مفهوم تحليل الانحدار

يعتبر العالم الإنكليزي Francis Galton (1822–1911) اول من استخدم مفهوم الانحدار في التطبيقات البيولوجية بهدف اكتشاف بعض العلاقات بين بعض المتغيرات البيولوجية.

ان موضوع تحليل الانحدار يهدف الى تقدير العلاقة الرياضية التي تربط متغيرين او اكثر مع بعض.

ان في تحليل الانحدار يوجد نوعين من المتغيرات، الأول هو المتغير التابع (Dependent variable) والثاني هو المتغير المستقل (او مجموع من المتغيرات المستقلة) (Independent variable).

والمتغير التابع هو المتغير الذي تتأثر قيمهُ في حالة تغير قيم المتغير المستقل ولا يتأثر بهِ.

على سبيل المثال انفاق الاسرة الشهري يتأثر بدخلها الشهري بينما العكس غير صحيح.

وغالباً ما يسمى المتغير المستقل بالمتغير المتنبأ بهِ او التنبؤي (Predictor Variable) او المتغير التوضيحي (Explained Variable).

تعريف تحليل الانحدار

ان تحليل الانحدار عبارة عن وسيلة إحصائية يستخدم لتحليل العلاقة بين متغير مستقل واحد او اكثر ومتغير تابع.

او يمكن تعريفهُ بشكل عام انهُ مقياس رياضي لمتوسط العلاقة بين متغيرين او اكثر بدلالة وحدات قياس المتغيرات الموجودة في العلاقة. وغالباً ما تسمى العلاقة من هذا النوع بنماذج الانحدار.

استعمالات تحليل الانحدار

1- وصف البيانات: يمكن تلخيص ووصف مجموعة من البيانات لدى الباحث لايجاد معادلة الانحدار التي تصف تلك البيانات.
2- تقدير المعلمات: يمكن تقدير معلمات النموذج المجهولة والتي يمكن من خلالها الاستدلال على أهمية وقوة واتجاه العلاقة بين المتغيرات.
3- التنبؤ: يمكن تقدير الاستجابة والتنبؤ بها بما يفيد كثيراً في التخطيط واتخاذ القرارات.
4- السيطرة: بعد إيجاد النماذج التي تصف العلاقة بين المتغيرات المستقلة والمتغير التابع يمكن السيطرة على قيم المتغير التابع من خلال التغيير في قيم المتغيرات المستقلة.

الانحدار والعلاقة السببية

المقصود بالعلاقة السببية هو ان اذا اثبت تحليل الانحدار وجود علاقة بين X و y فان ذلك لا يعني ان X هو المسبب لـ y، الا اذا تم اثبات ذلك بتجارب أخرى واخذ متغيرات أخرى بنظر الاعتبار.

أنواع الانحدار

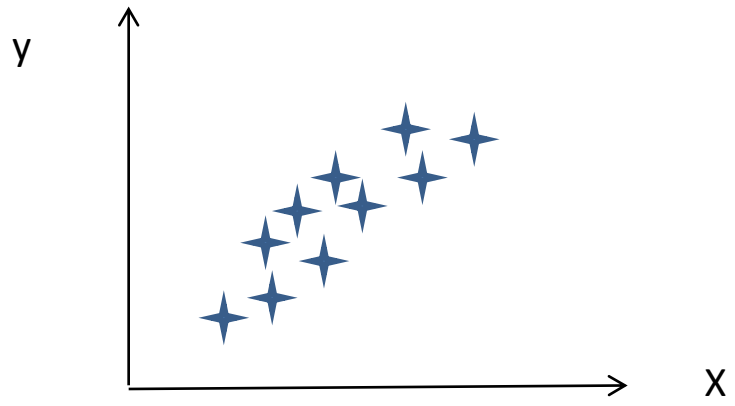
1- الانحدار الخطي البسيط Simple Linear Regression (في حالة عدم وجود تكرار في قيم Xi)

يتضمن الانحدار الخطي البسيط متغيرين، الأول هو المتغير التابع والثاني هو المتغير المستقل. ومثال على ذلك العلاقة بين العمر X وضغط الدم y لخمسين شخص في مدينة معينة.

ويمكن تعريف الانحدار الخطي البسيط بأنهُ عملية تقدير العلاقة بين متغيرين احدهما مستقل والأخر تابع.

