الفصل الثامن

الاختبارات الإحصائية

أشرنا فيما مضى إلى أهمية الفرضيات العلمية في منهجية البحث العلمي، وقلنا أن الفرضية هي عبارة عن تفسير أولي للظاهرة المدروسة، غير أن ذلك لا يعني بالضرورة أن كل التصورات الأولية التي يضعها الباحثون ويعتقدون أنها تفسر ظواهرهم هي فرضيات علمية، إذ يشترط حتى يتحول التصور المبدئي إلى فرضية علمية أن يكون قابلاً لاختبار مدى الصدق والثبات، فإذا اختبرت الفرضية وثبت صدقها، فإنها تتحول إلى تعميم يمكن الرجوع إليه لتفسير الظواهر المشابهة. أما إذا اختبرت الفرضية ولم يثبت صدقها فيجب تركها والبحث عن فرضية بديلة، يكون أمامها فرصة أكبر للنجاح.

وتشكل الفرضيات العلمية ركناً هاماً من الأسلوب العلمي المتبع في الجغرافيا وغيرها من مباحث العلوم الاجتماعية الأخرى. وأكثر الفرضيات استخداماً هي التي يضعها الباحثون عندما يستخدمون خصائص العينات لتقدير معالم المجتمعات الإحصائية التي أخذت منها تلك العينات، أو عندما يوازنون بين المعالم الإحصائية لمجتمعات متعددة، مستخدمين عينات مختارة من تلك المجتمعات، وهذا النوع من الفرضيات يرتبط بالمفاهيم الإحصائية الخاصة بدرجات الثقة ومستويات المعنوية، كما يرتبط بطبيعة البيانات من حيث توزيع العينة ونوعها وحجمها وطريقة اختيارها.

هذه الفرضيات وأمثالها هي الفرضيات القابلة للاختبار الإحصائي، ويضع الباحث في العادة فرضية مبدئية تسمى بفرضية العدم Null Hy- (Ho) pothesis أو الفرضية الصفرية، وهي تمثل في العادة عكس ما يتوقعه الباحث. وهي التي يتم في العادة اختبارها إحصائياً. فإذا تمكن الباحث من رفض تلك الفرضية، فهو يقبل في العادة الفرضية البديلة، وهي التي تمثل توقعات الباحث بمشكلة البحث (HI) والتي تتفق مع الأساس النظري للموضوع. أما إذا لم يتمكن الباحث من رفض الفرضية المبدئية أو فرضية العدم فلا يستطيع أن يقبل الفرضية كما لا يستطيع أن يصدر حكماً بشأنها.

وهناك نسبة من الخطأ المسموح به عند قبول أو رفض الفرضية. ويجب أن تحدد هذه النسبة أولاً: وتتراوح نسبة الخطأ بين ١-١٠٪ في العادة، ويختلف قبول الخطأ بحسب الدقة المطلوبة في البحث، ويعرف ذلك بمستوى المعنوية level of sigificanse، ويستخدم الجغرافيون مستوى معنوية ٠,٠٥ وأحياناً ٠,١٠ وبعض البحوث يشترط نسبة متدنية لا تتجاوز ٠,٠١ وأكثر مستويات المعنوية استخداماً من قبل الجغرافيين هي (٠,٠١) و(٠,٠٥).

إن اتخاذ القرار المناسب في قبول أو رفض الفرضية يعتمد على الاختبارات الإحصائية التي يجريها الباحث. والاختبارات الإحصائية تمكن الباحث من الحصول على قيمة اختبارية يحصل عليها من تطبيق اختبار معين، وهذه القيمة الاختبارية لا بد من مقارنتها بحدود دنيا لقبول الفرض الإحصائي، وهذه الحدود الدنيا تعرف: بالقيم الحرجة Critical Value، وهي التي تحدد منطقة الرفض، فإذا كانت النتيجة المتحصلة من الاختبار (القيمة المحسوبة) أكبر من القيمة الحرجة المرصودة في جداول خاصة بمعظم الفرضيات ـ فإننا نرفض الفرضية الصفرية أو فرضية العدم، ونقبل الفرضية البديلة والعكس بالعكس.

أنواع الاختبارات الإحصائية:

هناك نوعان من الاختبارات :

١ - النوع الأول : الاختبارات البارامترية (المعلمية) Parametric Tests وتستخدم في البيانات الكمية ، ومن أشهر أنواعها ثلاثة هي : اختبار ت(t)، واختبار ز(z)، واختبار ف (f) .

٢-النوع الثاني : الاختبارات غير بارامترية (غير معلمية) Non Parametric test وتستخدم في البيانات النوعية أو التصنيفية وهي عديدة ومتنوعة ومن أشهرها مربع كاي ($\times^{2}$) .

الاختبارات البارامترية:

تشترط الاختبارات البارامترية والأساليب الكمية المرتبطة بها ، توفر الخصائص التالية في بيانات المجتمع قيد الدراسة :

١ - أن يكون توزيع البيانات توزيعاً معتدلاً (متماثلاً) .

٢ - أن تكون المفردات المشاهدة أو الحالات (Cases) مستقلة عن بعضها البعض .

٣- أن يكون هناك تجانس بين المجتمعات موضوع المقارنة (التباين متساوي)

٤ - أن تكون البيانات المقاسة من نوع الفترة Interval Data .

فإذا لم تتوفر هذه الشروط فإن نتائج الاختبارات تكون غير مناسبة ، ولا بد من اللجوء إلى اختبارات أخرى مناسبة أو إجراء تعديلات على البيانات المرصودة .

١- الاختبار ز (z) واختبارات (t):

١- سبق وأن ذكرنا أن الاختبارات الخاصة بالعينات تفترض أن تكون بيانات العينة موزعة توزيعاً معتدلاً (متماثلاً)، وفي الغالب لا يكون توزيع العينة كذلك؛ لذا فإن من الواجب إجراء بعض التعديلات في البيانات ليتسنى لها الاقتراب من التوزيع المعتدل، ومن الجدير بالذكر أن بيانات العينة مهما كانت متماثلة فإنها لن تعكس تماماً خصائص المجتمع الذي سحبت منه، وبالتالي توجد بعض الفروق بين خصائص العينة ومعالم المجتمع الذي تمثله، والتي يمكن تقديرها على أساس احتمالات الخطأ السابق ذكرها (مستويات المعنوية).

٢- من المعلوم أننا لا نستطيع دائماً قياس جميع المفردات في المجتمع لمعرفة متوسطه الحسابي، ويستعاض عن متوسط المجتمع أحياناً بمتوسط عينة حجمها كبير، وحتى يكون متوسط العينة ممثلاً ومماثلاً لمتوسط المجتمع يجب أن يكون الفرق بين المتوسطين صغيراً حتى يتسنى لنا قبول الفرضية القائلة: بأن متوسط العينة يساوي متوسط المجتمع، وهو ما يمكن معرفته عن طريق الاختبارات الإحصائية.

٣- وبالمثل عند إجراء البحوث والدراسات الجغرافية تقابلنا الكثير من المشاكل التي تتطلب المقارنة بين متوسطي عينتين لمعرفة ما إذا كانت هاتان العينتان مسحوبتين من مجتمعين مختلفين ولهما المتوسط نفسه، أم مسحوبتين من المجتمع نفسه. فمثلاً قد نجد من الأفضل عملياً عند اختبار مدى فاعلية عامل معين أن نسحب عينتين:

الأولى: لتمثل المجتمع قبل تأثير هذا العامل، والثانية: لتمثيل المجتمع بعد تأثير العامل.

ونختبر ما إذا كان الفرق بين متوسط العينتين فرق جوهري أو غير جوهري. فإذا كان الفرق جوهرياً نستنتج فاعلية هذا العامل، أما إذا كان الفرق غير جوهري، فإننا نستنتج عدم فاعلية هذا العامل وإن الفرق قد يكون راجعاً للصدفة أو ناتجاً عن خطأ المعاينة. ولتوضيح ذلك نقول: إنه عند استخدام نوعين مختلفين من السماد لمعرفة ما إذا كان لهذين النوعين تأثير واضح على نوع معين من التربة، وبالتالي على الإنتاج الزراعي فيهما في منطقتين لهما الظروف نفسها؛ لهذا نسحب عينة من المجتمع الأول وعينة من المجتمع الثاني، ثم يحسب المتوسط الحسابي لكل عينة على حدة، ثم يجري الاختبار على قيم المتوسطين، فإذا كانت م١، م٢ هما متوسطا المجتمعين الأول والثاني على التوالي، وكانت $\bar{س}$ ١ – $\bar{س}$ ٢ متوسطا العينتين المسحوبتين من المجتمعين السابقين، فإن الفرق ($\bar{س}$١ – $\bar{س}$٢)هو متغير عشوائي للفرق بين($\bar{م}$١ – $\bar{م}$٢)، ويحدد الفرق بين المتوسطين بواسطة الوحدات المعيارية، أي بتحويل القيم الأصلية إلى وحدات معيارية ومعايرتها بوحدات معيارية نظرية، حتى نستطيع الحكم على أن هذا الفرق هو فرق جوهري أم لا، وهناك مجالات أخرى مشابهة منها على سبيل المثال: اختبار وجود فروق جوهرية بين مستوى كفاءة عمال الإنتاج في مصنعين مختلفين، أو اختبار وجود فروق جوهرية بين درجة صلابة نوعين من الصخور، فإذا ثبت أن الفرق بين كل نوعين هو فرق جوهري، فإن ذلك يكون دليلاً على أن العينتين مسحوبتان من مجتمعين مختلفين، أما إذا

ثبت أن الفرق غير جوهري، فإن ذلك يعني أن الفرق بين متوسطي العينتين يرجع لخطأ الصدفة أو لخطأ المعاينة . وأن العينتين مسحوبتان من مجتمع واحد، أو من مجتمعين لهما المتوسط الحسابي نفسه[1] .

٤- يستخدم اختبار (Z) للمقارنة بين المتوسط الحسابي للعينة والمتوسط الحسابي للمجتمع، عندما يكون الانحراف المعياري للمجتمع معروفاً، ويزيد حجم العينة عن ٣٠، أما الاختبار (t) فيستخدم عندما لا يكون الانحراف المعياري للمجتمع معروفاً، ويقل حجم العينة عن ٣٠، ومن الجدير بالذكر أن قيم (z) و (t) تتشابه إذا زاد حجم العينة عن ١٢٠ مفردة، وعندها يمكن استخدام أي منهما للاستدلال على صحة فرضيات المتوسطات، وهناك جداول خاصة بتوزيع ز (z) وأخرى خاصة بتوزيع ت (t) (انظر الملحق) تستخدم لمعرفة القيم الحرجة بغرض مقارنتها مع القيم المحسوبة .

٥- لم نذكر المعادلات الخاصة بحساب قيم اختبارات (ت) و (ز) نظراً لأن هذه الاختبارات تظهر أوتوماتيكياً عند عمل أي تحليلات بالحاسوب، وأصبح من النادر عملها يدوياً، وسنذكر لاحقاً كيفية حساب هذه الاختبارات من خلال الحاسب .

٢- تحليل التباين ***Analysis Of Variance:***

تتنوع استعمالات تحليل التباين لتشمل كافة ميادين الأعمال التجريبية,

---

(١) فتحي أبو راضي - مرجع سابق ص ٥٣٩ - ٥٤٢ .

فعلى سبيل المثال قد يريد الباحث معرفة تأثير أنواع مختلفة من الأسمدة على زيادة المحصول الزراعي، أو معرفة تأثير أنواع مختلفة من المبيدات الحشرية على هذا المحصول، أو معرفة مدى تأثير استخدام طرق مختلفة للري على رفع إنتاجية المحاصيل الزراعية، أو معرفة التأثير الحقيقي لأنواع مختلفة من الغذاء على زيادة وزن الماشية أو الدواجن.

وفي مجال الصناعة قد يريد الباحث معرفة مدى إقبال الناس في مناطق مختلفة على شراء بعض أنواع المنتجات الصناعية، وقد يريد معرفة تأثير طريقة تغليف البضاعة على رغبة الناس في شرائها.

وفي مجال العقار قد يرغب الباحث في معرفة الفروق بين متوسط أسعار الأراضي المخصصة للاستعمالات المختلفة (استعمالات سكنية، صناعية، زراعية، إلخ..) أو يود معرفة العوامل المؤثرة في الغطاء النباتي في منطقة ما. أو يريد معرفة أثر التكوين الجيولوجي على عدد الوديان في مناطق متجاورة.

إن الأمثلة السابقة توضح بعض استعمالات تحليل التباين الذي طوره العالم فيشر Fisher والذي يعرف أحياناً باسم اختبار ف أو F- Ratio Test نسبة إلى أول أحرف من اسم العالم فيشر الذي طوره، والذي أدى إلى تقدم سريع في علم الإحصاء. فقد أصبح أداة فعالة في الأبحاث المختلفة، وشاع استعماله في البحوث الجغرافية والبيولوجية والنفسية والتجارب الزراعية. وقد استفاد منه التربويون فائدة كبيرة في دراسة الأنواع المختلفة

للأساليب التعليمية على قطاعات مختلفة من الطلاب[1].

ويعتمد التحليل أساساً على حساب التباين بين العينات والتباين داخل كل العينات مجتمعة.

والمقياس المستخدم هو قيمة ف، وهي قيمة نظرية خاصة تؤخذ من جداول خاصة موضوعة لهذا الغرض عند مستوى دلالة يحددها الباحث. وهناك نوعان من تحليل التباين هما:

١- التحليل الأحادي التصنيف One Way Analysis of Variance.

