخطأ الذي يمكن أن تتورط فيه أداته الرئيسية وأن يعمل على الإقلال من هذا الخطأ بالطرق العلمية الممكنة.

مع أصدق الأمنيات

أخوكم / ثابت

@thabet_18

المحاضرة الثالثة

المقاييس الإحصائية للبيانات المبوبة

أولا: مقاييس النزعة المركزية (المتوسط – الوسيط – المنوال)

(1) مقاييس النزعة المركزية

المتوسط هو قيمة نموذجية يمكن أن تمثل مجموعة من البيانات بحيث تعطي دلالة معينة لتلك البيانات ، بمعنى أنه عندما ينظر الباحث (أو القارئ لتلك البيانات) ويريد أن يبحث عن شيء يربط هذه البيانات فإن تلك المتوسطات يمكن أن تعطيه بعضاً مما يريده .

وحيث أن مثل هذه القيم (المتوسطات) تميل إلى الوقوع في المركز داخل مجموعة البيانات (عند ترتيبها حسب قيمها) ، فإن هذه المتوسطات تُسمى أيضاً بمقاييس النزعة المركزية .

وهناك صور عديدة من هذه المقاييس وإن كان الأكثر شيوعاً :

- الوسط الحسابي (أو باختصار الوسط) .
- الوسيط
- المنوال (أو الشائع)

وغيرها ، وكل منها له مميزاته وعيوبه وهذا يعتمد على البيانات والهدف من استخدامه .

وإلى جانب كونه ممثلاً لمجموعة البيانات يجب أيضاً أن تتوافر في المتوسط عدة شروط ، منها :

أن يمكن تحديد قيمته بالضبط وتكون عملية حسابه سهلة إلى حد كبير

أن يأخذ في الاعتبار جميع البيانات .

ومن الجدير بالذكر أن بعض هذه المقاييس يمكن تحديدها حسابياً بسهولة ، وبعضها يمكن تحديدها بيانياً بسهولة ، والبعض يمكن تحديده حسلبياً وبيانياً بسهولة ، لكننا في هذا المقرر سنكتفي بالطريقة الأبسط (للطالب) عند تحديد هذه المقاييس ، وهذه الطريقة الأبسط ستختلف من مقياس لآخر .

(2) أهمية حساب مقاييس النزعة المركزية

عند معرفتنا بتلك المتوسطات (مقاييس النزعة المركزية) يصبح أمامنا فرصة كبيرة لأن :

ننظر لمتوسط مجموعة من البيانات لنعرف الكثير عن خصائص تلك المجموعة .

نعقد مقارنة بين عدة مجموعات من البيانات في وقت واحد وذلك من خلال مقارنة متوسطات تلك المجموعات بعضها ببعض .

الوسط الحسابي

(1) تعريف الوسط الحسابي

يُعرف الوسط الحسابي [وسنرمز له بالرمز] لمجموعة من البيانات ..... $x_1 , x_2 ,$ [قيم المتغير x وعددها n] كالآتي:

الوسط الحسابي = مجموع قيم البيانات / عددها أي $\bar{x} = \frac{\sum x}{n}$

س ١ : درجات خمسة طلاب في مقرر ما [الدرجة العظمى 20] هي : 10 , 12 , 7 , 2 , 9 . أوجد الوسط الحسابي لدرجاتهم .

ج ١ :

$$\bar{x} = \frac{\sum x}{n} = \frac{9+2+7+12+10}{5} = \frac{40}{5} = 8$$

من هذا المثال البسيط يمكن ملاحظة الخصائص العامة التالية للوسط الحسابي :

- يمكن تحديد قيمة الوسط الحسابي بالضبط، كما أن طريقة تحديده سهلة .
- يأخذ في الاعتبار جميع البيانات .
- لا يتأثر بترتيب البيانات .
- لا يُشترط أن يكون الوسط الحسابي عدداً صحيحاً ولا يُشترط أن يكون إحدى قيم البيانات ولكنه قيمة تقع بين أقل قيمة في البيانات وأكبر قيمة فيها .

يتأثر بالقيم المتطرفة في البيانات [كما يتضح من السؤالين التاليين] .

س ٢ : احسب الوسط الحسابي للقيم :

10 , 15 , 12 , 13 , 9

ج ٢ :

$$\frac{10+15+12+13+9}{5} = \frac{59}{5} = 11.8$$

س ٣ : احسب الوسط الحسابي للقيم :

10 , 15 , 12 , 13 , 900

ج ٣ :

$$\frac{10+15+12+13+900}{5} = \frac{950}{5} = 190$$

حاصل ضرب قيمة الوسط الحسابي في عدد البيانات = مجموع قيم البيانات

وهذا واضح من تعريف الوسط الحسابي : $\bar{x} = \frac{\sum x}{n}$ تعني أن $n \times \bar{x} = \sum x$

فمثلاً

9 , 2 , 7 , 12 , 10

خمس درجات وسطها الحسابي 8

$5 \times 8 = 9+2+7+12+10$

40 = 40

إذا أضفنا عدد ثابت c لكل قيمة من قيم البيانات ، فإن :

الوسط الحسابي الجديد = الوسط الحسابي القديم + العدد الثابت c

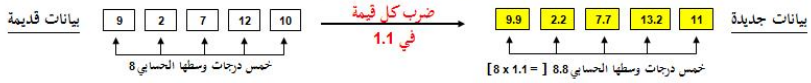
بيانات قديمة: 9 , 2 , 7 , 12 , 10 — خمس درجات وسطها الحسابي 8

إضافة 1 لكل قيمة

بيانات جديدة: 10 , 3 , 8 , 13 , 11 — خمس درجات وسطها الحسابي 9 [= 1 + 8]

إذا ضربنا كل قيمة من قيم البيانات في عدد ثابت c ، فإن :

الوسط الحسابي الجديد = الوسط الحسابي القديم x العدد الثابت c



اعتبر نفسك مدرساً للطلاب الخمسة المذكورين في س ١ [كانت درجاتهم (من 20) كالتالي : 9 , 2 , 7 , 12 , 10] وأردت أن تحسن من الوسط الحسابي لدرجاتهم ، أيهما أفضل : أن نزيد درجة كل طالب 5 درجات أم نزيد درجة كل طالب 50% من قيمتها ؟ علل إجابتك .

(2) حساب الوسط الحسابي لبيانات كمية متقطعة ذات تكرارت

س : أوجد الوسط الحسابي للأرقام :

5 , 5 , 5 , 5 , 5 , 5 , 3 , 3 , 6 , 6 , 4 , 4 , 4 , 4 , 2 , 2 , 8 , 8 , 8

ج : بتطبيق مباشر للتعريف :

$$\bar{x} = \frac{(5+5+5+5+5+5)+(3+3)+(6+6)+(4+4+4+4+4+4)+(2+2)+(8+8+8)}{20} = \frac{96}{20} = 4.8$$

لاحظ أن الرقم 5 متكرر 6 مرات ، الرقم 3 مرتان ، والرقم 6 مرتان ، والرقم 4 متكرر 5 مرات ، والرقم 2 مرتان ، والرقم 8 ثلاث مرات ، وبالتالي يمكن عمل العملية الحسابية السابقة كالآتي :

$$\bar{x} = \frac{(6 \times 5) + (2 \times 3) + (2 \times 6) + (5 \times 4) + (2 \times 2) + (3 \times 8)}{6 + 2 + 2 + 5 + 2 + 3}$$

$$= \frac{30+6+12+20+4+24}{20} = \frac{96}{20} = \underline{4.8}$$

وهذا يمكن إنجازه بيسر من خلال الجدول التكراري للبيانات كالتالي :

$$\frac{30+6+12+20+4+24}{20} = \frac{96}{20}$$

نعمل هذا العمود : حاصل ضرب كل قيمة في تكرارها

المتغير x	التكرار f	fx
5	6	30
3	2	6
6	2	12
4	5	20
2	2	4
8	3	24
	20	96

$\sum f = 20$ $\quad \sum fx = 96$

هذا الجدول التكراري المعطى لنا [مُعطى أو نعمله]

$$\therefore \quad \bar{x} = \frac{\sum fx}{\sum f} = \frac{96}{20} = 4.8$$

أي أنه في حالة البيانات الكمية المتقطعة ذات التكرارات يمكن حساب الوسط الحسابي من العلاقة :

$$\bar{x} = \frac{\sum fx}{\sum f}$$

حيث $\sum f$ هو مجموع التكرارات

$\sum fx$ هو مجموع حاصل ضرب كل قيمة في تكرارها

س: من مائة رقم يتكرر الرقم 4 عشرون مرة، والرقم 5 أربعون مرة، والرقم 6 ثلاثون مرة، والباقي كانوا الرقم 7 احسب الوسط الحسابي للمائة رقم .

ج : بتكوين الجدول التكراري للأرقام المذكورة ، ثم بضرب كل قيمة في تكرارها والتجميع [عمود fx] يكون الوسط الحسابي للأرقام المذكورة هو :

الجدول التكراري		
المتغير x	التكرار f	fx
4	20	80
5	40	200
6	30	180
7	10	70
	100	530

$\sum f = 100$ $\quad \sum fx = 530$

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{530}{100} = \underline{5.3}$$

(3) حساب الوسط الحسابي لبيانات كمية متصلة

عندما نتعامل مع بيانات متصلة تُعطى فيها قيم المتغير على صورة فترات، فيمكن اعتبار أن جميع القيم داخل الفترة مطابقة لمركز الفئة ، وبالتالي يمكن استخدام الصيغة السابقة لحساب الوسط الحسابي :

$$\bar{x} = \frac{\sum f x_0}{\sum f}$$

حيث $\sum f$ هو مجموع التكرارات ، $\sum f x_0$ هو مجموع حاصل ضرب مركز كل فئة في تكرار الفئة

ففي المثال التالي والذي يوضح اطوال سيقان الزهار بالسنتيمتر، يكون الوسط الحسابي لأطوال سيقان الأزهار هو:

الجدول التكراري

الفئة	المتغير x (الطول)	التكرار f	مركز الفئة x_0	fx_0
الأولى	$0 \le x < 20$	4	10	40
الثانية	$20 \le x < 30$	16	25	400
الثالثة	$30 \le x < 35$	12	32.5	390
الرابعة	$35 \le x < 40$	10	37.5	375
الخامسة	$40 \le x < 50$	6	45	270
السادسة	$50 \le x < 60$	2	55	110
		$\sum f = 50$		$\sum f x_0 = 1585$

$$\bar{x} = \frac{\sum f x_0}{\sum f} = \frac{1585}{50} = \underline{31.7}$$

مزايا وعيوب الوسط الحسابي

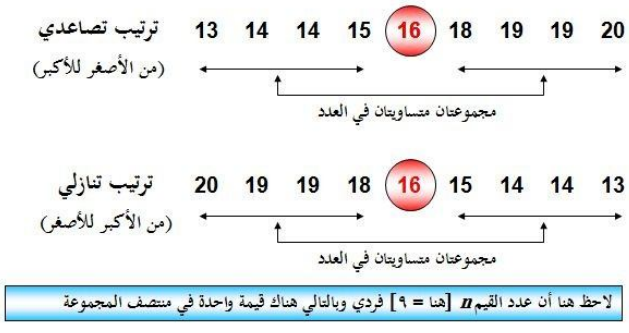
من كل ما سبق يمكن استعراض مزايا وعيوب الوسط الحسابي كالتالي :

- يمكن تحديد قيمة الوسط الحسابي بالضبط ، كما أن طريقة تحديده سهلة [ميزة].
- يأخذ في الاعتبار جميع البيانات [ميزة] .
- لا يتأثر بترتيب البيانات [ميزة] .
- يتأثر بالقيم المتطرفة في البيانات [عيب] .
- لا يمكن حسابه بالرسم ، أي بيانياً [عيب] .

تعريف الوسيط :

(ببساطة) يُعرف الوسيط [وسنرمز له بالرمز M] لمجموعة من القيم (المرتبة تصاعدياً أو تنازلياً حسب قيمها) على أنه القيمة التي تقسم مجموعة القيم إلى مجموعتين متساويتين في العدد ، أو بتعبير آخر هي القيمة التي في المنتصف

فمثلاً لمجموعة القيم : 13 , 14 , 19 , 16 , 20 , 15 , 18 , 14 , 19 [عددها ٩ قيم أي رقم فردي] ، إذا قمنا بترتيبها تصاعدياً أو تنازلياً

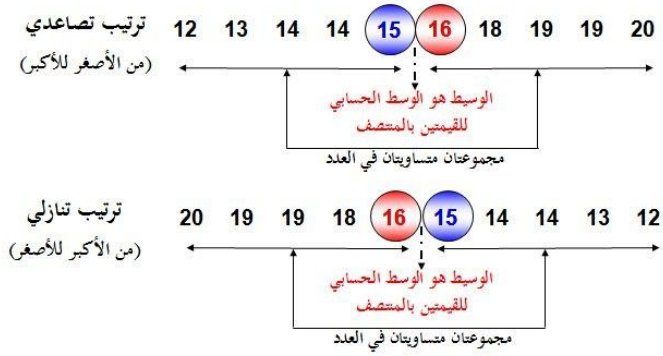


لاحظ هنا أن عدد القيم n [هنا = ٩] فردي وبالتالي هناك قيمة واحدة في منتصف المجموعة

يكون الوسيط هو العدد الخامس [رتبة الوسيط أي ترتيبه بين القيم] وقيمته 16

هام جداً — فرق بين رتبة الوسيط وقيمته

أما لمجموعة القيم : 12 , 13 , 14 , 19 , 16 , 20 , 15 , 18 , 14 , 19 [عددها ١٠ قيم (أي رقم زوجي) حيث أضفنا القيمة 12 للمجموعة السابقة]، إذا قمنا بترتيبها تصاعدياً أو تنازلياً



في هذه الحالة توجد قيمتان بالمنتصف وهما القيمة الخامسة والقيمة السادسة [وهما العددان 16 , 15] ، عندئذ يكون الوسيط هو الوسط الحسابي لهاتين القيمتين ، أي :

$$\frac{15+16}{2} = 15.5$$

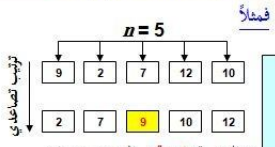
إذن من السابق يمكن استنتاج طريقة حساب الوسيط لمجموعة من القيم كالآتي :

- قم أولاً بترتيب البيانات تصاعدياً أو تنازلياً .
- حدد ما إذا كانت هناك قيمة واحدة بالمنتصف أم قيمتين ، وهذا يتوقف على n حيث n عدد القيم

فإذا كانت n فردية
كانت هناك قيمة واحدة في المنتصف رتبتها $\frac{n+1}{2}$
وتكون هذه القيمة هي الوسيط

وإذا كانت n زوجية
كانت هناك قيمتان في المنتصف رتبتهما $\frac{n}{2}$ ، $\frac{n}{2}+1$
ويكون الوسط الحسابي لهاتين القيمتين هو الوسيط

فمثلاً



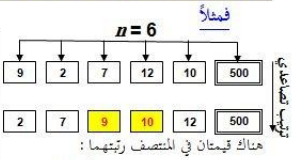
هناك قيمة واحدة في المنتصف رتبتها : $\frac{n+1}{2}=\frac{5+1}{2}=3$

أي القيمة الثالثة . وتكون تلك القيمة هي الوسيط . أي أن :

الوسيط = 9

تذكر :
الوسط الحسابي لهذه القيم هو :
$\frac{9+2+7+12+10}{5}=8$

فمثلاً



هناك قيمتان في المنتصف رتبتهما : $\frac{n}{2}=\frac{6}{2}=3$, $3+1=4$

أي القيمتان الثالثة والرابعة ، وتكون قيمة الوسيط هي الوسط الحسابي لهاتين القيمتين ، أي :

$\frac{9+10}{2}=9.5$

الوسط الحسابي لهذه القيم هو
$\frac{9+2+7+12+10+500}{6}=90$
وواضح تأثره كثيراً بالقيمة المتطرفة 500

هل لاحظت أن الوسيط لم يتأثر بالقيمة المتطرفة 500

لاحظ من الأمثلة السابقة أن كلاً من المتوسطين الوسط الحسابي و الوسيط من السهل حسابهما ومن الممكن أن يمثل كل منهما مقياساً للنزعة المركزية للبيانات ، لكن الأفضل (نسبياً هنا) أن نستخدم الوسط الحسابي كمقياس للنزعة المركزية للبيانات حيث أنه يأخذ في الاعتبار جميع قيم البيانات ، بينما يهتم الوسيط بقيم البيانات في المنتصف (وذلك بعد ترتيبها) .

مثال آخر : الأجر (بالريال) في الساعة لخمسة عاملين في مكتب هو : 37 , 92 , 32 , 39 , 25 . احسب الوسط الحسابي للأجور ووسيط هذه الأجور . أيهما تفضل كمقياس لمتوسط أجر الساعة ؟ ولماذا ؟

الوسط الحسابي للأجور هو : $\bar{x} = \frac{\sum x}{n} = \frac{25+39+32+92+37}{5} = \frac{225}{5} = 45$

أما لتحديد الوسيط ، فلابد أولاً من ترتيب القيم (تصاعدياً مثلاً) : 25 , 32 , 37 , 39 , 92

وحيث أن عدد القيم فردي ، إذن هناك قيمة واحدة في المنتصف [هي 37] وهي الوسيط

لاحظ في هذا السؤال أن الوسط الحسابي (بالرغم من عدم احتياجه لترتيب القيم وفي نفس الوقت يأخذ في الاعتبار جميع قيم البيانات) إلا أنه تأثر جداً بالقيمة المتطرفة 92 ، في حين لم يتأثر بها الوسيط لأنه يعتمد على البيانات في المنتصف . لذا يُفضل هنا استخدام الوسيط كمقياس للنزعة المركزية حيث يعطي دلالة أفضل لمتوسط الأجور من الوسط الحسابي .

الوسيط لبيانات كمية متصلة

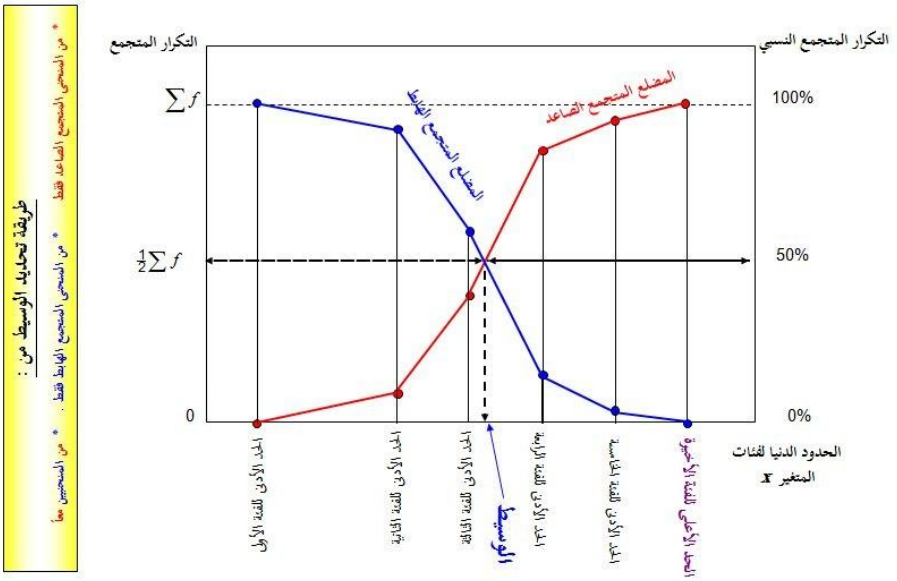
يمكن حساب الوسيط للبيانات الكمية المتصلة من خلال الرسم وكذلك من خلال المعادلات الاحصائية بسهولة

الوسيط من خلال الرسم البياني يتم كالتالي:

قيمة المتغير المناظر لنقطة تقاطع المنحنيين: المتجمع الصاعد والمتجمع الهابط للبيانات .

او القيمة التي يناظرها تكرار متجمع = نصف مجموع التكرارات

او القيمة التي يناظرها تكرار نسبي متجمع = 50%



الوسيط من خلال المعادلات الإحصائية

المساحة (بالكيلومتر)	عدد قطع الأراضي
1 -	14
3 -	29
5 -	18
7 - 10	9

مثال : في دراسة جغرافية لعدد من مساحات مجموعة من الأراضي لمنطقة سكنية بالرياض تبين أن التوزيع التكراري لها كما هو مبين .

المطلوب حساب الوسط الحسابي والوسيط لمساحة الأراضي

المتغير x هنا هو مساحة الأرض (بالكيلومتر) ، في حين يمثل عدد قطع الأراضي التكرار f .

أولاً : الوسط الحسابي : نستكمل الجدول التكراري كما هو مبين :

الجدول التكراري

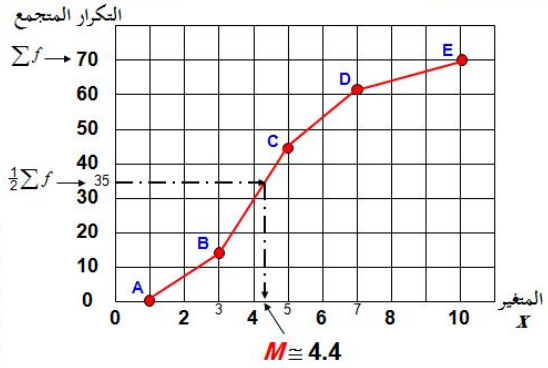
الفئة	المتغير (المساحة)	التكرار f	المركز x_0	fx_0
الأولى	$1 \le x < 3$	14	2	28
الثانية	$3 \le x < 5$	29	4	116
الثالثة	$5 \le x < 7$	18	6	108
الرابعة	$7 \le x < 10$	9	8.5	76.5
		$\sum f = 70$		$\sum fx_0 = 328.5$

$$\therefore \bar{x} = \frac{\sum f x_0}{\sum f} = \frac{328.5}{70} = 4.692857143 \cong 4.7$$

ثانياً : الوسيط : نكون الجدول التكراري المتجمع الصاعد [أو الهابط]

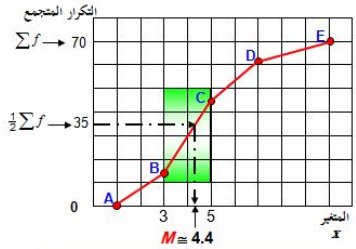
الجدول التكراري		
الفئة	المتغير (المساحة)	التكرار f
الأولى	$1 \le x < 3$	14
الثانية	$3 \le x < 5$	29
الثالثة	$5 \le x < 7$	18
الرابعة	$7 \le x < 10$	9
		$\sum f = 70$

الجدول التكراري المتجمع الصاعد		
المتغير x	التكرار	النقطة على المضلع
< 1	0	A (1 , 0)
< 3	14	B (3 , 14)
< 5	43	C (5 , 43)
< 7	61	D (7 , 61)
< 10	$\sum f = 70$	E (10 , 70)



هكذا انتهت الإجابة على السؤال، أي أن الوسط الحسابي للمساحة $\cong$ 4.7 والوسيط $\cong$ 4.4

لكن هناك ملاحظة هامة جداً يرجي ملاحظتها



الوسيط يقع بين النقطتين B(3 , 14) , C(5 , 43) [أي داخل الفئة $3 \le x < 5$] هذه الفئة تُسمى بـ الفئة الوسيطية

أي أن الفئة الوسيطية هي تلك الفئة التي يقع داخلها الوسيط

وهنا يتبادر إلى الذهن سؤالان هامان :

السؤال الأول : هل من الممكن تحديد الفئة الوسيطية من الجدول التكراري مباشرةً أم لازم نعمل الجدول التكراري المتجمع الصاعد ونرسم المضلع التكراري المتجمع الصاعد ؟ .

السؤال الثاني : هل من الممكن [بعد تحديد الفئة الوسيطية] تحديد الوسيط من الجدول التكراري مباشرة دون الحاجة للجدول التكراري المتجمع الصاعد أو المضلع التكراري المتجمع الصاعد ؟ .

والإجابة على السؤالين : نعم يمكن تحديد الفئةالوسيطية من الجدول التكراري مباشرةً ، ثم بعد ذلك يمكن أيضاً من هذا الجدول التكراري تحديد قيمة الوسيط دون أن نحتاج لعمل جدول تكراري متجمع صاعد ورسم المضلع التكراري المتجمع الصاعد ، وذلك كالتالي:

بالنسبة للسؤال الأول [تحديد الفئة الوسيطية]

(1) احسب أولاً نصف مجموع التكرارات .

(2) ابدأ بالرقم صفر في ذهنك وزود تكرارات الفئات على التوالي وكل مرة قارن بنصف مجموع التكرارات السابق . أول مايزيد الناتج عن نصف المجموع السابق أو يساويه تكون آخر فئة زودنا تكرارها تكون هي الفئة الوسيطية .

ويتم ذلك كالتالي

الجدول التكراري		
الفئة	المتغير (المساحة)	التكرار f
الأولى	$1 \le x < 3$	14
الثانية	$3 \le x < 5$	29
الثالثة	$5 \le x < 7$	18
الرابعة	$7 \le x < 10$	9
		$\sum f = 70$

- احسب $\frac{1}{2}\sum f$: $\frac{1}{2}\sum f = \frac{70}{2} = 35$ ⟵
- نبدأ بالصفر [في ذهننا]
- نزود على الصفر السابق تكرار الفئة الأولى [14] ينتج 14

14 أقل من 35 ، يبقى الفئة الأولى ليست الفئة الوسيطية

- نزود على الـ 14 الأخيرة تكرار الفئة الثانية [29] ينتج 43

43 أكبر من 35 ، يبقى الفئة الثانية هي الفئة الوسيطية

وبالنسبة للسؤال الثاني [تحديد الوسيط (بعد ما حددنا الفئة الوسيطية)]

(1) حدد الحد الأدنى للفئة الوسيطية وأيضاً طولها .

(2) احسب ما يُسمى بـ "التكرار المتجمع السابق" = مجموع تكرار الفئات السابقة للفئة الوسيطية

(3) احسب الوسيط من العلاقة :

$$\text{الوسيط } M = \text{الحد الأدنى للفئة الوسيطية} + \left[\frac{\text{نصف مجموع التكرارات} - \text{التكرار المتجمع السابق}}{\text{تكرار الفئة الوسيطية}} \times \text{طول الفئة الوسيطية}\right]$$

الجدول التكراري		
الفئة	المتغير (المساحة)	التكرار f
الأولى	$1 \le x < 3$	14
الثانية ← الفئة الوسيطية	$3 \le x < 5$	29
الثالثة	$5 \le x < 7$	18
الرابعة	$7 \le x < 10$	9
		$\sum f = 70$

- الفئة الوسيطية هي الفئة الثانية :

حدها الأدنى 3 وطولها 2 [= 3 – 5] وتكرارها 29

- التكرار المتجمع السابق :

يساوي مجموع تكرارات الفئات السابقة للفئة الوسيطية [أي تكرار الفئة الأولى فقط] = 14

- بالتعويض في القانون السابق :

$$M = 3 + \left[\frac{35 - 14}{29} \times 2\right] = 3 + \left[\frac{21}{29} \times 2\right] = 3 + 1.44827 = 4.44827 \cong 4.4$$

تُسمى الطريقة الحسابية السابقة (لحساب الوسيط) بـ "طريقة الاستكمال"

مثال: طُلب من ٣ مشرفين بإحدى المدارس تقسيم طلبة المدرسة إلى ٣ مجموعات متساوية على أن يقوم كل مشرف بتقديم بيان عن فئات العمر المختلفة لطلبة مجموعته وعدد الطلبة في كل فئة من فئات العمر ، فكانت الجداول التكرارية المبينة :

المجموعة (١)		
الفئة	العمر x	العدد f
الأولى	$x < 6$	20
الثانية	$6 \le x < 12$	25
الثالثة	$12 \le x < 15$	35
الرابعة	$15 \le x < 18$	18
		$\sum f = 98$

المجموعة (٢)		
الفئة	العمر x	العدد f
الأولى	$6 \le x < 12$	20
الثانية	$12 \le x < 15$	25
الثالثة	$15 \le x < 18$	35
الرابعة	$x \ge 18$	18
		$\sum f = 98$

المجموعة (٣)		
الفئة	العمر x	العدد f
الأولى	$x < 6$	20
الثانية	$6 \le x < 12$	25
الثالثة	$12 \le x < 15$	35
الرابعة	$x \ge 15$	18
		$\sum f = 98$

هل يمكن من خلال البيانات السابقة حساب الوسط الحسابي لعمر الطلبة في كل مجموعة ؟ علل إجابتك . وإذا كانت الإجابة بـ "لا" احسب مقياساً مناسباً يُعطي دلالة لمتوسط العمر في كل مجموعة .

المجموعة (١)		
الفئة	العمر x	العدد f
الأولى	$x < 6$	20
الثانية	$6 \le x < 12$	25
الثالثة	$12 \le x < 15$	35
الرابعة	$15 \le x < 18$	18
		$\sum f = 98$

المجموعة (٢)		
الفئة	العمر x	العدد f
الأولى	$6 \le x < 12$	20
الثانية	$12 \le x < 15$	25
الثالثة	$15 \le x < 18$	35
الرابعة	$x \ge 18$	18
		$\sum f = 98$

المجموعة (٣)		
الفئة	العمر x	العدد f
الأولى	$x < 6$	20
الثانية	$6 \le x < 12$	25
الثالثة	$12 \le x < 15$	35
الرابعة	$x \ge 15$	18
		$\sum f = 98$

الإجابة هي " لا " للمجموعات الثلاث [أي لا يمكن حساب الوسط الحسابي للعمر] ، والأسباب هي:

- في المجموعة الأولى : الحد الأدنى للفئة الأولى غير معروف [يُقال للجدول عندئذٍ أنه مفتوح من أسفل]
- في المجموعة الثانية : الحد الأعلى للفئة الأخيرة غير معروف [يُقال للجدول عندئذٍ أنه مفتوح من أعلى]
- في المجموعة الثالثة : الحد الأدنى للفئة الأولى غير معروف وأيضاً الحد الأعلى للفئة الأخيرة غير معروف [يُقال للجدول عندئذٍ أنه مفتوح من الطرفين]

مثل هذه الجداول تُسمى **جداول تكرارية مفتوحة** :

- من أسفل [إذا كان الحد الأدنى للفئة الأولى غير معروف]
- من أعلى [إذا كان الحد الأعلى للفئة الأخيرة غير معروف]
- من الطرفين [إذا كان الحد الأدنى للفئة الأولى والحد الأعلى للفئة الأخيرة غير معروفين]

تحديد الوسيط :

- الحد الأدنى للفئة الوسيطية = 12
- طول الفئة الوسيطية = 3 [= 12 – 15]
- تكرار الفئة الوسيطية = 35
- التكرار المتجمع السابق = مجموع تكرارات الفئات السابقة للفئة الوسيطية [أي الفئتين الأولى والثانية] = 20 + 25 = 45 . إذن

المجموعة (١)		
الفئة	العمر x	العدد f
الأولى	$x < 6$	20
الثانية	$6 \le x < 12$	25
الثالثة ← الفئة الوسيطية	$12 \le x < 15$	35
الرابعة	$15 \le x < 18$	18
		$\sum f = 98$

$$M = 12 + \left[\frac{(49-45)}{35} \times 3\right] = 12 + \left[\frac{4}{35} \times 3\right] = 12 + 0.342857 = 12.342857 \cong \underline{12.3}$$

وبنفس الطريقة يمكن التعامل مع المجموعتين (٢) ، (٣) ، وعليك التأكد من صحة الحل

المجموعة (٢)		
الفئة	العمر x	العدد f
الأولى	$6 \le x < 12$	20
الثانية	$12 \le x < 15$	25
الثالثة ← الفئة الوسيطية	$15 \le x < 18$	35
الرابعة	$x \ge 18$	18
		$\sum f = 98$

$$M = 15 + \left[\frac{(49-45)}{35} \times 3\right] \cong \underline{15.3}$$

المجموعة (٣)		
الفئة	العمر x	العدد f
الأولى	$x < 6$	20
الثانية	$6 \le x < 12$	25
الثالثة ← الفئة الوسيطية	$12 \le x < 15$	35
الرابعة	$x \ge 15$	18
		$\sum f = 98$

$$M = 12 + \left[\frac{(49-45)}{35} \times 3\right] \cong \underline{12.3}$$

المنوال

تعريف المنوال [الشائع] :

يُعرف المنوال لمجموعة من القيم على أنه القيمة التي تتكرر أكثر من غيرها أو القيمة الأكثر شيوعاً [لذا يُسمى في بعض الأحيان بال "الشائع"] . وأحياناً يُرمز للمنوال بالرمز $\hat{X}$

فمثلاً :

مجموعة القيم 18 12 11 10 10 (9 9 9) 7 5 2 2 لها منوال 9

ومجموعة القيم 16 15 12 10 8 5 3 9 ليس لها منوال [أو عديمة المنوال]

ومجموعة القيم 9 (7 7 7) 5 5 (4 4 4) 3 2 لها منوالان 4 , 7

أي أن مجموعة القيم قد تكون وحيدة المنوال [لها منوال واحد] ، وقد تكون عديدة المنوال [منوالان أو أكثر] وقد تكون عديمة المنوال [لا يوجد لها منوال]

أما مجموعة القيم 7 7 6 6 5 5 4 4 فقد تتسرع وتقول أنها رباعية المنوال ومناولها : 4 , 5 , 6 , 7

لكن [حيث أن جميع القيم لها نفس التكرار] هذه المجموعة الأخيرة عديمة المنوال

والمنوال [مقارنةً بالوسط الحسابي والوسيط] به العديد من العيوب منها :

- أنه لا يأخذ في الاعتبار جميع البيانات ولكنه يهتم فقط بالقيم الأكثر تكراراً .
- أنه قد لا يتواجد أو قد يكون هناك أكثر من منوال للبيانات .

إلا أنه أيضاً يتميز ببعض المزايا منها :

- أنه أسرع في تحديده من الوسط والوسيط
- من الممكن تحديده للتوزيعات التكرارية للبيانات المنفصلة سواء كانت تلك البيانات كمية متقطعة أو نوعية [والبيانات الأخيرة (النوعية) لا يمكن حساب الوسط الحسابي لها أو الوسيط]

بيانات منفصلة

درجات طلاب في مقرر الإحصاء

عدد الطلاب	درجة الطالب
28	12
24	14
39	16
9	18

بيانات كمية متقطعة
لها منوال وحيد وهو "الدرجة 16"

درجات طلاب في مقرر الإنجليزي

عدد الطلاب	درجة الطالب
23	12
30	14
30	16
17	18

بيانات كمية متقطعة
لها منوالان وهما "16 , 14"

درجات طلاب في مقرر الفقه

عدد الطلاب	درجة الطالب
25	12
25	14
25	16
25	18

بيانات كمية متقطعة
ليس لها منوال

سيارات في أحد المواقف

عدد السيارات	لون السيارة
10	أحمر R
23	أزرق B
12	أبيض W
5	أصفر Y

بيانات نوعية
لها منوال وهو "اللون الأزرق"

وماذا عن التوزيعات التكرارية للبيانات الكمية المتصلة

الموضوع في غاية البساطة :

- **حدد الفئة المنوالية** [وهي الفئة التي يناظرها أكبر كثافة تكرار]
- **حدد المنوال** [وهو (تقريباً) مركز الفئة المنوالية]

فمثلاً في التوزيع التكراري التالي:

	الجدول التكراري				
	المتغير x	التكرار f	طول الفئة c	مركز الفئة x_0	كثافة التكرار
الفئة الأولى	$0 \le x < 20$	4	20	10	0.2
الفئة الثانية	$20 \le x < 30$	16	10	25	1.6
الفئة الثالثة	$30 \le x < 35$	12	5	32.5	2.4
الفئة الرابعة	$35 \le x < 40$	10	5	37.5	2
الفئة الخامسة	$40 \le x < 50$	6	10	45	0.6
الفئة السادسة	$50 \le x < 60$	2	10	55	0.2
		$\sum f = 50$			

أكبر كثافة تكرار

إذن الفئة المنوالية هي الفئة الثالثة

والمنوال = 32.5

وبالرغم من هذه البساطة إلا أن هناك تنبيه و تحذير

بالنسبة للتنبيه : فالطريقة السابقة [اعتبار أن مركز الفئة المنوالية هو المنوال] هي طريقة تقريبية ، لكن هناك طرق أخرى حسابية وبيانية تعطي قيمة المنوال أكثر دقة ، لكننا لن نتطرق لهذه الطرق في هذا المكان وذلك للتبسيط

وبالنسبة للتحذير : فالفئة المنوالية هي الفئة التي يناظرها أكبر كثافة تكرار وليس أكبر تكرار ، ويتفق اللفظان "أكبر كثافة تكرار" و "أكبر تكرار" فقط إذا كانت أطوال الفئات واحدة .

