الفصل السابع عشر

تحليل التباين
Analysis of Variance "ANOVA"

17.1. مقدمة Introduction

تعرفنا في الأجزاء السابقة على أنه توجد طرق للمقارنة بين معدلي (متوسط) مجتمعين، ولكننا في كثير من الأوقات نحتاج إلى المقارنة بين متوسطات عدة مجتمعات، فبينما استعملنا اختبار Z أو اختبار T لاختبار تساوى معدلي مجتمعين معينين، فإننا نحتاج إلى طرق أكثر شمولاً لنتمكن من اختبار تساوى متوسطات ثلاث مجتمعات أو أكثر، وتجيء طريقة تحليل التباين Analysis of Variance على رأس هذه الطرق، (حيث يعد صورة عامة لاختبار ت t- test للمقارنة بين ثلاثة متوسطات أو أكثر، وإذا استخدم للمقارنة بين متوسطين فإن النتيجة تكون مماثلة للناتج من اختبار ت) وهى أداة قوية تمكننا من تحليل وتفسير المشاهدات من عدة مجتمعات في وقت واحد.

يعتمد تحليل التباين ANOVA في تحديد العامل أو العوامل التي لها تأثير على الإستجابة Response، على تجزئة التباين الكلى Total Variation المشاهد فى البيانات التي نحصل عليها من التجربة أو من المسح إلى أجزاء مختلفة كل منها يمكن إرجاعه إلى مصدر (سبب أو عامل) معلوم، وبذلك يمكن تقييم المقدار النسبي للتباين من كل مصدر ثم تقدير ما إذا كان ذلك معنويا أى مؤثراً وملموساً وذا قيمة أم لا.

وكما فى الشكل 16-2 فإن تحليل التباين يستخدم لتحديد مدى ارتباط مؤثر أو متغير X أو أكثر بالنتيجة Y أو المخرجات Output، بشرط وجود علاقة فعلية وليست محض صدفة بين هذا المعامل Factor وبين الإستجاب Response ، وهذا ما نسميه العلاقة السببية بين العوامل أو المؤثرات Causal Effect Between The Factors، و تحليل التباين ANOVA أيضا وسيلة فعالة لتحويل العلاقات النوعية Qualitative Relationship أو الغير كمية إلى علاقات كمية Quantitative Relationship بمعني أنها لا تخبرنا فقط بوجود علاقة بين متغيرين، وإنما تخبرنا كذلك بقيمة قوة هذه العلاقة كنسبة مئوية.

17.2. خطوات تحليل التباين (الأنوفا) ANOVA Steps

كما ذكرنا فإن تحليل التباين ANOVA يعتمد في تحديد العلاقة بين المتغيرات علي تجزئة التباين الكلى المشاهد فى البيانات، إلى تباين داخل المجموعات Variation Within Groups ويتحدد بمدى انحراف كل عنصر عن متوسط باقى عناصر المجموعة، وإلى تباين بين المجموعات Variation Between Groups ، ويتحدد بمدى انحراف كل مجموعة عن متوسط المجموعات أو انحراف كل عينة عن الأصل الذي تنتسب إليه أى أن :

Total Variation = Variation between Groups + Variation within Groups

ثم يتم بعد ذلك مقارنة مدى اقتراب التباين الداخلي Variation Within Groups من التباين الخارجي Variation Between Groups أو مدى ابتعاده عنه، وذلك بحساب النسبة الفائية F Ratio ، وهى خارج قسمة القيمة الأكبر لأى من التباينين على القيمة الأصغر لهما، ثم نقوم بمقارنة هذه النسبة F Ratio بقيمة F الجدولية لتحديد مستوى الدلالة، فإذا كانت القيمة المحسوبة أكبر من القيمة الجدولية تكون الفروق بين المجموعات دالة ، وأمكننا أن نستنتج عدم تجانس المجموعات التي نحلل تباينها، وأمكننا كذلك أن نرجع هذه المجموعات إلى أصولها المختلفة التي تنسب إليها، أما إذا كانت القيمة المحسوبة أقل من القيمة الجدولية أمكننا أن نستنتج تجانس المجموعات التي نحلل تباينها، وأمكننا أن نرجع هذه المجموعات جميعا إلى أصل واحد.

وفى هذه الحالة لا نستطيع تحديد أى المجموعات أعلى من الاخرى (مثل ما يحدث فى إختبار ت) ، ويتطلب ذلك استخدام اسلوب آخر بعد تحليل التباين لاجراء المقارنات المتعددة بين المتوسطات multiple com[parison of means، ومن طرق المقارنات المتعددة : دنكان Duncan (أقل تحفظاً) ، و توكى Tuky، و شفيه Scheffe (أكثر تحفظاً من الطرق الاخرى)

وبذلك نستطيع مثلا دراسة تأثير التحسينات التي تم إجراؤها مؤخراً على عملية ما، وذلك بمقارنة عينات قبل التحسينات وعينات بعد التحسينات.

وقبل البدء في الحديث عن خطوات إجراء تحليل التباين ANOVA يجب الإشارة إلى أن تنفيذ هذه الخطوات سيتم عن طريق الكمبيوتر إلا أننا سنذكر هذه الخطوات لفهم ما يتم فقط وهى كما يلي:

1. يتم حساب التباين بين المجموعات Variation Between Groups.
2. يتم حساب التباين داخل المجموعات Variation Within Groups.
3. يتم حساب النسبة الفائية وذلك بقسمة التباين الكبير على التباين الصغير لمعرفة مدى تجانس أو اختلاف المجموعات.