٢- التحليل الثنائي التصنيف Two. Way Analysis Of Variance.

وفيما يلي شرح مفصل لكيفية حساب هذين النوعين:

تحليل التباين الأحادي التصنيف One Way Analysis of Variance :

في هذا النوع من التحليل يمكن تقسيم الاختلافات الكلية إلى مصدرين هما: الاختلافات التي ترجع إلى قياسات العينات، والأخرى ترجع إلى الأخطاء التجريبية. وتتكون البيانات اللازمة لهذا التحليل من عدد من العينات المستقلة، وبكل عينة من هذه العينات مجموعة من القياسات، فعلى سبيل المثال: إذا كان لدينا أربعة أنواع من بذور القمح المحسنة بطرق مختلفة ونريد معرفة هل يختلف إنتاج هذه الأنواع المختلفة. نخصص لكل نوع مجموعة من قطع الأراضي المتساوية ولتكن ٥ قطع، وبذا نقوم بزراعة ٢٠ قطعة من الأرض بالأنواع الأربعة السابقة، ثم نسجل مقدار الناتج لكل

(١) مختار محمود الهانسي: مقدمة في طريق التحليل الإحصائي، دار النهضة العربية، بيروت، ١٩٨٤ ص ٣٢٩-٣٣٠ وانظر أيضاً:

عبدالرزاق البطيحي وآخرون: الإحصاء الجغرافي، مطبعة جامعة بغداد ١٩٧٩، ص ٢٠١ وما بعدها.

قطعة من القطع السابقة . فإذا كان الناتج من كل قطعة من القطع المخصصة للزراعة على النحو الموجود في الجدول التالي رقم (٨-١)، وإذا أردنا معرفة ما إن كان هناك ثمة فروق في إنتاج أنواع البذور المحسنة أم لا، لا بد من إجراء تحليل التباين على النحو التالي :

١- ننشئ جدولاً يضم العينة السابقة الموجودة في جدول رقم (٨-١)، ثم نقوم بتربيع هذه القيم ونحسب المجموع الكلي لها . كما في جدول (٨-٢) .

جدول رقم (٨-١)

كمية الإنتاج بالكيلو غرام لأربعة أنواع من البذور المحسنة

النوع الأول من البذور العينة الأولى س ١	النوع الثاني من البذور العينة الثانية س٢	النوع الثالث من البذور العينة الثالثة س٣	النوع الرابع من البذور العينة الرابعة س٤
٢٦	١٨	١٢	١١
٣٣	٢٥	١٩	١٥
٣٢	٢٤	١٦	١٣
٢٦	٢٨	١٧	١٢
٣٣	٢٥	١٦	١٤

٢- لا بد من وضع فرض إحصائي لاختباره، والفرض الإحصائي في هذه الحالة هو فرض العدم الذي ينص على أنه «لا توجد فروق بين العينات الإحصائية الداخلة في التحليل » و في حالة مثالنا السابق يكون الفرض : « لا يوجد اختلاف في مقدار الإنتاج بين الأنواع الأربعة من بذور القمح المحسنة بطرق مختلفة» .

٣- لاختبار الفرضية السابقة لابد من عمل الآتي:

١- حساب عامل التصحيح Correction Factor وهو عبارة عن (مجـ س)^{٢} ÷ن ويتم ذلك بجمع مفردات العينات المختلفة مجـ س١ + مجـ س ٢ + مجـ س ٣ + مجـ س ٤ = ١٥٠ + ١٢٠ + ٨٠ + ٦٥ = ٤١٥

عامل التصحيح = (٤١٥)^{٢} ÷ ٢٠ = ٨٦١١,٢٥

جدول رقم (٨-٢)

مجموع قيم الإنتاج ومربعاتها لأربعة أنواع من البذور المحسنة من القمح

	س١	(س١)^{٢}	س٢	(س٢)^{٢}	س٣	(س٣)^{٢}	س٤	(س٤)^{٢}
	٢٦	٦٧٦	١٨	٣٢٤	١٢	١٤٤	١١	١٢١
	٣٣	١٠٨٩	٢٥	٦٢٥	١٩	٣٦١	١٥	٢٢٥
	٣٢	١٠٢٤	٢٤	٥٧٦	١٦	٢٥٦	١٣	١٦٩
	٢٦	٦٧٦	٢٨	٧٨٤	١٧	٢٨٩	١٢	١٤٤
	٣٣	١٠٨٩	٢٥	٦٢٥	١٦	٢٥٦	١٤	١٩٦
مجـ	١٥٠	٤٥٥٤	١٢٠	٢٩٣٤	٨٠	١٣٠٦	٦٥	٨٥٥
ن	٥		٥		٥		٥	
سَ	٣٠		٢٤		١٦		١٣	

٢- حساب المجموع الكلي للمربعات، ويساوي إجمالي مربعات كل القيم في الأربع عينات مطروحاً منه عامل التصحيح، أي يساوي:

= ٤٥٥٤ + ٢٩٣٤ + ١٣٠٦ + ٨٥٥ - ٨٦١١,٢٥

= ٩٦٤٩ - ٨٦١١,٢٥ = ١٠٣٧,٧٥

٣- حساب مجموع المربعات بين المجموعات (العينات) ويساوي متوسط مربع مجموع كل مجموعة مطروحاً منه عامل التصحيح.

$١ \div ٥ [(١٥٠)^{٢} + (١٢٠)^{٢} + (٨٠)^{٢} + (٦٥)^{٢}] - ٨٦١١,٢٥$

= ٩٥٠٥ - ٨٦١١,٢٥ = ٨٩٣,٧٥

٤- حساب مجموع المربعات داخل المجموعات (العينات) وتساوي الفرق بين رقم (٢) ورقم (٣)، أي يساوي المجموع الكلي للمربعات - مجموع المربعات بين العينات.

= ١٠٣٧,٧٥ - ٨٩٣,٧٥ = ١٤٤ .

٥- حساب درجات الحرية للمكونات السابقة:

درجات الحرية بين العينات = عدد العينات -١ = ٤-١ = ٣

درجة الحرية الكلية = عدد المفردات -١ = ٢٠-١=١٩

درجات الحرية داخل العينات = عدد المفردات - عدد العينات ٢٠-٤=١٦ وهي تساوي أيضاً درجات الحرية الكلية - درجات الحرية بين العينات.

٦- حساب التباين بين العينات $= \dfrac{\text{مجموع المربعات بين العينات}}{\text{درجة الحرية بين العينات}} = \dfrac{\text{٨٩٣,٧٥}}{\text{٣}} = \text{٢٩٧,٩٢}$

٧- حساب التباين داخل العينات $= \dfrac{\text{مجموع المربعات داخل العينات}}{\text{درجة الحرية داخل العينات}} = \dfrac{\text{١٤٤}}{\text{١٦}} = \text{٩}$

٨- حساب نسبة ف $= \dfrac{\text{التباين الأكبر}}{\text{الأصغر}} = \dfrac{\text{التباين بين العينات}}{\text{التباين داخل العينات}} = \dfrac{\text{٢٩٧,٩٢}}{\text{٩}} = \text{٣٣,١}$

٩- يمكن تلخيص البيانات السابقة على شكل جدول تحليل التباين على الصورة التالية :

جدول رقم (٨-٣)

تحليل التباين لمتوسط الإنتاج لأربعة أنواع من البذور المحسنة

مصدر التباين	مجموع المربعات	درجة الحرية	التباين	قيمة ف
بين العينات	٨٩٣,٧٥	٤ - ١ = ٣	٢٩٧,٩٢	
داخل العينات	١٤٤,٠٠	٢٠ - ٤ = ١٦	٩	٣٣,١
المجموع	١٠٣٧,٥	٢٠ - ١ = ١٩		

١٠- استخدام جدول «ف» (F Tables) وهو عبارة عن جدول لحساب قيمة التباين بدرجات الحرية بين العينات وداخل العينات وبمستوى معنوية ٠,٠٥ أو ٠,٠١ ، وفي هذا الجدول تكون درجات الحرية الأفقية خاصة بدرجات الحرية للتباين الأصغر، ودرجات الحرية الرأسية خاصة بدرجات التباين الأكبر.

وفي المثال الذي بين أيدينا نجد أن نسبة ف لدرجة الحرية (١٦) بين العينات (ذات التباين الأكبر)، و (٣) داخل العينات ذات التباين الأصغر عند مستوى دلالة ٠,٠٥ هي ٣,٢٤ .

١١- إذا كانت قيمة ف المحسوبة أقل من نظيرتها في الجدول حسب مستوى المعنوية أو درجة الثقة المطلوبة فإن الفرض يصبح مقبولاً ، بمعنى أنه لا توجد فروق بين متوسطات العينات. أما إذا كانت ف المحسوبة أكبر من نظيرتها في جدول التوزيع (ف) فإن الفرض يرفض، بمعنى أن هناك فروقاً جوهرية بين متوسطات العينات، وفي مثالنا السابق نجد أن قيمة ف المحسوبة = ٣٣,١ وف في الجدول = ٣,٢٤ وعلى هذا نرفض الفرضية القائلة بعدم وجود فروق في مقادير الإنتاج بين عينات القمح المحسنة، ونستدل على ذلك أن أنواع البذور المحسنة أعطت مقادير من الإنتاج متفاوتة.

مثال آخر:

يرغب أحد الباحثين في المقارنة بين متوسطات الزيادة في وزن نوع معين من الأبقار يتم تغذيتها بثلاثة أنواع مختلفة من الغذاء، أ، ب، جـ،

وقد قام هذا الباحث باختيار (١٥) رأساً بطريقة عشوائية من كامل القطيع لمعرفة أثر أنواع العلف الثلاثة عليها؛ لذا خصص الباحث خمسة أبقار وأعطاها النوع (أ)، واختار خمسة أخرى وأعطاها العلف من نوع (ب)، وقام بتغذية الأبقار الخمسة الأخرى بالنوع (جـ)، وبعد فترة من الزمن قام بتسجيل الزيادة في وزن كل منها على النحو الموضح بالجدول رقم (٨-٤)، ويريد الباحث معرفة ما إذا كانت هذه الأنواع المختلفة من العلف لها تأثيرات مختلفة على تسمين الأبقار أم لا؛ لذا فهدف الباحث هو مقارنة متوسطات الزيادة في الوزن لهذه المجتمعات الثلاثة ويستخدم في ذلك أسلوب تحليل التباين في اتجاه واحد. وفي هذه الحالة تسمى أنواع الغذاء بالمعالجات Treatments.

جدول رقم (٨-٤)

أوزان ثلاث مجموعات من الأبقار بعد إعطائها أنواعًا مختلفة من الغذاء

المجموعة الأولى أعطيت العلف (أ)	المجموعة الثانية أعطيت العلف (ب)	المجموعة الثالثة أعطيت العلف (جـ)
١٢	١٣	١٨
١٠	١٦	١٨
٩	١٥	١٧
١١	١٥	١٩
٨	١٤	١٩

المطلوب باستخدام تحليل التباين معرفة مدى وجود اختلافات بين أنواع الغذاء المختلفة على أوزان الأبقار الخاضعة للتجربة.

الحل :

نبدأ بفرض العدم وهو أنه لا توجد اختلافات جوهرية بين الأنواع المختلفة من العلف التي تقدم للأبقار .

لاختبار الفرضية السابقة لا بد من عمل الحسابات التالية بعد أن نكون الجدول التالي رقم (٨-٥) الذي يظهر القيم ومربعاتها والمجموع الكلي لها .

جدول رقم (٨-٥)

مجموع أوزان الأبقار ومربعات قيمتها

	س١	س١^{٢}	س٢	س٢^{٢}	س٣	س٣^{٢}
	١٢	١٤٤	١٣	١٦٩	١٨	٣٢٤
	١٠	١٠٠	١٦	٢٥٦	١٨	٣٢٤
	٩	٨١	١٥	٢٢٥	١٧	٢٨٩
	١١	١٢١	١٥	٢٢٥	١٩	٣٦١
	٨	٤٦	١٤	١٩٦	١٩	٣٦١
مجـ	٥٠	٥١٠	٧٣	١٠٧١	٩١	١٦٥٩
ن	٥		٥		٥	
سَ	١٠		١٤,٦		١٨,٥	

١- حساب معامل التصحيح ويساوي (مجـ س)^{٢} ÷ ن حيث مجـ س = مجـ س$_{١}$ + مجـ س$_{٢}$ + مجـ س$_{٣}$

مجـ س = ٥٠ + ٧٣ + ٩١ = ٢١٤

ن = ١٥

∴ معامل التصحيح = (٢١٤)^{٢} ÷ ١٥ = ٣٠٥٣ .

٢- حساب المجموع الكلي للمربعات ويساوي إجمالي مربعات كل القيم مطروحاً منه عامل التصحيح .

= ٥١٠ + ١٠٧١ + ١٦٥٩ - ٣٠٥٣

= ٣٢٤٠ - ٣٠٥٣ = ١٨٧

٣- حساب مجموع المربعات بين العينات ، ويساوي متوسط مربع مجموع كل مجموعة مطروحاً منه عامل التصحيح .