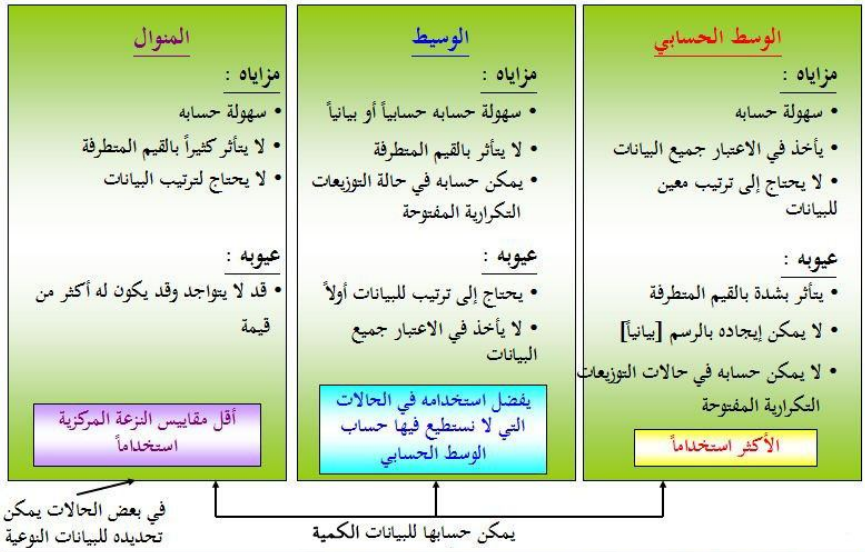
فمثلاً في المثال السابق ، إذا قلنا أن الفئة المنوالية هي الفئة المناظرة لأكبر تكرار فستكون تلك الفئة هي الفئة الثانية [وهذا خطأ جسيم] في حين أن الفئة المنوالية هي الفئة الثالثة [كما سبق وبينا] .

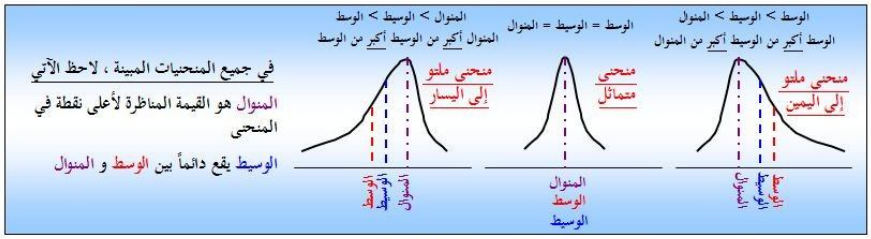
	الجدول التكراري				
	المتغير x	التكرار f	طول الفئة c	مركز الفئة x_0	كثافة التكرار
الفئة الأولى	$0 \le x < 20$	4	20	10	0.2
الفئة الثانية	$20 \le x < 30$	16	10	25	1.6
الفئة الثالثة	$30 \le x < 35$	12	5	32.5	2.4
الفئة الرابعة	$35 \le x < 40$	10	5	37.5	2
الفئة الخامسة	$40 \le x < 50$	6	10	45	0.6
الفئة السادسة	$50 \le x < 60$	2	10	55	0.2
		$\sum f = 50$			

الفئة المنوالية [خطأ جسيم]

الفئة المنوالية [صح]

مقارنة بين المتوسطات الثلاثة : الوسط ، الوسيط ، المنوال





والمنحنيات التكرارية **وحيدة المنوال والبسيطة الالتواء** تحقق العلاقة الاعتبارية التالية :

$$\text{الوسط} - \text{المنوال} = ٣ \times (\text{الوسط} - \text{الوسيط})$$

وهذه العلاقة يمكن وضعها على أي صورة من الصور التالية

$$\text{الوسط} = \frac{(٣ \times \text{الوسيط}) - \text{المنوال}}{٢}$$

وهذه الصورة مفيدة عندما يكون الوسيط و المنوال معلومان ونريد معرفة الوسط الحسابي

$$\text{المنوال} = (٣ \times \text{الوسيط}) - (٢ \times \text{الوسط})$$

وهذه الصورة مفيدة عندما يكون الوسط الحسابي و الوسيط معلومان ونريد معرفة المنوال

$$\text{الوسيط} = \frac{(٢ \times \text{الوسط}) + \text{المنوال}}{٣}$$

وهذه الصورة مفيدة عندما يكون الوسط الحسابي و المنوال معلومان ونريد معرفة الوسيط

- فمثلاً إذا كان **المنوال** لمجموعة من القيم = 95 ، **والوسيط** لها = 85 ، فإن :

$$\text{الوسط} = \frac{(٣ \times \text{الوسيط}) - \text{المنوال}}{٢} = \frac{(٣ \times ٨٥) - ٩٥}{٢} = \frac{٢٢٥ - ٩٥}{٢} = \frac{١٦٠}{٢} = ٨٠$$

- وإذا كان **الوسط الحسابي** لمجموعة من القيم = 80 ، **والوسيط** لها = 85 ، فإن :

$$\text{المنوال} = (٣ \times \text{الوسيط}) - (٢ \times \text{الوسط}) = (٣ \times ٨٥) - (٢ \times ٨٠) = ٢٥٥ - ١٦٠ = ٩٥$$

- وإذا كان **الوسط الحسابي** لمجموعة من القيم = 80 ، **والمنوال** لها = 95 ، فإن :

$$\text{الوسيط} = \frac{(٢ \times \text{الوسط}) + \text{المنوال}}{٣} = \frac{(٢ \times ٨٠) + ٩٥}{٣} = \frac{٩٥ + ١٦٠}{٣} = \frac{٢٥٥}{٣} = ٨٥$$

سؤال: المنحنى التكراري للبيانات المذكورة في أيٍ من الأمثلة السابقة :

◯ ملتو لليمين ◯ ملتو لليسار ◯ متماثل

المحاضرة الرابعة

المقاييس الإحصائية للبيانات المبوبة

ثانيا: مقاييس التشتت (المدى – الإنحراف المتوسط – التباين – الإنحراف المعياري)

تعريف التشتت

الدرجة التي تتجه بها البيانات الكمية **للانتشار حول قيمة متوسطة** (أحد مقاييس النزعة المركزية) تُسمى **تشتت** أو **تغير** البيانات

فمثلاً إذا كان لدينا ٣ مجموعات من الطلاب ، كل مجموعة مكونة من خمسة طلاب ، وكانت لها الدرجات التالية (من ١٠ درجات) في أحد المقررات

المجموعة الأولى	المجموعة الثانية	المجموعة الثالثة
5 , 5 , 5 , 5 , 5	3 , 4 , 5 , 6 , 7	1 , 2 , 5 , 8 , 9
وسطها الحسابي 5	وسطها الحسابي 5	وسطها الحسابي 5

المجموعات الثلاثة لها وسط حسابي 5 ، لكن في المجموعة الأولى : جميع القيم متساوية وتساوي الوسط 5 ، في حين تنتشر البيانات في المجموعة الثانية حول هذا الوسط بقدرٍ ما ، وفي المجموعة الثالثة تنتشر البيانات حول الوسط بقدرٍ آخر .

أي أن الوسط الحسابي وحده [وهو ممثل لمقياس نزعة مركزية ، أي قيمة نموذجية ممثلة للبيانات] ليس كافياً وحده لوصف البيانات ، ولكن لابد من وجود نوع آخر من المقاييس لرصد مدى تشتت البيانات عن تلك القيمة المتوسطة الممثلة للبيانات .
هذا النوع من المقاييس هو ما نسميه بـ **مقاييس التشتت** .

وهناك العديد من المقاييس التي يمكن استخدامها لقياس هذا التشتت ولكن أكثرها شيوعاً :

المدى – الانحراف المتوسط – الانحراف المعياري – الانحراف الربيعي

أولاً : المدى R :

مدى مجموعة من البيانات الكمية هو الفرق بين أكبر قيمة في البيانات وأقل قيمة فيها

فمثلاً لمجموعة القيم : 15 3 7 6 12 18 5 3 13 15 يكون المدى : $R = 18 - 3 = 15$

ولمجموعة القيم : 16 3 14 15 17 18 17 13 14 16 يكون المدى : $R = 18 - 3 = 15$

أي أن المدى واحد للمجموعتين في حين يبدو للعين المجردة أن هناك تشتت للبيانات أكبر في المجموعة الأولى عنه في المجموعة الثانية ، مما يعني أن المدى هنا لا يظهر هذا الفارق . لذا يُعد المدى **مقياساً للتشتت** لكنه **غير جيد** في كثير من الأحيان .

وبالرغم من بساطة تحديده إلا أن له بعض العيوب [مثل تأثره بالقيم المتطرفة كما اتضح من المثال السابق عند حسابه للمجموعة الثانية حيث تأثر بالقيمة المتطرفة 3] ، فإذا استبعدنا تلك القيمة يكون المدى مساوياً لـ : $R = 18 - 13 = 5$.

أيضاً من بين عيوبه أنه لا يمكن تحديده في حالة التوزيعات التكرارية المفتوحة .

الفئة	العمر x
الأولى	$x < 6$
الثانية	$6 \le x < 12$
الثالثة	$12 \le x < 15$
الرابعة	$15 \le x < 18$

مفتوح من أسفل

الفئة	العمر x
الأولى	$6 \le x < 12$
الثانية	$12 \le x < 15$
الثالثة	$15 \le x < 18$
الرابعة	$x \ge 18$

مفتوح من أعلى

الفئة	العمر x
الأولى	$x < 6$
الثانية	$6 \le x < 12$
الثالثة	$12 \le x < 15$
الرابعة	$x \ge 15$

مفتوح من الطرفين

لا يمكن تحديد مدى البيانات

الفئة	العمر x
الأولى	$2 \le x < 6$
الثانية	$6 \le x < 12$
الثالثة	$12 \le x < 15$
الرابعة	$15 \le x < 18$

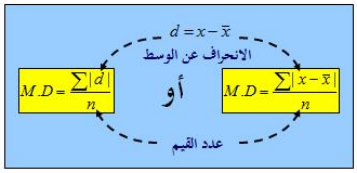
$R = 18 - 2 = 16$

الحد الأدنى للفئة الأولى — الحد الأعلى للفئة الأخيرة

ثانياً : الانحراف المتوسط [أو متوسط الانحرافات] $M.D$

يُعرف الانحراف المتوسط (أو متوسط الانحرافات) [وسنرمز له بالرمز $M.D$] على أنه متوسط القيم المطلقة للانحرافات عن قيمة متوسطة للبيانات [عادةً تكون الوسط الحسابي أو الوسيط] .

فإذا اعتبرنا أن القيمة المتوسطة للبيانات هي الوسط الحسابي ، فإن الانحراف المتوسط لمجموعة من البيانات عددها n يُعطى بـ



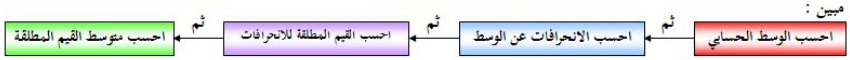
ملحوظة هامة : القيمة المطلقة لأي عدد y هي القيمة العددية له دون إشارة ، ونرمز له بنفس الرمز y لكن بين خطين رأسيين | | ، أي نكتب القيمة المطلقة لـ y على الصورة $|y|$. فمثلاً :

$|3| = 3$, $|-3| = 3$, $|2.5| = 2.5$, $|-3.25| = 3.25$

وهكذا .

حيث $d = x - \bar{x}$ هي انحراف القيمة x عن الوسط الحسابي ، $|d|$ هي القيمة المطلقة للانحراف d .

إذن لحساب الانحراف المتوسط (أو متوسط الانحرافات) $M.D$ لمجموعة من القيم يلزم حساب الوسط الحسابي أولاً ، ثم نحسب انحرافات كل قيمة من هذه القيم عن الوسط الحسابي ، ثم القيم المطلقة لهذه الانحرافات ، ثم متوسط هذه القيم المطلقة كما هو مبين :



فمثلاً : لمجموعتين من القيم يمكن حساب الانحراف المتوسط كالتالي:

لمجموعة القيم الأولى :

وسطها الحسابي : $\bar{x} = \frac{15+13+3+5+18+12+6+7+3+15}{10} = 9.7$

15	13	3	5	18	12	6	7	3	15
-9.7	-9.7	-9.7	-9.7	-9.7	-9.7	-9.7	-9.7	-9.7	-9.7
5.3	3.3	-6.7	-4.7	8.3	2.3	-3.7	-2.7	-6.7	5.3
5.3	3.3	6.7	4.7	8.3	2.3	3.7	2.7	6.7	5.3

الانحرافات عن الوسط — لاحظ أن مجموع الانحرافات = صفر

القيم المطلقة للانحرافات

إذن الانحراف المتوسط هو متوسط القيم المطلقة للانحرافات : $M.D = \frac{5.3+3.3+6.7+4.7+8.3+2.3+3.7+2.7+6.7+5.3}{10} = \underline{4.9}$

ولمجموعة القيم الثانية :

وسطها الحسابي : $\frac{16+14+13+17+18+17+15+14+3+16}{10} = 14.3$

16	14	13	17	18	17	15	14	3	16
-14.3	-14.3	-14.3	-14.3	-14.3	-14.3	-14.3	-14.3	-14.3	-14.3
1.7	-0.3	-1.3	2.7	3.7	2.7	0.7	-0.3	-11.3	1.7
1.7	0.3	1.3	2.7	3.7	2.7	0.7	0.3	11.3	1.7

الانحرافات عن الوسط — لاحظ أن مجموع الانحرافات = صفر

القيم المطلقة للانحرافات

إذن الانحراف المتوسط هو متوسط القيم المطلقة للانحرافات : $M.D = \frac{1.7+0.3+1.3+2.7+3.7+2.7+0.7+0.3+11.3+1.7}{10} = \underline{2.64}$

أي أن الانحراف المتوسط للمجموعة الثانية من القيم أقل من الانحراف المتوسط للمجموعة الأولى من القيم مما يعني أن المجموعة الثانية أقل تشتتاً من المجموعة الأولى وهذا الأمر لا يمكن ملاحظته عند استخدام المدى كمقياس للتشتت

ويمكن أن يتم حل السؤال السابق وذلك بتنظيم خطواتنا من خلال جداول كالتالي :

المجموعة الثانية [$n = 10$]

x	$\bar{x}$	$d = x - \bar{x}$	$\lvert d \rvert$
16	14.3	16 – 14.3 = 1.7	1.7
14	14.3	14 – 14.3 = -0.3	0.3
13	14.3	13 – 14.3 = -1.3	1.3
17	14.3	17 – 14.3 = 2.7	2.7
18	14.3	18 – 14.3 = 3.7	3.7
17	14.3	17 – 14.3 = 2.7	2.7
15	14.3	15 – 14.3 = 0.7	0.7
14	14.3	14 – 14.3 = -0.3	0.3
3	14.3	3 – 14.3 = -11.3	11.3
16	14.3	16 – 14.3 = 1.7	1.7
143	143	0	26.4
$\sum x = n\bar{x}$		$\sum d$	$\sum \lvert d \rvert$

مجموع الأعمدة

$\bar{x} = \frac{\sum x}{n} = \frac{143}{10} = 14.3$

$M.D = \frac{\sum |d|}{n} = \frac{26.4}{10} = 2.64$

المجموعة الأولى [$n = 10$]

x	$\bar{x}$	$d = x - \bar{x}$	$\lvert d \rvert$
15	9.7	15 – 9.7 = 5.3	5.3
13	9.7	13 – 9.7 = 3.3	3.3
3	9.7	3 – 9.7 = -6.7	6.7
5	9.7	5 – 9.7 = -4.7	4.7
18	9.7	18 – 9.7 = 8.3	8.3
12	9.7	12 – 9.7 = 2.3	2.3
6	9.7	6 – 9.7 = -3.7	3.7
7	9.7	7 – 9.7 = -2.7	2.7
3	9.7	3 – 9.7 = -6.7	6.7
15	9.7	15 – 9.7 = 5.3	5.3
97	97	0	49
$\sum x = n\bar{x}$		$\sum d$	$\sum \lvert d \rvert$

مجموع الأعمدة

$\bar{x} = \frac{\sum x}{n} = \frac{97}{10} = 9.7$

$M.D = \frac{\sum |d|}{n} = \frac{49}{10} = 4.9$

ويمكن الاستغناء عن هذا العمود

• **وفي حالة البيانات الكمية المتقطعة ذات التكرارات :**

يمكن تحديد الانحراف المتوسط ***M.D*** من العلاقة :

$$M.D=\frac{\sum f|d|}{\sum f}$$

أي نضرب القيمة المطلقة لانحراف كل قيمة [عن الوسط] في تكرارها ، ثم نقسم الناتج على مجموع التكرارات

فمثلاً : إذا كان المطلوب حساب الانحراف المتوسط للبيانات المبينة بالجدول التكراري :

الجدول التكراري

المتغير x	التكرار f
4	20
5	40
6	30
7	10

هذا هو السؤال

وإليك الجواب

الجدول التكراري

المتغير x	التكرار f	fx
4	20	80
5	40	200
6	30	180
7	10	70
	100	530

$\sum f=100$ $\sum fx=530$

$$\bar{x}=\frac{\sum fx}{\sum f}=\frac{530}{100}=5.3$$

خاص بحساب الوسط الحسابي

$d=x-\bar{x}$	$\lvert d\rvert$	$f\lvert d\rvert$
4 – 5.3 = -1.3	1.3	20 × 1.3 = 26
5 – 5.3 = -0.3	0.3	40 × 0.3 = 12
6 – 5.3 = 0.7	0.7	30 × 0.7 = 21
7 – 5.3 = 1.7	1.7	10 × 1.7 = 17
		$\sum f\lvert d\rvert=76$

$$M.D=\frac{\sum f|d|}{\sum f}=\frac{76}{100}=\underline{0.76}$$

انتبه :

مجموع الانحرافات هنا [والذي يجب أن يساوي صفراً] هو $\sum fd$ وليس $\sum d$ [حقق ذلك بنفسك]

وهذا الجزء يُضاف إذا كان مطلوباً حساب الانحراف المتوسط

وبالتالي يكون الحل [بصورة ملخصة] كالتالي :

الجدول التكراري

المتغير x	التكرار f	fx	$d=x-\bar{x}$	$\lvert d\rvert$	$f\lvert d\rvert$
4	20	80	4 – 5.3 = -1.3	1.3	20 × 1.3 = 26
5	40	200	5 – 5.3 = -0.3	0.3	40 × 0.3 = 12
6	30	180	6 – 5.3 = 0.7	0.7	30 × 0.7 = 21
7	10	70	7 – 5.3 = 1.7	1.7	10 × 1.7 = 17
	100	530			76
	$\sum f=100$	$\sum fx=530$			$\sum f\lvert d\rvert$

$$\bar{x}=\frac{\sum fx}{\sum f}=\frac{530}{100}=5.3$$

$$M.D=\frac{\sum f|d|}{\sum f}=\frac{76}{100}=0.76$$

• **وفي حالة البيانات الكمية المتصلة :** نستخدم نفس العلاقة السابقة لتحديد الانحراف المتوسط ***M.D*** ، أي يكون

$$M.D=\frac{\sum f|d|}{\sum f}$$ حيث $d=x_0-\bar{x}$ ، x_0 تمثل مراكز الفئات

أي أنه عند حساب الانحرافات نعتبر أن مركز أي فئة يمثل جميع القيم الموجودة في تلك الفئة

فمثلاً لو كان لدينا البيانات التالية والموزعة في جدول تكراري ذو فئات وطلب من حساب الإنحراف المتوسط:

ثالثاً : التباين s^2 والانحراف المعياري s

يُعرف متوسط مربعات الانحرافات عن الوسط الحسابي على أنه تباين مجموعة البيانات [ويُرمز له بالرمز s^2] ، ويُعرف الجذر التربيعي للتباين على أنه الانحراف المعياري للبيانات [ويُرمز له بالرمز s] ، أي أن :

التباين $s^2=\frac{\sum d^2}{n}$ ← ومنه يكون ← **الانحراف المعياري** $s=\sqrt{s^2}=\sqrt{\frac{\sum d^2}{n}}$

فمثلاً يتم حساب التباين والإنحراف المعياري كالتالي:

المجموعة الأولى [$n=10$]

x	$d=x-\bar{x}$	d^2
15	15 – 9.7 = 5.3	28.09
13	13 – 9.7 = 3.3	10.89
3	3 – 9.7 = -6.7	44.89
5	5 – 9.7 = -4.7	22.09
18	18 – 9.7 = 8.3	68.89
12	12 – 9.7 = 2.3	5.29
6	6 – 9.7 = -3.7	13.69
7	7 – 9.7 = -2.7	7.29
3	3 – 9.7 = -6.7	44.89
15	15 – 9.7 = 5.3	28.09
$\sum x$ 97		274.1 $\sum d^2$

$\bar{x}=\frac{\sum x}{n}=9.7$

$s^2=\frac{\sum d^2}{n}=\frac{274.1}{10}=27.41$

$s=\sqrt{s^2}=\sqrt{27.41}\cong\underline{5.24}$

المجموعة الثانية [$n=10$]

x	$d=x-\bar{x}$	d^2
16	16 – 14.3 = 1.7	2.89
14	14 – 14.3 = -0.3	0.09
13	13 – 14.3 = -1.3	1.69
17	17 – 14.3 = 2.7	7.29
18	18 – 14.3 = 3.7	13.69
17	17 – 14.3 = 2.7	7.29
15	15 – 14.3 = 0.7	0.49
14	14 – 14.3 = -0.3	0.09
3	3 – 14.3 = -11.3	127.69
16	16 – 14.3 = 1.7	2.89
$\sum x$ 143		164.1 $\sum d^2$

$\bar{x}=\frac{\sum x}{n}=14.3$

$s^2=\frac{\sum d^2}{n}=\frac{164.1}{10}=16.41$

$s=\sqrt{s^2}=\sqrt{16.41}\cong\underline{4.05}$

• **وفي حالة البيانات الكمية المنقطعة ذات التكرارات :**

يمكن تحديد التباين s^2 والانحراف المعياري s من :

التباين $s^2 = \dfrac{\sum fd^2}{\sum f}$ ← ومنه يكون ← **الانحراف المعياري** $s = \sqrt{s^2} = \sqrt{\dfrac{\sum fd^2}{\sum f}}$

فمثلاً : إذا كان المطلوب حساب الانحراف المعياري للبيانات المبينة بالجدول التكراري :

المتغير x	التكرار f
4	20
5	40
6	30
7	10

الجدول التكراري

هذا هو السؤال

وإليك الجواب

الجدول التكراري

المتغير x	التكرار f	fx	$d = x - \bar{x}$	d^2	fd^2
4	20	80	4 – 5.3 = -1.3	1.69	20 × 1.69 = 33.8
5	40	200	5 – 5.3 = -0.3	0.09	40 × 0.09 = 3.6
6	30	180	6 – 5.3 = 0.7	0.49	30 × 0.49 = 14.7
7	10	70	7 – 5.3 = 1.7	2.89	10 × 2.89 = 28.9
	100	530			81
	$\sum f = 100$	$\sum fx = 530$			$\sum fd^2$

$\bar{x} = \dfrac{\sum fx}{\sum f} = \dfrac{530}{100} = 5.3$

$s^2 = \dfrac{\sum fd^2}{\sum f} = \dfrac{81}{100} = 0.81$

$s = \sqrt{s^2} = \sqrt{0.81} = \underline{\underline{0.9}}$

• **وفي حالة البيانات الكمية المتصلة :** تُستخدم نفس العلاقة السابقة لتحديد الانحراف المعياري s ، أي يكون :

التباين $s^2 = \dfrac{\sum fd^2}{\sum f}$ ← ومنه يكون ← **الانحراف المعياري** $s = \sqrt{s^2} = \sqrt{\dfrac{\sum fd^2}{\sum f}}$ حيث $d = x_0 - \bar{x}$

أي أنه عند حساب الانحرافات نعتبر أن مركز أي فئة يمثل جميع القيم الموجودة في تلك الفئة

فمثلاً في المثال التوضيحي (٢–٤) بالباب الثاني [المحاضرة ٤/شريحة ٤] والذي قمنا بحساب الوسط الحسابي والانحراف المتوسط لبياناته ، يمكن حساب **الانحراف المعياري** لهذه البيانات كالتالي :

الفئة	المتغير x	التكرار f	المركز x_0	fx_0	$d = x_0 - \bar{x}$	d^2	fd^2
الأولى	$0 \le x < 20$	4	10	40	10 – 31.7 = -21.7	470.89	4 × 470.89 = 1883.56
الثانية	$20 \le x < 30$	16	25	400	25 – 31.7 = -6.7	44.89	16 × 44.89 = 718.24
الثالثة	$30 \le x < 35$	12	32.5	390	32.5 – 31.7 = 0.8	0.64	12 × 0.64 = 7.68
الرابعة	$35 \le x < 40$	10	37.5	375	37.5 – 31.7 = 5.8	33.64	10 × 33.64 = 336.4
الخامسة	$40 \le x < 50$	6	45	270	45 – 31.7 = 13.3	176.89	6 × 176.89 = 1061.34
السادسة	$50 \le x < 60$	2	55	110	55 – 31.7 = 23.3	542.89	2 × 542.89 = 1085.78
		50		1585			5093
		$\sum f$		$\sum fx_0$			$\sum fd^2$

$\therefore \bar{x} = \dfrac{\sum f x_0}{\sum f} = \dfrac{1585}{50} = \underline{\underline{31.7}}$

$s^2 = \dfrac{\sum fd^2}{\sum f} = \dfrac{5093}{50} = 101.86$

$s = \sqrt{s^2} = \sqrt{101.86} \cong \underline{\underline{10.09}}$

وبنفس الأسلوب يمكن التعامل مع المثال السابق لحساب التباين والإنحراف المعياري

مثال (٢–٦)

الفئة	المتغير x	التكرار f	المركز x_0	fx_0	$d = x_0 - \bar{x}$	d^2	$f \times d^2$
الأولى	$50 \le x < 60$	6	55	330	55 – 83.75 = -28.75	826.56	4959.38
الثانية	$60 \le x < 70$	9	65	585	65 – 83.75 = -18.75	351.56	3164.04
الثالثة	$70 \le x < 80$	15	75	1125	75 – 83.75 = -8.75	76.56	1148.4
الرابعة	$80 \le x < 90$	12	85	1020	85 – 83.75 = 1.25	1.56	18.72
الخامسة	$90 \le x < 100$	9	95	855	95 – 83.75 = 11.25	126.56	1139.04
السادسة	$100 \le x < 120$	6	110	660	110 – 83.75 = 26.25	689.06	4134.36
السابعة	$120 \le x < 180$	3	150	450	150 – 83.75 = 66.25	4389.06	13167.18
		60		5025			27731.12
		$\sum f$		$\sum fx_0$			$\sum fd^2$

$\bar{x} = \dfrac{\sum f x_0}{\sum f} = \dfrac{5025}{60} = \underline{\underline{83.75}}$

$s^2 = \dfrac{\sum fd^2}{\sum f} = \dfrac{27731.12}{60} \cong 462.19$

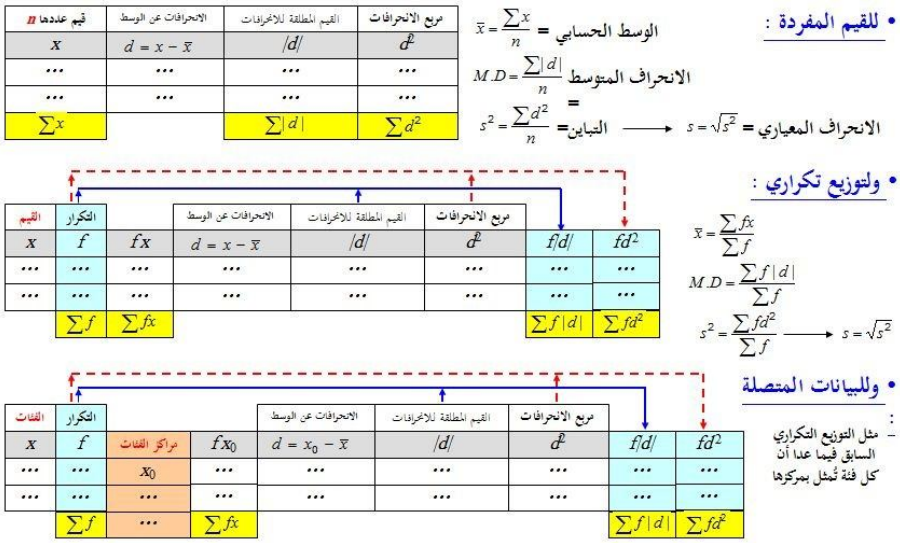
$s = \sqrt{s^2} = \sqrt{462.19} \cong \underline{\underline{21.5}}$

من السابق يتضح أن كلاً من الانحراف المتوسط والانحراف المعياري يعتمدا تماماً في حساباتهما على الوسط الحسابي ، وبالتالي فلهما نفس مزايا وعيوب الوسط الحسابي . أي :

المزايا : من السهل حسابهما – يأخذا في الاعتبار جميع البيانات – لا يحتاجا لترتيب معين للبيانات

العيوب : يتأثرا بشدة بالقيم المتطرفة – لا يمكن إيجادهما بالرسم (بيانياً) – لا يمكن حسابهما للتوزيعات التكرارية المفتوحة

ويمكن تلخيص كل ما يخص الوسط الحسابي والانحراف المتوسط والانحراف المعياري في الآتي :



- **للقيم المفردة :**

قيم عددها n	الانحرافات عن الوسط	القيم المطلقة للانحرافات	مربع الانحرافات
x	$d = x - \bar{x}$	$\lvert d \rvert$	d^2
...	...	...	...
...	...	...	...
$\sum x$		$\sum \lvert d \rvert$	$\sum d^2$

الوسط الحسابي $= \bar{x} = \frac{\sum x}{n}$

الانحراف المتوسط $= M.D = \frac{\sum |d|}{n}$

التباين $= s^2 = \frac{\sum d^2}{n}$ ⟵ الانحراف المعياري $= s = \sqrt{s^2}$

- **ولتوزيع تكراري :**

القيم	التكرار		الانحرافات عن الوسط	القيم المطلقة للانحرافات	مربع الانحرافات		
x	f	fx	$d = x - \bar{x}$	$\lvert d \rvert$	d^2	$f\lvert d \rvert$	fd^2
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
	$\sum f$	$\sum fx$				$\sum f\lvert d \rvert$	$\sum fd^2$

$\bar{x} = \frac{\sum fx}{\sum f}$

$M.D = \frac{\sum f|d|}{\sum f}$

$s^2 = \frac{\sum fd^2}{\sum f}$ ⟵ $s = \sqrt{s^2}$

- **وللبيانات المتصلة :**

– مثل التوزيع التكراري السابق فيما عدا أن كل فئة تُمثل بمركزها

الفئات	التكرار			الانحرافات عن الوسط	القيم المطلقة للانحرافات	مربع الانحرافات		
x	f	مراكز الفئات	fx_0	$d = x_0 - \bar{x}$	$\lvert d \rvert$	d^2	$f\lvert d \rvert$	fd^2
...	...	x_0	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
	$\sum f$	...	$\sum fx$				$\sum f\lvert d \rvert$	$\sum fd^2$

خاصيتان هامتان للانحراف المتوسط والانحراف المعياري :

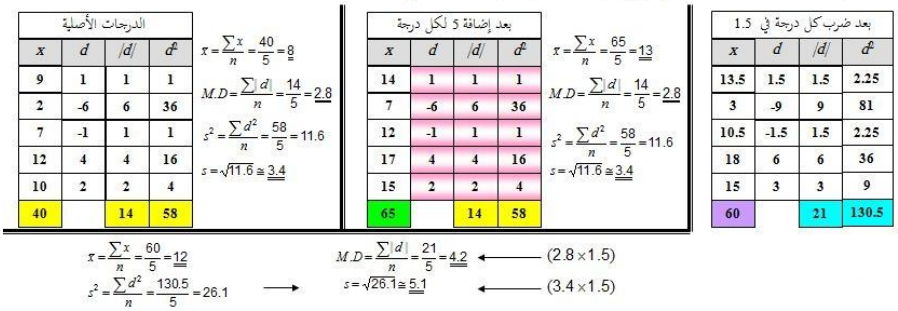
الخاصية الأولى : إضافة عدد ثابت C لكل قيمة من قيم البيانات لا يؤثر على قيمة الانحرافين المتوسط والمعياري .

الانحراف المتوسط (أو المعياري) الجديد = الانحراف المتوسط (أو المعياري) القديم

الخاصية الثانية : ضرب كل قيمة من قيم البيانات في عدد ثابت C يجعل :

الانحراف المتوسط (أو المعياري) الجديد = الانحراف المتوسط (أو المعياري) القديم × القيمة المطلقة للثابت C

فمثلاً، لو كانت لدينا البانات التالية والتي توضح درجات مجموعة من الطلاب كالتالي :



الدرجات الأصلية			
x	d	$\lvert d \rvert$	d^2
9	1	1	1
2	-6	6	36
7	-1	1	1
12	4	4	16
10	2	2	4
40		14	58

$\bar{x} = \frac{\sum x}{n} = \frac{40}{5} = 8$

$M.D = \frac{\sum |d|}{n} = \frac{14}{5} = 2.8$

$s^2 = \frac{\sum d^2}{n} = \frac{58}{5} = 11.6$

$s = \sqrt{11.6} \cong 3.4$

بعد إضافة 5 لكل درجة			
x	d	$\lvert d \rvert$	d^2
14	1	1	1
7	-6	6	36
12	-1	1	1
17	4	4	16
15	2	2	4
65		14	58

$\bar{x} = \frac{\sum x}{n} = \frac{65}{5} = 13$

$M.D = \frac{\sum |d|}{n} = \frac{14}{5} = 2.8$

$s^2 = \frac{\sum d^2}{n} = \frac{58}{5} = 11.6$

$s = \sqrt{11.6} \cong 3.4$

بعد ضرب كل درجة في 1.5			
x	d	$\lvert d \rvert$	d^2
13.5	1.5	1.5	2.25
3	-9	9	81
10.5	-1.5	1.5	2.25
18	6	6	36
15	3	3	9
60		21	130.5

$\bar{x} = \frac{\sum x}{n} = \frac{60}{5} = 12$

$s^2 = \frac{\sum d^2}{n} = \frac{130.5}{5} = 26.1$ ⟵

$M.D = \frac{\sum |d|}{n} = \frac{21}{5} = 4.2$ ⟵ (2.8×1.5)

$s = \sqrt{26.1} \cong 5.1$ ⟵ (3.4×1.5)

مع أصدق الأمنيات

أخوكم / ثابت

@thabet_18

المحاضرة الخامسة

اختبار الفروض الإحصائية

يعتبر اختبارات الفروض الإحصائية واحدة من أهم التطبيقات التي قدمها علم الإحصاء كحل للمشاكل العلمية المختلفة بشتى فروع العلم ورغم أهمية موضوع تقدير المعالم إلا انه غالبا ما يكون الاهتمام مركز ليس علي مجرد تقدير المعالم ولكن علي عملية وضع قواعد تمكن من التوصل إلي قرار بقبول أو رفض خاصية أو بالمعني الإحصائي فرض عن معالم مجتمع واحد أو أكثر وهذا ما يسمي اختبارات الفروض الإحصائية، ومن خلالها يمكن لأي شخص أن يتخذ قرار برفض أو قبول فرض معين أو مجموعة من الفروض المتعلقة بمشكلة معينة موجودة في الحياة العامة.

وبصفه عامه فان قبول او رفض اى قرار يجب ان يمر بعدة مراحل وهى:

- سحب عينة من المجتمع بحيث تكون ممثله للمجتمع(عشوائية)
- تجميع البيانات المتعلقة بالمشكلة من العينة
- تطبيق قواعد معينه لاختبار الفروض الموضوعه عن طريق الباحث وهى مشكله تحتاج لخبره احصائية
- اتخاذ القرار بناء على ما توصل اليه الباحث من نتائج.

اختبارات الفروض الاحصائية

Testing Statistical Hypotheses

من المعروف ان اتخاذ اى قرار لا يتم الا من خلال اختبارات الفروض الاحصائية التى تعتمد بدورها كما سبق على الاحتمالات وتوزيعات المعاينه وهذا يؤكد اهمية الدور الذى تلعبه نظريه الاحتمالات فى التنبؤ والتخطيط واتخاذ القرارات بالاضافه الى اهميتها فى تقدير معالم المجتمع المجهوله والتى تعتبر احد اهتمامات الباحثين.

تبدأ مشكله التعرف على معالم المجتمع المجهوله بما يسمى بالاستدلال الإحصائى (**Statistical Inferences**) حيث ينقسم الى فرعين. الفرع الأول يهتم بتقدير (**Estimation**) معالم المجتمع والفرع الأخر يختص باجراء اختبارات فرو ض (**Testing Hypothess**) تدور حول معالم المجتمع المجهوله.

الاستدلال الاحصائى يتم باستخدام عينة عشوائية مسحوبه من المجتمع وذلك لاستحالة التعامل مع المجتمع ككل، فالاحصاءات التحليلية قدمت القوانين التى سهلت هذه العملية وجعلتها تتم بأقل ألاخطاء الممكنه.