ان الانحدار الخطي البسيط الذي يوضح العلاقة بين متغيرين يولد لنا زوج من القيم. فمثلاً كان المتغير X (المتغير المستقل) والمتغير y (المتغير التابع) هما المتغيران اللذان يمثلان زوج القيم (X و y) أي ان Xi تمثل القيم i من المتغير X والتي تقابل القيمة yi والتي تمثل القيمة i للمتغير y.

فاذا مثلنا زوج القيم (yi، Xi) سنحصل على رسم يدعى بالشكل الانتشاري (الشكل (1)). لاحظ الشكل ادناه الشكل (1)):



الشكل (1) الرسم الانتشاري لزوج القيم (yi، Xi)

ان الانحدار الخطي البسيط يمكن تمثيلهُ بمعادلة خطية تدعى معادلة خط الانحدار البسيط والتي تأخذ الشكل الاتي:

$$y_i = \beta_0 + \beta_1 x_{i1} + e_i \quad , \; i = 1,2,...,n$$

حيث ان:

y_i هي قيمة المتغير التابع او متغير الاستجابة.

x_i هي قيم المتغير المستقل.

β_0, β_1 هي ثوابت تدعى معلمات معادلة الانحدار Regression Parameters.

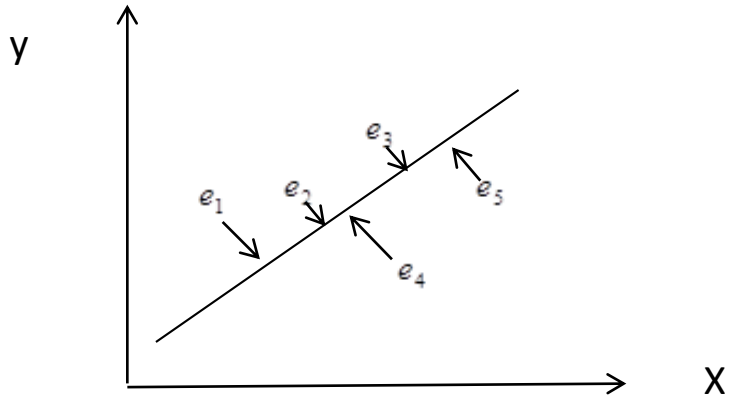
حيث ان

β_0 هي نقطة تقاطع خط الانحدار بالمحور y.

β_1 : معامل الانحدار لـ y على X_{i1} وهي تمثل مقدار التغير في y عند زيادة وحدة واحدة في X_{i1}.

e_i هو قيمة الخطأ العشوائي او المتبقي.

ويمكن توضيح الأخطاء العشوائية من خلال الشكل ادناه (الشكل (2)):



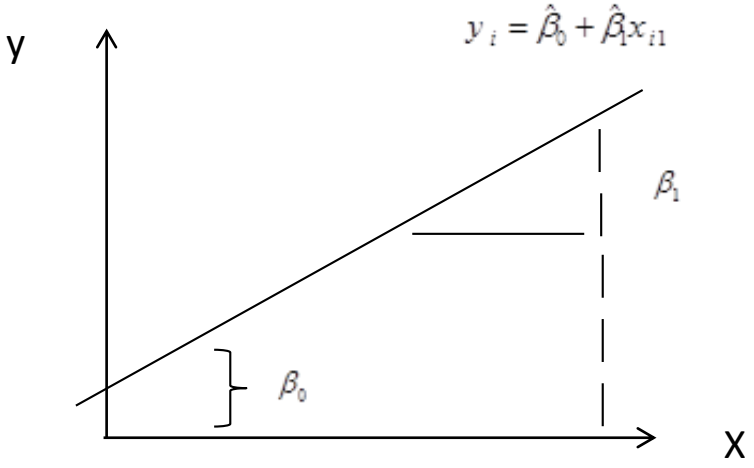
الشكل (2) يوضح الأخطاء العشوائية

من خلال معادلة خط الانحدار البسيط أعلاه (المعادلة (1)):

في حالة كانت قيمة β_1 تساوي صفر فان β_0 تمثل مقدار الزيادة او النقصان في قيمة y .

نلاحظ من الشكل أعلاه انهُ ومن خلال رسم زوج القيم (yi، Xi) فانهُ يمكن إيجاد معادلة خط الانحدار والتي تتخلل هذه النقاط بحيث يكون مجموع مربعات هذه الأخطاء اقل ما يمكن.

ان معادلة خط الانحدار يمكن ان توضح بشكل اكثر من خلال رسم هذه المعادلة وتحديد قيم المعلمات β_1 و β_0 ، لاحظ الشكل ادناه (الشكل (3)) :



الشكل (3)

يوضح معادلة خط الانحدار والمعلمات β_1 و β_0

نلاحظ ان β_0 هي نقطة تقاطع معادلة خط الانحدار مع المحور العمودي.