= ١ ÷ ٥ [(٥٠)^{٢} + (٧٣)^{٢} + (٩١)^{٢}] - ٣٠٥٣

= ٣٢٢٢ - ٣٠٥٣ = ١٦٩

٤- مجموع المربعات داخل المجموعات = المجموع الكلي للمربعات - مجموع المربعات بين العينات

= ١٨٧ - ١٦٩ = ١٨

٥- درجات الحرية

درجة الحرية بين العينات = عدد العينات -١ = ٣ - ١ = ٢

درجة الحرية الكلية = عدد المفردات - ١ = ١٥-١ = ١٤

درجة الحرية داخل العينات = عدد المفردات - عدد العينات

= ١٥ - ٣ = ١٢

٦ - حساب التباين بين العينات = $\frac{\text{مجموع المربعات بين العينات}}{\text{درجة الحرية بين العينات}}$ = $\frac{١٦٩}{٢}$ = ٨٤,٥

٧- حساب التباين داخل العينات = $\frac{\text{مجموع المربعات داخل العينات}}{\text{درجة الحرية داخل العينات}}$ = $\frac{١٨}{١٢}$ = ١,٥

٨- حساب نسبة ف = $\frac{\text{التباين الأكبر}}{\text{التباين الأصغر}}$ = $\frac{٨٤,٥}{١,٥}$ = ٥٦,٣

٩- يمكن تلخيص البيانات في الجدول التالي :

جدول رقم (٨-٦)

تحليل التباين لمتوسط أوزان الأبقار لثلاثة أنواع من الغذاء

مصدر التباين	مجموع المربعات	درجة الحرية	التباين	قيمة ف
بين العينات	١٦٩	٢	٨٤,٥	
داخل العينات	١٨	١٢	١,٥	٥٦,٣
المجموع	١٨٧٥	١٤		

١٠- من خلال جدول قيم ف وبدرجة حرية ٢ و ١٢ فإن قيمة ف = ٣,٨٨ في مستوى معنوية ٠,٠٥ . ولما كانت قيمة ف المحسوبة أكبر من القيمة الجدولية إذاً نرفض فرضية العدم بأنه لا توجد فروق بين أنواع الغذاء المختلفة، أي أن هناك فروقاً جوهرية بين الطرق المختلفة للتغذية ظهرت آثارها على أوزان الأبقار المختلفة خلال مدة التسمين.

تحليل التباين الثنائي ***Two - way Analysis of Variance***

كنا نفترض عند تحليل التباين الأحادي أن مفردات العينات متماثلة تماماً، والواقع أنها ليست كذلك، ففي حالة المثال السابق الخاص بتغذية الأبقار بأنواع مختلفة من الأعلاف نجد أن هناك اختلافات داخل عينة الأبقار التي تتغذى على نوع معين من العلف، فبعضها مثلاً يأكل أكثر من الآخر، وبعضها يتوعك أثناء التجربة ولو لفترة قصيرة. وهذا يعني أن المشاهدات تقع فيها أخطاء غير مقصودة وربما غير ظاهرة. وكذلك الحال في أنواع الأعلاف المقدمة فإننا نفترض التماثل التام في نوع العلف الواحد، والواقع إن هناك بعض الأخطاء التي تحدث أثناء التجارب والتي تؤثر بدورها على النتائج.

لقد حاول المختصون تدارك هذه المشكلات من خلال تحليل التباين الثنائي الذي يأخذ في الاعتبار الأخطاء التجريبية داخل العينات أو بين العينات، وبذلك يمكن تقسيم الاختلافات الكلية إلى ثلاثة مصادر هي:

١- اختلافات ترجع إلى العينات (القطاعات) ***Sum of Squares for Blocks (SSB)***.

٢- اختلافات ترجع إلى المعاملات (المشاهدات) ***Sum of Squares for Treatments (SSTR)***.

٣- اختلافات ترجع إلى الأخطاء التجريبية ***Sum of Squares for Errors (SSE)***

وتتكون البيانات هنا من عدد من العينات المستقلة لكل منها عدد من المفردات (المعاملات أو المشاهدات) ويوضح المثال التالي كيفية تطبيق تحليل التباين الثنائي وخطوات العمل اللازمة معها :

نفترض أننا سحبنا عينة عشوائية طبقية مكونة من ثلاث مجموعات من الأحواض الزراعية في زمام أحد المراكز الإدارية بهدف مقارنة الإنتاجية الزراعية بين أنواع التربة التي تمثلها هذه المجموعات وهي : التربة الطينية السوداء، والتربة الطينية الصفراء، والتربة الرملية . فإذا كانت كل مجموعة مكونة من عشرة أحواض تم اختيارها عشوائياً، ثم رصدت لها كميات الإنتاج لمحاصيل الحبوب بالطن كما في الجدول (٨-٧) فهل نقبل الفرض القائل بعدم وجود فروق بين متوسطات الإنتاج لهذه الأنواع الثلاثة من التربة [١].

١- مجـ س = مجـ س$_{١}$ + مجـ س$_{٢}$ + مجـ س$_{٣}$ = ٢٤٣ + ٢٢٢ + ٢١٠ = ٦٧٥

$$\text{٢- عامل التصحيح} = \frac{(\text{مجـ س})^{٢}}{\text{ن}} = \frac{(٦٧٥)^{٢}}{٣٠} \; ١٥١٨٧,٥$$

---

(١) فتحي عبدالعزيز أبو راضي : مقدمة الأساليب الكمية في الجغرافيا، دار المعرفة الجامعية، ١٩٨٣، ص ٥٦٥ - ٥٦٨ .

جدول رقم (٨-٧)

الإنتاجية الزراعية لمحاصيل الحبوب (بالطن) لمجموعة من الأحواض الزراعية في ثلاثة أنواع من التربة

التربة السوداء س ١	التربة الصفراء س ٢	التربة الرملية س ٣
٢٤	١٧	١٩
٢٧	٢٥	١٨
٢١	٢٤	٢٢
٢٢	١٩	٢٤
٢٦	٢٨	٢٣
١٩	٢١	١٨
٢٥	٢٠	٢١
٢٩	٢٥	١٩
٢٦	١٩	٢٥
٢٤	٢٤	٢١
مجـ س١ = ٢٤٣	مجـ س٢ = ٢٢٢	مجـ س٣ = ٢١٠

٣- مجموع المربعات بين العينات (المجموعات) ويساوي إجمالي مربع جميع القيم مقسوماً على عدد العينات (المجموعات) مطروحاً منه عامل التصحيح .

$$= \frac{٤٥٨٥٣}{٣} - ١٥١٨٧,٥ = ٩٦,٨٣$$

٤- مجموع المربعات بين المعاملات .

$$= ١ \div ١٠ = [(٢٤٣)^{٢} + (٢٢٢)^{٢} + (٢١٠)^{٢}] - ١٥١٨٧,٥$$

$$= ١٥٢٤٣,٣ - ١٥١٨٧,٥ = ٥٥,٨$$

٥- المجموع الكلي للمربعات ويساوي إجمالي مربعات كل القيم في العينات الثلاثة مطروحاً منه عامل التصحيح .

$$= ٥٩٨٥ + ٥٠٣٨ + ٤٤٦٦ - ١٥١٨٧,٥$$

$$= ١٥٤٨٩ - ١٥١٨٧,٥ = ٣٠١,٥$$

٦- نحسب الخطأ التجريبي لمجموع المربعات [البواقي] وهو عبارة عن الفرق بين المجموع الكلي للمربعات ومجموع المربعات بين العينات ومجموع المربعات بين المعاملات كما يلي :

$$٣٠١,٥ - ٩٦,٨٣ - ٥٥,٨ = ١٤٨,٨٧$$

٧- تحسب درجات الحرية كما يلي :

أ- درجات الحرية بين المعاملات = عدد المعاملات (الصفوف) -١

$$= ١٠ - ١ = ٩$$

ب- درجات الحرية بين العينات (الأعمدة) = عدد العينات - ١

$$= ٣ - ١ = ٢$$

جـ- درجات الحرية للمفردات الكلية ن = عدد المفردات - ١

= ٣٠-١=٢٩ .

د- درجات الحرية للخطأ (الفرق) = درجات الحرية للمفردات مطروحاً منه درجات حرية التباين بين العينات، وكذلك درجات الحرية بين المعاملات (الصفوف). = ٢٩ - ٢ - ٩ = ١٨

هـ- تقاس درجات الحرية للخطأ أيضاً عن طريق درجات الحرية بين العينات مضروباً في درجات الحرية بين المعاملات أو ٩ × ٢ = ١٨

ويمكن تلخيص النتائج السابقة في الجدول التالي :

جدول رقم (٨-٨)

تحليل التباين الثنائي

مصدر التباين	مجموع المربعات	درجة الحرية	التباين	قيمة ف
بين العينات	٩٦,٨٣	٢	٤٨,٤١	٥,٨٥
بين المعاملات	٥٥,٨٠	٩	٦,٢	
الخطأ	١٤٨,٨٧	١٨	٨,٢٧	
المجموع	٣٠١,٥٠	٢٩		

وبالبحث في جداول (ف) لإيجاد قيمتها النظرية عند درجات الحرية بين ٢ بين المجموعات و ١٨ للخطأ عند مستوى دلالة ٠,٠٥ نجد أنها=٣,٥٥، وحيث أن قيمة ف المحسوبة في المثال ٥,٨٥ أكبر من القيمة

النظرية عند مستوى الدلالة٠٥ , ٠ فإننا يجب أن نرفض الفرض القائل : إنه لا توجد فروق جوهرية بين متوسطات المجموعات الثلاثة من التربة من حيث الإنتاجية الزراعية للحبوب، وهذا يعني أن هناك احتمال ٩٥ , ٠ من أن الفروق بين الإنتاجية الزراعية لهذه الأنواع من التربة لم تحدث بفعل الصدفة، ولكنها فروق حقيقية لها دلالات إحصائية .

ولو أردنا معرفة ما إذا كان هناك فروق داخل العينات(أوبين المعاملات أو المشاهدات)نقوم بحساب التباين بين المعاملات الذي يساوي . . .

مجموع المربعات داخل العينات ــــــ الخطأ التجريبي = ٥٥ , ٨ ــــــ ٩ = ٦ , ٢

أما قيمة ف = التباين بين المعاملات ــــــ تباين الخطأ = ٦ , ٢ ــــــ ٨ , ٢٧ = ٠ , ٧٥

وبالبحث في جداول ف لإيجاد القيمة النظرية عند درجات حرية ٩ من المعاملات و ١٨ للخطأ نجد أنها ٤٦ , ٢ عند مستوى دلالة ٠٥ , ٠ . وبما أن القيمة المحسوبة لـ (ب) أقل من القيمة في الجدول نقبل الفرضية القائلة بأنه لا توجد فروق جوهرية بين المعاملات أو المشاهدات أو قطع الأراضي المخصصة لكل نوع من أنواع الحبوب .

٣- تحليل مربع كاي أو Chi square (X^2):

يستخدم مربع كاي في الأصل لاختبار مدى الأهمية الإحصائية للنتائج، غير أنه يمكن الانتفاع به لقياس مدى التركز أو التبعثر في

التوزيعات الجغرافية المختلفة. ويستعمل لقياس التطابق بين التوزيع الحقيقي لقيم الظاهرة المدروسة والتوزيع المتوقع لها وذلك من خلال استخدام صيغة رياضية إحصائية. وتتسع مجالات استخدام هذا التحليل لأنها تشمل اختبار العلاقات بين البيانات الاسمية Nominal Data مثل: صفة التدخين، أو الأمية، أو المرض، حيث لا يشترط مربع كاي البيانات الفئوية أو بيانات النسبة؛ ولذا يكثر استعماله في الدراسات التي تستخدم استمارات الاستبيان والتي تستعمل المتغيرات التي يطلق عليها Dummy Variables[١].

ويمكن إجمال أغراض هذا التحليل الإحصائي فيما يلي:

١- الحكم على مدى ملاءمة النموذج النظري للتوزيع الفعلي للبيانات. وبعبارة أخرى تحديد مدى انحراف التكرارات الفعلية عن التكرارت النظرية، وهنا يعالج مربع كاي التصنيفات الأحادية التي تم تقسيمها حسب طبيعة واحدة من الصفات.

٢- تحديد دلالة العلاقة بين مجموعتين أو أكثر من البيانات، استناداً إلى صفات معينة بينهما، وهذا ما يعرف بالتصنيف الثنائي.

فمثلاً إذا أخذنا عينة عشوائية من ١٠٠ شخص من عمال أحد المصانع ووجه إليهم سؤال فيما إذا كانوا يدخنون أم لا، ووجد أن ٤١ شخصاً يقبل على التدخين و ٥٩ شخصاً لا يدخنون، فهذا هو التصنيف الأحادي وفقاً

---

(١) مختار حمود الهانسي: مقدمة في طرق التحليل الإحصائي، دار النهضة العربية، بيروت، ١٩٨٤م ص ٢٨٥ - ٣٢٨ وانظر أيضاً:
بول هويل ترجمة بدرية شوقي عبدالرهاب ومحمد كامل الشربيني: المبادئ الأولية في الإحصاء، الطبعة الرابعة، دار جون وايلي، نيويورك، ١٩٨٤ ص ٢٤١ - ٢٥٥.

لظاهرة واحدة هي التدخين .