اختبارات الفروض ترتكز اساسا على فكرة ان هناك عينه اخرى غير مسحوبه من المجتمع المسحوب منه العينات فان الفرق بين الوسط الحسابى المحسوب من هذه العينه وبين المعلمه المجهوله قد يكون فرقا معنويا **Significant** غير راجع للصدفه. واشتقت اسمها منها حيث عن طريقها نستطيع ان نحدد وبسهوله هل الفروق بين المعلومات المحسوبة من العينة وبين المعلومات المفروضه لمجتع معين فرقا يرجع الى الصدفه ام فرق حقيقى، وبأسلوب أخر هل هو فرق معنوى او فرق غير معنوى؟ وبذلك سميت هذه الاختبارات باسم اختبارات المعنويه **Test of Significant**

مفهوم الاختبارات الإحصائية المعلمية واللامعملية

الاختبارات الاحصائيه قد تدور حول معالم المجتمع المجهوله مثل الفروض المتعلقة بالوسط الحسابى، النسبه، التباين، معامل الارتباط،... وفى هذه الحاله يطلق على هذه الاختبارات اسم الاختبارات المعلميه **Parametric Tests**

وقد تكون فروضا لا تتعلق بمعالم المجتمع ولكن تتعلق بأشياء اخرى قد تكون وصفية مثل العلاقة بين التعليم والتدخين، خضوع نتائج معينه لنظريه معينه، العلاقة بين لون العينين ولون الشعر،.... وفى هذه الحالة يسمى الاختبار باسم الاختبار اللامعلمى **Non Parametric Test**

فالإحصاء المعلمي: اساليب احصائية تتطلب افتراضات محددة عن طبيعة التوزيعات الاحتمالية للمجتمع

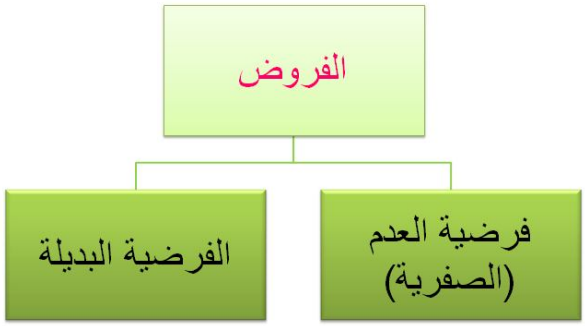
(مثل حساب الوسط الحسابي كمقياس للنزعة المركزية)

أما الإحصاء اللامعلمي: اساليب احصائية تتطلب افتراضات أقل عن طبيعة التوزيعات الاحتمالية للمجتمع

(مثل حساب الوسيط كمقياس للنزعة المركزية)

الفروض الإحصائية

تعتبر الفروض الإحصائية بمثابة اقتراح عن معالم المجتمع موضوع الدراسة، والتي ما زالت غير معلومة للباحث، فهي حلول ممكنة لمشكلة البحث



الفرضية الصفرية (فرضية العدم) HO (Null Hypothesis) :

هي الفرضية حول معلمة المجتمع التي نجري اختبار عليها باستخدام بيانات من عينة والتي تشير أن الفرق بين معلمة المجتمع والإحصائي من العينة ناتج عن الصدفة ولا فرق حقيقي بينهما. وهي الفرضية التي ننطلق منها ونرفضها عندما تتوفر دلائل على عدم صحتها، وخلاف ذلك نقبلها وتعني كلمة **Nul** انه لا يوجد فرق بين معلمة المجتمع والقيمة المدعاة (إحصائية العينة).

الفرضية البديلة Alternative Hypothesis (H_a) :

هي الفرضية التي يضعها الباحث كبديل عن فرضية العدم و نقبلها عندما نرفض فرضية العدم باعتبارها ليست صحيحة بناء على المعلومات المستقاة من العينة.

وفي اختبار الفروض يمكن أن نرتكب نوعين من الخطأ:

الخطأ من النوع الأول **Type I error**: الخطأ من النوع الأول هو "رفض الفرض العدمي بينما هو صحيح ". أي أنه على الرغم من أن الفرض العدمي في الواقع صحيح وكان من الواجب قبوله فقد تم أخذ قرار خاطئ برفضه. وباختصار شديد فإن الخطأ من النوع الأول هو : " رفض فرض صحيح". ويرمز له بالرمز α .

الخطأ من النوع الثاني **Type II error**: وفي المقابل فإن الخطأ من النوع الثاني يعني " قبول الفرض العدمي بينما هو خاطئ " أي أنه على الرغم من أن الفرض العدمي خاطئ وكان من الواجب رفضه فقد تم أخذ قرار خاطئ بقبوله وباختصار شديد فإن الخطأ من النوع الثاني هو " قبول فرض خاطئ ". ويرمز له بالرمز β .

ويمكن أن أن نمثل ذلك في الجدول التالي:

الفرضية / القرار	(H_o) صحيحة	(H_a) خاطئة
قبول (H_o)	صواب	خطأ ٢ بيتا (B)
رفض (H_o)	خطأ ١ ألفا (α)	صواب

فرضية صحيحة نتائج العينة تؤيد صحتها. (قبول صواب)

فرضية صحيحة نتائج العينة غير مؤيدة لصحتها. (رفض خطأ) وهذا يعطينا خطأ من النوع الأول ألفا (α) ويمكن أن يقلل برفع مستوى الدلالة

فرضية خاطئة نتائج تؤيد صحتها (قبول خطأ) وهذا يعطينا خطأ من النوع الثاني بيتا (B) ويمكن أن يقلل بزيادة حجم العينة

فرضية خاطئة نتائج غير مؤيدة صحتها (رفض صواب)

مستوى الدلالة الإحصائية ومنطقة الرفض

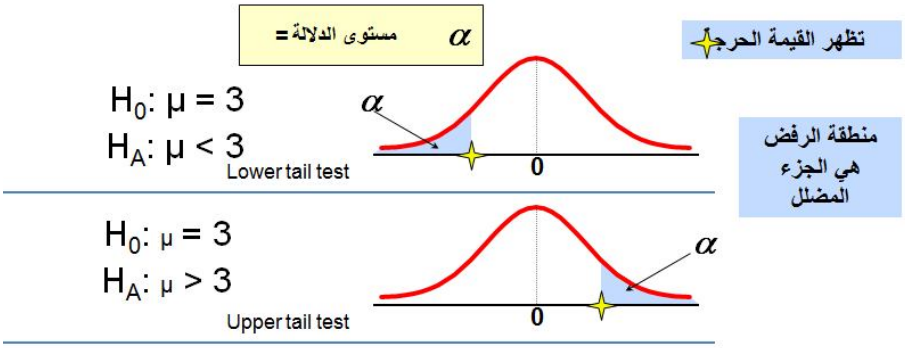
عندما نقبل الفرضية الصفرية (فرضية العدم) فإننا نقبلها بنسبة دقة ٩٠% أو ٩٥% أو ٩٩% أو غير ذلك وتسمى مستويات الدلالة أو الثقة **Significance Levels** أي يوجد نسبة خطأ معين في قبولنا للفرضية المبدئية بمعنى أننا نقبل صحة الفرضية الصفرية وهي خاطئة وهذا الخطأ هو α ويسمى مستوى المعنوية .

أي إذا كان مستوى الثقة ٩٥% $1-\alpha$ فان مستوى المعنوية α تساوي ٥% وهي عبارة عن مساحة المنطقة التي تقع تحت منحنى التوزيع والتي تمثل منطقة الرفض، وتكون أما على صورة ذيل واحد جهة اليمين أو اليسار أو ذيلين متساويين في المساحة واحد جهة اليمين والثاني جهة اليسار.

اختبار الفرضيات من طرف واحد:

هو الاختبار الذي تبين فيه الفرضية البديلة بأن معلمة المجتمع اكبر أو اصغر من معلمة المجتمع المفترضة.

مستوى الدلالة الإحصائية ومنطقة الرفض من طرف واحد



اختبار الفرضيات من طرفين:

هو الاختبار الذي لا تبين فيه الفرضية البديلة بأن معلمة المجتمع اكبر أو اصغر من معلمة المجتمع المفترضة بل مجرد أنها تختلف.

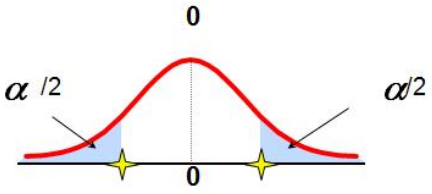
مستوى الدلالة الإحصائية ومنطقة الرفض من طرفين

مستوى الدلالة = α

اختبار من طرفين

$H_0: \mu = 3$

$H_A: \mu \neq 3$



مستوى المعنوية

Level of Significance

يعتبر مصطلح "مستوى المعنوية" واحدا من أهم المصطلحات المستخدمة في دراسة نظرية اختبارات الفروض. والمقصود بمستوى المعنوية α هو "احتمال حدوث الخطأ من النوع الأول". أو نسبة حدوثه "أي احتمال رفض الفرض العدمي بينما هو صحيح".

وعادة ما يرمز إلى مستوى المعنوية بالرمز اللاتيني ألفا α وأشهر قيمتين لمستوى المعنوية α هما 5%، 1 %، ولكن ليس هناك ما يمنع من أن يأخذ قيما أخرى.

ومن الملاحظات المهمة هنا هو أن "مستوى المعنوية" والذي يسمى أحياناً "مستوى الدلالة" هو المكمل لدرجة الثقة، بمعنى أن مجموعهما يساوي ١٠٠% أو واحد صحيح. فإذا كانت درجة الثقة ٩٥% فإن مستوى المعنوية يساوي ٥%. والعكس صحيح فإذا كان مستوى المعنوية ٥% فإن هذا يعني أن درجة الثقة ٩٥ %.

ولعل من أهم الملاحظات هنا هو استخدام تعبير "مستوى المعنوية" في حالات اختبارات الفروض، بينما يستخدم مصطلح "درجة أو مستوى الثقة" في حالات التقدير.

ويمكننا ضبط أو تحديد احتمال ارتكاب خطأ من النوع الأول α ولكن إذا خفضنا α فسوف نضطر إلى قبول احتمال أكبر لارتكاب خطأ من النوع الثاني β ، اللهم إلا إذا رفعنا حجم العينة .

وتمسى α مستوى المعنوية، و 1−α مستوى الثقة للاختبار.

والفكرة الأساسية في اختبار الفرض هي تقسيم المساحة تحت المنحنى إلى منطقتين: أحداهما تسمى "منطقة القبول" أي منطقة قبول الفرض العدمي. والأخرى تسمى "منطقة الرفض" أي منطقة رفض الفرض العدمي والتي تسمى أحيانا " بالمنطقة الحرجة Critical region ".

والنقطة الجديرة بالملاحظة هنا هي أن منطقة القبول تمثل درجة الثقة، بينما تمثل منطقة الرفض مستوى المعنوية.

خطوات إجراء الاختبار الإحصائي

الاختبار الاحصائى قد يكون متعلقا بعينة واحدة او عينتين او اكثر وقد يكون اختباراً معلمياً او غير معلمياً ويجب أن يمر الاختبار أياً كان نوعه بعدة خطوات وهذه الخطوات يمكن إيجازها في التالي:

١) صياغة الفرض الصفري H_o : والذي يأخذ – عادة – شكل " يساوي" فمثلاً إذا كان المطلوب هو اختبار ما إذا كان متوسط عمر الناخب هو 20 سنة فإن هذا الفرض يصاغ كما يلي: $Ho: \mu = 20$

٢) صياغة الفرض البديل H1 : والذي يأخذ أحد أشكال ثلاثة إما :

"لا يساوي " أو "أكبر من" أو "أقل من"

وبالرموز فإن الفرض البديل قد يأخذ شكل أحد الصيغ التالية:

والذي يحدد شكل الفرض البديل هو مدى اقتناع الباحث بذلك أو مدى توفر المعلومات الأولية

فمثلاً: إذا كانت وجهة نظر الباحث أن متوسط عمر الطالب موضع الدراسة لا يمكن أن يقل عن 20 سنة فإنه يختار الفرض البديل "أكبر من" والعكس صحيح إذا كان يعتقد أن متوسط عمر الطالب موضع الدراسة لا يزيد عن 20 سنة فإنه يختار الفرض البديل "أقل من" أما إذا لم يكن لديه أي تصور أو أي معلومات فإنه يختار الفرض البديل "لا يساوي"

٣) إحصائية الاختبار: وهي الإحصائية التي يتم حسابها من بيانات العينة بافتراض أن الفرض العدمي صحيح. ويتوقف شكل الإحصائية على العوامل التالية:

أ– توزيع المجتمع، وهل هو طبيعي أم لا، وهل تباينه معروف أم لا.

ب– حجم العينة، وهل هو كبير أم صغير.

جـ– الفرض العدمي المراد اختباره وهل هو عن الوسط أو النسبة أو التباين أو الارتباط... الخ.

والفكرة الأساسية (غالباً) في إحصائية الاختبار هي: حساب الفرق بين قيمة المعلمة التي نفترضها للمجتمع (في الفرض العدمي) والقيمة المقابلة لها في العينة أي التابع الإحصائي، ثم نقسم (أو ننسب) هذا الفرق إلى الخطأ المعياري للتابع الإحصائي.

٤) تحديد منطقتي القبول والرفضوذلك بناءً على الجداول الإحصائية والتي تعتمد على:

أ– توزيع المعاينة (وهل هو طبيعي أو t أو...)

ب– والفرض البديل (وهل هو لا يساوي أو أكبر من أو أقل من أي هل يستخدم اختبار الطرفين أو الطرف الأيمن أو الأيسر).

جـ– ومستوى المعنوية (وهل هو 1% أو 5% أو غير ذلك).

٥) المقارنة والقرار: بمعنى أن نقارن قيمة الإحصائية (المحسوبة من الخطوة الثالثة) بحدود منطقتي القبول والرفض (والتي حددناها في الخطوة الرابعة). فإذا وقعت قيمة الإحصائية داخل منطقة القبول فإن القرار هو: قبول الفرض الصفري. أما إذا وقعت قيمة الإحصائية في منطقة الرفض فإن القرار هو رفض الفرض الصفري، وفي هذه الحالة نقبل الفرض البديل. مع ملاحظة أن القرار مرتبط بمستوى المعنوية المحدد. بمعنى أن القرار قد يتغير إذا تغير مستوى المعنوية المستخدم (وفي بعض الحالات قد لا يتغير القرار، فهذا يتوقف على قيمة الإحصائية وما إذا كانت تقع في منطقة القبول أو منطقة الرفض).

أي أنه توجد طريقتين لاتخاذ القرار فى الاختبارات الاحصائية وهي:

أ- حساب احصاء الاختبار ومقارنته بقيمة جدوليه وتحدد القيمة الجدوليه بناء على نوع الاحتبار ذو طرف واحد One Tail Test أو ذو طرفين Two Tail Test

ب- حساب ما يسمى بالقيمة الاحتماليه P-value ويرمز لها فى بعض البرامج الإحصائية بالرمز .Sig فاذا كان الاختبار ذو طرف واحد تقارن .Sig بالقيمة α لكن اذا كان الاختبار ذو طرفين تقارن بالقيمة $\alpha/2$

المحاضرة السادسة

الاختبارات الإحصائية لعينة واحدة (اختبار t–test)

يهدف الباحث من اختبار الفرضيات حول المتوسط إلى اتخاذ قرار حول ما إذا كانت هذه الفرضية مقبولة أم مرفوضة، ويتم ذلك من خلال استخدام مختبر إحصائي مناسب، والمختبر الإحصائي هو متغير عشوائي ذو توزيع احتمالي يصف العلاقة بين القيم النظرية للمعلم والقيم المحسوبة من العينة، وفي العادة تقارن قيمة المختبر الإحصائي المحسوب من العينة مع قيمته المستخرجة من توزيعه الاحتمالي (باستخدام جداول خاصة) ومنها نتخذ القرار برفض أو قبول الفرضية الصفرية .

استخدام جدول t

يعتمد استخدام جدول t على مفهوم درجات الحرية df ، ودرجات الحرية (v) وتنطق (نيو) تساوي (n–1) وتعتبر درجات الحرية المعلمة الوحيدة لتوزيع t حيث يعتمد شكل منحنى t اعتمادا كاملا عليها، أي أن توزيع t يعتمد اعتمادا كاملا على حجم العينة n .

ولغرض استخدام جدول t نجد أن الصف العلوي من الأول يحوي Q على مساحة الطرف الأيمن (أو مساحة الطرف الأيسر) لتوزيع t بدرجات حرية محددة (v) ، بينما يحتوي الصف العلوي الثاني على Q٢ على مساحة الطرفين الأيمن والأيسر لتوزيع t بدرجات حرية محددة (v) ، حيث يستخدم هذا الصف عندما يكون الاختبار ذو طرفين، بينما يستخدم الصف الأول عندما يكون الاختبار ذو طرف واحد فقط.

ويحتوي العمود الأول بأقصى يسار الجدول على قيم لدرجات الحرية (v) ، ويمثل الرقم الموجود أما صف معين وتحت احتمال محدد قيمة t الحرجة التي تحدد منطقة الرفض لتوزيع t بدرجات حرية (v) ، ونتيجة لتماثل توزيع t حول الصفر، فإن جدول توزيع t يحتوي على القيم الموجبة فقط.

مثال:

افترض أن مستوى المعنوية في مشكلة معينة يساوي ٠.٠٥ ، وأن حجم العينة يساوي ٢٠ ، أوجد قيمة T الحرجة التي تناظر التالي:

١– اختبار ذو طرف أيمن.

٢– اختبار ذو طرف أيسر.

٣– اختبار ذو طرفين.

الحل:

١– عندما تكون $\alpha = 0.05$ و $v = (20-1) = 19$ ، نجد أن القيمة الموجودة أمام الصف ١٩ وتحت الإحتمال ٠.٠٥ الموجود في الصف العلوي الأول من الجدول هو ١.٧٢٩ ، أي أن قيمة t الحرجة بدرجات حرية ١٩ ومستوى معنوية ٠.٠٥ هو ١.٧٢٩ (لاختبار ذو طرف أيمن)، ويبين الجدول التالي جزء مستقطع من جدول t :

Q V درجات الحرية		0.05	
19		1.729	

٢- ونتيجة لأن توزيع t متماثل حول الصفر، فإن قيمة t الحرجة بدرجات حرية ١٩ والتي تكون المساحة إلى يسارها مساوية ٠.٠٥ هي -١.٧٢٩ ، وهذا في حالة الاختبار ذو الطرف الأيسر.

٣- إذا كان الاختبار ذو طرفين فإن قيمة α هي قيمة Q٢ الموجودة في الصف العلوي الثاني من جدول t ، وبالنظر إلى الجدول نجد أن القيمة الحرجة لـ t = 2.093 وهي القيمة الموجودة أمام الصف ١٩ وتحت العمود ٠.٠٥ في صف Q٢ العلوي الثاني، ويبين الجدول التالي جزء مستقطع من جدول t :

2Q V		0.05	
19		2.093	

t Table

cum. prob	$t_{.50}$	$t_{.75}$	$t_{.80}$	$t_{.85}$	$t_{.90}$	$t_{.95}$	$t_{.975}$	$t_{.98}$	$t_{.995}$	$t_{.999}$	$t_{.9995}$
one-tail	0.50	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
two-tails	1.00	0.50	0.40	0.30	0.20	0.10	0.05	0.02	0.01	0.002	0.001
df											
1	0.000	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	318.31	636.62
2	0.000	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.000	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.000	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.000	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.000	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.000	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.000	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.000	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.000	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.000	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.000	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.000	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.000	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.000	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.000	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.000	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.000	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.000	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.000	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.000	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.000	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	0.000	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.000	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.000	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.000	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.000	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.000	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	0.000	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.000	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.000	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	0.000	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	0.000	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.195	3.416
100	0.000	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.174	3.390
1000	0.000	0.675	0.842	1.037	1.282	1.646	1.962	2.330	2.581	3.098	3.300
z	0.000	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.090	3.291
	0%	50%	60%	70%	80%	90%	95%	98%	99%	99.8%	99.9%
	Confidence Level										

اختبار t-test لعينة واحدة :

ولكن إذا كان حجم العينة صغيراً (n<30) فإن قيمة (S2)تتغير كثيراً من عينة إلى أخرى وبالتالي لا يمكننا هنا أن نستخدم اختبار (Z)، مما دفع كثيراً من الإحصائيين للبحث عن البديل المناسب .

ففي سنة ١٩٠٨م استطاع العالم الايرلندي وليم كوسيت W.S. Gosset من نشر بحث تحت اسم مستعار بسبب ظروف خاصة هو (استيودنت، Student) استطاع من خلاله أن يشتق معادلة للتوزيع الاحتمالي (t) الذي قيمته هي :

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

وهذا الاختبار يشبه اختبار التوزيع الطبيعي (Z)، بيد أنه يختلف عنه في تضمنه للانحراف المعياري (S) للعينة بدلاً من الانحراف المعياري (σ) للمجتمع .

يستخدم هذا الصنف من اختبار (ت) للحكم على معنوية الفروق بين متوسط عينة ومتوسط المجتمع الذي سحبت منه. ولغرض توضيح ذلك إليك هذا العرض الموجز لخطوات اختبار (ت) حول متوسط حسابي واحد على افتراض أن تباين المجتمع $\sigma 2$ غير معلوم وأن حجم العينة صغير .

والجدول التالي يوضح خطوات اختبار حول متوسط حسابي واحد باستخدام المختبر الإحصائي (ت) t–test على افتراض أن تباين المجتمع $\sigma 2$ غير معلوم وأن حجم العينة صغير

خطوات الاختبار	اختبار ذو طرفين	اختبار ذو طرف واحد	
		طرف يمين	طرف يسار
١ – الفرضية الصفرية H0 ٢ – الفرضية البديلة H1	$\mu = \mu_0$ $\mu \neq \mu_0$	$\mu = \mu_0$ $\mu > \mu_0$	$\mu = \mu_0$ $\mu < \mu_0$
٣ – مستوى الدلالة	α		
٤ – منطقة الرفض	ت ≤ ت(df; $\alpha\frac{1}{2}$) أو ت ≥ ت(df; $\alpha\frac{1}{2}$)	ت ≥ ت(df; α)	ت ≤ ت(df; α)
٥ – المختبر الإحصائي	$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$ هذا المختبر الإحصائي يستخدم لتوضيح أهمية الفروق بين الوسط الحسابي للعينة والوسط الحسابي للمجتمع		
٦ – القرار	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت (df; $\alpha\frac{1}{2}$)] الجدولية أي أن : ت ≤ ت(df; $\alpha\frac{1}{2}$) أو ت ≥ ت(df; $\alpha\frac{1}{2}$)	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت (df; α)] الجدولية أي أن: ت ≤ ت(df; α)	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [-ت(df; α)] الجدولية أي أن: ت ≥ - ت(df; α)

مثال على اختبار t لعينة واحدة :

لو كانت لدينا عينة عشوائية تتكون من ٢٥٠ طالب ، وجـــد أن الوسط الحسابي لأطوال طلاب العينة ١٥٥.٩٥سم، والانحراف المعياري = ٢.٩٤ سم، علما بأن الوسط الحسابي لأطوال طلاب الجامعة يبلغ ١٥٨ سم، اختبر أهمية الفرق المعنوي بين الوسط الحسابي لأطوال طلاب العينة والوسط الحسابي لأطوال طلاب الجامعة .

الحل :

سيتم اختبار الفرضيات التالية :

الفرضية الصفرية : لا توجد فروق ذات دلالة إحصائية بين متوسط أطوال الطلاب في العينة ومتوسط أطوال الطلاب في الجامعة ($\mu = \mu_0$)

الفرضية البديلة : توجد فروق ذات دلالة إحصائية بين متوسط أطوال الطلاب في العينة ومتوسط أطوال الطلاب في الجامعة ($\mu \quad \mu_0$)

مستوى الدلالة : $\alpha = 0.05$

منطقة الرفض : قيمة (ت) الجدولية عند مستوى دلالة α 0.05 ودرجات حرية ٢٤٩ = ١.٩٦٠

المختبر الإحصائي :

$\overline{X}$ = ١٥٥.٩٥سم ، n= ٢٥٠ طالب ، S = ٢.٩٤ سم

μ = ١٥٨سم

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}} = \frac{155.95 - 158}{2.94/\sqrt{250}} = -11.006$$

القرار :

٠.٠ قيمة ت المحسوبة (– ١١.٠٠٦) أكبر من قيمة ت المجدولة (١.٩٦) عند مستوى دلالة $\alpha = 0.05$.

∴ نرفض الفرضية الصفية ونقل البديلة .

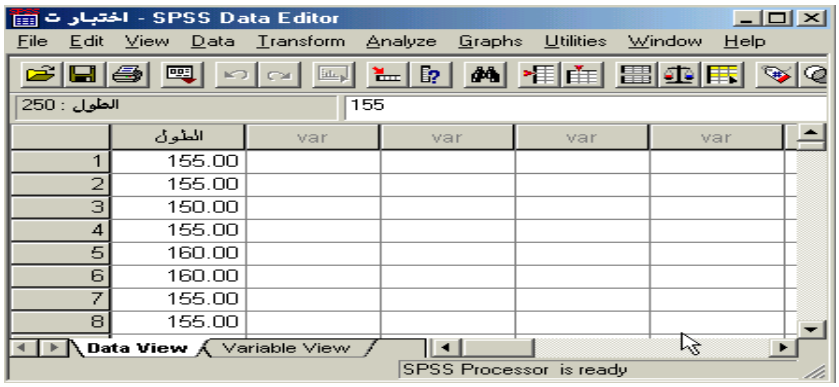
أي أنه توجد فروق ذات دلالة إحصائية بين الوسط الحسابي للعينة والوسط الحسابي لمجتمع البحث .

حساب اختبار (ت) لعينة واحدة

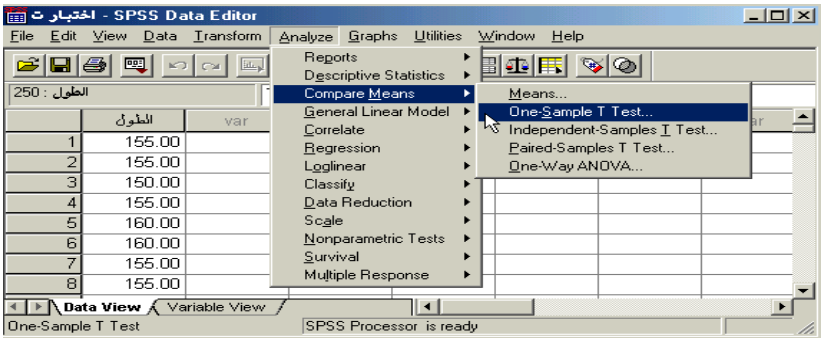
One Sample T-Test من خلال الـ SPSS

لغرض حساب قيمة (ت) لنفس المثال السابق من خلال استخدام برنامج الـ SPSS اتبع الخطوات التالية :

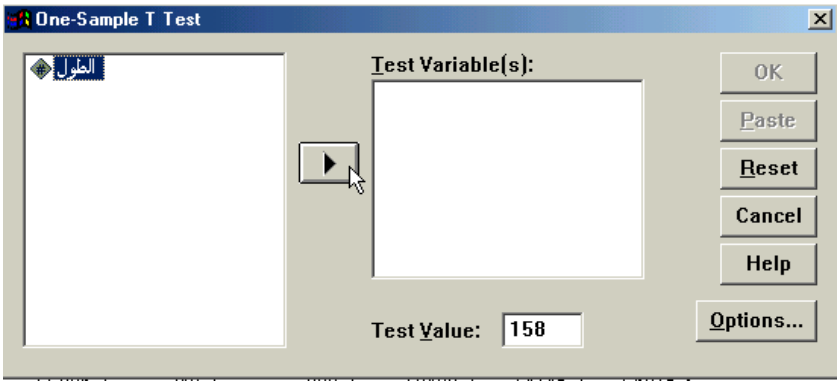
قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات Data Editor بالطريقة المناسبة كالتالي :



✓ من القائمة "تحليل" **Analyze** اختر الأمر "مقارنة المتوسطات" **Compare Means** فتظهر قائمة أوامر فرعية اختر منها "اختبار (ت) لعينة واحدة" **One-Sample T Test** كالتالي:



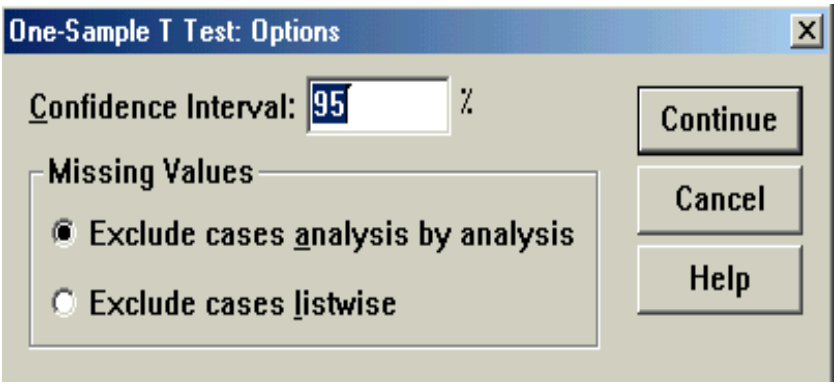
✓ بعد اختيار الأمر "اختبار (ت) لعينة واحدة" **One-Sample T Test** سوف يظهر لك صندوق الحوار التالي :



✓ من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار انقر نقرا مزدوجا على المتغير "الطول" (أو أنقر على السهم الذي يظهر في صندوق الحوار بعد التظليل على المتغير المرغوب نقله إلى الجهة الأخرى) ستلاحظ انتقاله مباشرة في المستطيل "متغيرات الاختبار" **Test Variable(s)**.

✓ في الحقل الخاص بـ "القيمة المختبرة" **Test Value** أكتب القيمة التي تريد أن تقارن بها متوسط العينة موضع الدراسة (في هذا المثال يتم كتابة الرقم ١٥٨ والذي يمثل متوسط أطوال الطلاب في الجامعة) .

✓ قم بالنقر على زر "خيارات" **Options** في الجهة السفلية اليمنى من صندوق الحوار السابق وذلك عند الرغبة في تغيير قيمة "فترة الثقة" **Confidence Interval** حيث يظهر لك صندوق الحوار التالي والذي يتيح إمكانية تغيير فترة الثقة المختبرة (بشكل تلقائي سوف تظهر القيمة ٩٥%) ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" **Continue** .



✓ أنقر بعد ذلك على زر "موافق" **OK** سيؤدي ذلك إلى تنفيذ الاختبار، وستلاحظ ظهور النتائج في شاشة المخرجات كالتالي :

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
الطول	250	155.9520	2.9422	.1861

One-Sample Test

	Test Value = 158					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
الطول	-11.006	249	.000	-2.0480	-2.4145	-1.6815

يتضح من النتائج أن قيمة (ت) المحسوبة **t–test = –11.006** ، ودرجات الحرية **df = 249** ، وقيمة **Sig.** (**2–tailed**) =0.000 ، وبما أن قيمة **Sig.** (**2–tailed**) في الجدول (٠.٠٠٠) أصغر من قيمة **α = 0.05** فإننا بالتالي نرفض الفرضية الصفرية، أي أنه توجد فروق ذات دلالة إحصائية بين متوسط أطوال العينة ومتوسط أطوال طلاب الجامعة .

والأوامر المستخدمة لحساب اختبار (ت) لعينة واحدة **One Sample T–Test** من خلال برنامج الـ **SPSS** (مع اختلاف المتغيرات موضع الدراسة) هي :

```
T–TEST
/TESTVAL=158
/MISSING=ANALYSIS
VARIABLES =الطول
/CRITERIA=CIN (.95)
```

مع أصدق الأمنيات

أخوكم / ثابت

@thabet_18

المحاضرة السابعة

الاختبارات الإحصائية لعينتين مستقلتين

(اختبار t–test لعينتين مستقلتين)

يستخدم هذا النوع من اختبار (ت) للحكم على معنوية Significance الفروق بين متوسطي عينتين غير مرتبطتينIndependent ولغرض توضيح ذلك إليك هذا العرض الموجز لخطوات اختبار (ت) حول متوسطين على افتراض أن تباين المجتمع σ_1^2 و σ_2^2غير معلومين ولكنهما متجانسين وأن حجم العينة صغير.

ويستند هذا الاختبار إلى توفر عدد من الافتراضات وهذه الافتراضات هي :

- **مستوى القياس: يشترط لاستخدام هذا الاختبار أن تكون البيانات فئوية (فترية) أو نسبية.**
- **أن يكون حجم العينة صغيراً: يقتضي هذا الافتراض أن يكون حجم العينة أقل من (٣٠) وأكبر من (٥) (وإذا كان حجم العينة أكبر من ٣٠ فلا بأس من ذلك) ، بحيث أن حجم ( ) لا يسمح بتقدير جدي لـ (σ) ، لذا ولتجنب الخطأ نستعيض عن (σ) بالمعلمة (S) . ويجب كذلك أن يكون الفرق بين حجم عينتي البحث صغيراً ، لأنه كلما زاد الفرق بين حجم العينتين أثر ذلك على قيمة (t) المحسوبة.**
- **التوزيع الطبيعي: ويقضي هذا الافتراض أن المشاهدات (س١) في المجتمع الأول تتخذ شكل التوزيع الطبيعي لوسط يساوي (μ_1) وكذلك الأمر بالنسبة للمشاهدات (س٢) في المجتمع الثاني يفترض فيها أن تتخذ شكل التوزيع الطبيعي لوسط يساوي (μ_2) ، وإن مخالفة هذا الافتراض ليس لها تبعات تذكر .**
- **تجانس التباين في المجتمعين: وبموجب هذا الافتراض يكون لتباين المشاهدات في كل من المجتمعين نفس القيمة (σ^2) وبذلك تكون القيمة المتوقعة للتباين في كل العينتين مساوية للمقدار (σ^2) أي يكون كل من (S_1^2 ، S_2^2)تقديراً مستقلاً لنفس المقدار (σ^2) وإن هناك اتفاقاً بين الإحصائيين على أن افتراض تجانس التباين هذا يمكن مخالفته إذا تساوت العينات في أعداد أفرادها، أي إذا كانت ($n_1 = n_2$) .**

أما عندما تكون ($n_1 \neq n_2$) فإن هناك وضعين هما :

١. **عندما تكون العينية الكبيرة الحجم منتمية للمجتمع ذي التباين الكبير، والعينة صغيرة الحجم منتمية للمجتمع ذي التباين الصغير، فإن احتمال ارتكاب الخطأ من النوع الأول تكون أقل من (α) أي لا يصل الأمر إلى رفض الفرضية الصفرية وهي صحيحة وذلك بسبب مخالفة هذا الافتراض. ومن ذلك يكون الباحث في هذه الحالة في الجانب الأمين، أي إن n_2 مخالفة تجانس التباين لا تؤثر على نتائج البحث.**
٢. **أما عندما تكون العينة كبيرة الحجم منتمية للمجتمع ذي التباين الصغير والعينة صغيرة الحجم منتمية للمجتمع ذي التباين الكبير، فإن احتمال ارتكاب الخطأ من النوع الأول يزداد (رفض الفرضية الصفرية وهي صحيحة) ، وفي هذه الحالة ليس هناك حاجة للقلق من مخالفة هذا الافتراض وذلك عندما تكون الفرضية الصفرية مقبولة، ولكن يجب أن يساورنا القلق إذا رفضناها ، وذلك لازدياد احتمال رفضها وهي صحيحة (أي ازدياد احتمال ارتكاب الخطأ من النوع الأول)، ولتفادي ذلك نستخدم اختبار ولتيش Welchبالعلاقة :**

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

– الاستقلالية: ويقتضي هذا الافتراض أن (n_1) من المشاهدات قد تم الحصول عليها عشوائيا من المجتمع الأول بشكل مستقل عن (n_2) من المشاهدات والتي تم الحصول عليها عشوائياً من المجتمع الثاني. وإن الاستقلالية هنا لا تعني استقلالية البيانات بين المجتمعين فقط ، بل تعني استقلالية المشاهدات ضمن المجتمع الواحد أيضاً (مثل عملية تطبيق اختبار قبلي واختبار بعدي على مجموعة واحدة) .

والجدول التالي يوضح خطوات اختبار الفرق بين متوسطين باستخدام المختبر الإحصائي (ت) على افتراض أن تباين المجتمع σ_1^2 و σ_2^2 غير معلومين ولكنهما متجانسين ، وأن كل من n_1 و n_2 صغير

خطوات الاختبار	اختبار ذو طرفين	اختبار ذو طرف واحد	
		طرف يمين	طرف يسار
١ – الفرضية الصفرية H0 ٢ – الفرضية البديلة H1	$\mu1 = \mu2$ $\mu1 \neq \mu_2$	$\mu1 = \mu2$ $\mu1 > \mu_2$	$\mu1 = \mu2$ $\mu1 < \mu_2$
٣ – مستوى الدلالة	α		
٤ – منطقة الرفض	ا ت اكـ ت$(df; \alpha\frac{1}{2})$	ت كـ ت $(df; \alpha)$	ت ≥ ت $(df; \alpha)$
٥ – المختبر الإحصائي	$t = \frac{\overline{X_1} - \overline{X_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ هذا المختبر الإحصائي يستخدم لتوضيح أهمية الفروق بين متوسطين للعينات المستقلة وعند تجانس التباين		
٦ – القرار	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت ($df; \alpha\frac{1}{2}$] الجدولية أي أن : ا ت ا كـ ت($df; \alpha\frac{1}{2}$)	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت ($df; \alpha$)] الجدولية أي أن: ت كـ ت$(df; \alpha)$	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [–ت $(df; \alpha)$] الجدولية أي أن: ت ≥ – ت$(df; \alpha)$

ولتوضيح ما ورد في الجدول السابق دعنا نتناول هذا المثال :

أراد باحث أن يعرف أثر استخدام نظم مساندة القرارات على كفاءة القرارات التي تتخذها الإدارة بمساعدة تلك النظم، فوزع ٥٠ مديرا لمنشآت صناعية عشوائيا في مجموعتين، ثم عين أحدهما بطريقة عشوائية لتكون مجموعة تجريبية والأخرى ضابطة، وفي نهاية التجربة وزع على المجموعتان استقصاء يقيس درجة فاعلية القرار وكفاءته عندما يتم اتخاذه باستخدام نظم مساندة القرارات بدلا من الطريقة التقليدية فكانت النتائج كما يلي:

المجموعة التجريبية	المجموعة الضابطة
n_1 = ٢٥	n_2 = ٢٥
$\overline{X_1}$ = ٧.٦٠	$\overline{X_2}$ = ٦.٠
S_1^2 = ٢.٢٧	S_2^2 = ١.٧٨

فهل تدل هذه البيانات على أن أداء المجموعة التجريبية كان أفضل من أداء المجموعة الضابطة عند مستوى $\alpha = 0.05$ ؟

الحل :

سيتم اختبار الفرضيات التالية :

الفرضية الصفرية : لا توجد فروق ذات دلالة إحصائية بين متوسط المجموعة التجريبية ومتوسط المجموعة الضابطة ($\mu1=\mu2$).