ومما سبق يتضح أنه عند الحديث عن تحليل الانحدار Regression فيجب التأكيد على أنه في حالة وجود مجموعة قراءات توضح العلاقة بين عدة متغيرات مستقلة Independent ومتغير تابعDependent ، وأمكن استنتاج المعادلة النموذج Model المعبرة عن هذه العلاقة وهذا الارتباط باستخدام الانحدار Regression، ووصفنا هذه العلاقة بمعادلة التنبؤ Prediction، وفي هذه المعادلة نجد أننا حددنا بصورة تقريبية مدى تأثير كل متغير من المتغيرات المستقلة على المتغير التابع، ولكن يبقى سؤال هام هل الاختلاف الحادث في المتغير التابع يرجع فقط إلى تأثير المتغيرات المستقلة؟ أم أنه يرجع إلى عوامل أخرى عشوائية وغير مرئية أو محسوسة للدارس؟

والإجابة على هذا التساؤل تتطلب تحليلا إحصائيا لهذا التباين والاختلاف في المتغير التابع أو النتيجة إلى جزأين مختلفين، أحدهما يعتمد على المتغيرات المستقلة ونطلق عليه الاختلاف المفسر Explained Variation لأن مصدره معلوم، والآخر لا يعتمد على هذه المتغيرات المستقلة ويعزى إلى عوامل مجهولة ونطلق عليه الاختلاف الغير مفسر أو الاختلاف المتبقي Residual Variation،وغالبا ما نرجعه إلى الأسباب العامة Common Causes، وبعد تجزئة هذا الاختلاف والتباين الكلي يكون من السهل تقييم واختبار ما إذا كان التباين المفسر كبير كبرا جوهريا بالنسبة إلى التباين الغير مفسر أم لا، ويتم ذلك من خلال المعامل الفائي F Factor والذي يحسب من المعادلة:

$$\text{F factor} = \frac{\text{Explained Variation}}{\text{Unexplained Variation}}$$

17.3. شروط تطبيق الأنوفا Assumptions of ANOVA

قبل البدء في إجراء تحليل التباين ANOVA يجب التأكد من تحقق الشروط التالية:

1- أن تكون العينة المنتقاة ممثلة للمجتمع ، وأن تخضع لمنحنى التوزيع الطبيعي أو تكون قريبة منه.

2- أن تكون العملية قيد الدراسة في حالة تحكم إحصائى لضمان ثبوت ظروف المقارنة.
3- أن يكون التباين للمجموعات قيد الدراسة و التي تسحب منها العينات متجانساً ومتساويا أو متقاربا.
4- أن يكون توزيع الاختلاف المتبقى Residuals خاضعا للتوزيع الطبيعى.

وجدير بالذكر أن أسلوب تحليل التباين يعطى نتائج لا بأس بها حتى فى حالة عدم توافر الشرطين الأخيرين.

17.4. تحليل تباين القياس المتكرر Repeated Measurement ANOVA

ويستخدم هذا النوع من تحليل التباين عند إجراء دراسات تجريبية تهتم بقياس مستوي أداء عملية ما عدة مرات متتالية تحت شروط مختلفة ، مثل قياس مستوى العملية قبل التحسين ثم بعد التحسين، وكذلك القياس بعد انتهاء عملية التحسين بفترة زمنية محددة (ستة أشهر مثلاً) ويسمى قياس المتابعة.

وفى هذه الحالة تم قياس مستوى أداء العملية ثلاث مرات متتالية ولذلك يسمى قياساً متكرراً، أما إذا قمنا باجراء قياسين فقط قبلى وبعدى فإنه لا يعد قياساً متكرراً.

ويوجد عدد من التصميمات التجريبية للقياس المتكرر ، فهناك قياس متكرر لمجموعة واحدة، وقياس متكرر لعدة مجموعات ، وغيرها من التصميمات الأكثر تعقيداً، ويتم التعامل معها وحلها عن طريق الكمبيوتر .. (صلاح مراد- فوزية هادى 2002 بتصرف)

17.5. تحليل التغاير Analysis of Covariance

وهو نوع آخر من تحليل التباين يستخدم لتتبع/ضبط أثر متغير دخيل بطريقة إحصائية

فقد نجد أنه أثناء الدراسة أن متغيراً خارجياً له أثر على المتغير التابع ولم نستطع وضعه فى التصميم، وفى هذ الحالة نقوم بقياس هذا المتغير ونجرى تحليل لعزل أثره من المتغير التابع.

فإذا كان الهدف من الدراسة تقويم فعالية جهود التحسين لعملية التدريب ، والمتغير التابع هو مستوى التحصيل ، أما المتغير الخارجى الذى يؤثر فى مستوى التحصيل هو نسبة ذكاء المتدربين، وغذا قمنا باختيار عدة مجموعات من المتدربين لتجريب طرق التدريب المختلفة، فإننا نقوم بقياس نسبة ذكاء المتدربين قبل إجراء التجربة ، ونقيس مستوى التحصيل بعد انتهاء التجربة، ثم نجرى تحليل تغاير لعزل أثر نسبة الذكاء من التحصيل قبل بحث الفروق بين المجموعات التجريبية فى التحصيل.