وإذا سألناهم أيضاً عن درجة التعليم ووجدنا أن ٧٠ منهم من فئة المتعلمين، و ٣٠ آخرين من فئة غير المتعليمن، كان ذلك أيضاً تصنيفاً أحادياً وفقاً لظاهرة التعليم، أما إذا ربطنا ظاهرة التعليم مع ظاهرة التدخين، فإن العلاقة هنا ثنائية، فإذا وجدنا أنه من بين الـ ٧٠ عامل المتعلمين هناك ١٥ عاملاً يقبلون على التدخين، ومن بين العمال غير المتعلمين وجدنا (٤) فقط غير المدخنين، تصبح العلاقة لظاهرتين هما التدخين والتعليم . وهنا يبدو السؤال هل هناك علاقة بين التعليم وانخفاض نسبة التدخين؟

يمكن تصوير المعلومات السابقة في الجداول التالية :

١- بالنسبة لظاهرة التدخين يمكن ترتيب البيانات على الشكل التالي :

جدول (٨-٩) تصنيف العمال بموجب التدخين

الصفة	العدد الفعلي
يدخن	٤١
لا يدخن	٥٩
الإجمالي	١٠٠

٢- بالنسبة لظاهرة التعليم يمكن ترتيب البيانات على النحو التالي :

جدول (٨-١٠) تصنيف العمال بموجب التعليم

الصفة	العدد
متعلم	٧٠
غير متعلم	٣٠
الإجمالي	١٠٠

٣- بالنسبة لظاهرة التعليم والتدخين يكون ترتيب البيانات على النحو التالي :

(٨-١١) تصنيف العمال بموجب التدخين والتعليم

الحالة التعليمية / التدخين	متعلم	غير متعلم	المجموع
مدخن	١٥	٢٦	٤١
لا يدخن	٥٥	٤	٥٩
الإجمالي	٧٠	٣٠	١٠٠

حساب مربع كاي :

١- تحسب قيمة مربع كاي من خلال المعادلة التالية :

$$\chi^{2} = \frac{\text{مجـ}\,(\text{أ}-\text{ب})^{2}}{\text{ب}}$$

حيث إن :

χ^{2} = مربع كاي .

أ= التكرار الحقيقي أو العدد الفعلي لقيم الظاهرة المدروسة .

ب- التكرار المتوقع أو العدد النظري المتوقع لقيمة الظاهرة المدروسة .

مجـ = مجموع القيم .

٢- لمربع كاي فرض إحصائي وفرض بديل، والفرض الإحصائي هو نقطة البداية في معظم الاختبارات الإحصائية، ويصاغ لغرض تطبيق اختبار إحصائي معين حول المشكلة تحت التحليل. ومتوقع عموماً عدم قبول الفرض الإحصائي ليحل محلها الفرض البديل .

الفرض الإحصائي : لا يوجد اختلاف مهم وجوهري بين التوزيع الحقيقي المشاهد وبين التوزيع النظري المتوقع (فرضية العدم Null Hypothesis) .

الفرض البديل : يوجد اختلاف مهم وجوهري بين التوزيع الحقيقي المشاهد وبين التوزيع النظري المتوقع .

يرفض الفرض الإحصائي إذا كانت قيمة مربع كاي المحسوبة أكبر من قيمة مربع كاي الموجودة في الجداول الخاصة بمربع كاي (انظر الملحق) أما إذا كانت قيمة مربع كاي المحسوبة أقل قيمة من قيمة مربع كاي الموجودة في الجدول، فتقبل الفرضية الإحصائية (فرضية العدم) .

٣- لابد عند تحديد قيم مربع كاي النظرية من الجداول الخاصة بها من إيجاد ما يعرف بدرجات الحرية، التي تحدد على أساسها قيم مربعات كاي، وهي تحسب على أساس ن - ١ حيث ن عدد المجموعات . فإذا كان عدد المجموعات ٥ فإن درجة الحرية تبلغ ٥-١ = ٤، وفي حالة التصنيفات الثنائية تحسب درجات الحرية كالتالي :

(عدد الأعمدة - ١) (عدد الأسطر - ١)

٤- بجانب درجات الحرية لا بد من تحديد مستوى المعنوية أو درجة الدقة المطلوبة أو نسبة الخطأ المسموح بها، ويتعامل الجغرافيون في الغالب مع مستوى المعنوية (∝) = ٠٥ , ٠، أو ٠١ , ٠، أو ١ , ٠ ، أي أن درجة الثقة تكون ٩٥٪ أو ٩٩٪ أو ٩٠٪ على التوالي .

التصنيف الأحادي:

١- في حالة التصنيف الأحادي لو كانت توقعاتنا أنَّ نصف عدد العمال هم من المدخنين والنصف الآخر من غير المدخنين فهل هناك فروق جوهرية بين ما توقعناه، وما تمت مشاهدته مع العلم بأن مستوى الدلالة ٠٥ , ٠؟

٢- يمكن ترتيب البيانات في هذا المثال كما هو موضح في الجدول التالي رقم (٨-١٢) .

جدول رقم (٨-١٢)

مربع كاي لعدد من المدخنين

الصفة	العدد الفعلي (أ)	العدد المتوقع (ب)	$(أ - ب)^2$	$\frac{(أ - ب)^2}{ب}$
مدخن	٤١	٥٠	٨١	١,٦٢
لا يدخن	٥٩	٥٠	٨١	١,٦٢
				٣,٢٤

٣- درجة الحرية = عدد الفئات - ١

= ٢ - ١ = ١

قيمة مربع كاي من الجدول = ٣,٨٤١ بمستوى دلالة (٠,٠٥).

٤- بما أن قيمة مربع كاي المحسوبة ٣,٢٤ أقل من القيمة الموجودة في الجدول تقبل فرضية العدم وهي أنه لا توجد فروق جوهرية بين التوزيع القائم والتوزيع المتوقع في صفة التدخين بين عمال المصنع المذكور.

٥- يمكن أن نتبع الخطوات السابقة نفسها لإظهارإن كان هناك فروق جوهرية في نسبة التعليم في المصنع أم لا (احسب هذه النسبة بنفسك).

مثال آخر:

أراد باحث جغرافي دراسة مقدار إنتاج الذرة في خمسة أنواع معينة من التربة. وفي نهاية الموسم حصل المزارع على كميات الذرة الموجودة في الجدول رقم (٨-١٣) فإذا كانت توقعاتنا أن مقادير الإنتاج ستكون متماثلة فهل نقبل فرضية العدم بأنه لا توجد فروق جوهرية بين أنواع التربة المختلفة؟ مع العلم أن مستوى الدلالة المطلوب هو ٠,٠٥

جدول رقم (٨-١٣)

حساب مربع كاي لمعرفة إنتاجية خمسة أنواع من الترب

نوع التربة	(أ) كمية الإنتاج الفعلي	كمية الإنتاج المتوقع (ب)	(أ-ب)٢	(أ-ب)٢ / ب
الرملية	٥٠	٥٠	صفر	صفر
الطينية السوداء	٧٠	٥٠	٤٠٠	٨
الجيرية	٤٥	٥٠	٢٥	٠,٥
الطينية الحمراء	٥٠	٥٠	صفر	صفر
تربة البودزول	٣٥	٥٠	٢٢٥	٤,٥
المجموع				١٣

وللإجابة على سؤال الباحث نطبق صيغة مربع كاي كما في الجدول السابق و تعادل هذه القيمة ١٣ .

وبما أن عدد الفئات ٥ فدرجات الحرية = ٥ - ١ = ٤ .

وللحكم على الفرضية نقارن قيمة مربع كاي المحسوبة مع القيمة النظرية المناظرة لأربع درجات حرية ومستوى دلالة ٠,٠٥ فنجد ف = ٤٨٨, .

وبما أن قيمة مربع كاي المحسوبة أكبر من القيمة الموجودة في الجدول

نرفض فرضية العدم، وهذا يعني أن هناك اختلافات واضحة بين إنتاجية الترب المختلفة.

التصنيف الثنائي:

في حالة التصنيف الثنائي يكون الهدف من مربع كاي إثبات أو نفي العلاقة بين التصنيفين القائمين على الصفتين المذكورتين، ففي مثال التدخين والتعليم تكون فرضية العدم هي أنه لا توجد علاقة بين التعليم من جهة والتدخين من جهة أخرى. ويطلق على ذلك اسم اختبار الاستقلال.

١ - فإذا رجعنا إلى مثال التدخين والتعليم الموجود في جدول رقم (٨-١١) نجد أن الذين يدخنون في المصنع عددهم ٤١ شخصاً منهم ١٥ متعلمين والباقي غير متعلمين (٢٦). أما الذين لا يدخنون فعددهم ٥٩ شخصاً منهم ٥٥ من المتعلمين والباقي من غير المتعلمين (٤). فإذا أراد الباحث معرفة ما إذا كان هناك علاقة بين إقبال العمال على التدخين وحالتهم التعليمية مع العلم أن مستوى الدلالة المطلوب ∝ = ٠,٠٥ .

٢ - إذا كان التعليم ليس له تأثير على التدخين يجب أن تكون نسبة المتعلمين مشابهة لنسبة المدخنين بمعنى أن من بين الـ ٧٠ شخصاً المتعلمين ٤١٪ منهم من المدخنين و ٥٩٪ من غير المدخنين. وكذلك الحال بالنسبة لغير المتعلمين فالمتوقع أن من مجموع الـ ٣٠ شخصاً غير متعلم ٤١٪ منهم من المدخنين و ٥٩٪ من غير المدخنين.

أي أن القيمة النظرية تحسب كالتالي:

بالنسبة للمتعلمين تكون القيمة المتوقعة أو النظرية:

$$\frac{٤١ \times ٧٠}{١٠٠} = ٢٨,٧٠$$

$$\frac{٥٩ \times ٧٠}{١٠٠} = ٤١,٣٠$$

بالنسبة لغير المتعلمين تكون القيمة النظرية للمدخنين

$$\frac{٤١ \times ٣٠}{١٠٠} = ١٢,٣$$

$$\frac{٥٩ \times ٣٠}{١٠٠} = ١٧,٧$$

وعلى هذا الأساس تكون القيمة المتوقعة على النحو الوارد في الجدول التالي رقم (٨-١٤)

جدول رقم (٨-١٤)

القيم المتوقعة بحسب التدخين والتعليم

	متعلم	غير متعلم	المجموع
مدخن	٢٨,٧	١٢,٣	٤١,٠
لا يدخن	٤١,٣	١٧,٧	٥٩,٠
المجموع	٧٠,٠	٣٠,٠	١٠٠

وهذا يعني أن القيمة الأولى في هذا الجدول رقم (٨-١٤) والتي تمثل المدخنين المتعلمين والتي تعادل ٢٨,٧ هي القيمة المتوقعة للقيمة الفعلية الموجودة في جدول رقم (٨-١١) الذي يمثل القيم الفعلية ومقدارها ١٥ .

وكذلك الحال مع بقية قيم هذا الجدول التي تمثل القيم المتوقعة لقيم الجدول السابق رقم (٨-١١) ذي القيم الحقيقية.

ولحساب مربع كاي نطرح القيمة المتوقعة من القيم الفعلية ثم تربع قيمة الناتج وتقسم على القيمة المتوقعة لكل خلية من خلايا الجدولين السابقين وبهذا تكون قيمة مربع كاي ($\times^{٢}$).

$$\frac{(١٥-٢٨,٧)^{٢}}{٢٨,٧} + \frac{(٢٦-١٢,٣)^{٢}}{١٢,٣} +$$

$$\frac{(٥٥-٤١,٣)^{٢}}{٤١,٣} + \frac{(٤-١٧,٧)^{٢}}{٧١,٧} = ٣٦,٩٤٧$$

وحيث إن درجة الحرية = (ع-١) (س-١).

أي إن عدد الصفوف -١ مضروباً في عدد الأعمدة -١

أي = (٢-١)(٢-١)=١

فإن قيمة مربع كاي من توزيعها في الجدول لدرجة حرية واحدة ويمستوى دلالة ٠,٠٥ = ٣,٨٤١

وبما أن قيمة مربع كاي المحسوبة أكبر من القيمة الموجودة في الجدول إذاً نرفض فرضية العدم التي تقول إن التعليم ليس له علاقة بالتدخين وتثبت الفرضية البديلة القائلة بوجود علاقة واضحة بين المستوى التعليمي وعدم الإقبال على التدخين.

الاختبارات الخاصة بتحديد حجم العينة:

لا يهتم الباحث بالعينة لذاتها، وإنما هي وسيلة لمعرفة مجتمع الدراسة من خلال هذه العينة. وبما أن الهدف هو مجتمع العينة، فإن تمثيل العينة لمجتمعها أمر هام جداً. فحجم العينة لا يرتبط بحجم المجتمع فقط بل يرتبط أيضاً بالدقة المطلوبة في المعلومات عن المجتمع. ويميز المختصون بين أرقام العينة وأرقام مجتمعها، فهم يستخدمون الحروف الرومانية للعينة، بينما يستخدمون الحروف اليونانية للمجتمع. ويطلقون على العمليات الإحصائية المرتبطة بتقديرات المجتمع اسم الإحصاءات الاستدلالية؛ لأنه يستدل منها في الغالب على خصائص المجتمع؛ ولهذا فنتائجها تقديرية وهناك احتمالات للخطأ لا يمكن تجاوزها، وكل ما يمكن عمله هو تقليل هذا الخطأ، وجعله معلناً.