الفرضية البديلة : توجد فروق ذات دلالة إحصائية بين متوسط المجموعة التجريبية ومتوسط المجموعة الضابطة لصالح المجموعة التجريبية ($\mu1>\mu2$)

مستوى الدلالة : $\alpha = 0.05$

منطقة الرفض : ٠.٠ قيمة مستوى الدلالة $\alpha = 0.05$ والاختبار بذيل واحد ، ودرجات الحرية = ٢٥ + ٢٥ – ٢ = ٤٨ ، بذلك تكون قيمة (ت) الجدولية = ١.٦٨

المختبر الإحصائي :

$$t = \frac{\overline{X_1} - \overline{X_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

ولتطبيق هذه العلاقة يلزمنا حساب قيمة الانحراف المعياري (S) من خلال العلاقة التالية:

$$S^2 = \frac{\left[(n_1 - 1)(S_1^2)^2\right] + \left[(n_2 - 1)(S_2^2)^2\right]}{(n_1 + n_2) - 2}$$

إذا التباين يساوي:

$$S^2 = \frac{\left[(25-1)(2.27)^2\right] + \left[(25-1)(1.78)^2\right]}{(25+25) - 2} = 4.16$$

إذن الانحراف المعياري يساوي :

$$S = \sqrt{S^2} = \sqrt{4.16} = 2.04$$

ثم نحسب قيمة (ت) من خلال تطبيق العلاقة التالية :

$$t = \frac{\overline{X_1} - \overline{X_2}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{7.60 - 6.0}{2.04\sqrt{\frac{1}{25} + \frac{1}{25}}} = 2.77$$

القرار :

٠.٠ قيمة (t) المحسوبة (٢.٧٧) أكبر من قيمة (ت) المجدولة (١.٦٨) عند مستوى دلالة α = ٠.٠٥ .

∴ نرفض الفرضية الصفرية ونقبل البديلة

أي أن المجموعة التي خضعت للتجربة يصبح أداؤهم أفضل في عملية اتخاذ القرار من الذين لم يخضعون للتجربة وذلك عند مستوى دلالة α ٠.٠٥ .

أما في حالة عدم تجانس التباين فإننا نستخدم طريقة وليتش **Weltch** لحساب قيمة (**t**) وذلك من خلال تطبيق العلاقة :

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

وللتحقق من تجانس التباين يتم ذلك من خلال تطبيق العلاقة التالية:

$$F = \frac{S_g^2}{S_l^2}$$

حيث:

S_g^2 تعني التباين الأكبر

S_l^2 تعني التباين الأصغر.

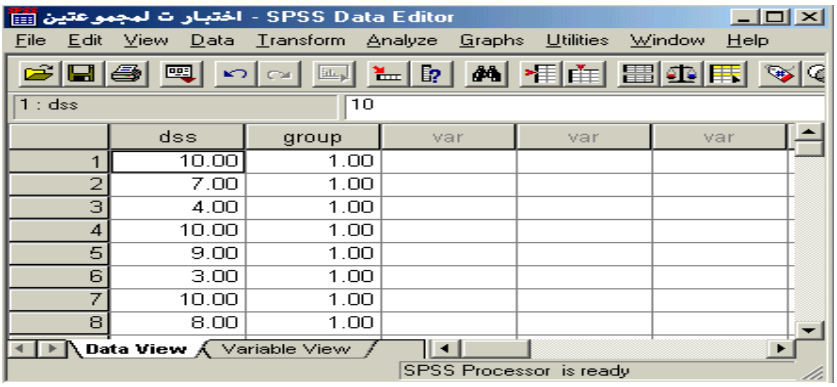
ومن ثم مقارنة قيمة (**F**) المحسوبة (لقياس تجانس التباين) بقيمة (**F**) المجدولة عند درجة حرية (n_1-1 و n_2-1)

حساب اختبار (ت) للعينات المستقلة

Independent Samples T-Test من خلال الـ **SPSS**

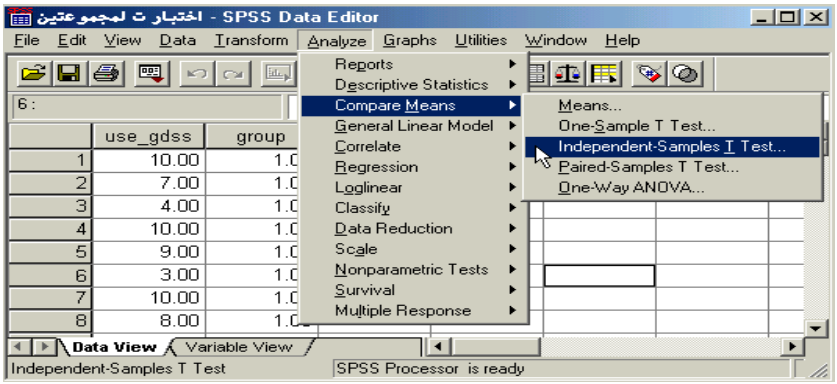
لغرض حساب قيمة (ت) لنفس المثال السابق من خلال استخدام برنامج الـ **SPSS** نتبع الخطوات التالية :

✓ قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات **Data Editor** بالطريقة المناسبة كالتالي :

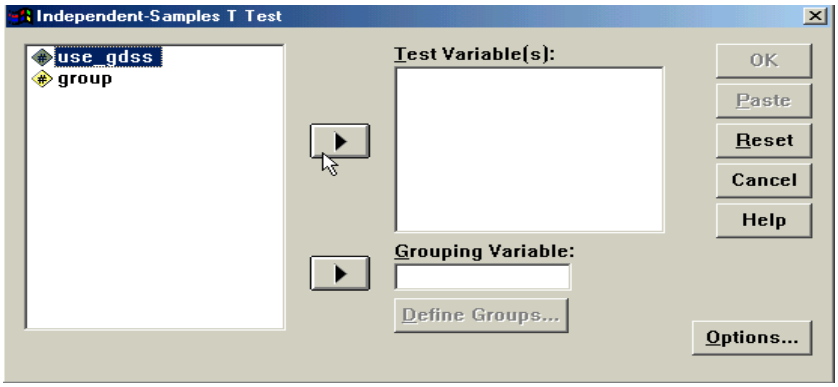


لاحظ أنه تم إدخال النتائج المتحصل عليها تحت متغير واحد باسم **dss** ، وتم إنشاء متغير آخر بمسمى **group** ليحوي رمز للمجموعة التجريبية والمجموعة الضابطة .

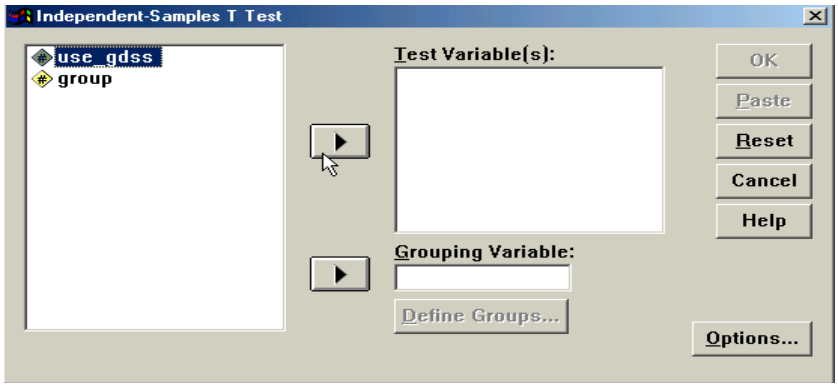
✓ من القائمة "تحليل" **Analyze** اختر الأمر "مقارنة المتوسطات" **Compare Means** فتظهر قائمة أوامر فرعية اختر منها "اختبار (ت) للعينات المستقلة" **Independent Samples T-Test** كالتالي:



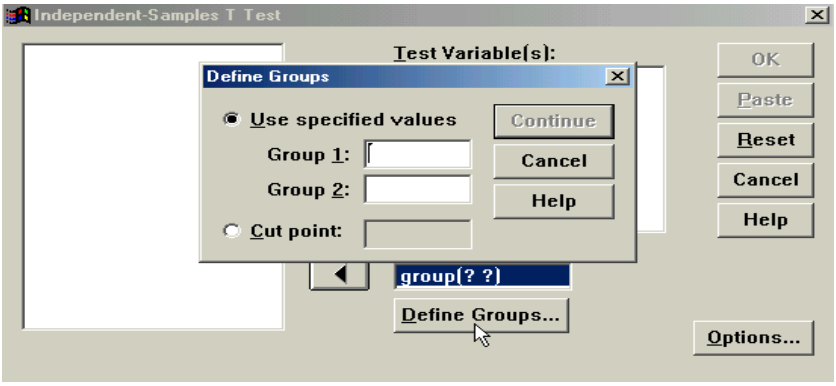
✓ بعد اختيار الأمر "اختبار (ت) للعينات المستقلة" **Independent Samples T-Test** سوف يظهر لك صندوق الحوار التالي :



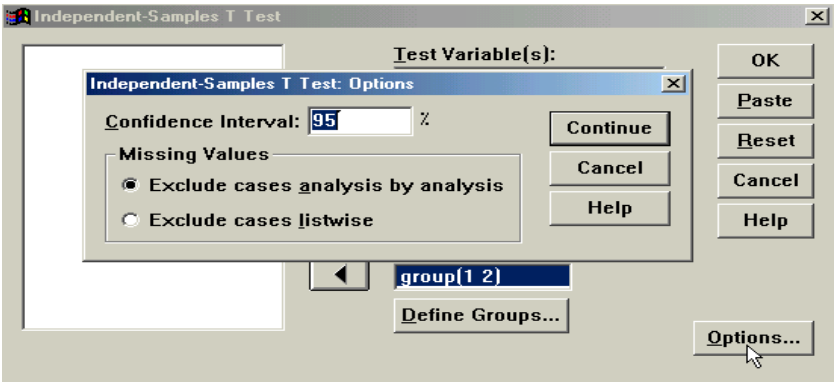
✓ من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار حدد المتغير المراد نقله إلى المستطيل الخاص بـ "متغيرات الاختبار" **Test Variable(s)** ومن ثم أنقر على السهم الذي يظهر مقابل المستطيل الخاص بـ "متغيرات الاختبار" ، ستلاحظ انتقال المتغير مباشرة في المستطيل "متغيرات الاختبار" **Test Variable(s)** ، واعمل نفس الشيء مع المتغير **group** لنقله إلى الحقل الخاص بـ "مجاميع المتغيرات" **Grouping Variable** .



✓ أنقر على زر "تعريف المجموعات" **Define Groups** في أسفل صندوق الحوار السابق ، سيؤدي ذلك إلى فتح صندوق حوار صغير يتيح الفرصة للمستخدم لتعريف قيم المجموعة الأولى (والتي يمثلها الرقم ١) وقيم المجموعة الثانية (والتي يمثلها الرقم ٢) .



✓ أنقر على زر "خيارات" **Options** في الجهة السفلية اليمنى من صندوق الحوار السابق وذلك عند الرغبة في تغيير قيمة "فترة الثقة" **Confidence Interval** حيث يظهر لك صندوق الحوار التالي والذي يتيح إمكانية تغيير فترة الثقة المختبرة (بشكل تلقائي سوف تظهر القيمة ٩٥%) ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" **Continue** .



✓ أنقر بعد ذلك على زر "موافق" **OK** سيؤدي ذلك إلى تنفيذ الاختبار، وستلاحظ ظهور النتائج في شاشة المخرجات كالتالي :

T-Test

Group Statistics

	GROUP	N	Mean	Std. Deviation	Std. Error Mean
USE_GDSS	USE DSS	25	7.6000	2.2730	.4546
	NOT USE DSS	25	6.0000	1.7795	.3559

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. E Differe
USE_GDSS	Equal variances assumed	1.095	.301	2.771	48	.008	1.6000	.5
	Equal variances not assumed			2.771	45.386	.008	1.6000	.5

يتضح من النتائج أن قيمة (**F**) = ١.٠٩٥ ومستوى دلالتها ٠.٣٠١ وهذه القيمة أكبر من ٠.٠٥ ، مما يدل على أنها غير دالة (وهذا يعني أن هناك تجانس بين تباين المجموعتين)، وهذا يدفعنا إلى قراءة نتائج اختبار (ت) المقابلة للعبارة "افتراض تساوي التباين" **Equal variances assumed**، من هذه النتائج نلاحظ أن قيمة (ت) المحسوبة **t-test** = ٢.٧٧١ ، ودرجات الحرية **df** = ٤٨ ، وقيمة (**Sig. (2-tailed**) = ٠.٠٠٨ ، وبما أن قيمة (**Sig. (2-tailed**) في الجدول (٠.٠٠٨) أصغر من قيمة **α** = ٠.٠٥ فإننا بالتالي نرفض الفرضية الصفرية، أي أنه توجد فروق ذات دلالة إحصائية بين متوسط المجموعة التجريبية ومتوسط المجموعة الضابطة لصالح المجموعة الضابطة (وذلك بسبب حصولها على متوسط حسابي أكبر = ٧.٦٠).

والأوامر المستخدمة لحساب اختبار (ت) للعينات المستقلة Independent Samples T–Test من خلال برنامج الـ SPSS (مع اختلاف المتغيرات موضع الدراسة) هي:

```
T–TEST
GROUPS=group(1 2)
/MISSING=ANALYSIS
/VARIABLES=use_gdss
/CRITERIA=CIN(.95) .
```

المحاضرة الثامنة

الاختبارات الإحصائية لعينتين مرتبطتين

(اختبار **t–test** لعينتين مرتبطتين)

يستخدم هذا النوع للحكم على دلالة الفروق ومعنويتها **Significance** بين متوسطي عينتين مرتبطتين **Correlated Data**، مثل اختبار دلالة الفروق بين متوسط أداء الموظفين قبل التدريب وبعد التدريب. ولغرض توضيح ذلك إليك هذا العرض الموجز لخطوات اختبار (ت) حول متوسطين مرتبطين على افتراض أن تباين المجتمع و غير معلومين وأن حجم العينة صغير .

والجدول التالي يوضح خطوات اختبار الفرق بين متوسطين مرتبطين باستخدام المختبر الإحصائي (ت) على افتراض أن تباين المجتمع وغير معلومين، وأن حجم العينة صغير

خطوات الاختبار	اختبار ذو طرفين	اختبار ذو طرف واحد	
		طرف يمين	طرف يسار
١ – الفرضية الصفرية H0 ٢ – الفرضية البديلة H1	$\mu1 = \mu2$ $\mu1 \neq \mu_2$	$\mu1 = \mu2$ $\mu1 > \mu_2$	$\mu1 = \mu2$ $\mu1 < \mu_2$
٣ – مستوى الدلالة	α		
٤ – منطقة الرفض	ت أكبر من $(df: \alpha\frac{1}{2})$	ت أكبر من $(df: \alpha)$	ت أقل من $(df: \alpha)$
٥ – المختبر الإحصائي	$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} - 2r\left(\frac{S_1}{\sqrt{n_1}}\right)\left(\frac{S_2}{\sqrt{n_2}}\right)}}$ هذا المختبر الإحصائي يستخدم لتوضيح أهمية الفروق بين متوسطين للعينات المرتبطة		
٦ – القرار	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت $(df: \alpha\frac{1}{2})$] الجدولية أي أن : ت أكبر من $(df: \alpha\frac{1}{2})$	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [ت $(df: \alpha)$] الجدولية أي أن : ت أكبر من $(df: \alpha)$	أرفض الفرضية الصفرية إذا كانت قيمة المختبر الإحصائي (ت) المحسوبة تساوي أو أكبر من قيمة [–ت $(df: \alpha)$] الجدولية أي أن : ت أقل من – ت $(df: \alpha)$

ولتوضيح ما ورد في الجدول السابق دعنا نتناول هذا المثال :

أراد باحث أن يعرف أثر برنامج التدريب الصيفي في الميدان على أداء الطلاب وتحصيلهم في كلية العلوم الإدارية، ولغرض تحقيق ذلك قام الباحث باختبار الطلاب قبل وبعد البرنامج التدريبي، ولكون نفس الطلاب أخذوا الاختبارين، فإن الباحث يتوقع معامل ارتباط موجب بين تحصيل الطلبة في كلا القياسين. ولغرض اختبار مدى دلالة الفروق بين الاختبار القبلي والاختبار البعدي، لا بد على الباحث أن يتأكد من قيمة الارتباط بين الاختبارين والتي كانت $r = 0.46$ ، وقد كانت النتائج التي تم التوصل إليها كما يلي :

الإختبار القبلي	الإختبار البعدي
n_1 = ١٠٠	n_2 = ١٠٠
$\overline{X_1}$ = ٥٤.٢٨	$\overline{X_2}$ = ٥٨.٦٦
S_1^2 = ٤٩	S_2^2 = ٦٤

فهل تدل هذه البيانات على أن أداء الطلاب التحصيلي في الكلية بعد أخذ البرنامج التدريبي كان أفضل من أدائهم قبل أخذ البرنامج التدريبي عند مستوى α = ٠.٠٥ ؟

الحل :

سيتم اختبار الفرضيات التالية :

الفرضية الصفرية : لا توجد فروق ذات دلالة إحصائية بين متوسط تحصيل الطلاب قبل وبعد البرنامج التدريبي ($\mu_1=\mu_2$).

الفرضية البديلة : توجد فروق ذات دلالة إحصائية بين متوسط تحصيل الطلاب قبل وبعد البرنامج التدريبي لصالح بعد البرنامج ($\mu_1 \geq \mu_2$)

مستوى الدلالة : α = ٠.٠٥

منطقة الرفض : قيمة مستوى الدلالة α = ٠.٠٥ والاختبار بذيل واحد، ودرجات الحرية = ١٠٠-١= ٩٩ ، بذلك تكون قيمة (ت) الجدولية = ١.٦٦٠

المختبر الإحصائي :

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} - 2r\left(\frac{S_1}{\sqrt{n_1}}\right)\left(\frac{S_2}{\sqrt{n_2}}\right)}}$$

إذا قيمة (ت) تساوي :

$$t = \frac{58.66 - 54.28}{\sqrt{\frac{64.0}{100} + \frac{49.0}{100} - 2(0.46)\left(\frac{8}{\sqrt{100}}\right)\left(\frac{7}{\sqrt{100}}\right)}} = 5.57$$

القرار :

قيمة (ت) المحسوبة (٥.٥٧) أكبر من قيمة (ت) المجدولة (١.٦٦٠) .عند مستوى دلالة α = ٠.٠٥ .

∴ نرفض الفرضية الصفرية ونقبل البديلة، أي أن للبرنامج التدريبي تأثير إيجابي على تحصيل الطلاب وأدائهم في الكلية وذلك عند مستوى دلالة ٠.٠٥α

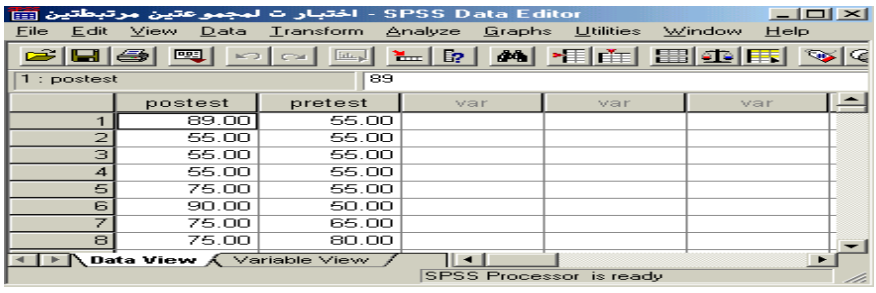
في هذه المعادلة ليس هناك مانع من الابتداء بـ X_1 أو X_2 في الترتيب ، لأن الإشارة ليس لها أي تأثير على النتيجة المتحصلة

حساب اختبار (ت) لعينتين غير مستقلتين (العينات المرتبطة)

Paired Samples T-Test من خلال الـ SPSS

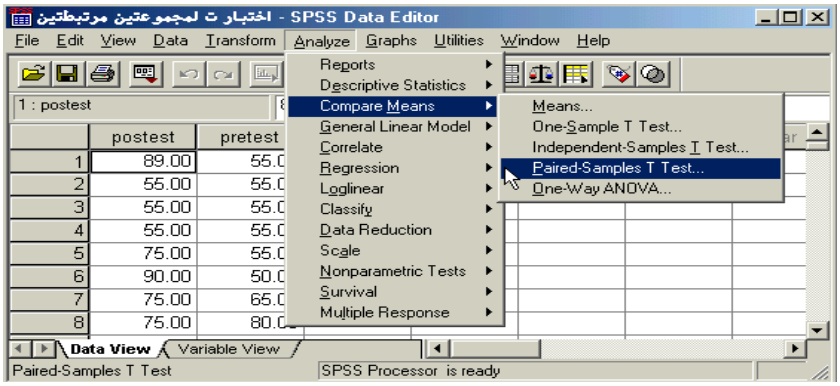
لغرض حساب قيمة (ت) لنفس المثال السابق من خلال استخدام برنامج الـ SPSS نتبع الخطوات التالية:

- ✓ **قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات Data Editor بالطريقة المناسبة كالتالي :**

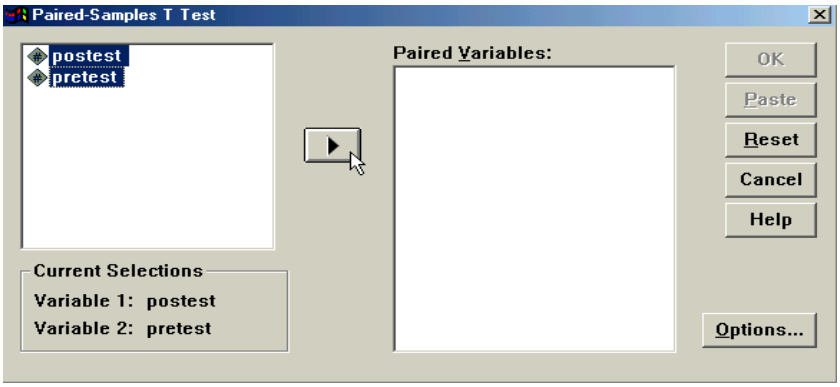


لاحظ أنه تم إدخال البيانات بطريقة مختلفة عن ما تم اتباعه في حالة العينتين المستقلتين، هنا لابد من إدخال بيانات كل متغير في عمود منفصل عن الآخر، وقد تم إعطاء كل متغير اسم مختلف عن الآخر **posttest** و **pretest**

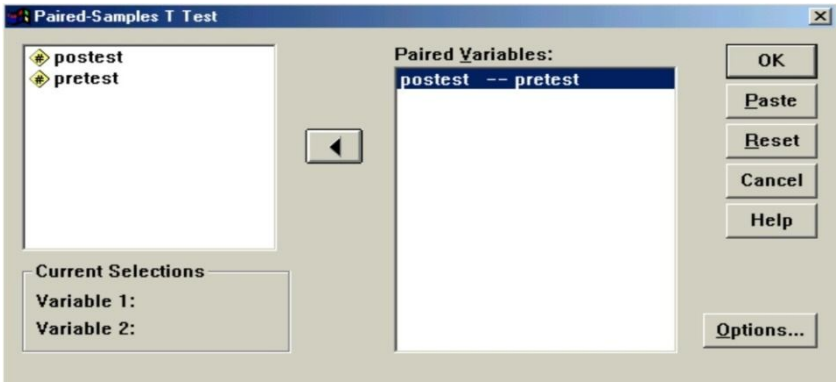
- ✓ من القائمة "تحليل" **Analyze** اختر الأمر "مقارنة المتوسطات" **Means Compare** فتظهر قائمة أوامر فرعية اختر منها "اختبار (ت) للعينات المرتبطة" **Paired-Samples T-Test** كالتالي :



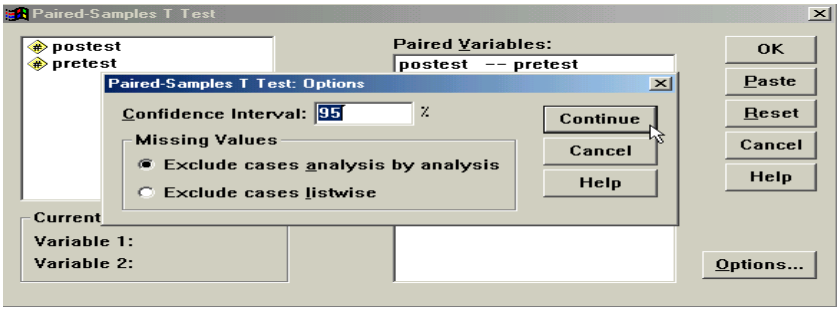
- ✓ بعد اختيار الأمر "اختبار (ت) للعينات المرتبطة" **Paired-Samples T-Test** سوف يظهر لك صندوق الحوار التالي :



- ✓ من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار حدد المتغيرين المرتبطين مع بعضها لتحليلها كأزواج، ونقلها إلى المستطيل الخاص بـ "المتغيرات الزوجية" **Paired Variables** (سوف تلاحظ أثناء التحديد ظهور اسم المتغير الأول واسم المتغير الثاني بعد كل عملية تحديد في المربع اسفل قائمة المتغيرات)، ثم بعد ذلك أنقر على السهم الذي يظهر مقابل المستطيل الخاص بـ "متغيرات الاختبار" ، ستلاحظ انتقال المتغير مباشرة في المستطيل "المتغيرات الزوجية" **Paired Variable(s)**، كرر نفس الإجراء مع المتغيرات الزوجية الأخرى والمراد تحليلها .



✓ أنقر على زر "خيارات" **Options** في الجهة السفلية اليمنى من صندوق الحوار السابق وذلك عند الرغبة في تغيير قيمة "فترة الثقة" **Confidence Interval** حيث يظهر لك صندوق الحوار التالي والذي يتيح إمكانية تغيير فترة الثقة المختبرة (بشكل تلقائي سوف تظهر القيمة ٩٥%) ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" **Continue** .



✓ أنقر بعد ذلك على زر "موافق" **OK** سيؤدي ذلك إلى تنفيذ الاختبار، وستلاحظ ظهور النتائج في شاشة المخرجات كالتالي :

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	POSTEST	58.6600	100	8.0000	.8000
	PRETEST	54.2800	100	7.0000	.7001

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	POSTEST & PRETEST	100	.458	.000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	POSTEST - PRETEST	4.3800	7.8570	.7857	2.8210	5.9390	5.575	99	.000

نلاحظ أن برنامج الـ **SPSS** قام مباشرة بحساب الإحصاءات الأساسية للبيانات مثل المتوسط الحسابي للمتغير **Posttest** (٥٨.٦٦٠) والانحراف المعياري لنفس المتغير (٨.٠٠) ، أما المتغير **Pretest** فقد كان المتوسط الحسابي (٥٤.٢٨٠) والانحراف المعياري (٧.٠٠) . بالإضافة إلى ذلك تم حساب معامل ارتباط بيرسون للمتغيرات موضع الدراسة **Paired Sample Correlation** وقد كانت قيمته (٠.٤٥٨) .

ثم بعد ذلك قام البرنامج بحساب قيمة (ت) للمتغيرات موضع الدراسة في الجدول المعنون بـ "اختبار العينات المرتبطة" **Paired Sample Test** ، ومن هذه النتائج نلاحظ أن قيمة (ت) المحسوبة **t-test** = ٥.٥٧٥ ، ودرجات الحرية **df** = ٩٩ ، وقيمة **Sig. (2-tailed)** = ٠.٠٠٠ ، وبما أن قيمة **Sig. (2-tailed)** في الجدول (٠.٠٠٠) أصغر من قيمة **α** = ٠.٠٥ فإننا بالتالي نرفض الفرضية الصفرية ، أي أن أداء الطلاب في الكلية بعد أخذ البرنامج التدريبي كان أفضل من أدائهم قبل أخذ البرنامج التدريبي عند مستوى **α** = ٠.٠٥ .

والأوامر المستخدمة لحساب اختبار (ت) للعينات المرتبطة **T-Test Paired Samples** من خلال برنامج الـ **SPSS**(مع اختلاف المتغيرات موضع الدراسة) هي:

```
T-TEST
PAIRS= postest WITH pretest (PAIRED)
/CRITERIA=CIN(.95)
/MISSING=ANALYSIS.
```

مع أصدق الأمنيات

أخوكم / ثابت

@thabet_18

المحاضرة التاسعة

الاختبارات الإحصائية لأكثر من عينتين

(اختبار تحليل التباين)

ناقشنا في المحاضرة السابقة طرق الاستدلال الإحصائي عن متوسط المجتمع والفرق بين متوسطين، وسنناقش في هذه المحاضرة طرق الاستدلال الإحصائي للفرق بين ثلاث متوسطات أو أكثر وذلك من خلال توزيع **F**.

سمي توزيع **F** بهذا الاسم تخليدا للعالم رونالد فيشر **R.A. Fisher** الذي يعتبر أول من اشتق هذا التوزيع ووصفه وذلك في العشرينات والثلاثينات من القرن العشرين لذلك تعرف أحيانا بتحليل فيشر للتباين.

ويستخدم هذا التوزيع أساسا لاختبار تساوي تباييني مجتمعين، ومن المثير للإنتباه ملاحظة أن اختبار تساوي التباينين يستخدم لاختبار تساوي ثلاث متوسطات أو أكثر. وتسمى طريقة الاستدلال الاحصائي عن تساوي ثلاث متوسطات أو أكثر بتحليل التباين **Analysis of Variance** .

تحليل التباين هو عملية يقصد بها تقسيم مجموع مربعات الانحرافات عن المتوسط الحسابي إلي مكوناته إرجاع كل من هذه المكونات إلي مسبباتها . وطريقة تحليل التباين تفيد في مقارنة عدد من المعاملات يزيد عن أثنين كما تمتاز طريقة تحليل التباين بأنه يمكن فيها استعمال كل البيانات المأخوذة من التجربة في حساب قيمة واحدة للانحراف القياسي يمكن بها مقارنة المجموعات أو المعاملات التجريبية.

فهي مجموعة من النماذج الإحصائية (**statistical model**) مع إجرءات مرافقة لهذه النماذج تمكن من مقارنة المتوسطات لمجتمعات إحصائية مختلفة عن طريق تقسيم التباين **variance** الكلي الملاحظ بينهم إلى أجزاء مختلفة.

تتلخص طريقة تحليل التباين في:

١- حساب المجموع الكلي لمربعات انحرافات كل المفردات في التجربة عن المتوسط العام .

٢- تقسيم هذا المجموع الكلي لمربعات الانحرافات **Total Sum Squares** إلي مكوناته طبقا للمصادر المسببة لها والذي يختلف عددها طبقا للتصميم المستعمل في التجربة.

٣- تقسم درجات الحرية الكلية طبقا للمصادر السابقة أيضا.

٤- تدون النتائج في جدول يسمي جدول تحليل التباين **ANOVA** ترتب فيه مصادر الاختلافات حسب التصميم الإحصائي المستعمل ويسهل هذا الجدول عمل اختبار معنوية المعاملات.

سبب الحاجة لاختبار تحليل التباين

فالاختبار الآنف الذكر (**t**) يهتم باختبار الفرضيات حول اختلاف وسطي مجتمعين إحصائيين، ولكن عندما نرغب في اختبار الفروق بين أكثر من مجتمعين (ثلاثة مثلاً) فهل من الممكن أن نستخدم اختبار (**Z** أو **t**) ؟ .

إنه من الممكن استخدام هذه الاختبارات وذلك بين كل زوجين من المتوسطات مع مراعاة خطأ "بانفروني" **Bonferroni**، أي إنه يمكن أن نختبر الفرق بين كل متوسطين معاً ، ومن ثم نقسم قيمة على عدد المقارنات ، أي انه إذا كان لدينا ثلاث متوسطات ونريد أن نقارن بينها باستخدام اختبار (**t**) فإن المقارنات ستكون كالتالي :

م١ ، م٢ ثم م١ ، م٣ ثم م٢ ، ، م٣

وإذا كان مستوى الدلالة الدلالة = ٠.٠٥ فإننا نقوم بقسمته على ٣ ومن ثم نختبر قيمة t المحسوبة على مستوى الدلالة الجديد وذلك لتحاشي الخطأ من النوع الأول (رفض الفرضية الصفرية عندما تكون صحيحة)

إلا أننا عندما نُقدْم على ذلك (أي نستخدم اختبار Z أو t بدلا من اختبار F) فإننا نكون قد ابتعدنا عن طريق الصواب لأسباب منها:

١. كلما ازداد عدد المتوسطاتكلما ازداد عدد الأزواج التي يجب أن نجري عليها الاختبار ، وقد يصبح العدد كبيراً جداً ، مثال ذلك أننا رغم وجود ثلاثة أزواج من المتوسطات لثلاث عينات، إلا أن هذا العدد من الأزواج يرتفع إلى (١٠) أزواج من المتوسطات إذا أصبح عدد العينات خمس ، أما إذا أصبح عدد العينات (٨) فإننا نجد أن لدينا (٢٨)زوجاً من المتوسطات ، وبذلك تتزايد كمية الجهد المبذول في إجراء العملياتالحسابية المختلفة تزايداً سريعاً كلما ازداد عدد العينات ، وبشكل عام إذا كان عدد الأوساط ( ) فإن عدد المقارنات المختلفة بين أزواج هذه الأوساط = $\frac{x(x-1)}{2}$

٢. عند المقارنة بين كل زوج من الأوساط فإننا نستخدم فقط المعلومات عن المجموعتين المقارنتين ونهمل المعلومات المتوفرة في باقي المجموعات والتي تجعل المقارنة أقوى فيما لو استعملت .

٣. إن هذا الاستخدام المتعدد لاختبار (t)مثلاُ يزيد من الخطر في ارتكاب الخطأ من النوع الأول (رفض الفرضية الصفرية عندما تكون صحيحة) . فإذا كان عدد المقارنات التي تستخدم فيها اختبار (t) = ك ، وكان مستوى الدلالة المستخدم في المقارنة = α فإن احتمال ارتكاب خطأ واحد أو أكثر من النوع الأول في هذه المقارنات يكون بالعلاقة = [1 – (1 – α)^{ك}] حيث "ك" تعني عدد المقارنات .

فإذا كان عدد المقارنات = ٣ ، وكان مستوى الدلالة α = ٠.٠٥ فإن احتمال ارتكاب خطأ واحد من النوع الأول = [١ – (١ – ٠.٠٥)^{٣}] = ٠.١٤ . أما إذا كان عدد الأوساط = ٢٠ وسطا، فإن عدد المقارنات = ١١٠ مقارنة ، واحتمال ارتكاب الخطأ من النوع الأول = ١ تقريباً . وهذه الزيادة في احتمال ارتكاب الخطأ من النوع الأول تعني ازدياد الاحتمال في ادعاء وجود فروق ذات دلالة إحصائية في حالة مقارنة واحدة على الأقل بين هذه الأوساط، في حين أن هذه الفروق ليست في الحقيقة ذات دلالة .

٤. إن الباحث كثيرا ما يرغب في اختبار مشكلة أوسع من مجرد معرفة ما إذا كانت أزواج المتوسطات مختلفة عن بعضها البعض .

وفي ضوء هذه الأسباب ، فإن التوجه يجب أن يكون نحو استخدام اختبار أفضل من اختباري (Z أو t) وأقوى منهما ، وذلك لإجراء المقارنة بين ثلاث أوساط فأكثر، وتحليل التباين هو أكثر هذه الاختبارات شهرة واستخداماً لمثل هذا الغرض، وهذا الاختبار يعتبر أداة فعالة وقوية في يد الباحث يفيده في بناء تصميم تجاربه بكفاءة، كما ويساعده في فحص أثر التفاعل بين متغيرات الدراسة .

تحليل التباين الأحادي

هو طريقة لاختبار معنوية الفرق بين المتوسطات لعدة عينات بمقارنة واحدة، ويعرف أيضاً بطريقة تؤدي لتقسيم الاختلافات الكلية لمجموعة من المشاهدات التجريبية لعدة أجزاء للتعرف على مصدر الاختلاف بينها ولذا فالهدف هنا فحص تباين المجتمع لمعرفة مدى تساوى متوسطات المجتمع ولكن لا بد من تحقيق ثلاثة أمور قبل استخدامه وهي:

١) العينات عشوائية ومستقلة.