أما إذا كانت الدراسة تتضمن قياساً قبلياً وبعدياً لعدة مجموعات تجريبية (وهو لا يعد قياساً متكرراً) ، فاننا نقوم بإجراء تحليل تباين أحادى (بين المجموعات) لدرجات القياس القبلى، ويكون امامنا حالتين:

1- اذا لم نجد فروقاً دالة بين المجموعات فى درجات القياس القبلى ، فإننا نقوم بإجراء تحليل تباين أحادى لبحث الفروق بين المجموعات فى درجات القياس البعدى.
2- إذا وجدنا فروقاً دالة بين المجموعات فى درجات القياس القبلى ، فإننا نقوم بإجراء تحليل تغاير لعزل أثر درجات القياس القبلى من القياس البعدىقبل التعرف على الفروق بين المجموعتين فى القياس البعدى.

(صلاح مراد- فوزية هادى 2002 بتصرف)

17.6. أساليب الإحصاء الاستدلالى اللابارامترى None parametric Statistical Inference

تستخدم الاساليب الاحصائية اللابارامترية فى حال البيانات الاسمية أو الترتيبية، كما تستخدم أيضاً فى حال العينات الصغيرة (أقل من 15) مهما كان مستوى القياس ، أما إذا كان حجم العينة يتراوح بين 15 و 20 فيفضل إجراء تحليل البيانات باستخدام الاساليب البارامترية واللابارامترية، وذلك للتأكد من صحة القرار الناتج عن التحليل، ونذكر فيما يلى بعض الاساليب واللابارامترية:

1- اختبار مان ويتنى Mann-Whiteny test

هو اختبار لابارامترى بديل لاختبار ت t- test، ويستخدم لمقارنة البيانات الترتيبية Ordinal أو الفترية Interval أو النسبية Ratio ، حيث نضع درجات المجموعتين معاً (إذا كانت البيانات فترية أو نسبية) ونرتبهم كأنهم مجموعة واحدة ، ثم نحسب قيمة إحصائية نرمز لها بالرمز (ى) ، ونختبر دلالة هذه القيمة باستخدام جداول خاصة باختبار مان ويتنى، والمنطق فى هذا الاختبار هو : إذا كانت المجموعتين متشابهتان فإن مجموع رتب كل منهما يتساوى، أما إذا اختلف مجموع الرتب اختلافاً ملحوظاً فإن هذا الفرق يكون دالاً.

مثال : اذا كانت آراء مجموعتن من المتدربين فى النظام التدريبى على النحو التالى:

المجموعة الاولى: 7، 8، 10، 12، 15

المجموعة الثانية: 9، 14، 16، 17

ولبحث الفرق بين المجموعتين نرتب درجاتهما معا كما يلى:

7، 8، 9، 10، 12، 14، 15، 16، 17، ثم نحسب قيمة (ى) وهى عدد درجات المجموعة الأولى (الكبرى) الأقل من كل درجة من درجات المجموعة الثانية (الصغرى) وهى (ى)=2+4+5+5=16.

وباستخدام جداول مان ويتنى وحيث (ن$_1$)=4، (ن$_2$)=5 ، فإن قيمة (ى) الجدولية =11 عند مستوى 0.05 ، 15 عند مستوى 0.01، وعليه فإن قيمة (ى)المحسوبة =16 تكون دالة عند مستوى 0.01 ، ونستنتج من ذلك وجود فرق دال بين رتب المجموعتين لصالح المجموعة الثانية.

2- اختبار ولكوكسون Wilcoxon test

هو اختبار لابارامترى بديل لاختبار ت t- test، ويستخدم فى حالة المجموعات المرتبطة أو المجموعة الواحدة مع قياسين قبلى وبعدى، حيث يتم حساب الفروق بين القياسين ، ويتم ترتيب هذه الفروق مع إهمال الفروق الصفرية، ثم نحسب قيمة إحصائية يرمز لها بالرمز (و) وهى مجموع رتب الفروق الاقل (بأى إشارة) ثم نقارن قيمة (و) المحسوبة بالقيم الجدولية من جداول ولكوكسون.

3- اختبار كروسكال واليس لتحليل التباين الأحادى Kruscak-Walis one way ANOVA

وهو يستخدم عندما يكون لدينا أكثر من مجموعتين مستقلتين، وهو يعد تعميماً لاختبار مان ويتنى، حيث يتم ترتيب درجات كل المجموعات معاً كأنها مجموعة واحدة، ثم يتم ترتيب كل مجموعة على حده وحساب قيمة إحصائية معينة ، يتم مقارنتها بجدول مربع كاى (كا^2) بدرجات حرية تساوى عدد المجموعات مطروحاً منه واحد، وتشترط هذه الطريقة أن لا يقل عدد كل مجموعة عن 5.

4- اختبار فريدمان لتحليل التباين الثنائى Friedman Two way ANOVA

إذا تضمنت الدراسة أكثر من مجموعتين مرتبطتين ، أو مجموعة واحدة تم قياسها عدة مرات فإننا نستخدم اختبار فريدمان لتحليل التباين الثنائى للرتب ، حيث يتم ترتيب درجات كل فرد فى القياسات المتكررة على حده ، وجمع رتب كل مجموعة ، ثم نحسب قيمة مربع كاى (كا2) ثم نقارنها بالقيم الجدولية باستخدام جدول مربع كاى بدرجات حرية تساوى عدد القياسات مطروحاً منه واحد

5- اختبار مربع كاى X^2 test

ويستخدم هذا الاختبار إذا كانت البيانات إسمية للمقارنة بين التكرارات الفعلية (الملاحظة) والمتوقعة، فاذا كانت الملاحظة مشابهة للمتوقعة فغننا نستنتج أن المجموعتين غير مختلفتين، أو أن المجموعات ليست مختلفة (فى حالة عدة مجموعات) ، ويشترط اختبار مربع كاى X2 test ألا يقل التكرار المتوقع عن خمسة فى 20% من الخلايا، ولا يوجد أى تكرار متوقع أقل من واحد.