اما β_1 فهي تمثل ميل خط الانحدار (Slop) $\frac{\Delta y}{\Delta X}$ أي مقدار التغير الحاصل في y نتيجة زيادة وحدة واحدة في X.

كما ان إشارة المعلمة β_1 تمثل العلاقة بين X و y ، فاذا كانت موجبة فان العلاقة طردية واذا كانت سالبة فان العلاقة عكسية.

بالنسبة الى الخطأ e_i فانهُ يمثل الفرق بين القيمة المشاهدة y_i والقيمة المقدرة $\hat{y}_i$ أي ان:

$$e_i = y_i - \hat{y}_i$$

حيث ان

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1}$$

حيث ان

$\hat{\beta}_0$: هي المعلمة المقدرة للمعلمة β_0 .

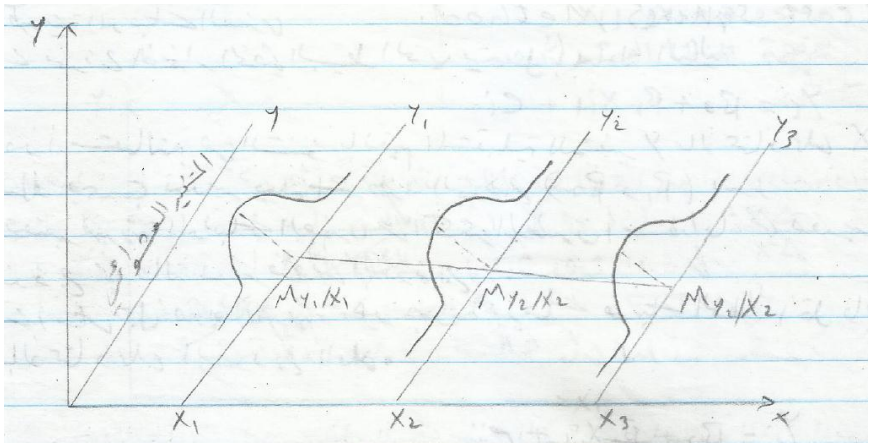
$\hat{\beta}_1$: هي المعلمة المقدرة للمعلمة β_1 .

$\hat{y}_i$: هي القيمة المقدرة للقيمة y_i .

فروض التحليل Assumptions of analysis

1- ان y هو متغير عشوائي (Random variable) وقيمهُ مستقلة احصائياً الواحدة عن الأخرى وتتوزع توزيعاً طبيعياً Normal Distribution بوسط حسابي قدرهُ $\mu_{y/x}$ وتباين $\sigma^2 = \sigma^2_{y/x}$.

ان y يعد متغيراً عشوائياً لان قيمهُ تتغير من قيمة الى أخرى من قيم X. ويرمز عادتاً للمتغير العشوائي y التابع لقيمة ثابتة من X بالرمز y(xi) ولتوزيعهُ الاحتمالي بالرمز F(y/x) لاحظ الشكل ادناه:



التوزيعات المختلفة لـ y المقابلة لـ X

ان متوسط قيم y (أي $\mu_{y/x}$) هي دالة خط مستقيم أي ان قيمة $\mu_{y/x}$ تقع على خط مستقيم.

ان تباين y ($\sigma^2_{y/X}$) لهُ قيمة واحدة من أي قيمة من قيم X أي ان $\sigma^2_{y/X} = \sigma^2$ وهذه الخاصية تدعى Homosccdasticity أي ثبات او تجانس تباين الخطأ.

2- ان e_i هو خطأ عشوائي Random error ويتوزع توزيعاً طبيعياً بوسط حسابي قدرهُ صفر وتباين σ^2 أي ان $e_i \sim N(0, \sigma_e^2)$

3- ان قيم e_{is} غير مرتبطة uncorrelated أي ان التباين المشترك بينهم يساوي صفراً أي ان $E(ee) = 0$ أي ان قيمة e_i في فترة ما غير مرتبطة مع قيمتها في فترة أخرى.

4- ان قيم المتغير X_{i1} ثابتة وتقاس بدون خطأ وقيمهُ غير مرتبطة بقيم الخطأ أي ان $E(x_i' e_i) = 0$.