لتوضيح ذلك نسوق مثالاً تطبيقاً: لنفرض أن لدينا مجتمعاً مكوناً من ٧٥ حالة (انظر جدول رقم ٨ : ١٥). فإذا فرضنا أن هذا المجتمع يمثل كمية الأمطار السنوية في مدينة ما لمدة ٧٥ عاماً. وقد قام ثلاثة باحثين مختلفين في دراسة هذه الظاهرة. فأخذ الأول عينة عشوائية لعشر سنوات والثاني درس عينة عشوائية مقدارها ٢٠ سنة. أما الثالث فقد اختار عينة تمثل ٥٠ سنة. والسؤال أي من هؤلاء الباحثين سيحصل على تقديرات صحيحة عن هذه الظاهرة؟ علماً بأن متوسط هذه الظاهرة هو ٢٣ , ٥١ ملم وقيمة الانحراف المعياري لهذه الأرقام يساوي ١٤ , ٢٤

جدول رقم (٨:١٥)

متوسط الأمطار السنوية في مدينة ما لمدة ٧٥ عاماً

٥٢	٣٦	٢٤	٩٦	٣٥	٣٥	٢٩	٧٨	٨٧	٩٦
٤٧	٢٣	٥٤	٤٥	٦٣	٢١	٥٠	٦٠	٤٧	٤٥
٦٢	٣٢	٥٢	١٤	٦٥	٩٨	٨٥	٢٣	٥٦	٥٤
٨٥	٣٦	٧٤	١٧	٩٦	١٢	٤٥	٧٨	٩٨	٦٥
٣٢	٣٢	٥٦	٤١	٩٥	٢١	٦٠	٢٥	٨٥	٤٦
١٣	٣٠	٤٥	٧٢	٤٢	٤٣	٦١	٩٤	٣٧	٣٧
٢٩	٧٤	٦٤	١٦	٤٢	٦٢	٤٥	١٨	٤٦	٧٤
٥٦	١٣	٤٦	٨٤	٣٥					

جدول رقم (٨-١٦)

المتوسط والانحراف المعياري والخطأ المعياري لثلاث عينات عشوائية

حجم العينة	المتوسط	الانحراف المعياري	الخطأ المعياري
١٠	٤٣	١٩,٣٥	٦,١٢
٢٠	٥٩,٢	٢٠,١	٤,٥
٥٠	٥٢,١	٢١,٨٤	٣,٠٩

إن الجدول رقم (٨-١٦) يظهر العديد من النقاط الهامة وهي :

١- إن متوسطات العينات جميعها لا تتساوى مع متوسط مجتمع الدراسة، فمتوسط العينة الأولى (٤٣) ، ومتوسط العينة الثانية (٥٩,٢)، أما الثالثة فهو (٥٢,١) وكل هذه المتوسطات لا تتساوى مع متوسط المجتمع المدروس والذي يبلغ ٥١,٢٣ . وهذا يعني أن إحصاءات العينة ليست قيماً صحيحة غير أن العينات الكبيرة تكون أقرب في متوسطها من المجتمع الأصلي من باقي العينات الأخرى.

٢- إن الانحراف المعياري للعينات يختلف عن الانحراف المعياري لمجتمع الدراسة فهو للعينة الأولى ١٩,٣٥ ، والثانية ٢٠,١٠، والثالثة ٢١,٨٤ ، وهذه الأرقام تختلف عن أرقام الانحراف المعياري لمجتمع الدراسة والبالغة ٢٤,١٤ .

٣- يرتبط بالانحراف المعياري قيمة الخطأ المعياري؛ لأن الانحراف المعياري هو الوسيلة الوحيدة التي تساعدنا في حساب قيمة الخطأ المعياري.

والخطأ المعياري يحسب من خلال المعادلة التالية

$$\text{الخطأ المعياري} = \frac{\text{الانحراف المعياري}}{\sqrt{\text{حجم العينة}}}$$

$$\text{خ} = \frac{\text{ع}}{\sqrt{\text{ن}}}$$

حيث:

خ = الخطأ المعياري

ع = الانحراف المعياري .

ن = حجم العينة .

وبحساب الخطأ المعياري للعينات الثلاثة من خلال المعادلة السابقة نجد أن الخطأ المعياري للعينة الأولى ٦,١٢ والثانية ٤,٥ والثالثة ٣,٠٩ ونلمس أنه كلما كان حجم العينة أكبر، كلما كان الخطأ المعياري أقل . وهذا يقودنا إلى القاعدة العامة أنه كلما كان حجم العينة كبيراً كانت ممثلة لمجتمعها بصورة أفضل . غير أن هذا الأمر ليس على إطلاقه فهناك حجم أفضل للعينة، وزيادة حجم العينة عن هذا الحجم قد يكون مضيعة للجهد والوقت والمال .

٤- الخطأ المعياري الثنائي التوزيع: يرتبط التوزيع الثنائي بالحالات التي لها نتيجتين فقط . فعند رمي قطعة نقود تكون النتيجة إما وجه (أ) أو وجه (ب) فقط . فإذا كان هدفنا الحصول على الوجه (أ)، فإنه عند ظهور (ب) يكون فشل (عدم حدوث الحدث) أما إذا كان الظاهر هو الوجه (أ) فإن النتيجة هي النجاح (حدوث الحدث). ومعنى ذلك أننا معنيون في الأساس باحتمالات النجاح .

تكون عملية اشتقاق الخطأ المعياري في هذا التوزيع بالصورة التالية :
الخطأ المعياري = الجذر التربيعي (لاحتمال أ × احتمال ب) مقسوماً على حجم العينة . ولا بد من تحويل الاحتمالات إلى نسبة مئوية، وذلك بضربها في ١٠٠، وبعد إيجاد قيمة الخطأ المعياري نعتمد ذات الأسس السابقة في تقدير معدل مجتمع العينة .

مثال:

في دراسة أجريت على عينات طلبة قسم الجغرافيا، وجد أن ٢٥٪ منهم يفضل التخصص في فرع من فروع الجغرافية الطبيعية. وقد كان حجم العينة ١٠٠ طالب ما هو الخطأ المعياري للعينة؟

$$\text{خ م} = \sqrt{\frac{\text{احتمال أ} \times \text{احتمال ب}}{\text{حجم العينة}}}$$

نحسب احتمال (أ) وهو نسبة الراغبين في الفروع الطبيعية $= \frac{٢٥}{١٠٠} = ٢٥٪$

نحسب احتمال (ب) وهو نسبة الراغبين في الفروع غير الطبيعية$= \frac{٧٥}{١٠٠} = ٧٥٪$

$$\therefore \text{خ م} = \sqrt{\frac{٢٥ \times ٧٥}{١٠٠}} = ٤,٣$$

نخلص إلى القول إنه ليس هناك قيماً حقيقة لمجتمع الدراسة من خلال عينتها، بل هناك تقديرات تحتمل الخطأ بنسب متفاوتة. و مهمة الباحث أن يقلل من قيمة هذا الخطأ عن طريق زيادة حجم العينة، وزيادة نسبة العينة إلى مجتمعها من جهة ثانية، والدقة المتناهية في اختبار مفردات العينة من جهة ثالثة.

تقدير حجم العينة: من المشكلات التي تعترض الباحث تقدير حجم العينة المطلوبة التي تعطي أفضل النتائج. وبذات الوقت تكون ممثلة لمجتمع

الدراسة تمثيلاً صادقاً بعيداً عن التحيز. ويمكننا الاستفادة من الموضوعات التي سبقت الإشارة إليها في تحديد الحد الأدنى المقبول للعينة الممثلة لمجتمعها، ويمكن ذكر الطرق التالية لتقدير حجم العينة.

١- تقدير حجم العينة من خلال تقديرات الخطأ المعياري: الخطأ المعياري هو مدى تباعد متوسط العينة عن متوسط المجتمع المأخوذة منه. وكلما كان متوسط العينة أقرب إلى متوسط المجتمع كلما كان الخطأ المعياري أصغر. وقد ذكرنا سابقاً أن الخطأ المعياري يعتمد على الانحراف المعياري للمجتمع. فإذا كان الانحراف المعياري للمجتمع معروفاً فليست ثمة مشكلة في معرفة الخطأ المعياري. غير أنه في الكثير من الحالات يكون الانحراف المعياري للمجتمع غير معروف، ويلجأ الاختصاصيون إلى اعتبار أن متوسط العينة يساوي متوسط المجتمع المأخوذة منه. وبالرغم من أن هذا ليس دقيقاً لأن متوسط العينة لا يساوي متوسط المجتمع، إلا أنه يتغاضى عن ذلك الخطأ أحياناً حينما لا توجد بدائل أخرى لتحديد حجم العينة ومن خلال معرفة متوسط العينة يمكن حساب الانحراف المعياري لها واعتباره الانحراف المعياري للمجتمع المدروس. ومن هذا المنطق لا بد أن تكون العينة أكثر من (٣٠)، لأنه إذا قلت عن ذلك لابد من استخدام أساليب إحصائية أخرى.

أ) تقدير حجم العينة من خلال الخطأ المعياري أحادي التوزيع: عرفنا أن معادلة الخطأ المعياري الأحادي التوزيع هي خ م $= \frac{ع}{\sqrt{ن}}$ ويمكن تحوير كتابة هذه المعادلة بالصيغة الجبرية التالية: هي

$$ن = \left(\frac{ع}{خ\ م}\right)^{٢}$$

حيث ن = حجم العينة

ع = الانحراف المعياري

خ م = الخطأ المعياري

مثال:

إذا كان الانحراف المعياري لمجموعة مساكن مدينة ما هو ٤,٨ وكان الخطأ المعياري يعادل ٠,٤ أوجد حجم العينة الممثلة لهذه المساكن.

الحل:

$$ن = \left(\frac{ع}{خ\ م}\right)^{٢} = \left(\frac{٤,٨}{٠,٤}\right)^{٢} = (١٢)^{٢} = ١٤٤$$

أي أن حجم العينة المطلوبة = ١٤٤ مسكناً

ب- تقدير حجم العينة من خلال الخطأ المعياري ثنائي التوزيع: لتقدير حجم العينة المطلوبة لا بد للباحث من القيام بعمل استكشافي عن طريق استبانة أولية للاستفادة منها في تقدير حجم العينة المناسبة. ومن الضروري أن يزيد حجم العينة الاستكشافية عن (٣٠) وذلك من أجل حساب قيمة الخطأ المعياري وتعتمد المعادلة التالية لحساب حجم العينة:

$$\text{حجم العينة} = \frac{(\text{احتمال أ} \times \text{احتمال ب})}{\text{مربع الخطأ المعياري}}$$

$$ن = \frac{أ \times ب}{(خ\ م)^{٢}}$$

حيث ن = حجم العينة

أ = احتمال وجود الظاهرة الأولى

ب = احتمال وجود الظاهرة الثانية

خ م = الخطأ المعياري

مثال:

درس أحد الباحثين مصادر المياه في المساكن هل هي من الشبكة العامة أم من خلال شراء التنكات (الوايتات)؟ واعتمد عينة أولية استكشافية مكونة من ٥٠ منزلاً، فوجد أن ٣٠ منزلاً يشربون من مياه الشبكة العامة. ما هو حجم العينة التي يجب اعتمادها بالدراسة الموسعة؟ علماً بأن قيمة الانحراف المعياري كانت تساوي ١٫١٤ .

الحل:

أولاً: نستخرج نسبة من يشربون من مياه الشبكة العامة وهي :

$$\frac{٣٠ \times ١٠٠}{٥٠} = ٪٦٠ \qquad \therefore \text{ من يشربون من التنكات } = ٪٤٠$$

ثانياً: نحسب الخطأ المعياري من خلال المعادلة

$$\text{خ م} = \frac{\text{ع}}{\sqrt{\text{ن}}}$$

$$\text{خ م} = \frac{\text{١٤,١}}{\sqrt{\text{٥٠}}} = \frac{\text{١٤,١}}{\text{٧,٠٧}} = \text{١,٩٩}$$

ثالثاً: نحسب حجم العينة ثنائية التوزيع

$$\text{ن} = \frac{\text{أ} \times \text{ب}}{(\text{خ م})^{٢}} = \frac{\text{٦٠} \times \text{٤٠}}{(\text{١,٩٩})^{٢}} = \frac{\text{٢٤٠٠}}{\text{٣,٩٦}} = \text{٦٠٦}$$

أي إن الحجم المطلوب للعينة التي تحقق أغراض الدراسة هو في حدود ٦٠٠ مسكن.

٢- تقدير حجم العينة من خلال استخدام مستويات الثقة (مستوى المعنوية):

في كثير من الأحيان لا يتمكن الباحث من معرفة الخطأ المعياري للعينة؛ ولذا لا بد له في هذه الحالة من تقدير مستوى معين من درجة الثقة في البيانات، ومن الجدير بالذكر أن معظم الباحثين في العلوم الاجتماعية ولا سيما الجغرافيين منهم يركزون على مستويات الثقة ٩٠٪، ٩٥٪، ٩٩٪، أي أن هناك نسبة خطأ مسموح بها. واحتمال هذا الخطأ هو ١٠٪، ٥٪، ١٪ على التوالي، ومعايير هذه المستويات من الثقة (معيار Z) هي:

± ١,٦٥، ± ١,٩٦، ± ٢,٥٨ . وتستخدم هذه المعايير لاختبار قدرة تمثيل العينة المختارة لمجتمع الدراسة. ومن ثم تحديد حجم العينة التي تمثل

مجتمع الدراسة ضمن حدود ثقة معينة وضمن نسبة معينة من الخطأ مسموح بها.

وفي هذه الحالة لسنا بحاجة إلى حساب الخطأ المعياري، وكذلك لسنا بحاجة إلى معرفة الانحراف المعياري، حيث يستعاض عنهما بمستويات الثقة التي حددها الباحث وبقيمة (Z) لهذه المستويات، فإذا اخترنا نسبة خطأ مسموح بها مقدارها ١٠٪ فإن قيمة (Z) المناظرة لها تعادل ١,٦٥ .

$$\text{ن} = \left(\text{ز} \times \frac{\sqrt{\text{ب}\,(\text{أ} - \text{ب})}}{\text{ي}} \right)^{٢}$$

حيث ن = حجم العينة

ز = قيمة معيار (Z) لمستوى التقدير المطلوبة.

ب = نسبة وجود الظاهرة

(أ−ب) =نسبة عدم وجود الظاهرة

ي = حد الثقة المرغوب الوصول إليه أو نسبة الخطأ المسموح به.