17.7. تفسير مخرجات الانحدار و الأنوفا Interpretation of Regression & ANOVA Outputs

سنتناول في الجزء التالى تعريفا لبعض القيم التي نحصل عليها عند إجراء الانحدار Regression أو تحليل التباين ANOVA باستخدام الكمبيوتر

أ- معامل التحديد R^2 (Pearson's Correlation) Coefficient of Determination

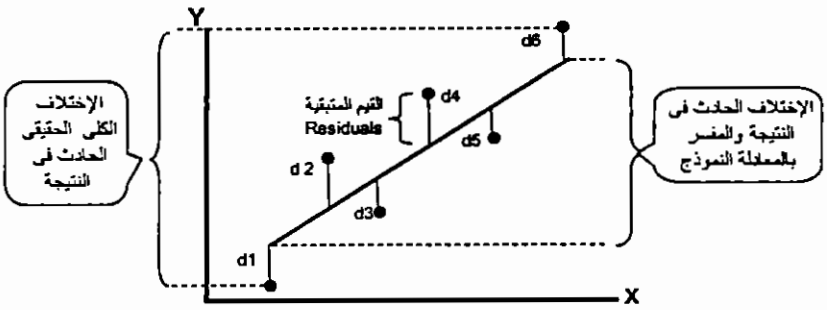
معامل للتعبير عن قوة العلاقة بين المتغير المستقل Dependant X أو المؤثر Predictor، والمتغير التابع Independent Y أو الإستجابة Response، ويساوى النسبة بين الاختلاف المفسر إلى الاختلاف الكلى، وبالتالي فهو يبين مدي صلاحية ودقة النموذج الخطى أو المعادلة المستنتجة، وتتراوح قيمته بين الصفر والواحد وكلما زادت قيمته كان أفضل، وتتحدد قيمته من المعادلة :

$$R^2 = \frac{\text{Explained Variation}}{\text{Total Variation}} = \frac{S_{xy}^{\ 2}}{S_{xx}S_{yy}} = \frac{SSR}{SST} = \frac{SST - SSE}{SST}$$

حيث إن SSR هو التباين المحسوب من خط الانحدار Variability accounted for by regression line ويسمى مجموع مربعات الانحدار Regression Sum of Squares ويحسب من العلاقة $SSR = \sum_{i=1}^{n}\left(\hat{y_i} - \bar{y}\right)^2$ ، وكذلك SSE التباين الغيرمحسوب من خط الانحدار Variability not accounted for by regression line ويسمى مجموع مربعات الخطأ Error Sum of Squares ويحسب من العلاقة

$$SSE = \sum_{i=1}^{n}\left(y_i - \hat{y}\right)^2$$

والشكل 17-1 يعرض تصورا يسهل من فهم العلاقة بين الاختلاف المفسر والاختلاف الكلى.



شكل رقم 17- 1 الاختلاف الكلى و الاختلاف المفسر

وعلى ذلك فان:

Total Error = Error Due To Regression + Error Due To Residua's

SSE+SST=SSR

فإذا كان 0.7= R2 فمعني ذلك أن 70% من التغيير الحادث في النتيجة Y كان بسبب المتغير X، وال30% الباقية قد تكون ناتجة عن متغيرات أخري مجهولة غير المتغير X، وكذلك إذا كان 0.68 = R^2 معني ذلك أن المعادلة المستنتجة عن طريق الانحدار Regression تشير إلى أن 68% من التباين الحادث في النتيجة Y هو بسبب المتغير X، وإذا كانت قيمة R^2 مساوية للصفر فإن ذلك يعنى انعدام تأثير المتغير X على النتيجة Y، أما إذا كانت قيمة R^2 مساوية للواحد الصحيح فإن ذلك يعنى تأثر النتيجة Y تأثرا تاما بالمتغير X ووقوع كافة النقاط على خط الانحدار.

ب- معامل التحديد المعدل Adjusted (R^2)

هو نفس المعامل السابق بعد زيادة دقة حسابه ودائما يكون أقل من R^2، فإذا كانت قيمة معامل التحديد المعدل Adjusted R^2 = 0.75 فإن ذلك يعنى أنه بعد إعادة حساب R^2 بدقة فإن 75% من التغيير الحادث في النتيجة Y هو بسبب المؤثر أوالمتغير X، و لكي يكون المتغير X من المتغيرات المؤثرة و نسميها RED X's، فيجب أن تكون قيمة معامل التحديد المعدل أكبر من 85% أى R^2 Adjusted > 0.85.

ج- مستوى الدلالة /معنوية القيمة بى P Value of Significance

وهو من أحد أهم العوامل التي نبحث عنها عند إجراء الانحدار أو تحليل التباين، ويعرف بأنه احتمال الخطأ فى رفض الفرض الصفرى الصحيح (ويطلق عليه اسم خطأ النوع الأول) ، وقيمته هي قيمة المساحة الطرفية تحت المنحنى بدءً من النقطة محل الدراسة،(ويشير إلى مدي التأكد من عدم وجود علاقة) ويتمتع هذا العامل بالخواص التالية:

1-تتراوح قيمته من صفر حتى الواحد الصحيح.