تقدير معلمات الانحدار

1- طريقة المربعات الصغرى Least Squares Method

ان نموذج الانحدار الخطي البسيط الذي يتمثل بالمعادلة الاتية:

$$y_i = \beta_0 + \beta_1 x_{i1} + e_i \quad , \quad i = 1, 2, ..., n$$

احد استعمالاتهِ هو التنبؤ بالقيم المستقبلية للمتغير y بالاعتماد على X بذلك يجب تقدير معلمات هذا النموذج (β_1, β_0).

وتعتبر طريقة المربعات الصغرى من اكثر الطرق استخداماً في تقدير معلمات النموذج في حالة توفر فروض النموذج.

ان أساس عمل هذه الطريقة هو جعل مجموع مربعات الخطأ اقل ما يمكن. وبالاعتماد على النموذج أعلاه:

$$y_i = \beta_0 + \beta_1 x_{i1} + e_i$$

فان:

$$e_i = y_i - \beta_0 - \beta_1 x_{i1}$$

وبإيجاد مجموع مربعات الخطأ (او البواقي) وليكن Q فان:

$$Q = \sum e_i^2 = \sum (y_i - \beta_0 - \beta_1 x_{i1})^2$$

وبأخذ المشتقة الجزئية للمعادلة أعلاه بالنسبة الى (β_1, β_0) نحصل على:

$$\frac{\partial Q}{\partial \beta_0} = -2\sum (y_i - \beta_0 - \beta_1 x_{i1})$$

$$\frac{\partial Q}{\partial \beta_1} = -2\sum (y_i - \beta_0 - \beta_1 x_{i1}) x_{i1}$$

اذ ان $(\hat{\beta}_1, \hat{\beta}_0)$ هما المعلمتان المقدرتان.

وبمساواة كلا المعادلتان بالصفر نحصل على:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum X_{i1} = \sum y_i \quad \dots \quad (1)$$

$$\hat{\beta}_0 \sum X_{i1} + \hat{\beta}_1 \sum X_{i1}^2 = \sum X_{i1} y_i \quad \dots \quad (2)$$

تسمى المعادلتين أعلاه بالمعادلتين الطبيعيتين normal equations.

ومن خلال المعادلة (1):

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum X_{i1} = \sum y_i$$

نحصل على:

$$n\hat{\beta}_0 = \sum y_i - \hat{\beta}_1 \sum X_{i1}$$

$$\hat{\beta}_0 = \frac{\sum y_i}{n} - \hat{\beta}_1 \frac{\sum X_{i1}}{n}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}$$

نعوض عن $\hat{\beta}_0$ بالمعادلة (2) لنحصل على قيمة $\hat{\beta}_1$ وكما يلي:

$$\left(\frac{\sum y_i}{n} - \hat{\beta}_1 \frac{\sum X_{i1}}{n}\right) \sum X_{i1} + \hat{\beta}_1 \sum X_{i1}^2 = \sum X_{i1} y_i$$

$$\frac{(\sum y_i)(\sum X_{i1}) - \hat{\beta}_1(\sum X_{i1})^2}{n} + \hat{\beta}_1 \sum X_{i1}^2 = \sum X_{i1} y_i$$

$$(\sum y_i)(\sum X_{i1}) - \hat{\beta}_1(\sum X_{i1})^2 + n\hat{\beta}_1 \sum X_{i1}^2 = n\sum X_{i1} y_i$$

وبنقل الحدود التي تحتوي على المعلمة المقدرة $\hat{\beta}_1$ نحصل على:

$$\hat{\beta}_1 \left[n\sum X_{i1}^2 - (\sum X_{i1})^2 \right] = n\sum X_{i1} y_i - (\sum y_i)(\sum X_{i1})$$

$$\therefore \hat{\beta}_1 = \frac{n\sum X_{i1} y_i - (\sum y_i)(\sum X_{i1})}{n\sum X_{i1}^2 - (\sum X_{i1})^2}$$

وبالقسمة على n للبسط والمقام نحصل على:

$$\therefore \hat{\beta}_1 = \frac{\sum X_{i1} y_i - \frac{(\sum y_i)(\sum X_{i1})}{n}}{\sum X_{i1}^2 - \frac{(\sum X_{i1})^2}{n}} = \frac{S_{Xy}}{S_{XX}}$$

حيث ان:

$$\begin{aligned} S_{XX} &= \sum X_{i1}^2 - \frac{(\sum X_{i1})^2}{n} \\ &= \sum X_{i1}^2 - n\bar{X}^2 \\ &= \sum (X_i - \bar{X})^2 \\ &= \sum (X_i - \bar{X}) X_i \end{aligned}$$

وان:

$$\begin{aligned} S_{Xy} &= \sum X_{i1} y_i - \frac{(\sum y_i)(\sum X_{i1})}{n} \\ &= \sum X_{i1} y_i - n\bar{X}\,\bar{y} \\ &= \sum (X_i - \bar{X}) y_i \\ &= \sum (y_i - \bar{y}) X_i \\ &= \sum (X_i - \bar{X})(y_i - \bar{y}) \end{aligned}$$

خصائص معادلة خط الانحدار

1- مجموع القيم المشاهدة تساوي مجموع القيم التنبؤية.