٢) مجتمعات هذهالعينات كلاً لها توزيع طبيعي.

٣) تساوي تباين المجتمعات التي أخذت منها العينات العشوائية المستقلة.

ولتوضيح ما سبق بمقارنة متوسطات ثلاث مجتمعات باستخدام ثلاث عينات (تحقق فيها الشروط الثلاثة السابقة) موضحة بالجدول الآتي:

العينة الأولى	العينة الثانية	العينة الثالثة
40	27	33
41	28	32
40.5	26.5	33.5
38.5	26.5	31.5
$\bar{X}_1 = 40$ $S_1 = 1.08$	$\bar{X}_2 = 27$ $S_2 = 0.71$	$\bar{X}_3 = 32.5$ $S_3 = 0.91$

السؤال: هل في البيانات ما يكفي لوجود فرق بين المتوسطات؟

الجواب: نعم (بمجرد النظر) فالتشتت (التباين) ظاهر ٤٠، ٢٧، ٣٢.٥ (المتوسطات) بمقارنته بالتشتت بين العينات (وحداتها ٤٠ ، ٤١ ، ٤٠.٥، ٣٨.٥) فيبدو معدوماً.

فإذا أخذنا البيانات الآتية:

العينة الأولى	العينة الثانية	العينة الثالثة
40	50	10
15	20	60
65	11	27.5
$\bar{X}_1 = 40$ $S_1 = 25$	$\bar{X}_2 = 27$ $S_2 = 20.4$	$\bar{X}_3 = 32.5$ $S_3 = 25.4$

فالبيانات هنا لها نفس المتوسطات في البيانات السابقة ولكن التشتت (داخل العينات) كبيراً بما هو عليه في المتوسطات.

فالدليل على وجود الفرق بين متوسطات الجدول الأول واضح ولا يظهر ذلك بوضوح في بيانات الجدول الثاني بالرغم من تساوي المتوسطات في الحالتين

ولذا يتبين لنا القصد من تحليل التباين والذي يعني الفرق بين المتوسطات والذي يقاس بالتشتت داخل البيانات.

الافتراضات الأساسية لاختبار تحليل التباين

يستند اختبار تحليل التباين إلى توفر عدد من الافتراضات، ومن هذه الافتراضات ما يلي :

١– مستوى القياس :

يشترط لاستخدام هذا الاختبار أن تكون البيانات فترية (فئوية) أو نسبية

٢– حجم العينة :

يقتضي هذا الافتراض أن يكون حجم العينة كبيراً .

٣– التوزيع الطبيعي للمجتمع الإحصائي :

يقتضي هذا الافتراض أن تكون المشاهدات في كل مجتمع من المجتمعات موزعة بشكل طبيعي، ولكن يرى الإحصائيون أن اختبار (F) لا يتأثر كثيراً بعدمتوفر هذا الشرط وذلك عندما يكون حجم الخلية كبيراً والتوزيع ليس طبيعياً

٤– تجانس التباين :

أي أن يكون للمجتمعات في مستويات المعالجة المختلفة نفس التباين ( ) بالرغم من أن لها بالطبع أوساطاً مختلفة، وإن الإخلال بهذا الافتراض في حالة تساوي أحجام الخلايا لا يؤثر على النتائج بشكل كبير، أما عند عدم تساوي أحجام الخلايا وانتماء الخلايا ذات الحجم الصغير إلى مجتمعات ذات تباين كبير، فإن الإخلال بهذا الافتراض يؤدي إلى ارتفاع احتمال الخطأ من النوع الأول ، أما عندما تنتمي الخلايا ذات الحجم الكبير للمجتمعات ذات التباين الكبير فإنه يقلل من احتمال ارتكاب الخطأ من النوع الأول .

مثال : إذا كان لدينا ثلاث منتجات لإحدى الأسر المنتجة، وتم تقييمها من قبل مجموعة من المستهلكين وحصلنا على النتائج التالية :

المنتج (١) X_1	المنتج (٢) X_2	المنتج (٣) X_3
٧	٤	٢
١٠	٦	٢
١٠	٧	٣
١١	٩	٧
١٢	٩	٦
٥٠	٣٥	٢٠

المطلوب: هل هناك فروق ذات دلالة بين المنتجات الثلاثة ؟

الحل :

لكون لدينا ثلاث متغيرات فترية، ولرغبة الأسر المنتجة معرفة الفروق بين هذه المتغيرات موضع الدراسة، فإن أنسب أسلوب إحصائي هنا هو تحليل التباين الأحادي One Way ANOVA ، ولغرض حساب تحليل التباين الأحادي، علينا اتباع الخطوات التالية:

- ✓ **نجمع قيم كل متغير للحصول على $\sum X$ لكل متغير .**
- ✓ **نربع كل درجة في كل متغير للحصول على X^2 لكل متغير .**
- ✓ **نجمع قيم مربع كل درجة للحصول على $\sum X^2$ لكل متغير .**
- ✓ **نربع مجموع كل متغير للحصول على $\left(\sum X\right)^2$ لكل متغير .**
- ✓ **نحسب متوسط كل متغير من خلال العلاقة : $\overline{X} = \frac{\sum x}{n}$**

✓ نحسب مجموع المربعات الكلي **Total Sum of Squares** وذلك من خلال العلاقة التالية :

$$Total..SS = \sum X^2 - \frac{\left(\sum X\right)^2}{n}$$

حيث n تعني مجموع أعدد الأفراد في جميع المجموعات .

أو يمكن حساب مجموع المربعات الكلي من خلال العلاقة التالية : ***Total SS = Between SS + Within SS***

✓ نحسب مجموع المربعات بين المجموعات **Between Sum of Squares** وذلك من خلال العلاقة التالية :

$$Between..SS = \sum \frac{\left(\sum X_g\right)^2}{n_g} - \frac{\left(\sum X\right)^2}{n}$$

حيث :

n_g تعني عدد الأفراد في كل مجموعة .

n تعني مجموع أعدد الأفراد في جميع المجموعات .

$\sum \frac{\left(\sum X_g\right)^2}{n_g}$ تعني مربع مجموع قيم كل مجموعة مقسوما على عدد أفراد تلك المجموعة .

$\frac{\left(\sum X\right)^2}{n}$ تعني مربع مجموع قيم كل المجموعات مقسوما على مجموع أعداد الأفراد في جميع المجموعات .

✓ نحسب مجموع المربعات داخل المجموعات **Within Sum of Squares** وذلك من خلال العلاقة التالية :

$$Within..SS = \sum \left[\sum X_g^2 - \frac{\left(\sum X_g\right)^2}{n_g} \right]$$

أو يمكن حساب مجموع المربعات داخل المجموعات من خلال العلاقة التالية : ***Within SS = Total SS – Between SS***

✓ نحسب درجات الحرية بين المجموعات **Between groups degrees of freedom** من خلال العلاقة التالية :

$K-1$ حيث K تعني عدد المجموعات .

ودرجات الحرية داخل المجموعات **Within groups degrees of freedom** من خلال العلاقة التالية :

$n-K$ حيث n تعني عدد الأفراد أو الاستجابات في المجموعات موضع الدراسة ، وK تعني عدد المجموعات .

ودرجات الحرية الكلية **Total degrees of freedom** من خلال العلاقة التالية :

$n-1$ حيث n تعني عدد الأفراد أو الاستجابات في المجموعات موضع الدراسة .

✓ نحسب التباين بين المجموعات أو ما يسمى متوسط المربعات بين المجموعات **Between mean square** وذلك من خلال العلاقة التالية:

$$Beween \quad ..groups \quad ..mean \quad ..square \quad = \frac{Between \quad ..SS}{K \; - 1}$$

✓ نحسب التباين داخل المجموعات أو ما يسمى متوسط المربعات داخل المجموعات **Within mean square** وذلك من خلال العلاقة التالية :

$$Within \;..groups \;..mean \;..square = \frac{Within \;..SS}{(n-K)}$$

✓ نحسب قيمة F من خلال العلاقة التالية :

$$F = \frac{Between \;..groups \;..mean \;..square}{Within \;..groups \;..mean \;..square}$$

✓ نقارن بعد ذلك قيمة F المحسوبة بقيمة F المجدولة لاتخاذ القرار المناسب اتجاه الفرضية موضع الدراسة .

نقوم الآن بتطبيق جميع الخطوات السابقة لحساب تحليل التباين الأحادي على بيانات المثال السابق وذلك حتى يسهل علينا استخلاص بعض القيم المطلوبة لحساب هذا النوع من الاختبارات الإحصائية وتطبيق المعادلات السابقة .

المنتج (١) X_1		المنتج (٢) X_2		المنتج (٣) X_3	
X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
٧	٤٩	٤	١٦	٢	٤
١٠	١٠٠	٦	٣٦	٢	٤
١٠	١٠٠	٧	٤٩	٣	٩
١١	١٢١	٩	٨١	٧	٤٩
١٢	١٤٤	٩	٨١	٦	٣٦
٥٠	**٥١٤**	**٣٥**	**٢٦٣**	**٢٠**	**١٠٢**

❑ وضع فرض العدم والفرض البديل.

صياغة الفرضية الصفرية كالتالي: $\mu_1 = \mu_2 = \mu_3$:**H0**

في حين تفترض الفرضية البديلة التالي : متوسطان على الأقل غير متساويين : **HA**

❑ تحديد مستوى الدلالة (α): وتحدد مستويات المعنوية سلفاً وهي عادة **0.05** أو **0.01**

❑ حساب إحصائية الاختبار (**F**) وذلك من خلال اتباع الخطوات التالية:

✓ المتوسط الحسابي لـ X_1 = $\overline{X} = \frac{50}{5} = 10$

✓ المتوسط الحسابي لـ X_2 = $\overline{X} = \frac{35}{5} = 7$

✓ المتوسط الحسابي لـ X_3 = $\overline{X} = \frac{20}{5} = 4$

✓ مجموع المربعات الكلي = **Total Sum of Squares**

$$Total \quad ..SS = \sum X^2 - \frac{(\sum X)^2}{(n_g)(k)} = 879 - \frac{(105)^2}{15} = 144$$

حيث **ng** تعني عدد أفراد المجموعة المحددة و **K** تعني عدد المجموعات موضع الدراسة

✓ مجموع المربعات بين المجموعات = **Between Sum of Squares**

$$Between \; ..SS = \sum \frac{(\sum X_g)^2}{n_g} - \frac{(\sum X)^2}{(n_g)(k)} = \frac{(50)^2}{5} + \frac{(35)^2}{5} + \frac{(20)^2}{5} - \frac{(105)^2}{15} = 90$$

✓ مجموع المربعات داخل المجموعات = **Within Sum of Squares**

$$Within \; ..SS = \sum \left[\sum X_g^2 - \frac{(\sum X_g)^2}{n_g} \right]$$

$$\sum x_1^2 = 514 - \frac{(50)^2}{5} = 14$$

$$\sum x_2^2 = 263 - \frac{(35)^2}{5} = 18$$

$$\sum x_3^2 = 102 - \frac{(20)^2}{5} = 22$$

نقوم بعد ذلك بجمع نواتج هذه المعادلات لنحصل على مجموع المربعات داخل المجموعات كالتالي :

Within sum of squares **= 14+18+22=54**

✓ نحسب درجات الحرية:

درجات الحرية بين المجموعات **Between groups**

degrees of freedom

$$(K-1) = 3-1 = 2$$

درجات الحرية داخل المجموعات **Within groups**

degrees of freedom

$$(n-K) = 15-3 = 12$$

درجات الحرية الكلية **Total degrees of freedom**

$$(n-1) = 15-1 = 14$$

✓ التباين بين المجموعات أو ما يسمى متوسط المربعات بين المجموعات = **Between mean square**

$$Beween\ \ ..groups\ \ ..mean\ ..square\ \ = \frac{Between\ \ ..SS}{K-1}$$

$$Beween\ \ ..groups\ \ ..mean\ ..square\ \ = \frac{90}{2} = 45$$

✓ التباين داخل المجموعات أو ما يسمى متوسط المربعات داخل المجموعات = **Within mean square**

$$Within\ \ ..groups\ \ ..mean\ ..square\ \ = \frac{Within\ \ ..SS}{(n-K)}$$

$$Within\ \ ..groups\ \ ..mean\ ..square\ \ = \frac{54}{12} = 4.5$$

✓ قيمة F =

$$F = \frac{Between\ ..groups\ ..mean\ ..square}{Within\ ..groups\ ..mean\ ..square} = \frac{45}{4.5} = 10$$

✓ نقوم بعد ذلك بتفريغ ما تم الحصول عليه من معلومات في جدول تحليل التباين كالتالي :

مصدر التباين	مجموع المربعات SS	درجات الحرية df	متوسط المربعات Means	قيمة F
بين المجموعات Between groups	٩٠	٢	٤٥	١٠
داخل المجموعات Within groups	٥٤	١٢	٤.٥	
الكلي (المجموع) Total	١٤٤	١٤		

وبالرجوع إلى جدول توزيع F نجد أن القيمة الحرجة لـ F بدرجات حرية للبسط تساوي ٢ ودرجات حرية للمقام تساوي ١٢ وباستخدام مستوى= ٠.٠٥ نجد أن القيمة الحرجة (المجدولة) تساوي ٣.٨٨ ،

وحيث أن القيمة المحسوبة لـ F = ١٠ وهي بالتالي أكبر من القيمة الحرجة المجدولة، نستنتج أن الفرضية الصفرية تكون مرفوضة، أي يوجد اختلاف بين متوسطي مجتمعين على الأقل من المجتمعات التي قيّمة من المستهلكين ولمعرفة بين أي من المنتجات تكون الفروق ينبغي علينا اللجوء إلى أسلوب المقارنات المتعددة **Multiple Comparisons** .

Table F for Alfa = 0.10

ν	df1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
df2=1	39.86266	49.50000	53.59326	55.83296	57.24008	58.20442	58.90595	59.43898	59.85759	60.19498	60.70521	61.22034	61.74029	62.00205	62.26497	62.52905	62.79428	63.06066	63.32812
2	8.52632	9.00000	9.16179	9.24342	9.29263	9.32553	9.34908	9.36677	9.38054	9.39157	9.40813	9.42471	9.44131	9.44962	9.45793	9.46626	9.47456	9.48289	9.49122
3	5.53832	5.46238	5.39077	5.34264	5.30916	5.28472	5.26619	5.25167	5.24000	5.23061	5.21562	5.20031	5.18468	5.17676	5.16811	5.15972	5.15119	5.14251	5.13370
4	4.54477	4.32456	4.19086	4.10725	4.05058	4.00975	3.97897	3.95494	3.93567	3.91988	3.89553	3.87036	3.84434	3.83099	3.81742	3.80361	3.78957	3.77527	3.76073
5	4.06042	3.77972	3.61948	3.52020	3.45298	3.40451	3.36790	3.33928	3.31628	3.29760	3.26824	3.23801	3.20665	3.19052	3.17408	3.15732	3.14023	3.12279	3.10500
6	3.77595	3.46330	3.28876	3.18076	3.10751	3.05455	3.01446	2.98304	2.95774	2.93693	2.90472	2.87122	2.83624	2.81836	2.79996	2.78117	2.76195	2.74229	2.72216
7	3.58943	3.25744	3.07407	2.96053	2.88334	2.82739	2.78493	2.75158	2.72468	2.70251	2.66811	2.63223	2.59473	2.57533	2.55546	2.53510	2.51422	2.49279	2.47079
8	3.45792	3.11312	2.92380	2.80643	2.72645	2.66833	2.62413	2.58925	2.56124	2.53804	2.50196	2.46422	2.42466	2.40410	2.38302	2.36136	2.33910	2.31618	2.29257
9	3.36030	3.00645	2.81286	2.69268	2.61061	2.55086	2.50531	2.46941	2.44034	2.41632	2.37888	2.33962	2.29832	2.27683	2.25472	2.23196	2.20849	2.18427	2.15923
10	3.28502	2.92447	2.72767	2.60534	2.52164	2.46058	2.41397	2.37715	2.34721	2.32260	2.28405	2.24351	2.20076	2.17843	2.15543	2.13169	2.10716	2.08176	2.05542
11	3.22520	2.85951	2.66022	2.53619	2.45118	2.38907	2.34157	2.30400	2.27350	2.24822	2.20872	2.16709	2.12305	2.10001	2.07621	2.05161	2.02612	1.99965	1.97211
12	3.17655	2.80680	2.60552	2.48010	2.39402	2.33102	2.28278	2.24487	2.21352	2.18776	2.14746	2.10485	2.05968	2.03599	2.01149	1.98610	1.95973	1.93228	1.90361
13	3.13621	2.76317	2.56027	2.43271	2.34672	2.28298	2.23410	2.19525	2.16282	2.13563	2.09459	2.05116	2.00498	1.98072	1.95557	1.92947	1.90229	1.87391	1.84420
14	3.10221	2.72647	2.52222	2.39469	2.30896	2.24526	2.19513	2.15590	2.12195	2.09360	2.05171	2.00752	1.96045	1.93566	1.91192	1.88516	1.85722	1.82800	1.79728
15	3.07319	2.69517	2.48979	2.36143	2.27502	2.20908	2.15818	2.11853	2.08521	2.05732	2.01707	1.97222	1.92421	1.89906	1.87277	1.84539	1.81676	1.78672	1.75505
16	3.04811	2.66817	2.46181	2.33276	2.24576	2.17922	2.12800	2.08798	2.05423	2.02615	1.98529	1.93991	1.89127	1.86556	1.83879	1.81086	1.78156	1.75075	1.71817
17	3.02623	2.64464	2.43763	2.30775	2.21925	2.15229	2.10169	2.06126	2.02829	2.00096	1.95772	1.91169	1.86226	1.83626	1.80901	1.78052	1.75062	1.71909	1.68566
18	3.00698	2.62395	2.41601	2.28577	2.19582	2.12958	2.07856	2.03789	2.00467	1.97698	1.93326	1.88681	1.83685	1.81035	1.78269	1.75371	1.72322	1.69099	1.65671
19	2.98990	2.60561	2.39702	2.26620	2.17596	2.10926	2.05802	2.01710	1.98366	1.95572	1.91170	1.86471	1.81416	1.78721	1.75926	1.72979	1.69876	1.66587	1.63077
20	2.97465	2.58925	2.38009	2.24893	2.15823	2.09132	2.03970	1.99853	1.96485	1.93674	1.89236	1.84494	1.79384	1.76667	1.73822	1.70833	1.67678	1.64326	1.60738
21	2.96096	2.57457	2.36489	2.23334	2.14231	2.07512	2.02325	1.98186	1.94797	1.91967	1.87497	1.82715	1.77555	1.74807	1.71927	1.68896	1.65691	1.62278	1.58615
22	2.94858	2.56131	2.35117	2.21927	2.12796	2.06050	2.00840	1.96680	1.93273	1.90425	1.85925	1.81106	1.75899	1.73122	1.70208	1.67128	1.63865	1.60415	1.56678
23	2.93726	2.54929	2.33873	2.20651	2.11491	2.04723	1.99491	1.95312	1.91888	1.89025	1.84497	1.79663	1.74391	1.71588	1.68643	1.65535	1.62237	1.58711	1.54903
24	2.92712	2.53833	2.32729	2.19488	2.10302	2.03513	1.98262	1.94066	1.90625	1.87768	1.83196	1.78308	1.73015	1.70185	1.67210	1.64067	1.60726	1.57166	1.53270
25	2.91774	2.52821	2.31702	2.18426	2.09216	2.02406	1.97128	1.92925	1.89469	1.86578	1.82000	1.77082	1.71752	1.68898	1.65895	1.62718	1.59325	1.55703	1.51760
26	2.90913	2.51910	2.30749	2.17447	2.08218	2.01389	1.96104	1.91876	1.88407	1.85503	1.80901	1.75957	1.70589	1.67712	1.64682	1.61472	1.58050	1.54388	1.50360
27	2.90119	2.51061	2.29871	2.16566	2.07298	2.00452	1.95151	1.90909	1.87427	1.84511	1.79889	1.74917	1.69516	1.66616	1.63560	1.60320	1.56859	1.53129	1.49057
28	2.89385	2.50276	2.29060	2.15776	2.06447	1.99585	1.94270	1.90014	1.86520	1.83593	1.78951	1.73956	1.68519	1.65600	1.62519	1.59250	1.55762	1.51976	1.47841
29	2.88703	2.49548	2.28307	2.14961	2.05658	1.98781	1.93452	1.89184	1.85679	1.82741	1.78081	1.73060	1.67593	1.64655	1.61551	1.58253	1.54721	1.50899	1.46706
30	2.88069	2.48872	2.27607	2.14223	2.04925	1.98023	1.92692	1.88412	1.84896	1.81949	1.77270	1.72227	1.66721	1.63776	1.60668	1.57322	1.53757	1.49891	1.45636
40	2.83535	2.44037	2.22609	2.09095	1.99682	1.92688	1.87252	1.82886	1.79190	1.76169	1.71456	1.66261	1.60515	1.57411	1.54108	1.50562	1.46716	1.42676	1.37691
60	2.79107	2.39225	2.17761	2.04099	1.96571	1.87472	1.81939	1.77482	1.73802	1.70701	1.65743	1.60227	1.54249	1.51072	1.47556	1.43726	1.39520	1.34757	1.29166
120	2.74781	2.34734	2.12999	1.99230	1.89587	1.82381	1.76748	1.72196	1.68425	1.65228	1.60120	1.54500	1.48207	1.44722	1.40938	1.36760	1.32034	1.26657	1.19256
∞	2.70554	2.30259	2.08380	1.94586	1.84727	1.77411	1.71672	1.67010	1.63152	1.59872	1.54578	1.48714	1.42060	1.38218	1.34187	1.29913	1.25995	1.16860	1.00000

Table F for Alfa = 0.05

ν	df1=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
df2=1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817	243.9060	245.9499	248.0131	249.0518	250.0951	251.1432	252.1957	253.2529	254.3144
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959	19.4125	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.5720	8.5494	8.5266
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.7170	5.6877	5.6581	5.6281
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.3650
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600	3.9999	3.9381	3.8742	3.8415	3.8082	3.7742	3.7398	3.7047	3.6689
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.9130	2.8450	2.7740	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536	2.7876	2.7186	2.6464	2.6090	2.5705	2.5309	2.4901	2.4480	2.4045
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.3410	2.2962
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.4630	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.1040	2.0584	2.0107	1.9604
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.3080	2.2341	2.1555	2.1141	2.0712	2.0266	1.9795	1.9302	1.8780
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210	2.2504	2.1757	2.0960	2.0540	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0282	1.9842	1.9380	1.8894	1.8380	1.7831
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.0050	1.9605	1.9139	1.8648	1.8128	1.7570
24	4.2597	3.4028	3.0088	2.7762	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.9390	1.8920	1.8424	1.7896	1.7330
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.7110
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.9010	1.8533	1.8027	1.7488	1.6906
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926	1.9174	1.8364	1.7480	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.4290	1.3519	1.2539
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.3940	1.3180	1.2214	1.0000

F table for Alfa = 0.025

/	df_2=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
df_2=1	647.7890	799.5000	864.1630	899.5833	921.8479	937.1111	948.2169	956.6562	963.2846	968.6274	976.7079	984.8668	993.1028	997.2492	1001.414	1005.598	1009.800	1014.020	1018.258
2	38.5063	39.0000	39.1655	39.2484	39.2982	39.3315	39.3552	39.3730	39.3869	39.3980	39.4146	39.4313	39.4479	39.4562	39.465	39.473	39.481	39.490	39.498
3	17.4434	16.0441	15.4392	15.1010	14.8848	14.7347	14.6244	14.5399	14.4731	14.4189	14.3366	14.2527	14.1674	14.1241	14.081	14.037	13.992	13.947	13.902
4	12.2179	10.6491	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047	8.8439	8.7512	8.6565	8.5599	8.5109	8.461	8.411	8.360	8.309	8.257
5	10.0070	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811	6.6192	6.5245	6.4277	6.3286	6.2780	6.227	6.175	6.123	6.069	6.015
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234	5.4613	5.3662	5.2687	5.1684	5.1172	5.065	5.012	4.959	4.904	4.849
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232	4.7611	4.6658	4.5678	4.4667	4.4150	4.362	4.309	4.254	4.199	4.142
8	7.5709	6.0595	5.4160	5.0526	4.8172	4.6517	4.5286	4.4333	4.3572	4.2951	4.1997	4.1012	3.9995	3.9472	3.896	3.840	3.784	3.728	3.670
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260	3.9639	3.8682	3.7694	3.6669	3.6142	3.560	3.505	3.449	3.392	3.333
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790	3.7168	3.6209	3.5217	3.4185	3.3654	3.311	3.255	3.198	3.140	3.080
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879	3.5257	3.4296	3.3299	3.2261	3.1725	3.118	3.061	3.004	2.944	2.883
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358	3.3736	3.2772	3.1772	3.0728	3.0187	2.963	2.906	2.848	2.787	2.725
13	6.4143	4.9653	4.3472	3.9959	3.7667	3.6043	3.4827	3.3880	3.3120	3.2497	3.1532	3.0527	2.9477	2.8932	2.837	2.780	2.720	2.659	2.595
14	6.2979	4.8567	4.2417	3.8919	3.6634	3.5014	3.3799	3.2853	3.2093	3.1469	3.0502	2.9493	2.8437	2.7888	2.732	2.674	2.614	2.552	2.487
15	6.1995	4.7650	4.1528	3.8043	3.5764	3.4147	3.2934	3.1987	3.1227	3.0602	2.9633	2.8621	2.7559	2.7006	2.644	2.585	2.524	2.461	2.395
16	6.1151	4.6867	4.0768	3.7294	3.5021	3.3406	3.2194	3.1248	3.0488	2.9862	2.8890	2.7875	2.6808	2.6252	2.568	2.509	2.447	2.383	2.316
17	6.0420	4.6189	4.0112	3.6648	3.4379	3.2767	3.1556	3.0610	2.9849	2.9222	2.8249	2.7230	2.6158	2.5598	2.502	2.442	2.380	2.315	2.247
18	5.9781	4.5597	3.9539	3.6083	3.3820	3.2209	3.0999	3.0053	2.9291	2.8664	2.7689	2.6667	2.5590	2.5027	2.445	2.384	2.321	2.256	2.187
19	5.9216	4.5075	3.9034	3.5587	3.3327	3.1718	3.0509	2.9563	2.8801	2.8172	2.7196	2.6171	2.5089	2.4523	2.394	2.333	2.270	2.203	2.133
20	5.8715	4.4613	3.8587	3.5147	3.2891	3.1283	3.0074	2.9128	2.8365	2.7737	2.6758	2.5731	2.4645	2.4076	2.349	2.287	2.223	2.156	2.085
21	5.8266	4.4199	3.8188	3.4754	3.2501	3.0895	2.9686	2.8740	2.7977	2.7348	2.6368	2.5338	2.4247	2.3675	2.308	2.246	2.182	2.114	2.042
22	5.7863	4.3828	3.7829	3.4401	3.2151	3.0546	2.9338	2.8392	2.7628	2.6998	2.6017	2.4984	2.3890	2.3315	2.272	2.210	2.145	2.076	2.003
23	5.7498	4.3492	3.7505	3.4082	3.1835	3.0232	2.9022	2.8077	2.7312	2.6682	2.5699	2.4665	2.3567	2.2989	2.239	2.176	2.111	2.041	1.968
24	5.7166	4.3187	3.7211	3.3794	3.1548	2.9946	2.8738	2.7791	2.7027	2.6396	2.5411	2.4374	2.3273	2.2693	2.209	2.146	2.080	2.010	1.935
25	5.6864	4.2909	3.6943	3.3530	3.1287	2.9685	2.8478	2.7531	2.6766	2.6135	2.5149	2.4110	2.3005	2.2422	2.182	2.118	2.052	1.981	1.906
26	5.6586	4.2655	3.6697	3.3289	3.1048	2.9447	2.8240	2.7293	2.6528	2.5896	2.4908	2.3867	2.2759	2.2176	2.157	2.093	2.026	1.954	1.878
27	5.6331	4.2421	3.6472	3.3067	3.0828	2.9228	2.8021	2.7074	2.6309	2.5676	2.4688	2.3644	2.2533	2.1946	2.132	2.069	2.002	1.930	1.852
28	5.6096	4.2205	3.6264	3.2863	3.0626	2.9027	2.7820	2.6872	2.6106	2.5473	2.4484	2.3438	2.2324	2.1735	2.112	2.048	1.980	1.907	1.829
29	5.5878	4.2006	3.6072	3.2674	3.0438	2.8840	2.7633	2.6686	2.5919	2.5286	2.4295	2.3248	2.2131	2.1540	2.092	2.028	1.959	1.886	1.807
30	5.5675	4.1821	3.5894	3.2499	3.0265	2.8667	2.7460	2.6512	2.5746	2.5112	2.4120	2.3072	2.1952	2.1359	2.074	2.009	1.940	1.866	1.787
40	5.4239	4.0510	3.4633	3.1261	2.9037	2.7444	2.6238	2.5289	2.4519	2.3882	2.2882	2.1819	2.0677	2.0069	1.943	1.875	1.803	1.726	1.637
60	5.2856	3.9253	3.3425	3.0077	2.7863	2.6274	2.5068	2.4117	2.3344	2.2702	2.1692	2.0613	1.9445	1.8817	1.815	1.744	1.667	1.581	1.482
120	5.1523	3.8046	3.2269	2.8943	2.6740	2.5154	2.3948	2.2994	2.2217	2.1570	2.0548	1.9450	1.8249	1.7597	1.690	1.614	1.530	1.433	1.310
∞	5.0239	3.6889	3.1161	2.7858	2.5665	2.4082	2.2875	2.1918	2.1136	2.0483	1.9447	1.8326	1.7085	1.6402	1.566	1.484	1.388	1.268	1.000

Table F for Alfa = 0.10

\	df_2=1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
df_2=1	39.86306	49.50000	53.59324	55.83296	57.24008	58.20442	58.90595	59.43898	59.85759	60.19498	60.70521	61.22034	61.74029	62.00205	62.26497	62.52905	62.79428	63.06064	63.32812
2	8.52632	9.00000	9.16179	9.24342	9.29263	9.32553	9.34908	9.36677	9.38054	9.39157	9.40813	9.42471	9.44131	9.44962	9.45793	9.46626	9.47458	9.48289	9.49122
3	5.53832	5.46238	5.39077	5.34264	5.30916	5.28473	5.26619	5.25167	5.24000	5.23041	5.21562	5.20031	5.18448	5.17636	5.16811	5.15972	5.15119	5.14251	5.13370
4	4.54477	4.32456	4.19086	4.10725	4.05058	4.00975	3.97897	3.95494	3.93567	3.91988	3.89553	3.87036	3.84434	3.83099	3.81742	3.80361	3.78957	3.77527	3.76072
5	4.06062	3.77972	3.61948	3.52020	3.45298	3.40451	3.36790	3.33928	3.31628	3.29760	3.26824	3.23801	3.20665	3.19052	3.17408	3.15732	3.14023	3.12279	3.10500
6	3.77595	3.46330	3.28876	3.18076	3.10751	3.05455	3.01466	2.98304	2.95774	2.93693	2.90472	2.87122	2.83634	2.81824	2.79996	2.78117	2.76195	2.74229	2.72216
7	3.58943	3.25744	3.07407	2.96053	2.88334	2.82739	2.78493	2.75158	2.72468	2.70251	2.66811	2.63223	2.59473	2.57533	2.55566	2.53570	2.51542	2.49479	2.47079
8	3.45792	3.11312	2.92380	2.80643	2.72645	2.66833	2.62413	2.58925	2.56126	2.53806	2.50196	2.46422	2.42466	2.40410	2.38302	2.36136	2.33910	2.31618	2.29257
9	3.36030	3.00645	2.81286	2.69268	2.61061	2.55086	2.50531	2.46941	2.44034	2.41632	2.37888	2.33962	2.29832	2.27683	2.25472	2.23196	2.20849	2.18427	2.15923
10	3.28501	2.92447	2.72767	2.60534	2.52164	2.46058	2.41397	2.37715	2.34731	2.32260	2.28405	2.24351	2.20076	2.17843	2.15563	2.13169	2.10716	2.08176	2.05542
11	3.22520	2.85951	2.66023	2.53419	2.45118	2.38907	2.34157	2.30400	2.27350	2.24823	2.20873	2.16709	2.12305	2.10001	2.07621	2.05161	2.02612	1.99965	1.97211
12	3.17655	2.80680	2.60552	2.48010	2.39402	2.33102	2.28278	2.24487	2.21352	2.18776	2.14764	2.10485	2.05968	2.03599	2.01149	1.98610	1.95972	1.93228	1.90361
13	3.13621	2.76317	2.56027	2.43371	2.34672	2.28298	2.23410	2.19525	2.16382	2.13762	2.09659	2.05316	2.00698	1.98272	1.95787	1.93147	1.90429	1.87591	1.84620
14	3.10221	2.72647	2.52222	2.39469	2.30694	2.24256	2.19313	2.15390	2.12195	2.09560	2.05371	2.00953	1.96265	1.93766	1.91193	1.88516	1.85723	1.82800	1.79728
15	3.07319	2.69517	2.48979	2.36163	2.27302	2.20808	2.15818	2.11852	2.08621	2.05932	2.01707	1.97222	1.92431	1.89904	1.87277	1.84529	1.81676	1.78672	1.75505
16	3.04811	2.66817	2.46181	2.33276	2.24276	2.17722	2.12600	2.08598	2.05223	2.02315	1.98039	1.93392	1.88127	1.85556	1.82879	1.80084	1.77156	1.74075	1.70817
17	3.02623	2.64484	2.43743	2.30775	2.21625	2.15039	2.10169	2.06134	2.02839	2.00096	1.95772	1.91149	1.86236	1.83626	1.80901	1.78053	1.75063	1.71909	1.68566
18	3.00698	2.62395	2.41601	2.28577	2.19582	2.12958	2.07856	2.03789	2.00467	1.97698	1.93334	1.88681	1.83685	1.81035	1.78269	1.75371	1.72322	1.69099	1.65671
19	2.98990	2.60561	2.39702	2.26630	2.17596	2.10936	2.05802	2.01710	1.98366	1.95573	1.91170	1.86471	1.81416	1.78721	1.75916	1.72979	1.69876	1.66587	1.63077
20	2.97465	2.58925	2.38009	2.24893	2.15823	2.09132	2.03970	1.99852	1.96485	1.93674	1.89236	1.84496	1.79356	1.76667	1.73822	1.70833	1.67678	1.64326	1.60738
21	2.96096	2.57457	2.36489	2.23334	2.14231	2.07512	2.02325	1.98186	1.94797	1.91967	1.87497	1.82715	1.77555	1.74807	1.71927	1.68896	1.65691	1.62278	1.58615
22	2.94858	2.56121	2.35117	2.21927	2.12784	2.06050	2.00860	1.96680	1.93272	1.90425	1.85925	1.81106	1.75899	1.73122	1.70208	1.67128	1.63865	1.60415	1.56678
23	2.93736	2.54929	2.33873	2.20651	2.11491	2.04722	1.99492	1.95312	1.91888	1.89025	1.84497	1.79643	1.74392	1.71588	1.68643	1.65525	1.62227	1.58711	1.54903
24	2.92712	2.53833	2.32729	2.19488	2.10302	2.03512	1.98262	1.94066	1.90625	1.87768	1.83194	1.78308	1.73015	1.70185	1.67210	1.64067	1.60726	1.57166	1.53270
25	2.91774	2.52831	2.31702	2.18424	2.09216	2.02406	1.97128	1.92915	1.89469	1.86578	1.82000	1.77082	1.71752	1.68898	1.65895	1.62718	1.59325	1.55703	1.51760
26	2.90913	2.51910	2.30749	2.17447	2.08218	2.01389	1.96104	1.91876	1.88407	1.85503	1.80901	1.75957	1.70589	1.67712	1.64682	1.61472	1.58050	1.54368	1.50360
27	2.90119	2.51061	2.29871	2.16548	2.07298	2.00452	1.95151	1.90909	1.87427	1.84511	1.79889	1.74917	1.69514	1.66616	1.63560	1.60320	1.56859	1.53129	1.49057
28	2.89385	2.50276	2.29060	2.15714	2.06447	1.99585	1.94270	1.90014	1.86520	1.83592	1.78951	1.73954	1.68519	1.65600	1.62519	1.59250	1.55753	1.51976	1.47861
29	2.88703	2.49548	2.28307	2.14961	2.05658	1.98781	1.93452	1.89184	1.85679	1.82741	1.78081	1.73060	1.67593	1.64655	1.61551	1.58253	1.54721	1.50899	1.46706
30	2.88069	2.48872	2.27607	2.14253	2.04918	1.98033	1.92692	1.88412	1.84896	1.81949	1.77270	1.72227	1.66731	1.63774	1.60648	1.57323	1.53767	1.49891	1.45636
40	2.83535	2.44037	2.22609	2.09095	1.99882	1.92988	1.87252	1.82886	1.79290	1.76269	1.71456	1.66241	1.60515	1.57411	1.54108	1.50582	1.46816	1.42676	1.37691
60	2.79107	2.39225	2.17761	2.04099	1.94871	1.87872	1.81939	1.77483	1.73802	1.70701	1.65743	1.60327	1.54349	1.51072	1.47556	1.43734	1.39520	1.34787	1.29146
120	2.74781	2.34734	2.13999	1.99220	1.89887	1.82881	1.76768	1.72196	1.68425	1.65228	1.60120	1.54500	1.48207	1.44722	1.40928	1.36760	1.32034	1.26457	1.19256
∞	2.70554	2.30259	2.08380	1.94486	1.84727	1.77611	1.71672	1.67020	1.63152	1.59872	1.54678	1.48716	1.42060	1.38218	1.34187	1.29913	1.25395	1.16360	1.00000

كيف يمكن حساب المقارنات المتعددة (Post Hoc Comparisons) :

هي طريقة لإجراء عدد من الاختبارات الأولية لتحديد الفروق المعنوية للمتوسطات حال رفض فرضية العدم.