مثال: (١)

أراد باحث دراسة انتشار الهاتف الجوال في منطقة معينة واختار عينة استطلاعية عددها (٨٠) شخصاً، أجاب (٢٥) منهم بأنهم يملكون هاتف جوال، والباقي لا يملكون . ما هو الحجم الأمثل للعينة الموسعة التي يجب أن يختارها الباحث للخروج بنتائج إيجابية وبنسبة خطأ لا تتجاوز ٥٪؟

الحل:

إذا حدد الباحث نسبة الخطأ بـ ٥٪ فإن حد الثقة المرغوب فيه = ٠٥ , وقيمة (Z) المناظرة له تعادل ١,٩٦ .

$$\text{نسبة وجود الظاهرة (من يملكون الجوال)} = \frac{٢٥}{٨٠} = ٠,٣١٢٥$$

نسبة عدم وجود الظاهرة (من لا يملكون الجوال) = (١ − ٠,٣١٢٥) = ٠,٦٨٧٥

$$\text{حجم العينة} = \left(\text{ز} \times \frac{\sqrt{\text{ب}(١-\text{ب})}}{\text{ي}}\right)^{٢}$$

$$= \left(١,٩٦ \times \frac{\sqrt{٠,٣١٢٥ \times ٠,٦٨٧٥}}{٠,٠٥}\right)^{٢} = \left(\frac{١,٩٦ \times ٠,٤٦٣٥}{٠,٠٥}\right)^{٢} = ٣٣٠$$

إذن حجم العينة المطلوبة في حدود الثقة المقدرة يساوي (٣٣٠) شخصاً؛ ولهذا يجب أن لا يقل عدد الأشخاص الذين سيقوم باستجوابهم عن هذا العدد.

مثال (٢):

بلغ عدد المساكن في مدينة الرياض عام ١٩٩٢م ٣٥١٢٥٣ منزلاً. أرادت باحثة دراسة نوعية مياه الشرب في المدينة بالاعتماد على نوعية الشبكة (حديد، بلاستيك) ومدى شمولية شبكة الصرف الصحي باعتبار أن ذلك يؤثر في نوعية المياه، ما هو حجم العينة الممثلة في حدود مستوى ثقة ٩٠٪؟

الحل:

١- لابد من تقسيم المنازل تبعاً للمتغيرات المطلوبة على النحو التالي :

أ- منازل ذات شبكة حديد وليس بها صرف صحي .

ب- منازل ذات شبكة حديد وبها صرف صحي .

جـ- منازل ذات شبكة بلاستيك وليس بها صرف صحي .

د- منازل ذات شبكة بلاستيك وبها صرف صحي .

٢- تم تحديد عدد المنازل لكل فئة من الفئات السابقة فكانت على النحو التالي :

أ- الفئة الأولى تضم ٦,٥٦٧ منزلاً وتعادل ٠,٠١٨٧ من مجموع المنازل .

ب- الفئة الثانية تضم ١٤,٤٤٥ منزلاً وتعادل ٠,٠٤١١ من مجموع المنازل .

جـ- الفئة الثالثة تضم ١٣٥,٦٩٦ منزلاً وتعادل ٠,٣٨٦٣ من مجموع المنازل .

د- الفئة الرابعة تضم ١٩٤,٥٤٥ منزلاً وتعادل ٠,٥٥٣٩ من مجموع المنازل .

٣- حساب حجم العينة للفئة الأولى

قيمة ز لمستوى الثقة ٩٠٪ هو ١,٦٥

ب = نسبة وجود الظاهرة وتعادل ٠,٠١٨٧

(١-ب) نسبة عدم وجود الظاهرة تعادل ١ - ٠,٠١٨٧ = ٠,٩٨١٣

ي = ٠,١

من خلال المعادلة

$$ن = \left(\frac{ز \times \sqrt{ب\,(١-ب)}}{ي} \right)^{٢}$$

$$ن = \left(\frac{١,٦٥ \times \sqrt{٠,٠١٨٧ \times ٠,٩٨١٣}}{٠,١} \right)^{٢}$$

$$ن = \left(\frac{١,٦٥ \times ٠,١٣٥٥}{٠,١} \right)^{٢} = \left(\frac{٠,٢٢٣٥}{٠,١} \right)^{٢} = ٥ \text{ عينات}$$

٤- حساب حجم العينة للفئة الثانية

قيمة ز لمستوى الثقة ٩٠٪ هو ١,٦٥

ب= نسبة وجود الظاهرة وتعادل ٠,٠٤١١

(١-ب) نسبة عدم وجود الظاهرة ويعادل ١ - ٠,٠٤١١ = ٠,٩٥٨٩

ي = ٠,١

$$ن = \left(\frac{١,٦٥ \times \sqrt{٠,٠٤١١ \times ٠,٩٥٨٩}}{٠,١} \right)^{٢} = \left(\frac{١,٦٥ \times ٠,١٩٩}{٠,١} \right)^{٢} =$$

$$= \left(\frac{٠,٣٢٨}{٠,١} \right)^٢ = (٣,٢٨)^٢ = ١١ \text{ عينة}$$

٥- بالطريقة نفسها نحسب حجم العينة للفئة الثالثة والرابعة. وهي على النحو التالي :

أ- الفئة الثالثة = ٦٥ عينة.

ب- الفئة الرابعة = ٦٧ عينة.

٦- على هذا الأساس يتم تحديد العدد الكلي للعينات وهي ٥ + ١١ + ٦٥ + ٦٧ = ١٤٨ عينة.

إجراء الاختبارات الإحصائية باستخدام الحاسوب:

١- **اختبار ت** (t) : سنستخدم الحاسوب في تطبيق اختبار (ت) على نوعين من العينات هما العينات المستقلة والعينات المزدوجة.

أ- **اختبار (ت) للعينات المستقلة** : لدينا قائمة تمثل معدل المواليد في عينتين من الدول العربية : عينة تمثل تسعاً من الدول العربية في آسيا والأخرى تمثل تسعاً من الدول العربية في أفريقيا وهاتان العينتان مستقلتان ومسحوبتان من مجتمعين مختلفين ونريد أن نعرف إن كان هناك فرق في معدل المواليد بين العينتين وفي مستوى معنوية ٠,٠٥

قبل أن نبدأ التحليل نضع الفرضية الصفرية أو فرضية العدم، وهي أنه لا يوجد فرق بين معدل المواليد بين الدول العربية في آسيا والدول العربية في أفريقيا، والفرض البديل هو هناك فرق في معدل المواليد بين العينتين في مستوى دلالة ٠,٠٥ .

يمكن تلخيص خطوات إجراء هذا الاختبار باستخدام الحاسوب على النحو التالي :

١- ندخل البيانات الخاصة بمعدل المواليد في العينتين إلى ورقة العمل في Spss . حيث يظهر في العمود الأول متغيراً رمزياً . حيث أعطيت الدول العربية في آسيا الرمز (١) والدول العربية في أفريقيا الرمز (٢) أما العمود الثاني فيمثل معدل المواليد في كل دولة من الدول المختارة . وبهذه الطريقة يستطيع الحاسب تمييز البيانات الخاصة بكل عينة .

٢- نختار من شريط القوائم Statistics فتظهر قائمة منسدلة نختار منها Compare Means أي الموازنة بين المتوسطات الحسابية ، فتظهر قائمة ثالثة عنوانها means (انظر شكل ٨-١) وتشمل ثلاثةأساليب للموازنة بين المتوسطات الحسابية هي .

أ- اختبار (t) للموازنة بين المتوسطات الحسابية لعينتين مستقلتين

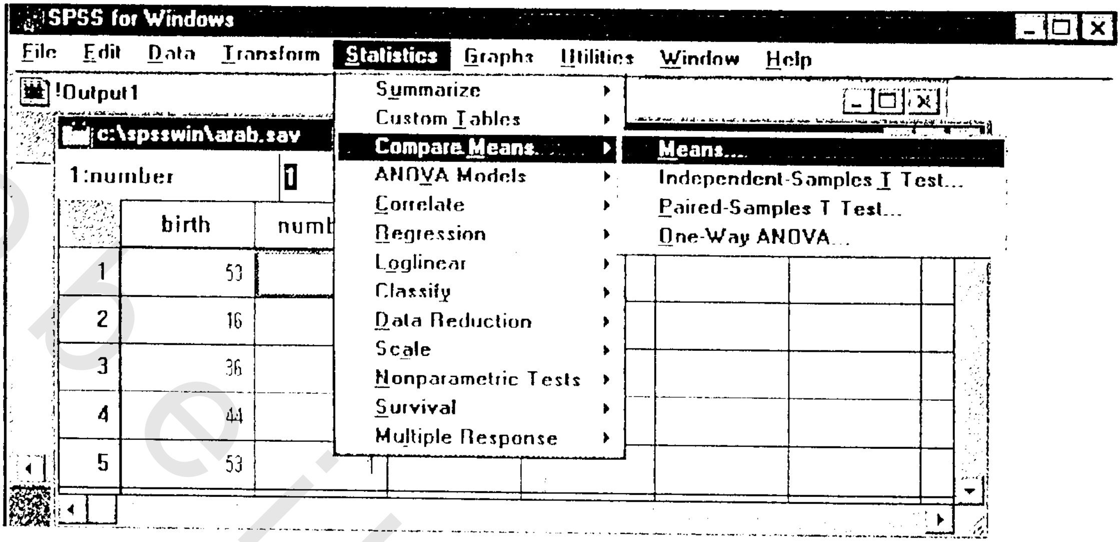
Independent Samples t Test

ب- اختبار (t) للموازنة بين المتوسطات الحسابية لعينتين مزدوجتين

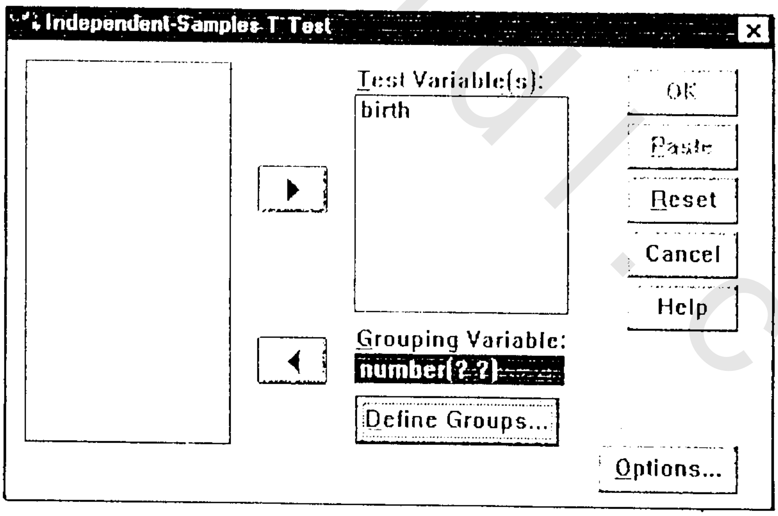
Paired Samples (t) Test

ج- تحليل التباين الأحادي One way (ANOVA)

٣- نختار من تلك القائمة اختبار (t) للعينات المستقلة وننقر عليه بالفارة فتظهر شاشة جديدة شكل رقم (٨ : ٢) عنوانها اختبار (t) للعينات المستقلة Independed Samples (t) Test فيها يظهر اسم المتغيرين Number و Birth ننقل المتغير Birth . أي المواليد إلى المكان المخصص بالمتغير الذي



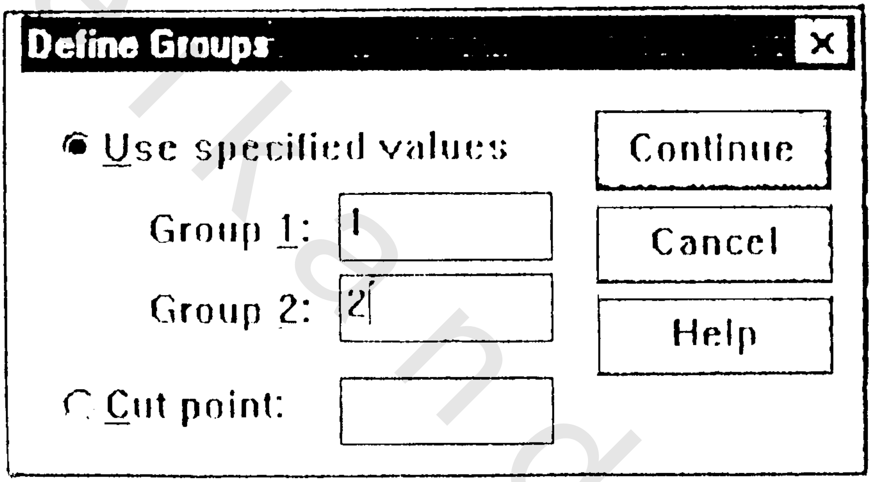
شكل (١:٨) الأوامر الخاصة بالدخول إلى اختيارات المتوسطات



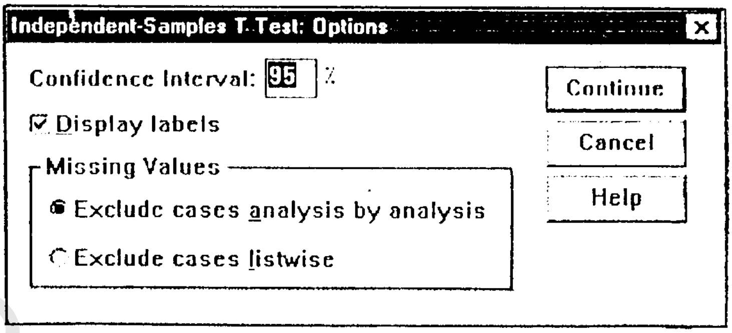
شكل (٢:٨) لوحة اختيار العينات المستقلة

نجري عليه الاختبار (Test Variables) ، والمتغير Number إلى المربع الخاص بالمتغير التصنيفي Grouping Variable شكل رقم (٨-٢)، ثم نحدد القيمتين اللتين يأخذهما المتغير التصنيفي (انظر شكل ٨-٣).