2-إذا كانت قيمته أقل من أو تساوى 5% ، فإننا نرفض فرضية العدم بعدم وجود علاقة، وبالتالي نقبل الفرضية البديلة بوجود علاقة بين المتغيرات، أما إذا كانت قيمته أكبر من 5% فإننا لا نستطيع أن نرفض الفرضية الصفرية، ونقول أنه ليس لدينا دليل كاف بعدم وجود علاقة There is no Evidence to Reject The Null Hypothesis (يرجى الرجوع إلى الفصل السابع عشر تحت عنوان اختبار الفرضيات).

3-كلما كانت قيمته صغيرة، كان احتمال رفض الفرض الصفرى وهو صحيح صغيراً

4-درجة تأكدنا من جود علاقة بين السبب X و النتيجة Y هي القيمة (1 – P)، فإذا كانت قيمة P = 0.03 ، فمعني ذلك أننا متأكدون بنسبة 3% فقط من عدم وجود علاقة، أي أننا متأكدون بنسبة 97% من وجود علاقة.

5-إذا كان لدينا أكثر من متغير X، مثلا ثمانية متغيرات X ونريد تحديد أكثرهم تأثيراً في النتيجة Y، فإننا نحسب P ، فتكون العوامل التي لها P أقل هي العوامل الأكثر تأثيراً في النتيجة Y.

6-فى نافذة النتائج فى برنامج مينيتاب ، إذا لم يحتوي العمود Lower 95% and Upper 95% علي قيمة " صفر " داخل مداه، أى إذا لم يحتوى المدى على قيم سالبة وأخرى موجبة، فإننا نكون متأكدين بنسبة 95% من عدم وجود علاقة، أى أننا لا نستطيع رفض فرضية العدم We Will Fail To Reject The Null Hypothesis.

د- الفرق بين معامل الارتباط و القيمة بى Difference between Coefficient of Correlation (r) and P value.

و بعد هذا الشرح والتوضيح نستطيع إجمال الفرق بين معامل الارتباط و القيمة بى فيما يلى:

1-المعامل P يحدد هل يوجد علاقة أم لا، وقيمته المطلقة تعبر عن قوة العلاقة وإشارته تعبر عن اتجاه ميل الخط.

2-المعامل (r) يحدد مدي الارتباط بين النقط والخط المستقيم، وقيمته المطلقة تعبر عن قرب النقط من الخط وإشارته تعبر عن نوع العلاقة هل هي عكسية أم طردية، ولكنها لا تعبر عن درجة ميل الخط.

هـ- الخطأ المعياري للفروق Standard Error of Residuals

كما قلنا أن هناك خطأ Error بين القيم الحقيقية للنقط والقيم المحسوبة من المعادلة فى شكل 16-9، والمعامل الذي بين أيدينا هو قيمة الانحراف المعياري لهذا الخطأ Error.

و- المشاهدات Observations

هي عدد القيم التي تم أخذها في الاعتبار عند إجراء دراسة الانحدار Regression (عدد النقط المرسومة)، بمعنى أنها حجم العينة Sample Size.

ز- درجات الحرية Degree of Freedom (DF)

هي عدد القيم التي تم أخذها في الاعتبار عند إجراء العملية الإحصائية مثل الانحدار Regression، مطروحا منها عدد العوامل التي تم حسابها، فلو كان حجم العينة Sample Size=10 وتم حساب المتوسط فإن درجات الحرية Degree of Freedom =10-1=9

ح- مجموع المربعات Sum of Squares (SS)

يعتبر مجموع المربعات Sum of Squares من القيم التي تعبر عن مدي قرب مجموعة النقاط المسجلة من الخط الممثل للعلاقة، وكلما زادت قيمتها دل ذلك علي وجود تبعثر و تباين بين النقاط وبين الخط، ومجموع المربعات ما هي إلا مجموع مربعات الفرق بين القيم الحقيقية والقيم المستنتجة للنقط، وفى بعض الأحيان نعبر عنها بمجموع المربعات للخطأ أو Sum of Squares Due to Error أو SSE، ولها أهمية كبرى في تقدير وحساب قيمة كل من معامل الارتباط، و تباين خط الانحدار.

ط- متوسط المربعات Mean of Squares (MS)

وهو مسمى آخر للتباين ، وقيمته تساوى قيمة مجموع المربعات Sum of Squares SS مقسوما على درجات الحرية Degree of Freedom DFF .

ي- المعامل الفائي F Factor

ويطلق عليه أيضا Computed F Statistics، وهو خارج قسمة القيمة الأكبر لأى من التباينين على القيمة الأصغر لهما، فإذا كانت هذه النسبة صغيرة وتقترب من الصفر أمكننا

أن نستنتج تجانس المجموعات التي نحلل تباينها، وأمكننا أن نرجع هذه المجموعات جميعا إلى أصل واحد، أما إذا كانت هذه النسبة كبيرة استنتجنا عدم تجانس المجموعات التي نحلل تباينها، وأمكننا أن نرجع هذه المجموعات إلى أصولها المختلفة التي تنسب إليها، ويطلق عليه أيضا القيمة الفائية المحسوبة F Computed، فإذا كانت قيمته تساوى 7.2518 فإن هذا يعنى أن الاختلاف بين المجموعات Variance Between Groups قد يكون أكبر من الاختلاف داخل المجموعة Variance Within Group أو أصغر منه ب 7.2518مرة، وهذا يشير إلى وجود اختلاف بين المجموعات.