حال رفض فرضية العدم عند مستوى معنوية فلا دليل على وجود فروق معنوية بين المتوسطات ولا نعرف أي منها يختلف عن متوسطات (العينات) ومن حيث قيمة F معنوية ففرق المشاهدات بين المتوسطات معنوي أيضاً وعلى العموم توجد عدة طرق إحصائية لعمل اختبار بهذا الخصوص مثل:

١) طريقة شيفيه Scheffe وتستخدم هذه الطريقة في إجراء جميع المقارنات بين الأوساط وهي الطريقة المفضلة في حالة كون حجوم الخلايا غير متساوية أو عند الرغبة في إجراء مقارنات معقدة كأن نقارن ثلاث مجتمعات بمجتمع واحد، أو مجتمعين مقابل مجتمعين أو غيرها من مثل هذه المقارنات .

٢) طريقة توكي Tukey وتستخدم هذه الطريقة لمقارنة جميع الأزواج الممكنة للأوساط موضع الدراسة سواء كانت حجوم الخلايا متساوية أو غير متساوية (في حالة عدم تساوي حجوم الخلايا يستخدم الوسط التوافقي لحجم الخلية)، ويعتبر هذا الاختبار أدق من اختبار شيفيه Scheffe لمقارنة أزواج الأوساط.

٣) طريقة نيومن–كولز Newman–Keuls (S–N–K) وتفيد هذه الطريقة في المقارنة بين أزواج الأوساط فقط، وهي تستند كما هي الحال في طريقة توكي على توزيع مدى ستيودنتايز Studentize range ، وهي طريقة جيدة وقوية للكشف عن الفروق بين الأوساط في حالة تساوي حجوم الخلايا أو عدم تساويها (في حالة عدم تساوي حجوم الخلايا يستخدم الوسط التوافقي لحجم الخلية كما هو الحال في اختبار توكي)، ويعتبر هذا الاختبار (نيومن–كولز) أدق الاختبارات البعدية للكشف عن الفروق بين أزواج الأوساط، يليه اختبار توكي ثم بعد ذلك اختبار شفيه .

فعندما تشير نتائج تحليل التباين إلى عدم وجود فرق ذا دلالة يعزى إلى مستويات المعالجة فانه لا يوجد مبرر منطقي لأجراء أية اختبارات إحصائية أخرى.

أما إذا أشارت نتائج تحليل التباين (اختبار F) إلى أن هناك فرقا ذا دلالة يعزى إلى مستويات المعالجة،

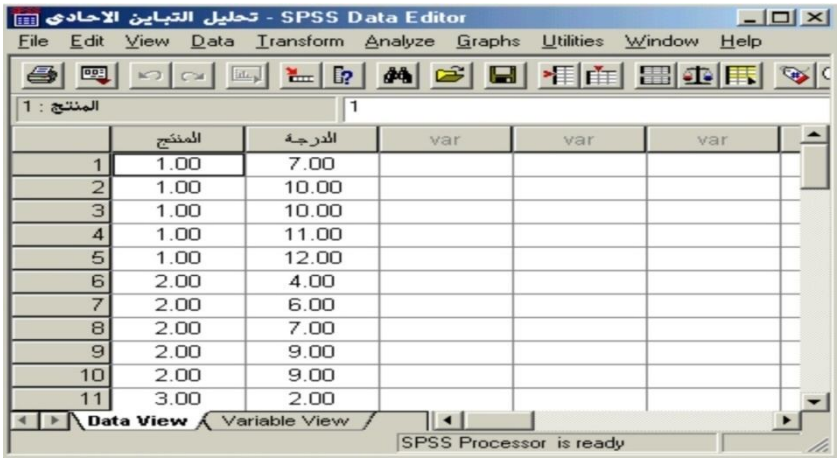
فان السؤال الذي يبقى قائما هو "أي مستوى من مستويات المعالجة يختلف عن الآخرين؟ أو بمعنى آخر أين توجد الفروق الحقيقية؟

حساب تحليل التباين الأحادي من خلال برنامج الـ SPSS

One Way Analysis of Variance

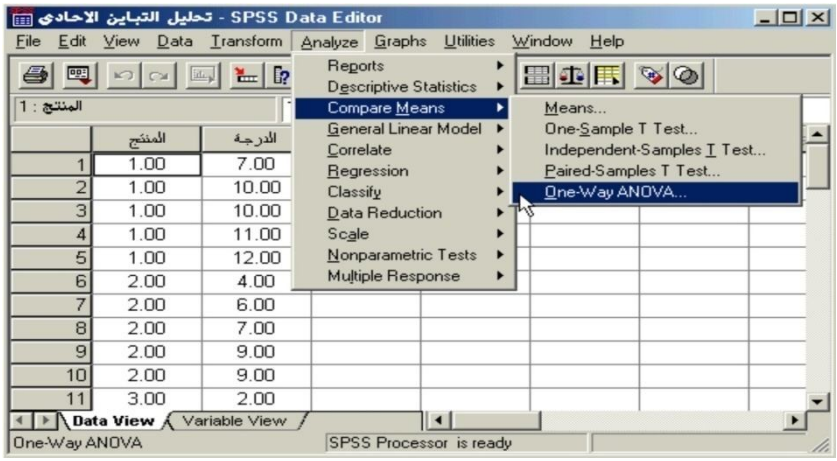
لغرض حساب قيمة تحليل التباين الأحادي One Way Analysis of Variance لنفس المثال السابق من خلال استخدام برنامج الـ SPSS نتبع الخطوات التالية :

- ✓ قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات Data Editor بالطريقة المناسبة كالتالي :

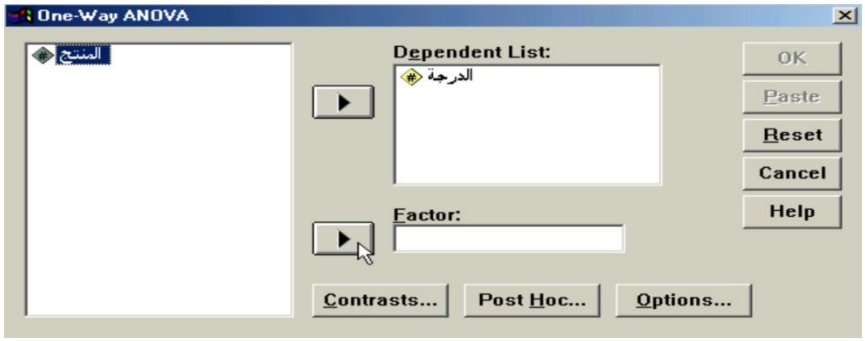


لاحظ أنه تم إدخال البيانات بطريقة مناسبة للتحليل الذي تم اختياره، حيث أدخلت مستويات المتغير المستقل في عمود وأطلق عليه اسم "المنتج"، وأدخلت درجات التقييم للمنتج تحت عمود آخر أطلق عليه اسم "الدرجة" .

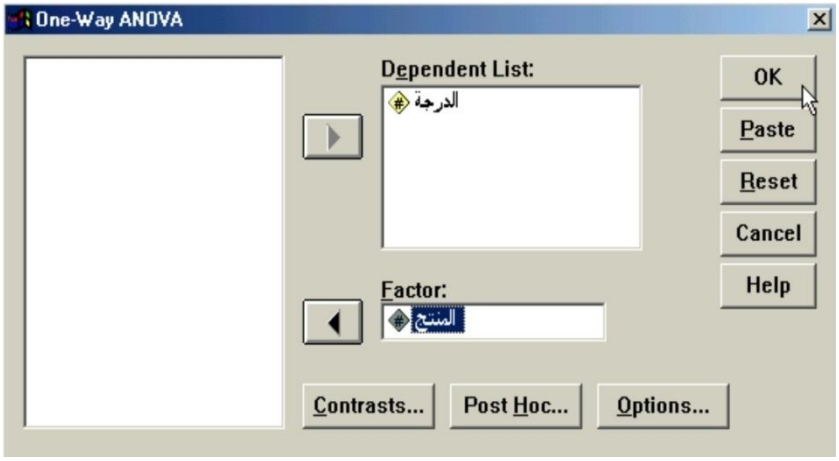
- ✓ من القائمة "تحليل" Analyze اختر الأمر "مقارنة المتوسطات" Compare Means فتظهر قائمة أوامر فرعية اختر منها "تحليل التباين الأحادي" One-Way ANOVA كالتالي :



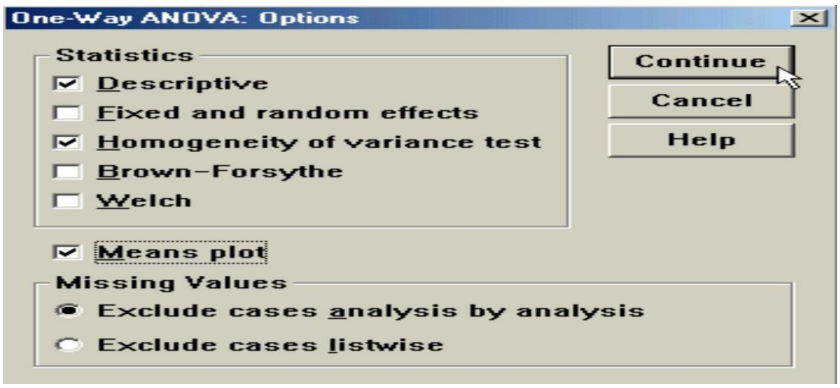
✓ بعد اختيار الأمر "تحليل التباين الأحادي" **One-Way ANOVA** سوف يظهر لك صندوق الحوار التالي :



✓ من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار حدد المتغير المستقل والمتغير التابع المراد إجراء تحليل التباين الأحادي لها، ونقلها إلى المستطيل الخاص بـ "المتغيرات التابعة" **Dependent List** من خلال النقر على السهم الذي يظهر مقابل المستطيل الخاص بـ "متغيرات الاختبار"، ستلاحظ انتقال المتغير مباشرة في المستطيل "المتغيرات التابعة" **Dependent List** (في هذا المثال المتغير التابع هو "الدرجة")، كرر نفس الإجراء مع المتغير المستقل **Independent Variable** وقم بنقله إلى المستطيل الخاص بـ "العامل" **Factor** (في هذا المثال المتغير المستقل هو "المنتج") كما يبدوا ذلك في الشكل التالي :



✓ أنقر على زر "خيارات" **Options** في الجهة السفلية اليمنى من صندوق الحوار السابق وذلك عند الرغبة في حساب الخصائص الأساسية للمتغيرات موضع الدراسة **Statistics** ، وعند الرغبة في عرض المتوسطات من خلال رسم بياني **Means Plot** ، وكذلك كيفية التعامل مع القيم المفقودة **Missing Values** ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" **Continue** .



✓ أنقر على زر "المقارنات البعدية المتعددة" **Post Hoc** في الجهة السفلية من صندوق الحوار السابق وذلك عند الرغبة في حساب المقارنات البعدية بين متوسطات المتغيرات موضع الدراسة والكشف عن مواقع الفروق وذلك في حالة كون قيمة ***F*** ذات دلالة إحصائية، وهناك العديد من الأساليب الإحصائية المعدة لهذا الغرض اشهرها اختبار توكي واختبار شفيه.

وفي المثال الحالي تم اختيار طريقة توكي **Tukey** للمقارنة البعدية بين أزواج الأوساط (ويمكن اختيار أكثر من طريقة في وقت واحد) كما يبدوا ذلك في الشكل التالي:

وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" **Continue** ، وستنتقل إلى صندوق الحوار الرئيسي، ثم أنقر بعد ذلك على زر "موافق" **OK** في صندوق الحوار الرئيسي سيؤدي ذلك إلى تنفيذ الاختبار، وستلاحظ ظهور النتائج في شاشة المخرجات كالتالي:

Oneway

Descriptives

الدرجة

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
المنتج١	5	10.0000	1.87083	.83666	7.6771	12.3229	7.00	12.00
المنتج٢	5	7.0000	2.12132	.94868	4.3660	9.6340	4.00	9.00
المنتج٣	5	4.0000	2.34521	1.04881	1.0880	6.9120	2.00	7.00
Total	15	7.0000	3.20713	.82808	5.2239	8.7761	2.00	12.00

ANOVA

الدرجة

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	90.000	2	45.000	10.000	.003
Within Groups	54.000	12	4.500		
Total	144.000	14			

Post Hoc Tests

Multiple Comparisons

Dependent Variable: الدرجة

Tukey HSD

(I) المنتج	(J) المنتج	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
المنتج١	المنتج٢	3.0000	1.34164	.105	-.5793	6.5793
	المنتج٣	6.0000*	1.34164	.002	2.4207	9.5793
المنتج٢	المنتج١	-3.0000	1.34164	.105	-6.5793	.5793
	المنتج٣	3.0000	1.34164	.105	-.5793	6.5793
المنتج٣	المنتج١	-6.0000*	1.34164	.002	-9.5793	-2.4207
	المنتج٢	-3.0000	1.34164	.105	-6.5793	.5793

*. The mean difference is significant at the .05 level.

نلاحظ أن برنامج الـ SPSS قام مباشرة بحساب الإحصاءات الأساسية للبيانات مثل المتوسط الحسابي والانحراف المعياري وغيرها من الإحصاءات ذات العلاقة.

ثم بعد ذلك قام البرنامج بحساب قيمة (F) للمتغيرات موضع الدراسة في الجدول المعنون بـ ANOVA ، ومن هذه النتائج نلاحظ أن قيمة (F) المحسوبة = ١٠ ، ودرجات الحرية df (٢ ، ١٢) ، والقيمة الحرجة .Sig=٠.٠٠٣ ، وبما أن القيمة الحرجة لـ .Sig F في الجدول (٠.٠٠٣) أصغر من قيمة α = ٠.٠٥

فإننا نستنتج أن الفرضية الصفرية تكون مرفوضة، أي يوجد اخلاف بين متوسطي مجتمعين على الأقل من المجتمعات التي قُيّمة من المستهلكين. ولمعرفة بين أي من المنتجات تكون الفروق قمنا بحساب اختبار المقارنات البعدية Post Hoc Comparisons لتحديد هذه الفروق، وقد أظهرت النتائج وجود فروق ذات دلالة إحصائية بين منتج (١) والذي متوسطه ١٠ ومنتج (٣) والذي متوسطه ٤ وذلك لصالح منتج (١).

المحاضرة العاشرة

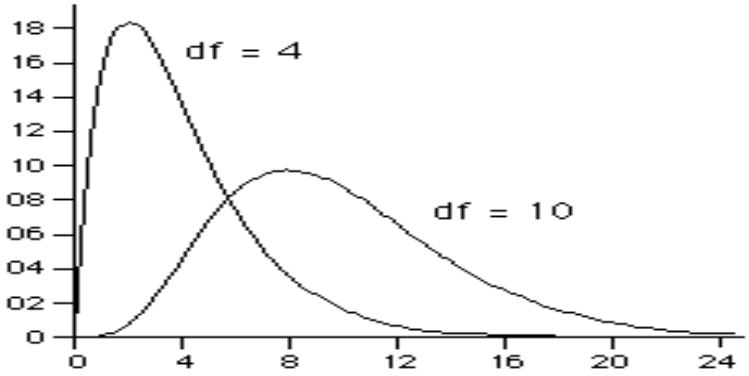
اختبار كا٢

يعتبر توزيع كاي تربيع χ^2 من التوزيعات الإحتمالية الشائعة الاستخدام حيث توجد له تطبيقات عديدة بدرجة يمكن معها القول أنه يأتي في المرتبة الثانية للتوزيع المعتدل من حيث كثرة تطبيقاته.

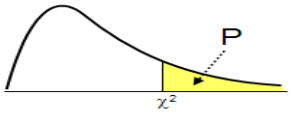
توزيع كاي تربيع χ^2 :

يعتمد توزيع χ^2 مثل توزيع t اعتمادا كاملا على درجات الحرية، فكلما زادت درجات الحرية كلما قل التواء التوزيع واقترب من التماثل.

شكل توزيع كاي تربيع χ^2



جدول توزيع χ^2



	P										
DF	0.995	0.975	0.20	0.10	0.05	0.025	0.02	0.01	0.005	0.002	0.001
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315
21	8.034	10.283	26.171	29.615	32.671	35.479	36.343	38.932	41.401	44.522	46.797
22	8.643	10.982	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.962	48.268
23	9.260	11.689	28.429	32.007	35.172	38.076	38.968	41.638	44.181	47.391	49.728
24	9.886	12.401	29.553	33.196	36.415	39.364	40.270	42.980	45.559	48.812	51.179
25	10.520	13.120	30.675	34.382	37.652	40.646	41.566	44.314	46.928	50.223	52.620
26	11.160	13.844	31.795	35.563	38.885	41.923	42.856	45.642	48.290	51.627	54.052
27	11.808	14.573	32.912	36.741	40.113	43.195	44.140	46.963	49.645	53.023	55.476
28	12.461	15.308	34.027	37.916	41.337	44.461	45.419	48.278	50.993	54.411	56.892
29	13.121	16.047	35.139	39.087	42.557	45.722	46.693	49.588	52.336	55.792	58.301
30	13.787	16.791	36.250	40.256	43.773	46.979	47.962	50.892	53.672	57.167	59.703
31	14.458	17.539	37.359	41.422	44.985	48.232	49.226	52.191	55.003	58.536	61.098

شكل مقطعي لجدول توزيع χ^2

P / DF	المساحة تحت المنحنى إلى يمين قيمة χ^2						
	0.995	.....	0.975	.....	0.05	.....	0.01
. . .							
10	2.156		3.247		18.307		23.209
. . .							

فنلاحظ من الجدول السابق لـ χ^2 أن:

الصف العلوي لجدول توزيع χ^2 يحوي على الاحتمالات الشائعة الاستخدام والتي تمثل مساحة الطرف الأيمن للمنحنى، ويحوي كل صف من الصفوف التي تلي الصف العلوي على القيم المختلفة المناظرة لتوزيع معين من توزيعات χ^2 بدرجات حرية محددة، وتوجد درجات الحرية بالعمود الأول بيسار الجدول، وبالتالي فإن القيمة الموجودة أمام درجات حرية معينة وتحت احتمال محدد هي قيمة المجدولة χ^2 لكاي تربيع بدرجات الحرية المحددة والتي تكون المساحة تحت المنحنى على يمينها مساوية للإحتمال المحدد.

فمثلا نجد أن قيم χ^2 بدرجات حرية **10** والتي تساوي المساحة على يمينها **0.01** , **0.05** من المساحة الكلية هي **18.307**, **23.209** على التوالي.

وقيم χ^2 جميعها موجبة أي أنها أكبر من أو تساوي الصفر. ويرمز لـ χ^2 بالرمز $\chi^2_{(v,\alpha)}$ حيث يمثل الدليل السفلي الأول درجات الحرية بينما يمثل الدليل السفلي الثاني المساحة على يمين % القيمة.

فمثلا $\chi^2_{(10,0.05)}$ تمثل قيمة χ^2 بدرجات حرية **10** والتي تساوي المساحة على يمينها **0.05** وبالنظر للجدول نجد أن

$$\chi^2_{(10,0.05)} = 18.307$$

استخدامات توزيع χ^2

اختبار تباين المجتمع

يستخدم توزيع χ^2 في إجراء العديد من الاختبارات الإحصائية مثل:

الاختبارات المتعلقة بتباين مجتمع ما (وذلك لاختبار المشاكل التي تتطلب اختبار تشتت مجتمع ما)، ويتم ذلك من خلال استخدام المعادلة التالية:

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

ويفترض في هذا الاختبار أن العينة مسحوبة من مجتمع معتدل وذلك من خلال مقارنة قيمة χ^2 المحسوبة من المعادلة بالقيمة الحرجة لـ χ^2 والمستخرجة من جداول χ^2.

مثال:

إذا عرف أن تباين قوة مقاومة الكسر للكابلات التي تنتجها إحدى الشركات لاتزيد عن 40000 ، وتستخدم الشركة الآن طريقة إنتاج جديدة يعتقد أنها ستزيد من تباين قوة مقاومة الكابلات للكسر، سحبت عينة عشوائية من عشرة كابلات فوجد تباينها يساوي 50000 .

بافتراض أن قوة مقاومة الكسر للكابلات تتبع التوزيع المعتدل، اختبر الفرض القائل بوجود زيادة معنوية في التباين عند مستوى معنوية $\alpha = 0.01$.

الحل:

- **وضع فرض العدم والفرض البديل.**

صياغة الفرضية الصفرية كالتالي:

$$H_0: \sigma^2 \leq 40000$$

في حين تفترض الفرضية البديلة التالي :

$$H_1: \sigma^2 > 40000$$

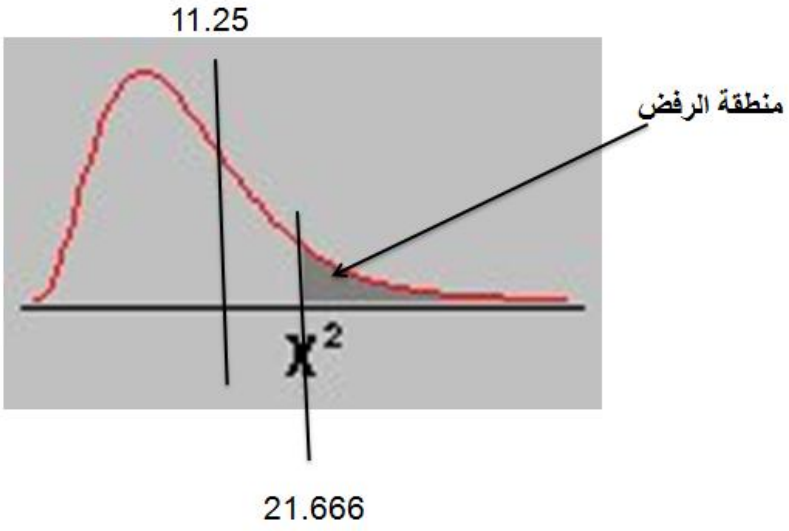
- **تحديد مستوى الدلالة (α): وهي 0.01 .**
- **درجات الحرية = 9 ، فإن قيمة χ^2 المجدولة هي 21.666**

لذا فإن قاعدة القرار هي أن يتم رفض الفرضية الصفرية H0 عندما تكون $\chi^2 \geq 21.666$

وحيث أن قيمة χ^2 لاختبار تباين المجتمع يتم حسابها كالتالي :

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{(10-1)50000}{40000} = \frac{(9)50000}{40000} = \frac{450000}{40000} = 11.25$$

وحيث أن قيمة χ^2 المحسوبة أقل من قيمة χ^2 المجدولة، فإننا بالتالي نرفض الفرضية الصفرية H0 عند مستوى دلالة 0.01 وبالتالي يمكننا القول أن بيانات العينة تدل على أن الزيادة الظاهرة في التباين ليست معنوية عند مستوى الدلالة المحدد ، والشكل التالي يوضح ذلك.



اختبار مربع كاي لجودة التوفيق

Testing of Goodness of Fit

ويهتم هذا النوع من الاختبارات الإحصائية باختبار ما إذا كانت مشاهدات عينة تم اختيارها من مجتمع له توزيع احتمالي معين أو نظرية معينة.

ويستخدم هذا الاختبار عندما تكون البيانات اسمية أو على شكل تكرارات ويقصد بجودة التوفيق هنا دراسة مدى تشابه تكرارات العينة والتي تسمى عادة بالتكرارات الملاحظة Observed مع التكرارات المتوقعة Expected للمتغير موضوع الدراسة في المجتمع الأصلي.

ويستخدم اختبار كا^2 كطريقة إحصائية للمقارنة بين التكرارين الملاحظ والمتوقع. فإذا كانت العينة ممثلة للمجتمع في تكراراتها ومتطابقة معه فإن قيمة كا^2 تكون عادة صفراً وتزداد هذه القيمة لتصبح أكثر من صفر كلما كان هناك فرق بين تكرارات العينة (الملاحظة) وبين تكرارات التوزيع النظري للمجتمع (المتوقعة).

الفروض الإحصائية:

H0 : مجموعة المشاهدات التي تم اختيارها تتبع توزيع احتمالي معين أو نظرية معينة.

HA : مجموعة المشاهدات التي تم اختيارها لا تتفق مع هذا التوزيع أو نظرية معينة.

إحصاء الاختبار:

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

حيث O_i تمثل التكرار المشاهد للنتيجة رقم i

E_i تمثل التكرار المتوقع المناظر للنتيجة رقم i حيث :

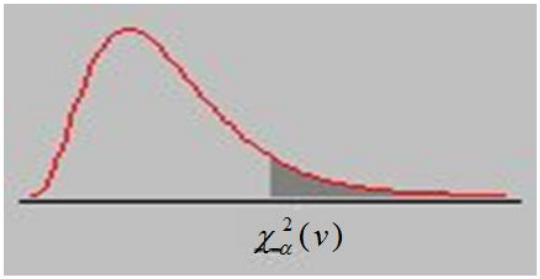
$$E_i = np_i$$

حيث $n = \sum_i O_i$ والقيمة P_i نحصل عليها من التوزيع الإحتمالي أو النظرية المعطاة في فرضية العدم.

ويجب أن يكون التكرار المتوقع في أية خلية لا يقل عن 5 حتى يتم حساب إحصائي الاختبار χ^2 بشكل صحيح.

مناطق الرفض والقبول:

سوف نستخدم جدول مربع كاي χ^2 لتعيين القيمة المجدولة (الحرجة) $\chi^2_\alpha(v)$ حيث $v = k^* - 1$ ، و k^* هي عدد المشاهدات النهائية بعد عملية الدمج إن وجدت.



القرار:

نقبل فرض العدم إذا كانت (قيمة كاي تربيع المحسوبة أصغر من القيمة المجدولة) أي أن :

$$\chi_0^2 < \chi_\alpha^2(v)$$

ونرفض فرض العدم إذا كانت (قيمة كاي تربيع المحسوبة أكبر من القيمة المجدولة) أي أن :

$$\chi_0^2 > \chi_\alpha^2(v)$$

مثال:

اختار أحد الباحثين عينة حجمها **n=800** شخصا من أحد المدن، وكان توزيعهم حسب فصيلة الدم كالتالي:

فصيلة الدم	A	B	AB	O
عدد الأشخاص (التكرار المشاهد)	200	150	100	350

هل يتفق هذا التوزيع مع توزيع أفراد مدينة أخرى كان توزيع فصيلة دمهم حسب النسب التالية:

فصيلة الدم	A	B	AB	O
النسب المئوية للأشخاص	25%	15%	15%	45%

استخدم مستوى معنوية $\alpha = 0.05$

الحل:

الفروض الإحصائية:

H_0 : توزيع فصيلة الدم في العينة يتفق مع التوزيع المناظر للمدينة الأخرى.

H_A : توزيع فصيلة الدم في العينة لا يتفق مع التوزيع المناظر للمدينة الأخرى.

مستوى المعنوية:

مستوى معنوية $\alpha = 0.05$

إحصاء الاختبار:

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

حيث O_i تمثل التكرار المشاهد للنتيجة رقم i ،

E_i تمثل التكرار المتوقع المناظر للنتيجة رقم i حيث $E_i = np_i$

$$E_1 = np_1 = 800(0.25) = 200$$
$$E_2 = np_2 = 800(0.15) = 120$$
$$E_3 = np_3 = 800(0.15) = 120$$
$$E_4 = np_4 = 800(0.45) = 360$$

ونلاحظ أن جميع المشاهدات المتوقعة أكبر من 5% وأيضا حجم العينة، لذا يمكن تعيين احصاء الاختبار كاي تربيع لاختبار هذه البيانات، وبكتابة كلا من المشاهدات والقيم المتوقعة معا في جدول واحد كالتالي:

فصيلة الدم	A	B	AB	O
عدد الأشخاص (التكرار المشاهد)	200	150	100	350
التكرار المتوقع	200	120	120	360

وبتطبيق معادلة كاي تربيع للحصول على قيمة كاي المحسوبة ويتم ذلك كالتالي:

$$\chi_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} = \left[\frac{(200-200)^2}{200} + \frac{(150-120)^2}{120} + \frac{(100-120)^2}{120} + \frac{(350-360)^2}{360} \right] = 11.11$$

مناطق الرفض والقبول:

عند استخدام جدول مربع كاي لتعيين القيمة المجدولة لـ $\chi_\alpha^2(v)$ حيث **v=4–1=3** ، وبالتالي فإن $\chi_{0.05}^2(3) = 7.815$

القرار:

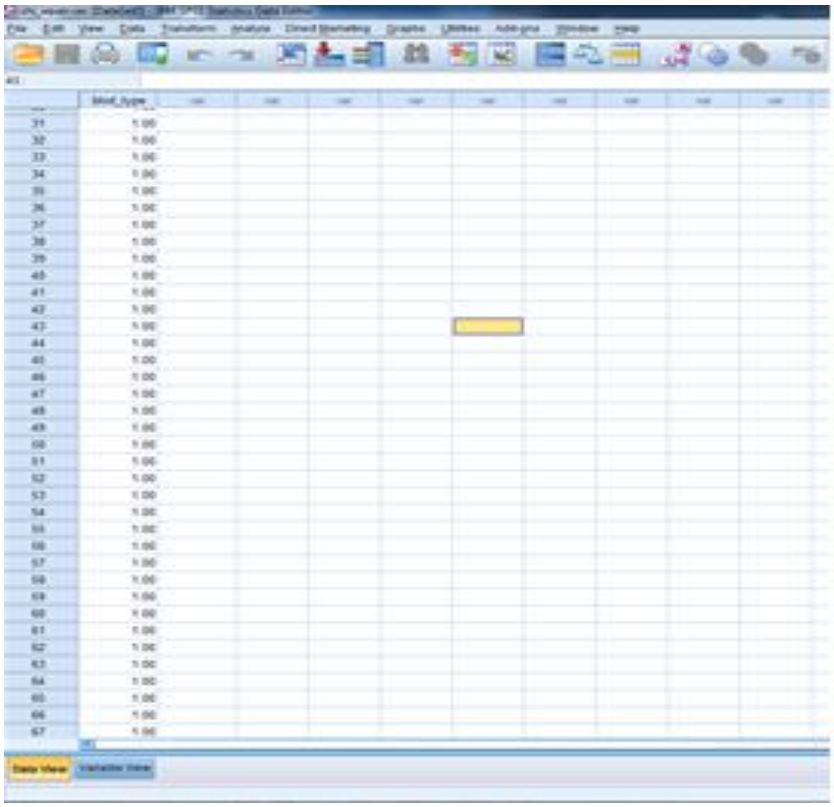
ونلاحظ أن $\chi_0^2 > \chi_\alpha^2(v)$ لذا سوف نرفض فرضية العدم وبالتالي فإن توزيع فصيلة الدم في المدينتين مختلف.

حساب اختبار مربع كاي (كا^{2}) لجودة التوافيق

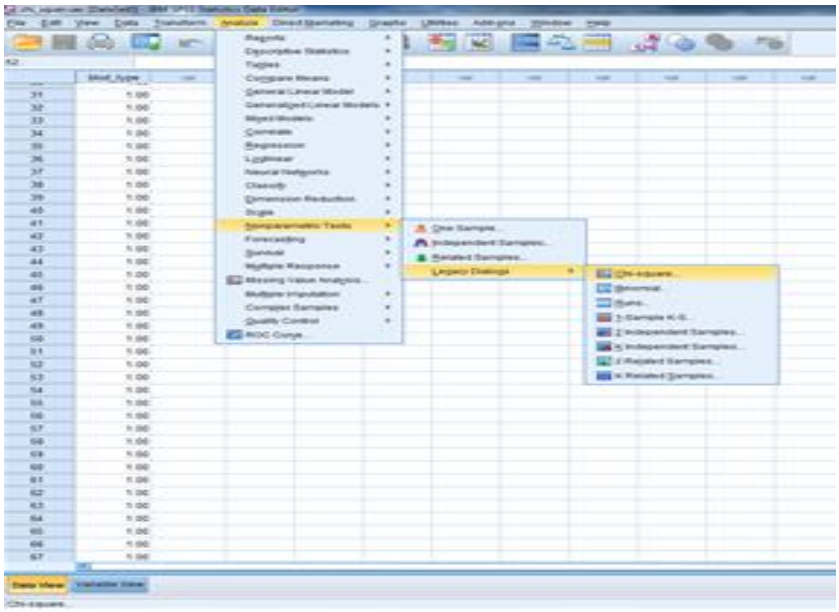
Chi Squre–Testing of Goodness of Fit

من خلال برنامج SPSS

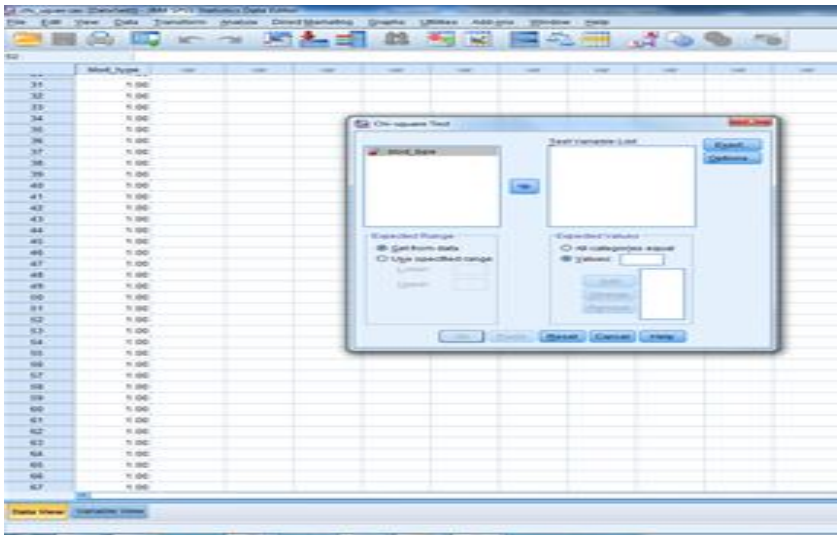
قم بإدخال البيانات على شكل تكرارات في عمود واحد كالتالي:



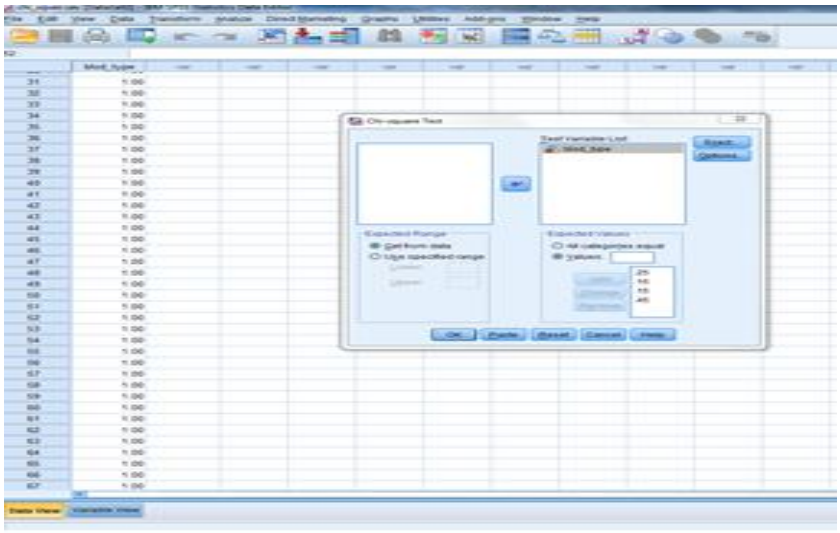
بعد إدخال البيانات قم باختيار Analyze ثم Nnparametric Tests ثم Legacy dialogs ثم ... Chi–Squre



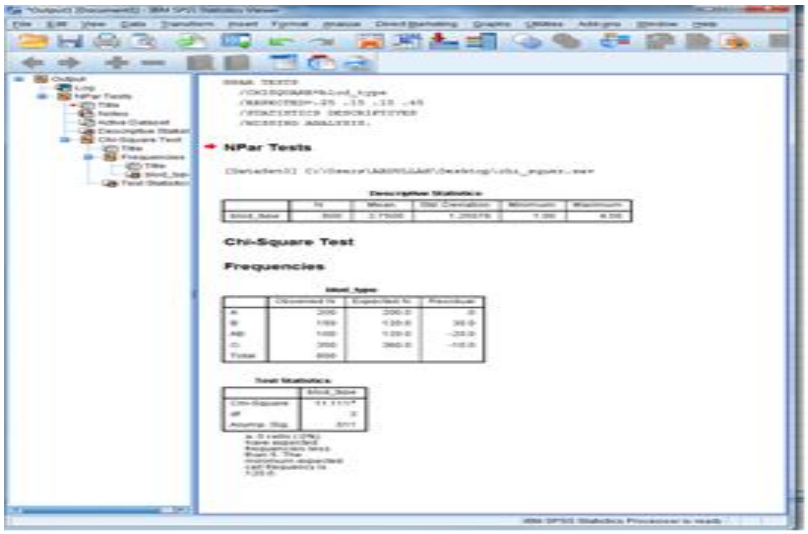
سيفتح لك صندوق حوار كالتالي:



قم بنقل المتغير موضع البحث إلى الصندوق المعنون **Test Variable list** وقم بعد ذلك بتحديد التكرار المتوقع من خلال إدخال الاعداد أو النسب المحددة للتكرار المتوقع بشكل متتابع حسب ترتيب القيم الملاحظة



بعد ذلك إنقر على زر **OK** فيقوم البرنامج بحساب قيمة **Chi_Squre** والجداول المحددة لذلك كالتالي:



يظهر لنا من خلال المخرجات أن قيمة كا^2 المحسوبة ١١.١١ ودرجات الحرية ٣ ومستوى الدلالة .Sig = ٠.٠١١ وهذه القيمة تعني أن قيمة كا^2 دالة إحصائيا أي توجد فروق ذات دلالة إحصائية بين المجموعتين

وهذه النتيجة ستدفعنا إلى أن نرفض فرضية العدم وبالتالي فإن توزيع فصيلة الدم في القبيلتين مختلف.