شكل (٣:٨) لوحة تحديد المجموعات في المتغيرات الرمزي



٤- بعد أن ننهي كتابة القيمتين الخاصتين بالمتغير التصنيفي وهما (١) للدول الآسيوية و (٢) للدول الأفريقية ننقر خيارات (Options) في الشكل السابق نفسه رقم (٨-٢) فتظهر شاشة جديدة تحوي العديد من الخيارات، وأهم هذه الخيارات هو تحديد درجة الثقة وهي ٩٥٪، أي أن مستوى المعنوية ٥٪ (شكل (٨-٤) وبهذا نكون قد أنهينا كافة الأوامر الخاصة بإجراء



شكل (٨:٤) تحديد مستوى المعنوية

الاختبار. وبمجرد أن ننقر الفأرة على المفتاح Continue في الشكل رقم (٨-٤) نعود إلى الشكل السابق رقم (٨-٢) وننقر OK فتخرج لنا الشاشة رقم (٨-٥) الخاصة بالنتائج.

٥- يظهر الشكل رقم (٨-٥) نتائج الاختبار ويشمل ما يلي:

أ- الجزء الأول يشمل اسم المتغيرين، وعدد المشاهدات في كل متغير والمتوسط الحسابي والانحراف المعياري، والخطأ المعياري الخاص بكل متغير.

ب- الجزء الثاني ويشمل اختبار (t) في حالتين.

١- الحالة الأولى عندما يكون تباين العينتين متساوياً.

٢- الحالة الثانية عندما لا يكون التباين متساوياً.

وتتضمن النتائج أيضاً قيمة (t)الاختبارية وتساوي - ١,٢٦ ودرجات الحرية، ومستوى المعنوية الذي يكون للفرق بين المتوسطين عند دلالة إحصائية، وحدود الثقة لذلك الفرق على مستوى معنوية ٠,٠٥ ومن الجدير بالذكر أن الذي يهمنا من هذا الاختبار هو قيمة (t) ودلالتها الإحصائية، فقيمة (t) = - ١,٢٦ وهي دالة عند مستوى معنوية ٠,٢٢٧ مما يدل على عدم وجود فرق في العينتين، وهذا يجعلنا نقبل فرضية العدم وهي عدم وجود فرق جوهري بين معدل المواليد بين الدول العربية في آسيا والدول العربية في أفريقيا.

شكل (٨-٥)

نتائج اختبار (ت) للعينات المستقلة

t-tests for independent samples of CODE code

Variable	Number of Cases	Mean	SD	SE of Mean
BIRTH birth				
CODE 1	9	31.8889	10.994	3.665
CODE 2	9	37.7778	8.772	2.924

Mean Difference = -5.8889

Levene's Test for Equality of Variances: F= .254 P= .621

t-test for Equality of Means

Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	-1.26	16	.227	4.688	(-15.830, 4.052)
Unequal	-1.26	15.25	.228	4.688	(-15.884, 4.106)

ب- اختبار (ت) للعينات المزدوجة:

العينات المزدوجة هي التي تقيس استجابة المفردة الإحصائية نفسها لتأثيرين مختلفين، فمثلاً حينما يقوم أحد أساتذة الخرائط بتقويم المهارات الجديدة التي يكتسبها طلبته في قراءة الخرائط؛ وذلك بأن يقوم في بداية الفصل بامتحان عينة من الطلبة، ويسجل الوقت اللازم لكل طالب لاستخراج بعض المعلومات من الخرائط، ثم يقوم في نهاية الفصل بإجراء الاختبار ذاته على عينة الطلبة نفسها لمعرفة درجة التحسن في مهاراتهم.

وقد تكون العينات المزدوجة أيضاً حينما تكون كل مفردة من مفردات عينة إحصائية تقابلها في العينة الثانية مفردة أخرى محددة تشترك معها في الخاصية الواحدة، مثال ذلك البيانات الخاصة بمعدلات انجراف التربة في سفحين متشابهين في الانحدار، ولكن أحدهما أجرد لا نبات فيه، والثاني مشجر. فهنا نجد التشابه في الخصائص الطبيعية للعينة من حيث الانحدار مع اختلاف واحد هو وجود الأشجار أو عدمها[1].

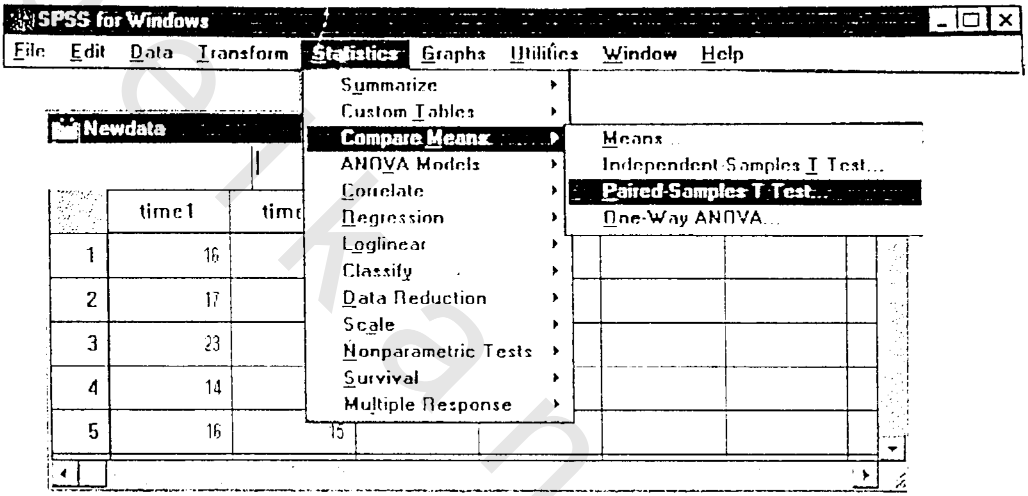
فإذا كانت لدينا بيانات تمثل الوقت اللازم لاستخراج بعض المعلومات من الخرائط لعينة من الطلبة في بداية الفصل الدراسي ونهايته، ونريد معرفة أخبار الدلالة الإحصائية للفرق بين المتوسط الحسابي للوقت الذي يحتاجه الطالب عند بداية ونهاية الفصل، لمعرفة ما إذا كان هناك تحسن في مهارات الطلاب أم لا.

١- ندخل البيانات إلى الحاسوب بحيث يمثل المتغير الأول (time 1) الوقت الذي استغرقه الطالب بالدقائق في بداية الفصل إلى العمود الأول من ورقة العمل، بينما يحتل المتغير الثاني (time 2) العمود الثاني من ورقة العمل.

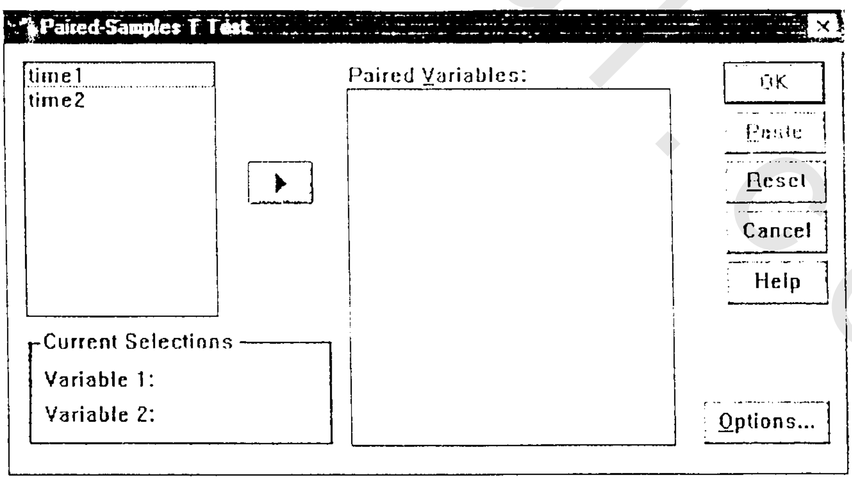
---

(١) نعمان شحادة: مرجع سابق، ص ٢٧٦ .

٢- تختار إحصاءات Statisties من شريط القوائم وثم نختار Compare means, فتظهر القائمة الثالثة نختار اختبار (ت للعينات المزدوجة - Paird samples test)فتظهر شاشة جديدة عنوانها Paired - samples t test (انظر شكل ٨-٦).



شكل (٨:٦) أوامر الدخول إلى اختبار (ت) للعينات المزدوجة



شكل (٨:٧) طريقة إدخال المتغيرات إلى تحليل (ت) للعينات المزدوجة

٣- ننقل في تلك الشاشة اسمي المتغيرين إلى المكان المخصص لهما وعنوانه Paired Variables (شكل ٨-٧) ثم ننقر OK فتظهر الشاشة الخاصة بالنتائج.

٤- يتضمن الجزء الأول من النتائج اسمي المتغيرين، وعدد مفردات كل متغير ومعامل الارتباط بينهما ويساوي ٧٠٢, ٠، ويبين ذلك الجزء المتوسط الحسابي لكل من المتغيرين والانحراف المعياري والخطأ المعياري.

٥- أما الجزء الثاني وهو الذي يهمنا فهو نتائج اختبار (ت) ويشمل:

- متوسط الفرق بين المتغيرين الأول والثاني ويبلغ ٩, ١ .
- الانحراف المعياري للفرق بين المتوسطين ويبلغ ٩٩٨, ٢ .
- الخطأ المعياري للفرق ويبلغ ٩٤٨, ٠ .
- قيمة (ت) الاختبارية وتساوي ٢ .
- درجة الحرية (عدد القيم - ١) = ٩ .
- مستوى معنوية الاختبار ٠٧٦٥, ٠ والمعنوية هنا ليست في الحدود المطلوبة وهي ٠٥, ٠ ؛ ولذا فإن الفرق بين المتغيرين لا يكون له دلالة إحصائية عند مستوى ٠٥, ٠ (انظر شكل ٨: ٨).

شكل (٨-٨) نتائج اختبار (ت) للعينات المزدوجة

Variable	Number of pairs	Corr	2-tail Sig	Mean	SD	SE of Mean
TIME1	10	.702	.024	19.5000	3.923	1.241
TIME2				17.6000	3.836	1.213

Paired Differences Mean	SD	SE of Mean	t-value	df	2-tail Sig
1.9000	2.998	.948	2.00	9	.076
95% CI (-.245, 4.045)					

٣- تحليل التباين:

سنجري اختبار التباين الأحادي الخاص بتحليل التباين للموازنة بين المتوسطات الحسابية لثلاث عينات فأكثر، سنوضح هذا الاختبار من خلال المثال الآتي الذي يعد توسعة للمثال السابق الخاص باختبار الاختلاف بين معدل المواليد لعينتين من الدول العربية في آسيا والدول العربية في أفريقيا. فقد أضفنا إلى هذا الجدول معدل المواليد العام لبعض الدول الإسلامية غير العربية وبذا يصبح على النحو التالي:

١- معدل المواليد للدول العربية في آسيا.

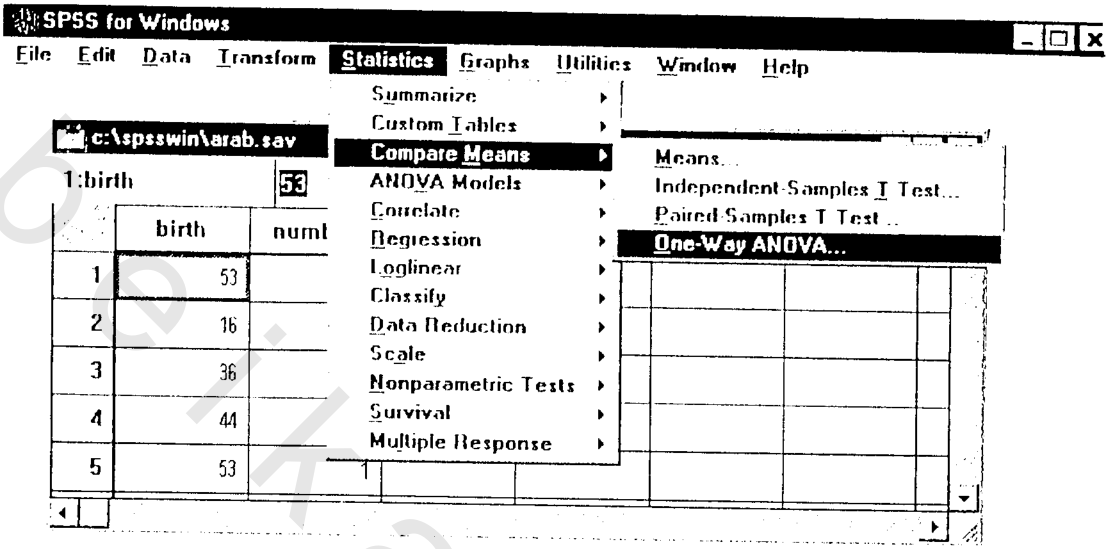
٢- معدل المواليد للدولة العربية في أفريقيا.

٣- معدل المواليد للدول الإسلامية غير العربية.