ك- القيمة الفائية الحرجة F Critical

والقيمة الحرجة Critical Value هي القيمة الموجبة التي تجعل مساحة المنطقة أسفل الذيل الملتوي للمنحنى وخارج الفترة من F إلى مالا نهاية مساوية لقيمة معينة α عند درجتي الحرية $n\&m$، حيث $n\&m$ تساوى 1، 2، 3،.....

وهى القيمة الحرجة في حالة إجراء اختبار الفرضيات Hypothesis Testingعلى البيانات المتاحة لمقارنة الاختلاف بين مجموعتين Variance of Two Groups باستعمال التوزيع F.

ل- النسبة الفائية F Significance

وهي نسبة توضح إلي أي مدي نحن متأكدون من عدم وجود علاقة وكلما كانت صغيرة دل ذلك على عدم وجود علاقة.

م- عمود الخطأ المعيارى Standard Error Column

يعطى الانحراف المعياري لكل عامل من العوامل المدروسة.

ن- عمود الحالة تى T State Column

وهو معامل لتحديد مدي صحة المعادلة المستنتجة ويقيس بعد المعامل Factor عن قيمة الفرضية (صفر) بعدد من الانحرافات المعيارية Standard Deviation.

مثال رقم 17- 1:

بالتعرض للمثال الذي تناولناه في الفصل الخامس عشر(مثال رقم 15-1)، والذى يتعرض للعلاقة بين دخل الأسرة ومساحة المنزل الذى تملكه، فإذا كانت هذه البيانات كما فى الجدول 17-1 ونريد حساب كل من :

3- معامل الارتباط Correlation Coefficient

4- إيجاد معادلة الخط المستقيم باستخدام الانحدار وحساب المعاملات B1 & B0 ، وكذلك حساب القيم المتبقية Residual

جدول رقم 17- 1 للعلاقة بين دخل الأسرة ومساحة المنزل الذى تملكه

40	60	34	56	50	28	37	45	26	22	Income الدخل بالالف جنيه
20	30	18	32	21	22	24	26	17	16	Square Meters (Y) المساحة بالمئة متر المربع

أولا: الحل اليدوي ، نكون الجدول 17-2.

جدول رقم 17- 2 حل المثال رقم 71-1

Family	X Income	Y Square Meters	X Y	X^2	Y^2
1	22	16	352	484	256
2	26	17	442	676	289
3	45	26	1170	2025	676
4	37	24	888	1369	576
5	28	22	616	784	484
6	50	21	1050	2500	441
7	56	32	1792	3136	1024
8	34	18	612	1156	324
9	60	30	1800	3600	900
10	40	20	800	1600	900
	398	226	9522	17330	5370

ومن هذا الجدول

$$SSx = \sum X^2 - \frac{(\sum X)2}{N} = 17330 - \frac{(398)2}{10} = 1489.6$$

$$SCPxy = \sum XY - \frac{(\sum x)(\sum y)}{N} = 9522 - \frac{398\ X\ 226}{10} = 527.2$$

$$R = \frac{SCPxy}{\sqrt{SSx} * \sqrt{SSy}} = \frac{527.2}{\sqrt{1489.6} * \sqrt{262.4}} = \frac{527.2}{625.2} = 0.843$$

$$COV\ (X.Y) = \frac{1}{N-1} * SCPxy\ = \frac{1}{9} * 5272\ = 58.58$$

$$B_1 = \frac{SCP\ \times y}{SS\ \times} = \frac{527.2}{1489.6} = 0.354$$

$$B_0 = \bar{y} - b_1 \bar{X} = \frac{226}{10} - 0.354\ X\ \frac{398}{10} = 8.51$$

$$\hat{Y} = B_0 + B_1 X \therefore \hat{Y} = 8.51 + 0.354X$$

ولحساب القيم المتبقية Residuals تقوم بتجهيز الجدول التالي:

جدول رقم 17- 3 حل المثال 17-1

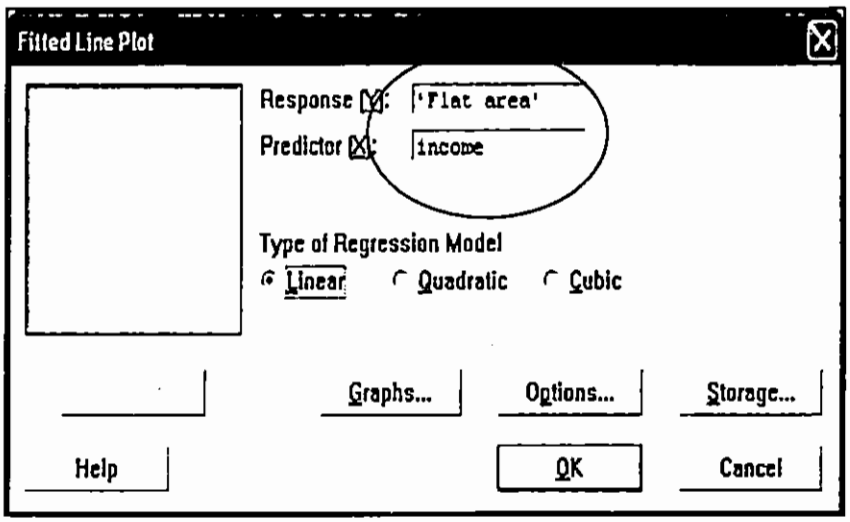
X Income	Y Square Meters	$\hat{Y}$	$(Y - \hat{Y})$	$(Y - \hat{Y})^2$
22	16	16.3	0.3	0.09
26	17	17.7	0.71	0.504
45	26	24.4	1.56	2.434
37	24	21.6	2.39	5.712
28	22	18.4	3.58	12.816
50	21	26.2	5.21	27.144
56	32	28.3	3.67	13.469
34	18	20.5	2.54	6.452
60	30	29.8	0.25	0.063
40	20	22.7	2.67	7.129