اختبار مربع كاي للاستقلالية (الاعتمادية)

Testing of Independence

كاي تربيع للاستقلالية (**Chi–Square test of independency**) هو اختبار بسيط يقوم به الباحث لمعرفة ما إذا كان هناك عــلاقة بين شيئين أو متغيرين. يجرى هذا الاختبار عن طريقة مقارنة قيمة يحددها الباحث مسبقا تعرف بمستوى المعنوية (الفا) بالقيمة المسماة **p–Value** تحسب من البيـانات التوفرة، حيث سيتضح عن طـريق المقارنة بين القيمتين ما إذا كانت هنالك علاقة بين الاثنين أم لا .

فرضية العدم (**Null hypothesis**): لا توجد أي علاقة بين المتغيرين ويرمز لهذه الفرضية **H0** والذي يتم افتراض صحته عند القيام بالاختبار.

عند القيام بالاختبار لمتغيرين، تكتب هذه الفرضية بهذه الطريقة: V_1مستقل عن V_2 ، حيث V_1 و V_2 تمثل المتغيرين تحت الدراسة. ويمكن كتابة فرض العدم الإحصائي بالشكل التالي:

$$H_0: V_1 \text{ is independent of } V_2$$

الفرض البديل (**Alternative hypothesis**): توجد علاقة بين المتغيرين تحت الدراسة ويرمز لهذه الفرضية H_A وتكتب الطريقة التالية: V_1 غير مستقل أو يتبع لـ V_2 ، حيث V_1 و V_2 المتغيرين تحت الدراسة. ويمكن كتابة الفرض البديل بالشكل التالي:

$$H_A: V_1 \text{ is dependent on } V_2$$

مستوى المعنوية (**Level of Significance**) الفا:

عند إجراء إختبار كاي تربيع فإن على الباحث اختيار قيمة تسمى **Level of Significance** أو مستوى المعنوية (الفا) وهذه القيمة يمكن القول بأنها تمثل احتمال الوقوع في خطأ في الاختبار يسمى الخطأ من النوع الأول وهو رفض فرض العدم H_0مع أنه صحيح. بمعنى أن يستنتج الباحث بناء على البيانات المتوفرة أن هنالك علاقة بين المتغيرين مع أنه لا توجد علاقة وهو استنتاج خاطئ.

هذه القيمة التي يحددها الباحث يقوم بمقارنتها بقيمة تسمى **p–value** والتي يمكن حسابها يدويا أو باستخدام أحد البرامج الإحصائية وذلك من البيانات التي جمعها الباحث.

غالبا في الأبحاث ما يتم استخدام قيمة الفا أو **Level of Significance** على أنها ٠،٠١ أو ٠،٠٥، و الاختيار يرجع للباحث ومدى مجال الخطأ الذي يود أن يسمح به، حيث في حالة إختيار الفا = ٠،٠١ فإن نتيجة الاختبار تكون أدق.

المختبر الإحصائي:

التكرار المتوقع:

$$\frac{\text{مجموع الصف} \times \text{مجموع العمود}}{\text{حجم العينة}}$$

نكرر تطبيق هذه المعادلة لجميع الصفوف والأعمدة لكلا المتغيرين

تحديد درجات الحرية:

درجات الحرية = (عدد الصفوف – ١) × (عدد الأعمدة – ١)

تحديد قيمة كا^{٢} المجدولة:

يتم بعد ذلك تحديد قيمة كا^{٢} المجدولة من خلال الرجوع إلى جدول كا^{٢} عند درجة حرية محددة وفقا لمعطيات الدراسة

القرار:

نقارن كا^{٢} المحسوبة بالجدولية، فعندما تكون قيمة كا^{٢} المحسوبة أكبر من قيمة كا^{٢} المجدولة فإننا نرفض الفرضية الصفرية أو فرض العدم والتي تنص على أنه لا توجد أي علاقة بين المتغيرين ونقبل الفرض البديل والتي تثبت وجود علاقة بين المتغيرين تحت الدراسة.

أما إذا كانت قيمة كا^{٢} المحسوبة أقل من قيمة كا^{٢} المجدولة فإننا نقبل الفرضية الصفرية أو فرض العدم

حساب اختبار مربع كاي (كا^{2}) للإستقلالية

Chi Squre–Test of Independence

من خلال برنامج SPSS

مثال:

في دراسة للعلاقة بين التقدير الذي يحصل عليه الطالب في الجامعة وجنسه أخذت عينة من نتائج الطلاب الذكور و الاناث وكانت كما يلي:

أولا: الإناث

راسب	راسب	راسب	جيد جدا	راسب	ممتاز	مقبول	ممتاز
جيد	جيد جدا	جيد جدا	جيد	مقبول	مقبول	مقبول	راسب
مقبول	راسب	مقبول	مقبول	مقبول	راسب	جيد جدا	جيد جدا
جيد	جيد	ممتاز	جيد جدا	ممتاز	جيد	جيد	جيد
					جيد جدا	ممتاز	جيد

ثانيا: الذكور

راسب	جيد	جيد	جيد	راسب	جيد جدا	راسب	جيد جدا
جيد جدا	جيد	راسب	راسب	راسب	راسب	راسب	مقبول
مقبول	ممتاز	ممتاز	راسب	راسب	مقبول	مقبول	ممتاز
ممتاز	جيد	جيد	مقبول	راسب	راسب	جيد	جيد
			جيد جدا	ممتاز	جيد	جيد جدا	ممتاز

والمطلوب:

هل توجد علاقة بين تقدير الطالب وجنسه عند مستوى الدلالة ٠.٠٥ ؟ = α

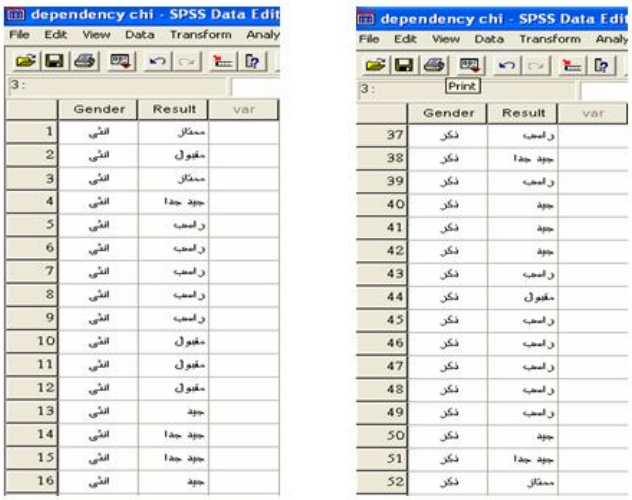
الحل:

الفرضية الصفرية: تقدير الطالب لا يعتمد على جنسه (متغير الجنس والتقدير مستقلان)

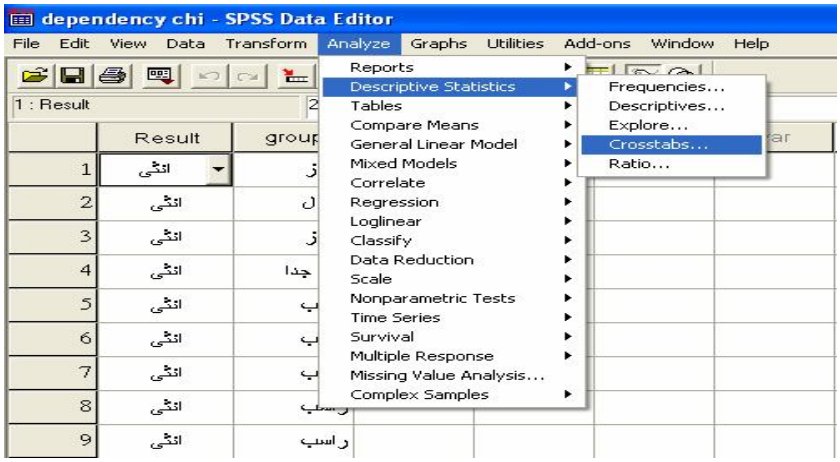
الفرضية البديلة: تقدير الطالب يعتمد على جنسه (توجد علاقة بين جنس الطالب وتقديره)

ثم نقـوم بتعريف متغيرين نوعيين هما (*Result*) و (*Gender*) في شـاشة تعريف المتغيرات بحيث يكون كـود متغير (*Result*) هو (٠= راسب، ١=مقبول، ٢=جيد، ٣ جيد جدًا، ٤=ممتاز) وكود المتغير (*Gender*) هو (١=ذكر، ٢=انثى)

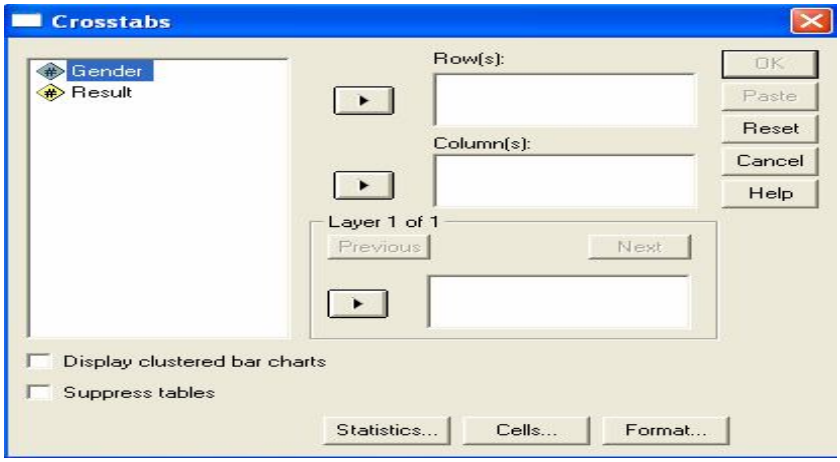
ندخل البيانات كما في الشكل التالي:



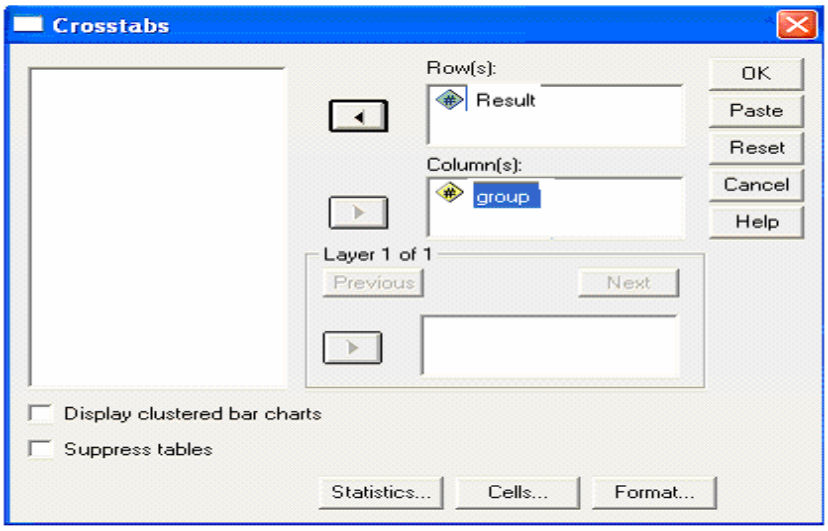
من قائمة التحليل ***Analyze*** نختار القائمة الفرعية للإحصاءات الوصفية ***Descriptive Statistics*** ومن ثم نختار الأمر ***Cross tabs*** كما في الشكل التالي:



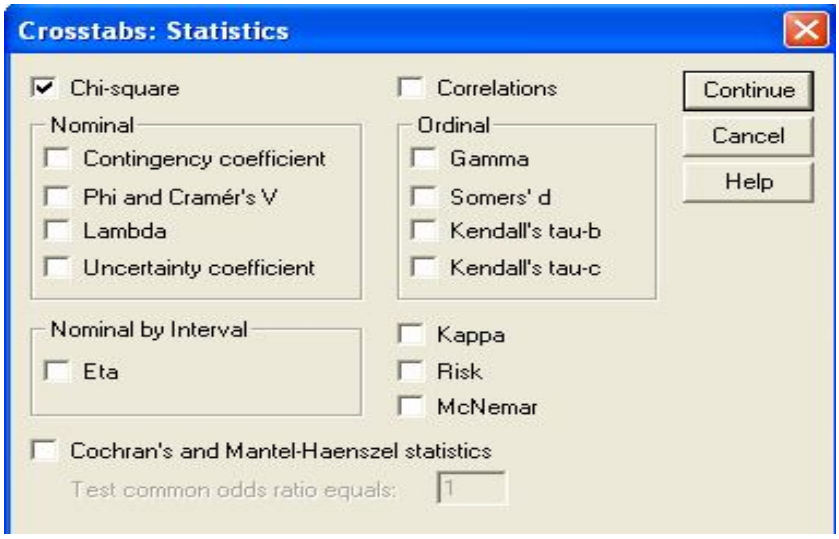
يظهر المربع الحواري التالي:



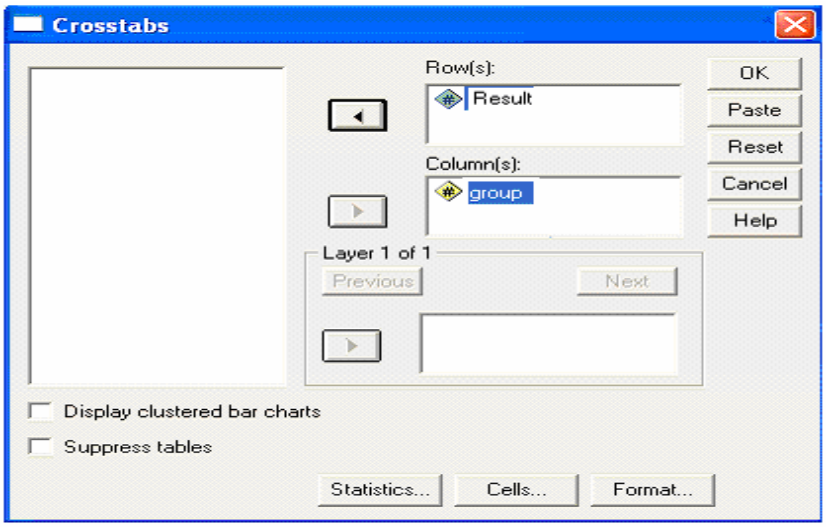
ننقل المتغير ***Result*** لخانة الصفوف ***Rows*** والمتغير ***Gender*** لخانة الأعمدة ***Columns*** باستخدام الأسهم .



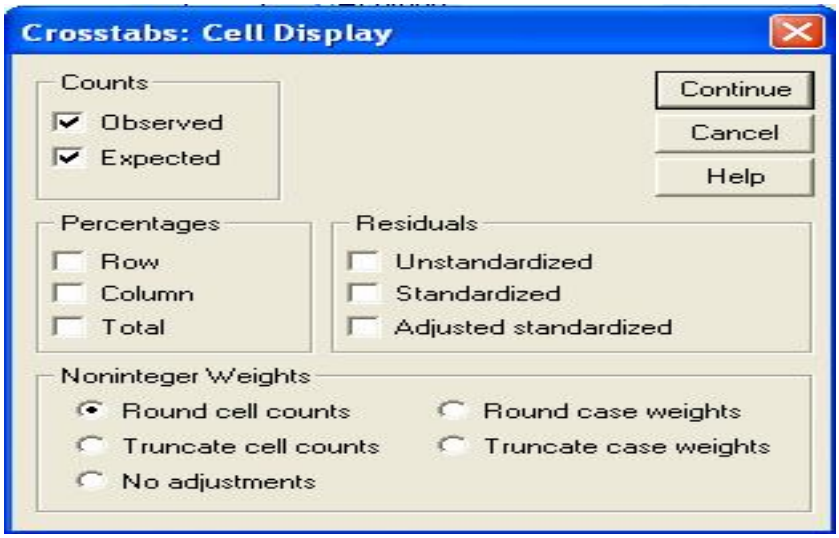
ومن ثم نضغط على ***Statistics*** للحصول على المربع الحواري التالي:



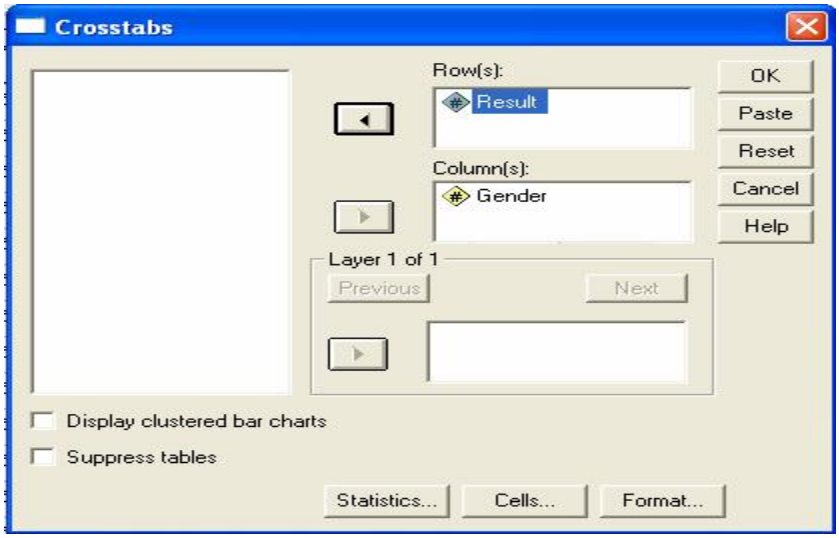
نضع علامة على خانة اختبار مربع كاي ***Chi-Square*** لحساب اختبار الاستقلالية ومن ثم نضغط على ***Continue*** للعودة للمربع الحواري السابق:



لإظهار جدول التوقعات نضغط على زر **Cell** ليظهر المربع الحواري التالي:



نختار الخيار ***Expected*** جدول توقعات ظهور البيانات ومن ثم نضغط ***Continue*** للعودة للمربع الحواري السابق.



نضغط على ***Ok*** للحصول على النتائج.

تتكون نتائج الأمر ***Cross tabulati*** من ثلاثة جداول:

الأول يصف حجم العينات المدخلة ونسب البيانات المفقودة كالتالي:

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
group * Result	72	100.0%	0	.0%	72	100.0%

الجدول الثاني يبين جدول توزيع العينة حسب المتغيرين والقيم المتوقعة حسب اختبار الاستقلالية كالتالي:

عدد الذكور الراسبين

group * Result Crosstabulation

			Result		Total
			ذكر	انثى	
group	راسب	Count	12.00	7.00	19
		Expected Count	9.76	9.24	19.0
	مقبول	Count	5.00	8.00	13
		Expected Count	6.68	6.32	13.0
	جيد	Count	9.00	8.00	17
		Expected Count	8.74	8.26	17.0
	جيد جدا	Count	5.00	7.00	12
		Expected Count	6.17	5.83	12.0
	ممتاز	Count	6.00	5.00	11
		Expected Count	5.65	5.35	11.0
Total		Count	37.00	35.00	72
		Expected Count	37.00	35.00	72.0

توقع الذكور الراسبين

يبين الجدول الثاني السابق أن عدد البيانات المدخلة ٧٢ ، عدد الذكور ٣٧ (منهم ١٢ راسب وقيمتها المتوقعة ٩.٧٦ ، ٥ مقبول وقيمتها المتوقعة ٦.٦٨ ، ٩ جيد وقيمتها المتوقعة ٨.٧٤ ، ٥ جيدجدا وقيمتها المتوقعة ٦.١٧ ، و ٦ ممتاز وقيمتها المتوقعة ٥.٦٥) والاناث ٣٥ (منهم ٧ راسب وقيمتها المتوقعة ٩.٢٤ ، ٨ مقبول وقيمتها المتوقعة ٦.٣٢ ، ٨ جيد وقيمتها المتوقعة ٨.٢٦ ، ٧ جيدجدا وقيمتها المتوقعة ٥.٨٣ ، و ٥ ممتاز وقيمتها المتوقعة ٥.٣٥)

الجدول الثالث يبين نتيجة اختبار مربع كاي كالتالي:

قيمة الاختبار

درجة الحرية

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.437[a]	4	.656
Likelihood Ratio	2.459	4	.652
Linear-by-Linear Association	.298	1	.585
N of Valid Cases	72		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5.35.

مستوى دلالة الاختبار

يبين الجدول الثالث السابق أن قيمة اختبار مربع كاي هي ٢.٤٣٧ بدرجة حرية مقادرها ٤

يتبين لنا من الجدول أن أقل قيمة لمستوى الدلالة هي ٠.٦٥٦ = *Asymp. Sig. (2-sided)* وهي اكبر من مستوى الدلالة α = ٠.٠٠٥ وبالتالي لا نستطيع رفض الفرضية الصفرية أي أن تقدير الطالب لا يعتمد على جنسه.

مع أصدق الأمنيات

أخوكم / ثابت

@thabet_18

المحاضرة الحادية عشر

معامل الارتباط

بالرغم من أن مقاييس العلاقة تختلف عما سبقها من مقاييس ، فهي تتعلق بدراسة العلاقة بين متغيرين، بينما المقاييس السابقة فتهتم بدراسة الفروق بين المتغيرات .

وعندما نقول مقاييس العلاقة نعني بذلك تلك المقاييس التي تبين درجة العلاقة والارتباط بين متغيرين أو أكثر مثلا، كأن يكون الهدف معرفة هل هناك علاقة بين مستوى الإنتاجية وجودة المنتج في مصنع ما؟، أي هل كلما زادت الإنتاجية تقل جودة المنتج أو العكس .

معامل الارتباط: هو تعبير يشير إلى المقياس الإحصائي الذي يدل على مقدار العلاقة بين المتغيرات سلبية كانت أم إيجابية، وتتراوح قيمته بين الارتباط الموجب التام (+١) وبين الارتباط السالب التام (−١) .

العلاقة الطردية بين المتغيرات: هو تعبير يشير إلى تزايد المتغيرين المستقل والتابع معا، فإذا كانت الإنتاجية مرتفعة، ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط موجب، وأعلى درجة تمثله هي (+١) .

العلاقة العكسية بين المتغيرات: هو تعبير يشير إلى تزايد في متغير يقابله تناقص في المتغير الآخر، فإذا كانت الإنتاجية منخفضة ومستوى الجودة مرتفع، يقال حينئذ أن بينهما ارتباط سالب، وأعلى درجة تمثله هي (−١) .

ومن الطبيعي ملاحظة أن الارتباط الكامل لا وجود له في الظواهر الطبيعية، وأن معامل الارتباط الناتج في الأبحاث والدراسات الإنسانية والاجتماعية يكون عادة كسرا موجبا أو سالبا. والجدول التالي يوضح أنواع العلاقات بين المتغيرات كما يصفها معامل الارتباط:

نوع العلاقة	قيمة معامل الارتباط
طردية كاملة	+١
طردية ناقصة	+ كسر(قيمة موجبة)
صفرية	صفر
عكسية ناقصة	− كسر (قيمة سالبة)
عكسية كاملة	−١

إن معامل الارتباط التام الموجب (+١) يعنى التغير في اتجاه واحد في كلا الظاهرتين مع بقاء الأوضاع النسبية لوحدات الظاهرة ثابتة، سواء كان هذا التغير في اتجاه الزيادة (أي زيادة قيم الظاهرة الأولى تتبعها زيادة في قيم الظاهرة الأخرى)، أو في اتجاه النقص (أي نقص قيم الظاهرة الأولى يتبعها نقص في قيم الظاهرة الأخرى) .

طرق التعرف على العلاقة بين متغيرين وحسابها

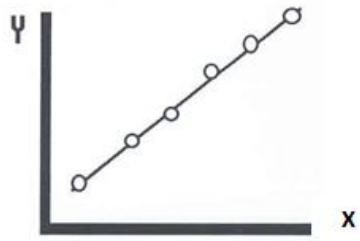
أولا: طريقة شكل الانتشار **Scatter Diagram** :

هناك وسيلة مبدئية يعرف الباحث من خلالها نوع الارتباط بين المتغيرين وما إذا كان الارتباط قوياً وضعيفاً أو منعدماً، وما إذا كانت العلاقة خطية أو غير خطية، موجبة أو سالبة. هذه الوسيلة هي " شكل الانتشار " والتي تصلح إذا كان المتغيران كميين. وجدير بالذكر أن هذه وسيلة مبدئية تساعدفقط في معرفة نوع الارتباط ولا تعتبر بديلاً عن الطرق الإحصائية التي سوف نتناولها بالتفصيل في هذه المحاضرة.

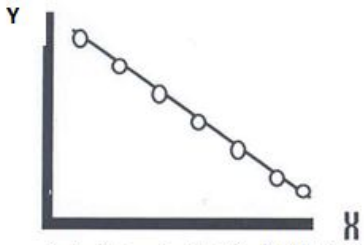
والمقصود بشكل الانتشار هو تمثيل قيم الظاهرتين بيانياً على المحورين، المتغير الأول X على المحور الأفقي، والمتغير الثاني Y على المحور الرأسي، حيث يتم تمثيل كل زوج Pair من القيم بنقطة، فنحصل على شكل يمثل كيفية انتشار القيم على المستوى، وهو الذي يسمى شكل الانتشار. وطريقة انتشار القيم تدل على وجود أو عدم وجود علاقة بين المتغيرين ومدى قوتها ونوعها. فإذا كانت تتوزع بشكل منتظم دل ذلك على وجود علاقة (يمكن استنتاجها)، أما إذا كانت النقط مبعثرة ولا تنتشر حسب نظام معين دل ذلك على عدم وجود علاقة بين المتغيرين أو أن العلاقة بينهما ضعيفة. والأشكال التالية تظهر بعض أشكال الانتشار المعروفة :

الشكل الأول :

إذا وقعت جميع النقاط على خط مستقيم، دل ذلك على أن العلاقة بينهما خطية وأنها ثابتة أو تامة. وهذه تمثل أقوى أنواع الارتباط بين المتغيرين " ارتباط تام ". فإذا كانت العلاقة طردية فإن " الارتباط طردي تام " كما في الشكل الأول (أ). ومثاله العلاقة بين الكمية المشتراه من سلعة والمبلغ المدفوع لشراء هذه الكمية. أما إذا كانت العلاقة عكسية (وجميع النقاط تقع على خط مستقيم واحد فإن "الارتباط عكسي تام" كما في الشكل الأول (ب). ومثال على ذلك العلاقة بين السرعة والزمن.



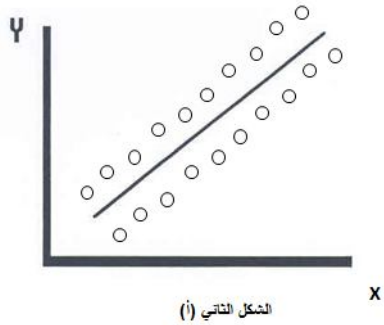
الشكل الأول (أ) ارتباط طردي تام (موجب)



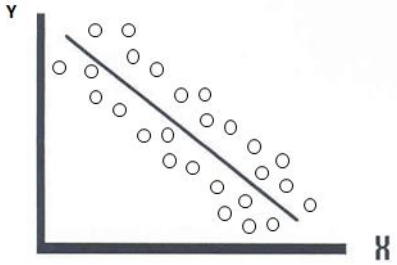
الشكل الأول (ب) ارتباط عكسي تام (سالب)

الشكل الثاني :

أما إذا كانت النقاط تأخذ شكل خط مستقيم ولكن لا تقع جميعها على الخط قيل أن العلاقة خطية (موجبة أو سالبة) كما في الشكل الثاني أ، ب.



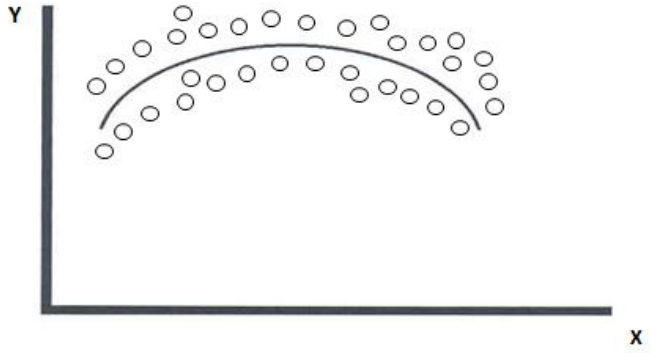
الشكل الثاني (أ)
ارتباط موجب قوي
(ارتباط خطي طردي)



الشكل الثاني (ب)
ارتباط سالب قوي
(ارتباط خطي عكسي)

الشكل الثالث :

وإذا كانت العلاقة تأخذ شكل منحنى فإن الارتباط لا يكون خطياً "ارتباط غير خطي " Non Linear Correlation كما في الشكل الثالث :

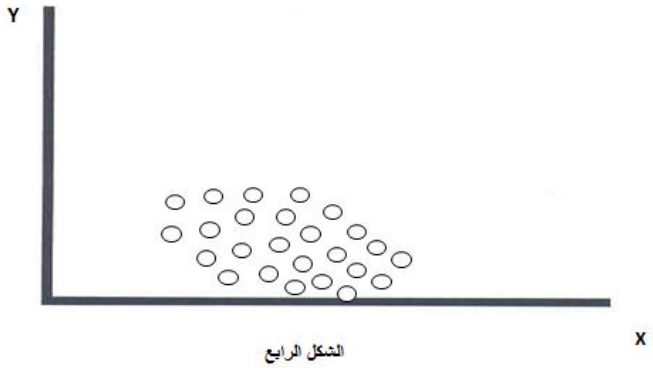


الشكل الثالث

(ارتباط غير خطي)

الشكل الرابع :

أما إذا كانت النقاط تتبعثر بدون نظام معين فإن ذلك يدل على عدم وجود علاقة بين المتغيرين (أو أن العلاقة بينهما ضعيفة جداً) كالعلاقة مثلاً بين دخل الشخص وطوله كما في الشكل الرابع :



الشكل الرابع

(لا توجد علاقة)

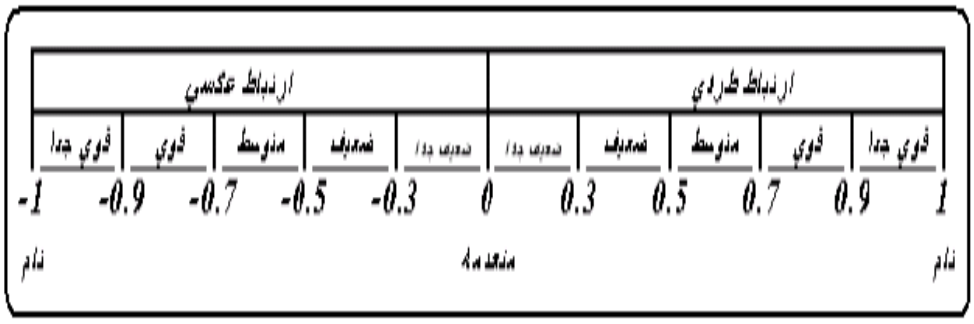
ثانيا: معامل الارتباط Correlation Coefficient :

يقاس الارتباط بين متغرين بمقياس إحصائي يسمى " معامل الارتباط " ويعكس هذا المقياس درجة أو قوة العلاقة بين المتغيرين واتجاه هذه العلاقة. وتنحصر قيمة معامل الارتباط بين + ١، − ١.

- **فإذا كانت قيمة معامل الارتباط تساوي + ١ فمعنى ذلك أن الارتباط بين المتغيرين طردي تام، وهو أقوى أنواع الارتباط الطردي بين متغيرين.**
- **وإذا كانت قيمة معامل الارتباط تساوي − ١ فمعنى ذلك أن الارتباط بين المتغيرين عكسي تام، وهو أقوى أنواع الارتباط العكسي بين متغيرين.**
- **وإذا كانت قيمة معامل الارتباط تساوي صفر، فمعنى ذلك أنه لا يوجد ارتباط بين المتغيرين.**
- **وكلما اقتربت قيمة معامل الارتباط من + ١ أو − ١كلما كان الارتباط قوياً، وكلما اقترب من الصفر كلما كان الارتباط ضعيفاً.**

والخلاصة :

أنه كلما كانت العلاقة قوية بين المتغيرين كلما اقترب معامل الارتباط من + ١ أو – ١ فإذا وصلت قيمة المعامل إلى + ١ أو – ١ كان الارتباط تاماً بين المتغيرين. وأنه كلما كانت العلاقة ضعيفة بين المتغيرين كلما اقترب معامل الارتباط من الصفر، فإذا وصلت قيمة المعامل إلى الصفر كان الارتباط منعدماً بين المتغيرين. ومعنى ذلك أيضاً أنه لا يوجد ارتباط بين متغيرين تكون قيمة المعامل فيه أكبر من + ١ ولا أصغر من – ١. ويمكن تمثيل قوة العلاقة بالشكل التالي:



معامل بيرسون للارتباط الخطى البسيط

Simple Correlation

يفترض بيرسون Pearson أن المتغيرين كميان، وأن العلاقة بينهما خطية (أي تأخذ شكل خط مستقيم، ويرى بيرسون أن أفضل مقياس للارتباط بين متغيرين قد يختلفان في وحدات القياس و / أو في مستواهما العام (مثل الارتباط بين العمر والدخل) حيث يقاس العمر بالسنوات ويقاس الدخل بالعملة، بالريال أو الدولار.. كما أن المستوى العام للعمر – أي متوسط العمر – قد يساوي أربعين عاماً. فبينما المستوى العام – أي متوسط – الدخل الشهري قد يكون خمسة آلاف ريال مثلاً.

وبالتالي فإن أفضل مقياس للارتباط بين مثل هذين المتغيرين – حسب رأي بيرسون – هو عن طريق حساب انحرافات كل من المتغيرين عن وسطه الحسابي وقسمة هذه الانحرافات على الانحراف المعياري لكل منهما، فنحصل على ما يسمى بالوحدات المعيارية لكل متغير.

ويكون معامل ارتباط بيرسون هو " متوسط حاصل ضرب هذه الوحدات المعيارية ". ومعامل الارتباط يكون بدون تمييز.

إن الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (رو)، وفي حالة العينة بالرمز r ، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة كتقدير لمعامل الارتباط في المجتمع.

ويتم حساب معامل ارتباط بيرسون من خلال العلاقة التالية:

$$r = \frac{\sum XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sqrt{\left(\sum X^2 - \frac{\left(\sum X\right)^2}{n}\right)\left(\sum Y^2 - \frac{\left(\sum Y\right)^2}{n}\right)}}$$

حيث :

$\sum XY$ **تعني مجموع حاصل ضرب كل قيمة من** $\mathbf{X}$ **في** $\mathbf{Y}$ **.**

$\left(\sum X\right)$ **تعني مجموع قيم المتغير** $\mathbf{X}$ **.**

$\left(\sum Y\right)$ **تعني مجموع قيم المتغير** $\mathbf{Y}$ **.**

$\sum X^2$ **تعني مجموع مربع قيم المتغير** $\mathbf{X}$ **.**

$\left(\sum X\right)^2$ **تعني مربع مجموع قيم المتغير** $\mathbf{X}$ **.**

$\sum Y^2$ **تعني مجموع مربع قيم المتغير** $\mathbf{Y}$ **.**

$\left(\sum Y\right)^2$ **تعني مربع مجموع قيم المتغير** $\mathbf{Y}$ **.**

n **عدد قيم الدراسة (عدد الأزواج المطلوب حساب الارتباط بينها) .**

مثال :

رغبة إحدى الشركات معرفة العلاقة بين عدد ساعات العمل لموظفيها ومستوى الإنتاجية لهم ، فقاموا بجمع معلومات عن هذا الموضوع وحصلوا على النتائج التالية :

الموظفين	أ	ب	ج	د	هـ	و	ز	ح	ط	ي
ساعات العمل $\mathbf{X}$	٨	٢	٨	٥	١٥	١١	١٣	٦	٤	٦
مستوى الإنتاجية $\mathbf{Y}$	٣	١	٦	٣	١٤	١٢	٩	٤	٤	٥

المطلوب: حساب معامل ارتباط بيرسون للبيانات السابقة .