١- لإجراء التحليل من خلال الحاسوب نقوم بإدخال المعلومات إلى الحاسب الآلي، بحيث ندخل في العمود الأول متغيراً تصنيفياً يميز بين المجموعات الثلاثة حيث يعطي للدول العربية في آسيا رقم (١)، وللدول العربية في أفريقيا رقم (٢)، وللدول الإسلامية غير العربية رقم (٣)، فيما يتضمن العمود الثاني القيم الخاصة بمتغير المواليد في كل دولة من الدول المختارة.

٢- نختار Statistics من شريط القوائم، ثم Compare means، ثم من القائمة الثالثة One Way Anova وهو أسلوب تحليل التباين المنفرد (انظر شكل ٨-٩) فتظهر لوحة جديدة تحمل اسم One Way Anova وتتضمن

شكل (٨-٩) أوامر الدخول إلى تحليل التباين الأحادي



اسمي المتغيرين وهما Number، Birth، نقوم بنقل اسم المتغير الأول Birth إلى المكان المعنون dependent list، والمتغير الثاني Number إلى المكان

اسمي المتغيرين وهما Number، Birth، نقوم بنقل اسم المتغير الأول Birth إلى المكان المعنون dependent list، والمتغير الثاني Number إلى المكان المعنون في Factor، ثم نحدد القيمة الأدنى للمتغير التصنيفي وهي ١، والقيمة العظمى ومقدارها ٣، تماماً كما فعلنا بالاختبار الثنائي، وبهذا ننتهي من إعداد البرنامج الخاص بتنفيذ اختبار تحليل التباين الأحادي وبهذا نقر OK فتظهر نتائج الاختبار.

٣- تتضمن نتائج اختبار تحليل التباين الجدول نفسه الذي ذكرناه في السابق والذي يحوي مصدر التباين بين العينات وداخل العينات ومجموع

المربعات، إضافة إلى درجات الحرية ومقادر ف ومستويات المعنوية. ويظهر من الجدول أن قيمة ف = ٠,٨٧٥ وأن مستويات المعنوية هو ٠,٤٢، وهذا يدل على أنه لا توجد فروق جوهرية بين معدل المواليد وبين الفئات الثلاثة؛ بمعنى أنها متماثلة من حيث معدلات المواليد. (شكل ٨-١٠).

شكل (٨-١٠) نتائج تحليل التباين

```
- - - - -  O N E W A Y  - - - - -

    Variable   BIRTH        birth
 By Variable   CODE         code

                          Analysis of Variance

                                   Sum of         Mean           F        F
      Source             D.F.      Squares        Squares        Ratio    Prob.

Between Groups             2       162.7407       81.3704        .8757    .4294
Within Groups             24      2230.0000       92.9167
Total                     26      2392.7407
```

٢- اختبار مربع كاي:

اختبار مربع كاي هو أحد الاختبارات غير البارامترية، وهو يستخدم للموازنة بين توزيع تكراري فعلي وآخر نظري. ومن المعلوم أن بيانات هذا الاختبار هي بيانات نوعية تصنيفية تكون على شكل تكرارات. وسنقوم بإجراء هذا التحليل على المثال المتعلق بالعينة المختارة من عمال أحد المصانع لدراسة أثر التعليم على التدخين (انظر جدول رقم ٨-١١).

١- الخطوة الأولى: هي تكوين مصفوفة معلومات بالمئة حالة المختارة للأشخاص الذين خضعوا للاختبار، وتدوين حالة كل شخص فيما يتعلق

بالتدخين (المتغير الأول) الذي أعطي اسم : (smoke) والتعليم (المتغير الثاني) الذي أعطي اسم : (Educ)، وقد رمز للمدخن برقم (١) وغير المدخن برقم (٢) أما المتغير الثاني وهو التعليم (Educ) فقد رمز للمتعلم بـ (١)، وغير المتعلم بـ (٢) وقد خصصت الخانة الأولى من المصفوفة إلى الرقم المتسلسل.

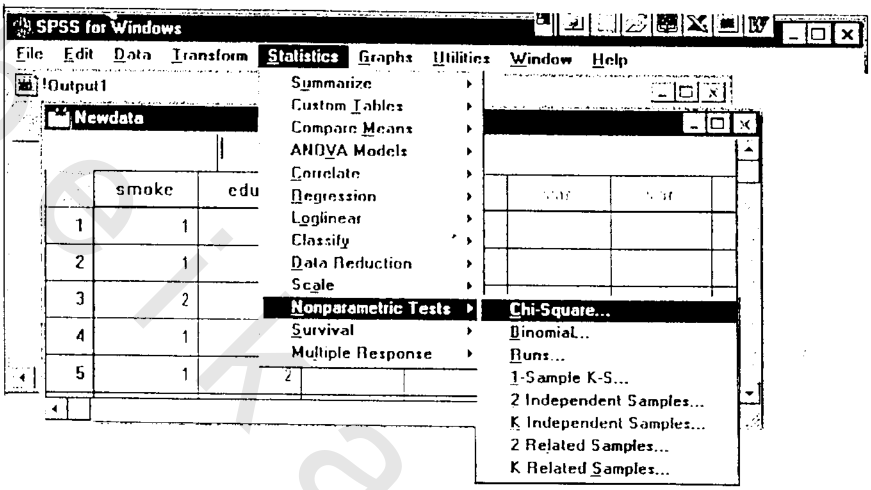
٢- من شريط القوائم نختار الأساليب الإحصائية Statistics فتظهر قائمة منسدلة، نختار منها الاختبارات غير البرامترية Nonparmetric Tests فتظهر قائمة ثالثة نختار منها مربع كاي (Chi Square) وهو الاختيار الأول في الشكل رقم (٨-١١).

٣- تظهر الشاشة الخاصة بتنفيذ تحليل مربع كاي (شكل٨-١٢) ويظهر بها أسماء المتغيرات الموجودة في الملف فنختار منها المتغير (التعليم). الذي سوف نجري عليه الاختبار، ننقل اسم ذلك المتغير إلى المكان المعنون بـ Test Variable List . ولما كانت التكرارت المتوقعة للتوزيع النظري موزعة بالتساوي نختار المفتاح All Categories equal، أما المدى المتوقع Expected Range فيؤخذ من البيانات Get from Data (انظر شكل ٨-١٢)

٤- ننقر الأمر OK فتظهر النتائج التكرارية للتوزيع الفعلي والتوزيع المتوقع، إضافة إلى مربع كاي، ومستوى المعنوية الخاص بها، ونرى من خلال جدول النتائج أن التوزيعات الفعلية هي ٧٠ عاملاً من فئة المتعلمين و ٣٠ من فئة غير المتعلمين. أما التوزيع المتوقع فهو ٥٠ لكل فئة، أما قيمة مربع كاي فهي ١٦ ومستوى المعنوية عالي مقداره ٠,٠٠٠١ أي أن احتمال الخطأ هو ١ من عشرة آلاف.

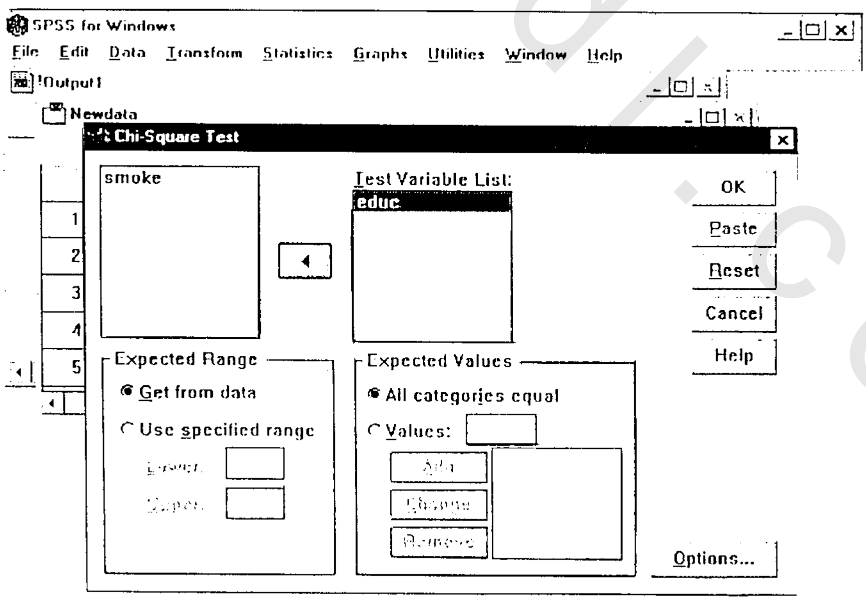
شكل (٨-١١)

أوامر الدخول إلى مربع كاي



شكل (٨-١٢)

ترتيب المتغيرات في مربع كاي



تمــــارين

س١ : أوجد قيمة ف لدرجة حرية (٤ , ١٧) مع العلم أن مستوى الدلالة المطلوب ٠,٠١ .

س٢ : ما هي قيمة ف إذا كان لديك ٣ عينات (متغيرات)، وكل عينة مكونة من خمس قيم (حالات) مع العلم أن مستوى الثقة المطلوب هو ٠,٩٥ ؟

س٣ : تم إجراء تجربة لدراسة تأثير الأوقات المختلفة للزراعة، وكذلك الطرق المختلفة للزراعة على إنتاجية محصول قصب السكر، وقد اختيرت عينة من ١٢ قطعة أرض زراعية متساوية المساحة، وسجلت إنتاجية هذه القطع من قصب السكر في الجدول الآتي. والمطلوب تحليل هذه البيانات بأسلوب تحليل التباين الثنائي وإظهار فيما إذا كان هناك فروق جوهرية بين أوقات الزراعة أو بين طرق الزراعة عند مستوى دلالة ٠,٠١ .

طرق الزراعة	وقت الزراعة			
	أكتوبر	نوفمبر	فبراير	مارس
الطريقة الأولى	٧,١٠	٣,٦٩	٤,٧٠	١,٩٠
الطريقة الثانية	١٠,٢٠	٤,٧٩	٤,٥٨	٢,٦٤
الطريقة الثالثة	٨,٣٠	٣,٥٨	٤,٩٠	١,٨٠

س٤ : إذا كان جدول تحليل التباين الثنائي هو :

مصدر الاختلاف	مجموع المربعات	درجة الحرية	متوسط مربعات ف	
المعالجات	٢٧,١	٣	٩,٠٣٣	
القطاعات	٧٤,٥	٥	١٤,٩٠	
الخطأ	٣٣,٤	١٥	٢,٢٢٧	

أ- احسب قيمة ف بالنسبة للمعالجات وقيمة ف بالنسبة للقطاعات .

ب- اختبر معنوية الفروق بين المعالجات عند مستوى دلالة ٠,٠٥ والفروق بين القطاعات عند مستوى دلالة ٠,٠١ .

س٥ : إذا كانت قيمة مربع كاي ($\times^{2}$) المحسوبة ٢,٥ وكانت درجات الحرية ٣ فهل تقبل فرضية العدم في مستوى دلالة ٠,٠٥ ؟

س ٦ : إذا كانت قيمة مربع كاي المحسوبة ١٦ وكانت درجات الحرية ٦ فهل تقبل فرضية العدم في مستوى دلالة ٠,٠١ ؟

٧- طبقاً لمتوسطات الإنتاج السابقة لثلاث قطع من الأراضي كانت نسبة الإنتاج ٢ : ١ : ١ أي أن المزرعة الأولى تعطي ضعف كل من المزرعتين الثانية والثالثة ، وفي أحد المواسم كانت كمية الإنتاج للمزارع الثلاثة ٤٠طن و٢٠ طن و ٣٠ طن للمزارع الثلاثة على التوالي ، هل تتفق مقادير الإنتاج الحالية مع المتوسطات السابقة عند مستوى دلالة ٠,٠٥ ؟

س ٨ : قام باحث بدراسة العلاقة بين الدخل ومستوى التعليم فأخذ عينة مكونة من ٥٠ شخصاً منهم ٣٠ شخصاً من حملة الدكتوراه، و ٢٠ شخصاً من حملة البكالوريوس، ووجد أن ٢٠ من أصل ٢٨ شخصاً من حملة الدكتوراه هم من ذوي الدخل العالي، كما وجد أن عشرة أشخاص فقط من حملة البكالوريوس هم من ذوي الدخل العالي كما في الجدول التالي :

الدخل / نوع الشهادة	دخل عالي	دخل متوسط	المجموع
بكالوريوس	٨	١٢	٢٠
دكتوراه	٢٠	١٠	٣٠
الإجمالي	٢٨	٢٢	٥٠

المطلوب : هو معرفة ما إذا كان هناك علاقة بين الدخل ومستوى التعليم عند مستوى الدلالة ٠,٠٥ .

٩- أخذت عينة لمن درسوا مادة الكمية والإحصائية مكونة من ١٠٠ طالبة، و ١٥٠ طالباً، فوجد أن من بين كل ١٠٠ طالبة دخلت اختبار الكمية نجح ٧٥ ورسب ٢٥ ، وأن من بين كل ١٥٠ طالب دخل الامتحان نجح ٦٠ ورسب ٩٠ ، فإن كانت العينة السابقة مأخوذة على أساس أنها تمثل قطاعاً لجميع الدارسين لمادة الكيمة والإحصائية على النحو الموضح في الجدول التالي . فهل تستطيع القول أن هناك اختلافات جوهرية بين مقدرة

كل من النساء والرجال في مادة الكمية والإحصائية بمستوى دلالة ٠,٠٥ ؟

نتيجة الامتحان / النوع	راسب	ناجح	المجموع
رجال	٩٠	٦٠	١٥٠
نساء	٢٥	٧٥	١٠٠
الإجمالي	١١٥	١٣٥	٢٥٠