$$SSE = \sum\left(Y - \hat{Y}\right) = 75.81$$

$$\text{OR}\ \ SSE = SSy - \frac{(SCPxy)^2}{SSx} = 262.4 - \frac{(527.2)^2}{1489.6} = 75.81$$

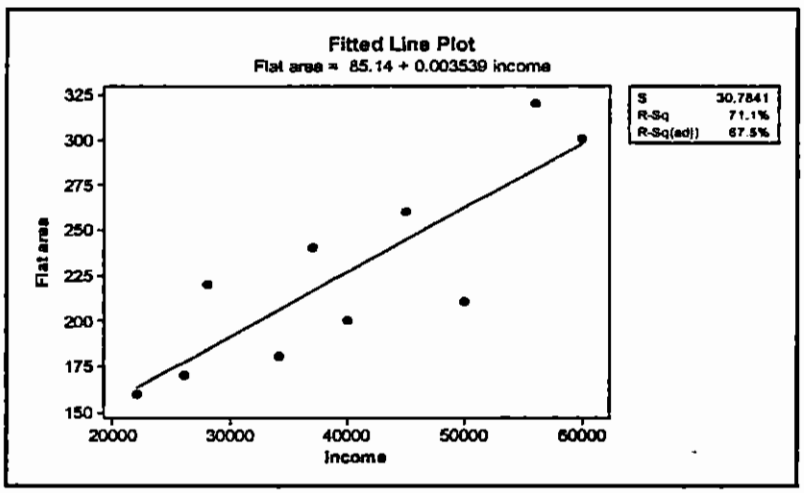
ثانيا: الحل باستخدام المينيتاب Minitab ومن القوائم:

2- Stat > Regression > Fitted Line Plot

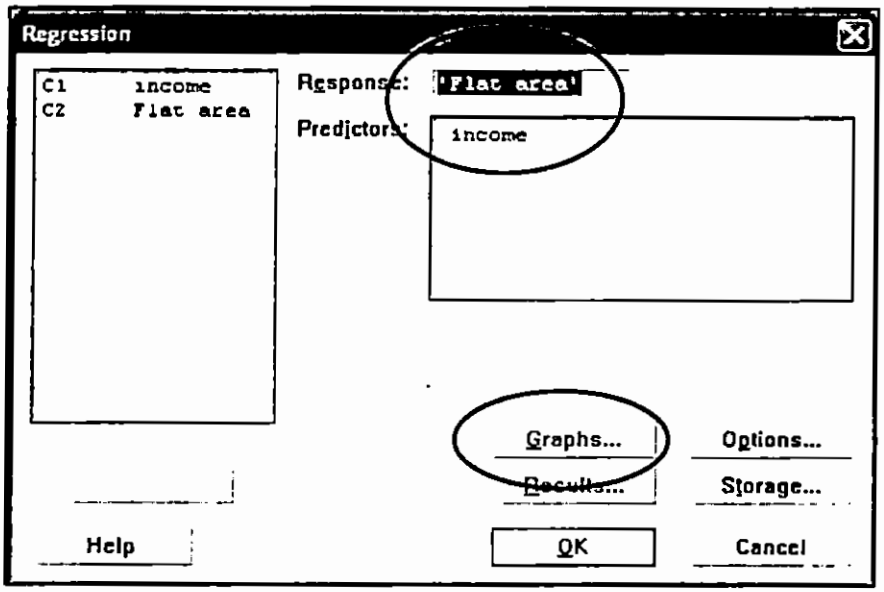


شكل رقم 17- 2 حل المثال 17-1 بالمينيتاب

فيظهر الشكل 17-2 ، فنختار المتغير المستقل وهو فى هذه الحالة "دخل الأسرة Income"، وكذلك نختار المتغير التابع وهو فى هذه الحالة "مساحة الشقة Flat rate" فيظهر الشكل 17-3 والذى يعكس العلاقة بين المتغيرين