الحل :

لكون بيانات كلا المتغيرين فترية، ولرغبة الشركة معرفة العلاقة بين المتغيرين موضع الدراسة، فإن أنسب أسلوب إحصائي هنا هو معامل ارتباط بيرسون، ولغرض حساب معامل ارتباط بيرسون من خلال الدرجات الخام وتطبيق المعادلة الآنفة الذكر، فإننا نتبع الخطوات التالية:

- ✓ **نربع كل درجة في المتغير** $\mathbf{X}$ **لكي نحصل على** $\mathbf{X^2}$ **.**
- ✓ **نربع كل درجة في المتغير** $\mathbf{Y}$ **لكي نحصل على** $\mathbf{Y^2}$ **.**
- ✓ **نضرب درجات المتغير** $\mathbf{X}$ **في درجات المتغير** $\mathbf{Y}$ **لكي نحصل على** $\mathbf{XY}$
- ✓ **نقوم بعد ذلك بعرض البيانات المتحصل عليها في جدول كالتالي وذلك حتى يسهل علينا استخلاص بعض القيم المطلوبة لحساب معامل ارتباط بيرسون من خلال الدرجات الخام .**

الموظفين	ساعات العمل $\mathbf{X}$	مستوى الإنتاجية $\mathbf{Y}$	$\mathbf{X^2}$	$\mathbf{Y^2}$	$\mathbf{XY}$
أ	٨	٣	٦٤	٩	٢٤
ب	٢	١	٤	١	٢
ج	٨	٦	٦٤	٣٦	٤٨
د	٥	٣	٢٥	٩	١٥
هـ	١٥	١٤	٢٢٥	١٩٦	٢١٠
و	١١	١٢	١٢١	١٤٤	١٣٢
ز	١٣	٩	١٦٩	٨١	١١٧
ح	٦	٤	٣٦	١٦	٢٤
ط	٤	٤	١٦	١٦	١٦
ي	٦	٥	٣٦	٢٥	٣٠
المجموع	٧٨	٦١	٧٦٠	٥٣٣	٦١٨

نطبق معادلة معامل ارتباط بيرسون كالتالي :

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

$$r = \frac{618 - \frac{(78)(61)}{10}}{\sqrt{\left(760 - \frac{(78)^2}{10}\right)\left((533) - \frac{(61)^2}{10}\right)}} = \frac{618 - 475.8}{\sqrt{(760 - 608.4)(533 - 372.1)}}$$

$$= \frac{142.2}{\sqrt{(151.6)(160.9)}} = \frac{142.2}{\sqrt{24392.44}} = \frac{142.2}{156.2} = 0.91$$

وهذه النتيجة توضح أن درجة الارتباط = ٠.٩١ وهذه النتيجة تعتبر مؤشر على علاقة إيجابية قوية بين ساعات العمل ومستوى الإنتاجية.

ملاحظة مهمة :

من خواص معامل بيرسون للارتباط الخطي أنه لا يتأثر بالعمليات الحسابية التي تجري على المتغيرين x , y. بمعنى أنه لا يتأثر بالطرح (أو الجمع)، ولا بالقسمة (أو الضرب). أي إذا طرحنا (أو جمعنا) قيمة معينة من كل قيم x وقيمة أخرى من كل قيم y، أو قسمنا (أو ضربنا) قيم x على قيمة معينة وكل قيم y على قيمة أخرى فإن قيمة معامل الارتباط لا تتغير أي نحصل على القيمة نفسها.

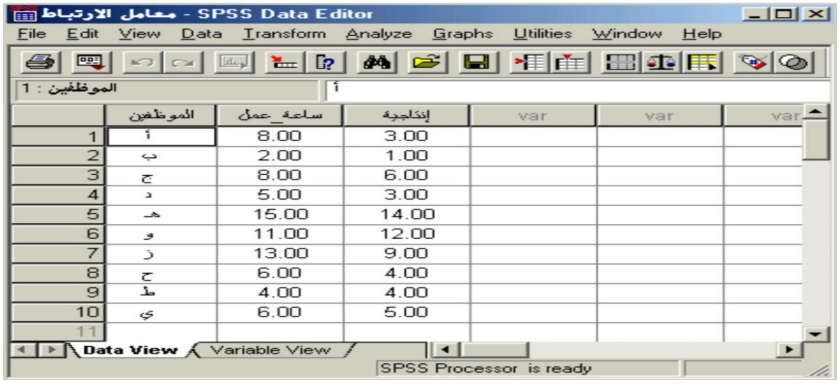
حساب معامل ارتباط بيرسون

Pearson Correlation Coefficient

من خلال برنامج الـ SPSS

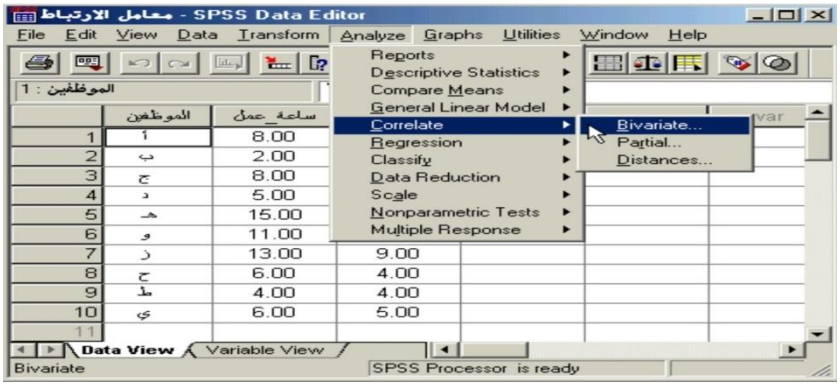
لغرض حساب قيمة معامل ارتباط بيرسون لنفس المثال السابق من خلال استخدام برنامج الـ SPSS نتبع الخطوات التالية :

قم بإدخال البيانات المراد تحليلها من خلال شاشة تحرير البيانات Data Editor بالطريقة المناسبة كالتالي :

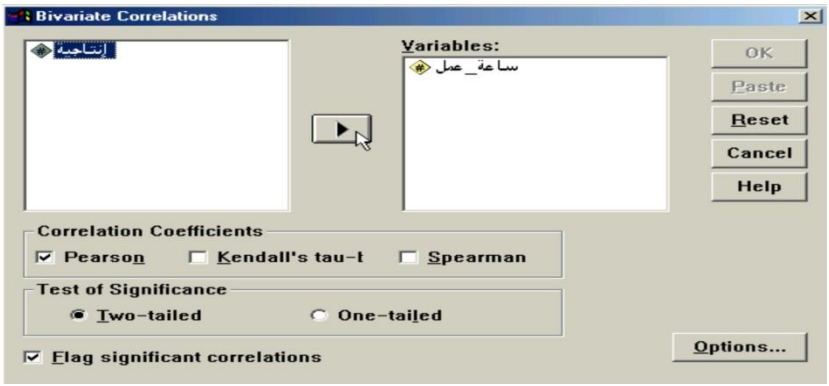


لاحظ أنه تم إدخال بيانات كل متغير في عمود منفصل عن الآخر، وقد تم إعطاء كل متغير الاسم المناسب له .

من القائمة "تحليل" **Analyze** اختر الأمر "الارتباط" **Correlate** فتظهر قائمة أوامر فرعية اختر منها "بسيط أو فردي" **Bivariate** كالتالي :



✓ بعد اختيار الأمر "بسيط أو فردي" **Bivariate** سوف يظهر لك صندوق الحوار التالي:



✓ من قائمة المتغيرات في الجهة اليسرى من صندوق الحوار حدد المتغيرين المراد حساب قيمة الارتباط لهما ، ونقلها إلى المستطيل الخاص بـ "المتغيرات " **Variables** ، ثم بعد ذلك أنقر على السهم الذي يظهر مقابل المستطيل الخاص بـ "متغيرات الدراسة" ، ستلاحظ انتقال هذه المتغيرات مباشرة إلى المستطيل "المتغيرات" **Variables** . بعد ذلك اختر نوع معامل الارتباط المراد حسابه (حيث يظهر في اسفل صندوق الحوار هذا ثلاث أنواع من معامل الارتباط وهي معامل ارتباط بيرسون **Pearson** ومعامل ارتباط كاندل **Kendall** ومعامل ارتباط سبيرمان **Spearman**) وسوف نختار هنا معامل ارتباط بيرسون **Pearson** . كذلك يستطيع المستخدم من خلال صندوق الحوار هذا تحديد ما إذا كان المختبر الإحصائي بذيل واحد **One-tailed** أو بذيلين **Tow-tailed**

✓ أنقر على زر "خيارات" **Options** في الجهة السفلية اليمنى من صندوق الحوار السابق وذلك عند الرغبة في حساب الخصائص الأساسية للمتغيرات موضع الدراسة **Statistics** ، وكذلك كيفية التعامل مع القيم المفقودة **Missing Values** ، وبعد الانتهاء من التعديل على هذا الصندوق الحواري أنقر على زر "استمرار" **Continue** .

✓ أنقر بعد ذلك على زر "موافق" **OK** سيؤدي ذلك إلى تنفيذ الاختبار، وستلاحظ ظهور النتائج في شاشة المخرجات كالتالي :

➡ **Correlations**

Descriptive Statistics

	Mean	Std. Deviation	N
ساعة_عمل	7.8000	4.10420	10
إنتاجية	6.1000	4.22821	10

Correlations

		ساعة_عمل	إنتاجية
ساعة_عمل	Pearson Correlation	1	.910**
	Sig. (2-tailed)	.	.000
	N	10	10
إنتاجية	Pearson Correlation	.910**	1
	Sig. (2-tailed)	.000	.
	N	10	10

**. Correlation is significant at the 0.01 level

نلاحظ أن برنامج الـ **SPSS** قام مباشرة بحساب الإحصاءات الأساسية للبيانات مثل المتوسط الحسابي والانحراف المعياري للمتغيرات ثم قام البرنامج بحساب معامل ارتباط بيرسون **Pearson Correlation** بين المتغيرين ساعات العمل ومستوى الإنتاجية والذي يساوي **0.91** وكذلك مستوى الدلالة لهذا الارتباط والذي يبدوا أنه دال عند مستوى دلالة **0.001** مما يدل على وجود علاقة إيجابية وقوية بين المتغيرين ساعات العمل في الشركة ومستوى إنتاجية الموظفين، فإننا بالتالي نرفض الفرضية الصفرية، أي توجد علاقة بين المتغيرين .

والأوامر المستخدمة لحساب معامل ارتباط بيرسون (**r**) **Pearson Correlation Coefficient** من خلال برنامج الـ **SPSS** (مع اختلاف المتغيرات موضع الدراسة) هي:

```
CORRELATIONS
/VARIABLES=ساعة_عمل إنتاجية
/PRINT=TWOTAIL NOSIG
/STATISTICS DESCRIPTIVES
/MISSING=PAIRWISE .
```

معامل ارتباط الرتب

Rank Correlation

قد يرغب الباحث في حساب معامل الارتباط بين رتب المتغيرين وليس بين القيم ذاتها، فقد يكون المتغيران وصفيين ترتيبين **Ordinal** أو أن يكون أحد المتغيرين كمياً بينما الأخر وصفياً ترتيبياً، أو أن يكون المتغيران كميين، ويكون اهتمام الباحث منصباً على الرتب أكثر من القيم. ففي انتخابات مجلس الشيوخ أو النواب الأمريكي مثلاً، يعتبر المرشح الأول هو من حصل على أعلى الأصوات بغض النظر عن عددها، والذي يحصل على عدد أصوات أقل منه مباشرة هو الثاني.. وهكذا.

فإذا كانت رتب المتغيرين تسير في الاتجاه نفسه: بمعنى أن الرتب الأعلى للمتغير الأول تناظرها رتب أعلى للمتغير الثاني كانت العلاقة طردية بينهما. وإذا كانت الرتب الأعلى للمتغير الأول تناظرها رتب أدنى للمتغير الثاني كانت العلاقة بينهما عكسية. ففي مثالنا السابق عن العلاقة بين دخل الناخب وعمره، كان الناخب الأكبر عمراً (بصفة عامة)هو الأعلى دخلاً، فمن الواضح أن العلاقة بينهما طردية، أما إذا كان الناخب الأكبر عمراً (بصفة عامة) هو الأقل مشاركة في العمل السياسي فإننا في هذه الحالة نكون أمام علاقة عكسية.

ولحساب معامل ارتباط الرتب هناك طرق مختلفة أهمها معاملي سبيرمان وكيندال.

حساب معامل سبيرمان لارتباط الرتب

Spearman rank Correlation Coefficient

لحساب معامل سبيرمان لارتباط الرتبةنوم بترتيب كل من المتغيرين ترتيباً تصاعدياً أو تنازلياً أما تصاعدياً لكلا المتغيرين أو تنازلياً لكليهما). وفي حالة الترتيب التصاعدي تأخذ أقل قيمة من قيم المتغير الرتبة رقم ١، والقيمة الأعلى منها مباشرة الرتبة رقم ٢ وهكذا (بالنسبة لكل من المتغيرين). أما في حالة الترتيب التنازلي تأخذ أكبر قيمة من قيم المتغير الرتبة رقم ١، والقيمة الأقل منها مباشرة الرتبة رقم ٢ وهكذا (بالنسبة لكل من المتغيرين). وعند تساوي قيمتين (أو أكثر) من قيم المتغير نعطي كل قيمة رتبة مختلفة (كما لو كانت القيم غير متساوية) ثم نحسب متوسط هذه الرتب، ويعطى هذا المتوسط لكل من هذه القيم المتساوية.

وبعد ترتيب المتغيرين نحسب الفروق بين رتب كل من المتغيرين (ونرمز للفروق بالرمز **d**) ثم نقوم بتربيع هذه الفروق ونحصل على مجموعها أي نحصل على $\sum d^2$ ثم نعوض في معامل سبيرمان لارتباط الرتب والذي يأخذ الشكل التالي :

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2 - 1)}$$

حيث : $\sum d^2$ هو مجموع مربعات الفروق بين رتب المتغيرين

n هي عدد أزواج القيم.

مما سبق نستطيع إجمال بعضاً من الملاحظات فيما يلي :

١ – مجموع الفروق بين الرتب يساوي صفر.

٢ – أن قيمة معامل ارتباط الرتب تنحصر بين – ١، + ١ فإذا كانت الرتبة رقم ١ للمتغير الأول تناظرها الرتبة ١ للمتغير الثاني، والرتبة ٢ للمتغير الأول تناظرها الرتبة رقم ٢ للمتغير الثاني، وهكذا.. فإن معامل ارتباط الرتب يساوي + ١ (ارتباط طردي تام بين الرتب).

وإذا كانت الرتبة رقم ١ (أقل رتبة) للمتغير الأول تناظرها أعلى رتبة للمتغير الثاني وهكذا.. فإن معامل ارتباط الرتب يساوي – ١ (ارتباط عكسي تام بين الرتب).

٣ – كذلك نلاحظ أن مجموع الرتب لكل من المتغيرين تساوي $\frac{n(n+1)}{2}$

مثال:

البيانات التالية تمثل إجابات عينة من سبعة أشخاص حول برامج الضمان الاجتماعي، ومدى ملاءمتها لحاجات الناس.

السؤال الأول	جيدة	مقبولة	ممتازة	جيدة	جيدة جداً	مقبولة	جيدة
السؤال الثاني	جيدة جداً	مقبولة	جيدة جداً	جيدة	جيدة	جيدة	ممتازة

٣

والمطلوب: حساب معامل سبيرمان لارتباط الرتب بين هذين السؤالين ؟

الحل :

ننظم البيانات في الجدول التالي مع ملاحظة ما يلي :

١ – بالنسبة للسؤال الأول، فإن التقدير الأعلى سيحصل على الرتبة رقم 1 والأقل منه مباشرة سيحصل على الرتبة رقم 2 وهكذا.. أي أن الترتيب تنازلي. ونكرر العمل نفسه مع السؤال الثاني.

٢ –عند حصول إجابتين أو أكثر على التقدير نفسه نعطي لكل إجابة مبدئياً رتبة كما لو كانوا مختلفين ثم نحسب متوسط هذه الرتب، وهذا المتوسط هو الذي يعطى لكل إجابة.

٣ – ثم نحسب الفروق بين رتب السؤالين ونرمز لها بالرمز d ثم نربع هذه الفروق فنحصل على d^2 ونعوض في القانون عن $\sum d^2$ مع ملاحظة أن n = 7.

السؤال الأول X	السؤال الثاني Y	رتب X	رتب Y	d الفرق بين الرتب	d^2 مربعات الفرق
جيدة	جيدة جداً	4	2.5	1.5	2.25
مقبولة	مقبولة	6.5	7	- 0.5	0.25
ممتازة	جيدة جداً	1	2.5	- 1.5	2.25
جيدة	جيدة	4	5	- 1.0	1.00
جيدة جداً	جيدة	2	5	- 3.0	9.00
مقبولة	جيدة	6.5	5	1.5	2.25
جيدة	ممتازة	4	1	3.0	9.00
المجموع				**Zero**	**26.0**

$$r_s = 1 - \frac{6\,(\sum d^2)}{n\,(n^2-1)} = 1 - \frac{6(26)}{7\,(49-1)}$$

$$= 1 - \frac{156}{336} = 1 - 0.46$$

$$r_s = 0.54$$

وهذا يعني أن الارتباط بين إجابات المبحوثين بالنسبة للسؤالين هو ارتباط طردي متوسط. وبالتالي فليس بالضرورة أن يكون رأي المجيبين في برامج الضمان الاجتماعي تعني ملاءمتها لحاجات الناس.

مثال:

البيانات التالية تمثل أعداد الساعات التي ذاكرها عشرة طلاب والدرجات التي حصلوا عليها في امتحان أحد المقررات :

عدد الساعات X	10	6	12	14	11	6	19	16	3	9
الدرجات y	60	48	83	76	74	58	98	89	37	69

المطلوب: أحسب معامل سبيرمان لارتباط الرتب.

الحل :

كما في المثال السابق ننظم الحل في الجدول التالي مع ملاحظة أن n = 10

عدد الساعات X	الدرجات Y	رتب X	رتب Y	الفروق d	d^2
10	60	6	7	- 1	1.00
6	48	8.5	9	- 0.5	0.05
12	83	4	3	1	1.00
14	76	3	4	- 1	1.00
11	74	5	5	0	0
6	58	8.5	8	0.5	0.25
19	98	1	1	0	0
16	89	2	2	0	0
3	37	10	10	0	0
9	69	7	6	1	1.00
المجموع				Zero	**4.50**

وبالتعويض في القانون حيث $\sum d^2 = 4.5$ **، n = 10 نحصل على :**

$$r_s = 1 - \frac{6(\sum d^2)}{n(n^2-1)} = 1 - \frac{6(4.5)}{10(1-99)} = 1 - \frac{27}{990} = 1 - 0.027$$

$$r_s = 0.973$$

مما يعني أننا أمام علاقة طردية قوية بين المتغيرين. فكلما زادت عدد الساعات التي يدرسها الطالب في هذا المثال، كلما زادت درجاته في الامتحانات قوة وذلك بما نسبته % 97.

مع أصدق الأمنيات

أخوكم / ثابت

@thabet_18

المحاضرة الثانية عشر

تحليل الانحدار

يعتبر تحليل الانحدار أكثر طرق التحليل الإحصائي استخداما، حيث يتم من خلاله التنبؤ بقيمة احد المتغيرات (المتغير التابع) عند قيمة محددة لمتغير أو متغيرات أخرى (المتغيرات المستقلة).

وتسمى العلاقة الرياضية التي تصف سلوك المتغيرات محل الدراسة والتي من خلالها يتم التنبؤ بسلوك احد المتغيرين عند معرفة الآخر بمعادلة خط الانحدار.

وهناك صورتان أساسيتان لمعادلة الانحدار وهما:

الصورة الأولى: معادلة انحدار **y | x** (التى يطلق عليها معادلة انحدار **y** على **x**)

الصورة الثانية: معادلة انحدار **x | y** (التى يطلق عليها معادلة انحدار **x** على **y**)

معادلة انحدار y على x

وهى تلك المعادلة التى يطلق عليها معادلة انحدار **y | x**. أى تتحدد قيمة المتغير **y** تبعا لقيمة المتغير **x** لذلك يمكن التعبير عن تلك العلاقة الخطية بالمعادلة التالية:

$$\hat{y} = b_0 + b_1 x$$

حيث يسمى b_0 ثابت الانحدار او الجزء الثابت او الجزء المقطوع من محور الصادات بينما b_1 يطلق عليها معامل الانحدار أو معدل التغير فى الدالة .

ولتحديد المعادلة الدالة على العلاقة بين المتغيرين **y** و **x** لابد من تقدير قيمة للثابتين b_0 و b_1 الذين يمكن تقديرهما من خلال تطبيق طريقة المربعات الصغرى كما يلى:

تقوم نظرية المربعات الصغرى على تدنية مجموع مربعات الأخطاء فى التقدير إلى اقل حد ممكن.

ويمكن استخدام المعادلات التالية لحساب معامل الانحدار باستخدام طريقة المربعات الصغرى:

$$b_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n}$$

$$= \bar{y} - b_1 \bar{x}$$

مثال

عند دراسة العلاقة بين عدد غرف المسكن وكمية الكهرباء المستهلكة بالألف كيلو وات فكانت كما يلي:

المطلوب أوجد:

عدد الغرف x	9	7	10	5	3	7	8	10	4	6
استهلاك كهرباء y	12	9	14	6	4	7	10	10	5	8

١. **معادلة انحدار y على x ؟**

٢. **تحديد معدل التزايد أو التناقص فى استهلاك الكهرباء؟**

٣. **ما هو الاستهلاك المتوقع لمسكن مكون من ٨ غرف؟**

الحل:

نقوم بعمل الجدول التالي:

x	y	xy	x^2	y^2
9	12	108	81	144
7	9	63	49	81
10	14	140	100	196
5	6	30	25	36
3	4	12	9	16
7	7	49	49	49
8	10	80	64	100
10	10	100	100	100
4	5	20	16	25
6	8	48	36	64
69	85	650	529	811

من خلال الجدول السابق يمكن تقدير معادلة انحدار x على y كما يلي:

اولاً – يتم تقدير قيمة معامل الانحدار b_1

$$b_1 = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{n\sum x^2 - \left(\sum x\right)^2} = \frac{10(650) - (69)(850}{10(529) - (69)^2} = 1.2003$$

وبالتالي يكون معدل التزايد في استهلاك الكهرباء هو لأنها موجبة ويساوى 1.2003 أي أن كل غرفة بالمسكن تعمل على زيادة استهلاك الكهرباء بمقدار 1200.3 كيلو وات.

ثانياً – تقدير قيمة b_0

$$b_0 = \frac{\sum y}{n} - b_1 \frac{\sum x}{n}$$

$$\bar{x} = \frac{\sum x}{n} = \frac{69}{10} = 6.9 \quad , \quad \bar{y} = \frac{\sum y}{n} = \frac{85}{10} = 8.5$$

$$b_0 = \bar{y} - b_1\bar{x} = 8.5 - (1.2003)(6.9) = 0.21793$$

معادلة انحدار x على y هي:

$$\hat{y} = b_0 + b_1 x$$

$$\hat{y} = 0.21793 + 1.2003x$$

الاستهلاك المتوقع لمسكن مكون من ٨ غرف:

يتم التعويض فى معادلة الانحدار التى سبق إيجادها عندما تكون x = ٨ كما يلي:

$$\hat{y} = 0.21793 + 1.2003(8) = 9.8203$$

أي أن الاستهلاك المتوقع لمسكن مكون من 8 غرف هو 9820.3 كيلو وات

معادلة انحدار x على y

وهى التى يطلق عليها معادلة انحدار x | y. أى تتحدد قيمة المتغير x تبعا لقيمة المتغير y لذلك يمكن التعبير عن تلك العلاقة الخطية بالمعادلة التالية:

$$\hat{x} = c_0 + c_1 y$$

حيث يسمى c_0 ثابت الانحدار او الجزء الثابت بينما c_1 يطلق عليها معامل الانحدار أو معدل التغير في الدالة

ولتحديد المعادلة الدالة على العلاقة بين المتغيرين x و y لابد من تقدير قيمة للثابتين c_0 و c_1 الذين يمكن تقديرهما من خلال تطبيق طريقة المربعات الصغرى فتكون النتيجة كما يلي:

$$c_1 = \frac{n\sum xy - \sum x \sum y}{n\sum y^2 - (\sum y)^2}$$

$$c_0 = \frac{\sum x}{n} - c_1 \frac{\sum y}{n}$$

$$= \bar{x} - c_1 \bar{y}$$

مثال

باستخدام بيانات المثال السابق لعدد الغرف واستهلاك الكهرباء، وعلى اعتبار هو المتغير المستقل وعدد الغرف هو المتغير التابع، أوجد التالي:

عدد الغرف x	9	7	10	5	3	7	8	10	4	6
استهلاك كهرباء y	12	9	14	6	4	7	10	10	5	8

المطلوب أوجد:

١. **معادلة انحدار x على y ؟**
٢. **ماهو عدد الغرف المتوقع لأستهلاك 25000 كيلو وات ؟**

الحل:

نقوم بعمل الجدول التالي:

y^2	x^2	xy	y	x
144	81	108	12	9
81	49	63	9	7
196	100	140	14	10
36	25	30	6	5
16	9	12	4	3
49	49	49	7	7
100	64	80	10	8
100	100	100	10	10
25	16	20	5	4
64	36	48	8	6
811	529	650	85	69

من خلال الجدول السابق يمكن تقدير معادلة انحدار x على y كما يلي:

أولا– يتم تقدير قيمة معامل الانحدار c_1

$$c_1 = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{n\sum y^2 - \left(\sum y\right)^2} = \frac{10((650) - (69)(85)}{10(811) - (85)^2} \cong 0.718$$

ثانياً – تقدير قيمة c_0

$$c_0 = \frac{\sum x}{n} - c_1 \frac{\sum y}{n}$$

$$c_0 = \bar{x} - c_1\bar{y} = 6.9 - 0.718(8.5) \cong 0.797$$

معادلة انحدار x على y هي:

وبالتالي تكون معادلة انحدار x **على** y

$$\hat{x} = c_0 + c_1 y$$

$$\hat{x} = c_0 + c_1 y \quad \Rightarrow \quad \hat{x} = 0.797 + 0.718y$$

إذا كان الاستهلاك للمنزل ٢٥٠٠٠ كيلوات

فإن عدد الغرف المتوقعة هو:

يتم التعويض في معادلة الانحدار التي سبق إيجادها عندما تكون y = ٢٥ كما يلي:

$$\hat{x} = 0.797 + 0.718(25) = 18.747 \cong 19$$

أي أنه إذا كان استهلاك الكهرباء في احد المنازل ٢٥٠٠٠ كيلوات فإن عدد الغرف المتوقع في هذا المنزل = 19غرفة تقريبا.

المحاضرة الثالثة عشر

قياس الثبات والصدق للاختبارات والمقاييس

أولا: قياس الثبات للاختبارات والمقاييس

معنى الثبات :

إذا أجرى اختبار ما على مجموعة من الأفراد ورصدت درجات كل فرد فى هذا الاختبار ثم أعيد إجراء نفس هذا الاختبار على نفس هذه المجموعة ورصدت أيضا درجات كل فرد ودلت النتائج على أن الدرجات التى حصل عليها الطلاب فى المرة الأولى لتطبيق الاختبار هى نفس الدرجات التى حصل عليها هؤلاء الطلاب فى المرة الثانية ، نستنتج من ذلك أن نتائج الاختبار ثابتة تماماً لأن نتائج القياس لم تتغير فى المرة الثانية بل ظلت كما كانت قائمة فى المرة الأولى

ثبات الاداة Reliability

- **درجة الاتساق في قياس السمة موضوع القياس من مرة لاخرى فيما لو اعدنا تطبيق الاداة عددا من المرات (يسمى دقة القياس)**
- **يعبر عن الثبات بصورة كمية يطلق عليها معامل الثبات تتروح بين صفر والواحد الصحيح (٠-١)**
- **كلما زادت قيمة المعامل دلت على ان الاداة تتمتع بثبات مرتفع والعكس صحيح)**

اخطاء تؤثر على الثبات بشكل اساسي:

- **اخطاء القياس المنتظمة والتي تعود الى اداة القياس كأن تكون صعبة جدا او سهلة جدا**
- **اخطاء القياس العشوائية والتي تعود للمفحوص نفسه كأن يكون مريض او غير مهتم**
- **الاختبار الصادق هو اختبار ثابت وليس كل اختبار ثابت هو اختبار صادق**

انواع الثبات

- **ثبات الاعادة Test – Retest Reliability**
- **ثبات الصورة المتكافئة Equivalent – Form Reliability**
- **الثبات بالطريقة النصفية Spilt-Half Reliability**
- **ثبات التكافؤ المنطقي Rationale – Equivalence Reliability**
- **ثبات المصححين Scorer / Rater Reliability**

ثبات الاعادة

- **يطبيق الاختبار على عينة ما**
- **يعطي الباحث مهلة اسبوع**
- **يعيد الباحث تطبيق نفس الاختبار على نفس العينة**
- **يقارن الباحث نتائج التطبيق الاول مع نتائج اعادة التطبيق**
- **اذا كانت متطابقة او متقاربة فان الاداة تتمتع بمعامل ثبات مرتفع**

ثبات الصور المتكافئة

- **اعداد صورتين متكافئتين لاداة ما**
- **يتم تطبيق الصورتين على عينة ما**
- **يتم حساب معامل الارتباط بين نتائج صورتي الاداة**
- **اذا كانت معامل الارتباط عالي فان الاداة تتمتع بمعامل ثبات مرتفع**

ثبات الطريقة النصفية (التجزءة النصفية)

- **يطبق الاختبار او الاداة مرة واحدة فقط**
- **تقسم فقرات الاختبار او اسئلته الى نصفين (الفقرات الفردية معا والزوجية معا)**
- **مثال : الفقرات ١،٣،٥،٧،٩،١١ معا ثم ٢،٤،٦،٨،١٠ تكون معا**
- **يقوم الباحث بحساب معامل الثبات باستخدام طريقة سبيرمان – براون Spearman – Brown أو معامل ارتباط بيرسون**
- **يمكن استخدام هذه الطريقة بسهولة في برنامج التحليل الاحصائي SPSS**
- **اذا كانت معامل الثبات عالي فان الاداة تتمتع بمعامل ثبات مرتفع**

ثبات التكافؤ المنطقي

- **حساب اتساق اجابات افراد العينة من فقرة الى اخرى**
- **تحسب من خلال استخدام معادلة كودر – ريتشاردسون KR-21**
- **اذا كانت معامل الثبات عالي فان الاداة تتمتع بمعامل ثبات مرتفع**

ثبات المصححين

- **حساب ثبات الاداة اذا كان هناك اكثر من مصحح او ملاحظ اشتركوا في التصحيح او جمع البيانات**
- **تحسب من خلال اعداد قائمة بدرجات كل مصحح على حده**
- **ثم يحسب معامل الارتباط بين قوائم المصححين هذه**
- **اذا كانت معامل الارتباط عالي فان الاداة تتمتع بمعامل ثبات مرتفع**

العوامل المؤثرة في الثبات

- **طول الاختبار او كثرة عدد فقراته : كلما زادت الفقرات زاد معامل الثبات (أن لا يزيد طول الأداة عن ٣٥ إلى ٤٥ فقرة)**
- **زمن الاختبار: كلما زاد زمن الاختبار زاد معامل الثبات (مع ملاحظة أن هذا الأمر قد يكون مناسبا للإختبارات التحصيلية لكن أدوات القياس فالأمر يختلف)**
- **تباين مجموعة الثبات(العينة): كلما كان افراد العينة متباينين كلما زاد معامل الثبات**
- **صعوبة الاختبار: يرتفع معامل الثبات اذا كان الاختبار متوسط الصعوبة (الاختبار الصعب او السهل يؤدي الى معاملات ثبات منخفضة)**

ما هو معامل الثبات المناسب

- **ان لا يقل معامل الثبات بشكل عام عن ٠.٦٠**
- **افضل معامل ثبات هو ما كان فوق الـ ٠.٩٠**
- **يلجأ الباحث احيانا الى التبرير اذا كان معامل الثبات لديه منخفض او ادراج دراسات سابقة استخدمت نفس الاداة وذكر معاملات الثبات التي احتسبوها.**

حساب معامل الثبات :

يحسب الثبات من خلال حساب معامل الارتباط وهو خير طريقة لمقارنة هذه الدرجات التى حصل علها الطلاب فى الاختبارين .

ويحسب معامل الثبات من العلاقة التالية :

$$\text{Re}liability = \frac{2(r)}{1+(r)}$$

وقيمة r لبيرسون يتم حسابها من العلاقة التالية :

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

طرق حساب معامل الثبات

طريقة إعادة الاختبار :

تقوم فكرة هذه الطريقة على إجراء الاختبار على مجموعة من الأفراد ثم إعادة إجراء نفس الاختبار على نفس مجموعة الأفراد بعد مضى فترة زمنية وهكذا يحصل كل فرد على درجة فى الإجراء الأول للاختبار وعلى درجة أخرى فى الإجراء الثاني للاختبار، وعندما نرصد هذه الدرجات ونحسب معامل ارتباط درجات المرة الأولى بدرجات المرة الثانية فأننا نحصل بذلك على معامل ثبات الاختبار .

مثال :

الجدول التالي يوضح درجات مجموعة من الطلاب فى اختبار تم إجراؤه على نفس الطلاب مرتين والمطلوب حساب قيمة معمل ثبات الاختبار ؟

درجة الاختبار الأول	٣	٥	٩	٨	٢
درجة الاختبار المعاد	٤	٦	٧	٤	٣

الحل :

نفترض أن درجات الاختبار الأول هى "x" ودرجات الاختبار الثانى هى "y" ثم نكون الجدول التالى :

x	y	xy	x2	y2
٣	٤	١٢	٩	١٦
٥	٦	٣٠	٢٥	٣٦
٩	٧	٦٣	٨١	٤٩
٨	٤	٣٢	٦٤	١٦
٢	٣	٦	٤	٩
٢٧	٢٤	١٤٣	١٨٣	١٢٦

حساب معامل الارتباط لبيرسون :

$$r = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{n}\right)\left(\sum Y^2 - \frac{(\sum Y)^2}{n}\right)}}$$

حساب معامل الارتباط لبيرسون :

$$r = \frac{[(5)(143)] - [(27)(24)]}{\sqrt{[(183) - \frac{(27)^2}{5}]}\sqrt{[(126) - \frac{(24)^2}{5}]}} = 0.668$$

إذا معامل الثبات يساوي:

$$\text{Re}liability = \frac{2(0.668)}{1 + (0.668)} = 0.8$$

ثانيا: قياس الصدق للاختبارات والمقاييس

صدق الاداة Validity

ان تقيس الاداة ما صممت لقياسه

انواع الصدق:

- **صدق المحتوى**
- **صدق المفهوم او صدق البناء**
- **الصدق العاملي**
- **الصدق التلازمي**
- **الصدق التنبؤي**

خطوات اجراء صدق المحتوى:

- **اعداد وتحليل محتوى الظاهرة محور الدراسة**
- **صياغة الفقرات**
- **عرض الفقرات ونتائج تحليلها على مجموعة من الخبراء في ميدان البحث لمعرفة مدى مناسبة الفقرات وسلامتها وانتمائها للظاهرة المقاسة**
- **احيانا يقوم الباحث باعداد كشف يتكون من درجات للخبراء لوضع تقيمهم عليه**

مثال : الفقرة مناسبة اللغة سليمة

١٠٩٨٧٦٥٤٣٢١ ١٠٩٨٧٦٥٤٣٢١

صدق المفهوم او البناء

قياس مفهوم افتراضي غير قابل للملاحظة مثل الذكاء او الدافعية

- **يضع الباحث فرضيات حول خصائص الافراد المقاسة**
- **يربط الباحث نتائج الاختبار بالفرضيات**
- **يبين فيما اذا كانت النتائج تدعم او ترفض فرضياته**
- **عندها يتضح صدق المفهوم او البناء**

الصدق العاملي Factor analysis

يستخدم في حساب صدق الأداة التي تنطوي على عوامل مرتبطة

مثل: الذكاء ينقسم إلى ذكاءات متعددة حسب جاردنر

الصدق التلازمي

مدى ارتباط الدرجات المحققة على الأداة بالدرجات المحققة على أداة اخرى تقيس نفس السمة

مثال: قام باحث بإعداد اختبار ذكاء ويريد حساب دلالات صدق هذا الاختبار

- **يقوم بتطبيق اختباره**
- **يقوم بتطبيق اختبار اخر من اختبارات الذكاء المعروفة**
- **يقوم بتبويب البيانات ووضعها في برنامج التحليل الاحصائي SPSS**
- **يقوم بحساب معامل ارتباط بيرسون بين الاختبارين**
- **اذا كان معامل الارتباط قوي بين الاختبارين وذو دلالة عندها نقول انه يوجد صدق تلازمي للاختبار**

الصدق التنبؤي

هو الدرجة التي يمكن من خلالها للمقياس ان يكون قادرا على التنبؤ بأداء معين (محك) في المستقبل

مثال: قدرة اختبارات الذكاء على التنبؤ بالتحصيل الأكاديمي المستقبلي للطلاب

مع أصدق الأمنيات

أخوكم / ثابت

@thabet_18