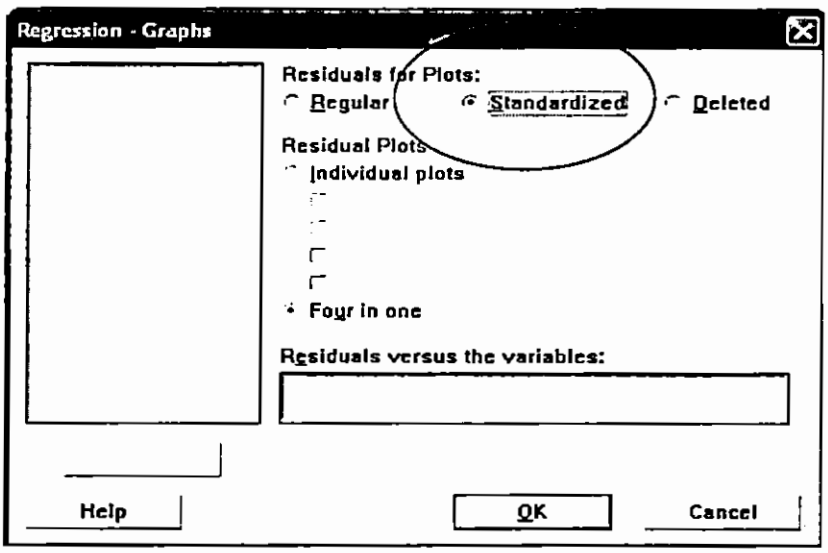


شكل رقم 17- 3 شكل العلاقة بين المتغيرين والناتج من المينيتاب



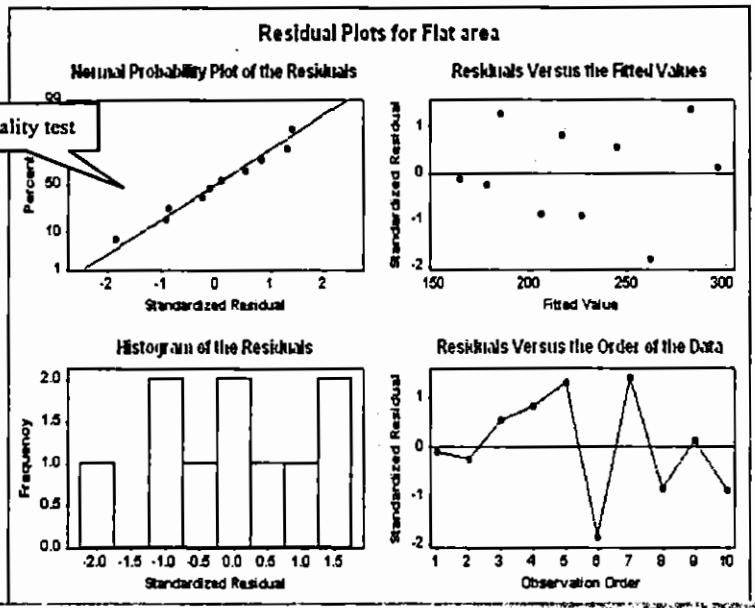
شكل رقم 17- 4 حل المثال 17-1 لايجاد معادلة الانحدار

فيظهر الشكل 17-5 :



شكل رقم 17- 5 صندوق حوارى فى المينيتاب لحل المثال 17-1

فنختار Standardized و Four in One ثم OK فيظهر الشكل 17-6:



شكل رقم 17- 6 المنحنيات الأربع التى تظهر فى نتائج المينيتاب

كذلك تظهر البيانات التالية فى نافذة النتائج Session Window :

```
Regression Analysis: Flat Area versus Income
The Regression Equation Is
Flat Area = 85.14 + 0.003539 Income
S = 30.7841   R-Sq = 71.1%   R-Sq (Adj) = 67.5%
Analysis Of Variance
Source       DF       SS       MS      F      P
Regression    1  18658.7  18658.7  19.69  0.002
Error         8   7581.3    947.7
Total         9  26240.0
```

شكل رقم 17- 7 نافذة النتائج للمثال 17-1

وما يهمنا عند التعليق على النتائج التي حصلنا عليها فى شكل 17- 6 وشكل 17-7 النقاط التالية:

1. يجب أن تكون قيمة معامل الارتباط R^2 أكبر من 85% لكى يكون الارتباط قويا، وحيث انها تساوى 71.1% فى مثالنا، فإن الارتباط يكون ضعيفاً .
2. يجب أن تكون قيمة P أقل من 0.05 حتى يكون النموذج معبراً عن البيانات.

وبالمقارنة بما كنا نفعله في خرائط المراقبة و التحكم Control Charts نجد أننا نكرر ما كنا نفعله تماماً (يرجى مراجعة الفصل الخامس تحت عنوان التحكم الإحصائى للعمليات) ،

كنا نختار عوامل معينة لنقيسها، ثم نحدد طريقة القياس، ثم ندرب القائمين على القياس، ونتأكد من كفاءة نظام القياس، ثم نضع برنامج وخطة للقياس، ثم نسجل قراءات ونرسم خرائط المراقبة والتحكم، ثم نحدد المؤثرات الخاصة Special Causes و المؤثرات العامة Common Causes، ونحاول إزالة أسبابها والتخلص منها، ثم نهمل القراءة المناظرة لها ثم نرسم خرائط المراقبة و التحكم Control Chart مرة أخرى، وهكذا حتى نحصل على خريطة خالية من أى أسباب خاصة Special Causes، وهذا ما سنفعله، أى سنقوم باختيار مجموعة من عوامل معينة لنحدد العلاقات بينها، ثم نحدد العوامل التي سنقيسها، ثم نحدد طريقة القياس، ثم ندرب القائمين على القياس، ونتأكد من كفاءة نظام القياس، ثم نضع برنامج وخطة للقياس، ثم نسجل قراءات ونرسم الانحدار، ثم من خلال الأربعة منحنيات السابقة ، نحاول تحديد النقط الغريبة/الشاذة المسببة لعدم تحقق الشروط الأربعة، ثم نحاول التخلص منها، أو التغلب عليها، ثم نهمل القراءة المناظرة لها، ثم نطبق الانحدار مرة أخرى، أو ندخل عوامل جديدة ونعيد الانحدار حتى نحصل على علاقات واضحة أو نجزم بعدم وجود علاقة أصلا.