Proof:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1}$$
$$\hat{y}_i = (\bar{y} - \hat{\beta}_1 \bar{X}) + \hat{\beta}_1 x_{i1}$$
$$\hat{y}_i = \bar{y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 x_{i1}$$
$$\hat{y}_i = \bar{y} + \hat{\beta}_1 (x_{i1} - \bar{X})$$

وبأخذ المجموع للطرفين:

$$\sum \hat{y}_i = n\bar{y} + \hat{\beta}_1 \sum (x_{i1} - \bar{X})$$

وبما ان: $\sum (x_i - \bar{X}) = 0$ بذلك فان:

$$\sum \hat{y}_i = n\bar{y}$$
$$\sum \hat{y}_i = \cancel{n} \frac{\sum y_i}{\cancel{n}}$$
$$\therefore \sum \hat{y}_i = \sum y_i$$

2- مجموع الأخطاء العشوائية يساوي صفر أي ان:

$$\sum e_i = 0$$

Proof:

بما ان:

$$e_i = y_i - \hat{y}_i$$

وبأخذ المجموع للطرفين نحصل:

$$\sum e_i = \sum (y_i - \hat{y}_i)$$
$$\sum e_i = \sum y_i - \sum \hat{y}_i$$

وبما ان $\sum \hat{y}_i = \sum y_i$ بذلك فان:

$$\sum e_i = 0$$

3- مجموع الأخطاء (او البواقي) الموزون بقيم Xi المناظرة، تساوي صفر.

اي ان:

$$\sum e_i X_i = 0$$

Proof:

$$\sum e_i X_i = \sum X_i (y_i - \hat{y}_i)$$

وبما ان:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1}$$

بذلك فان:

$$\sum e_i X_i = \sum X_i (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1}\ ))$$

وبما ان

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}$$

بذلك فان:

$$\sum e_i X_i = \sum X_i (y_i - ((\bar{y} - \hat{\beta}_1 \bar{X}) + \hat{\beta}_1 x_{i1}\ ))$$

بعد التبسيط واستخراج المعلمة المقدرة $\hat{\beta}_1$ كعامل مشترك نحصل على:

$$\begin{aligned}\sum e_i X_i &= \sum X_i (y_i - (\bar{y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 x_{i1})) \\ &= \sum X_i (y_i - (\bar{y} + \hat{\beta}_1 (x_{i1} - \bar{X}\ )) \\ &= \sum X_i ((y_i - \bar{y}) - \hat{\beta}_1 (x_{i1} - \bar{X}\ ))\end{aligned}$$

بعد ادخال $\sum X_i$ على القوس نحصل على:

$$\sum e_i X_i = \sum X_i (y_i - \bar{y}) - \hat{\beta}_1 \sum X_i (x_{i1} - \bar{X}\ ))$$

وبما ان:

$$S_{Xy} = \sum (y_i - \bar{y}) X_i$$
$$S_{XX} = \sum (X_i - \bar{X}\ ) X_i$$

نحصل على:

$$\sum e_i X_i = S_{Xy} - \hat{\beta}_1 S_{XX}$$

وبما ان:

$$\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}}$$

بذلك فان:

$$\sum e_i X_i = S_{Xy} - \frac{S_{Xy}}{S_{XX}} S_{XX}$$

$$\sum e_i X_i = S_{Xy} - S_{Xy}$$
$$\therefore \sum e_i X_i = 0$$

4- وزن الأخطاء بقيم y_i المناظرة ينتج مجموع مربعات الخطأ.

Proof:

$$\sum y_i e_i = \sum y_i (y_i - \hat{y}_i)$$
$$\because \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$
$$\therefore \sum y_i e_i = \sum y_i (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))$$
$$\because \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}$$
$$\therefore \sum y_i e_i = \sum y_i (y_i - ((\bar{y} - \hat{\beta}_1 \bar{X}) + \hat{\beta}_1 x_i)$$
$$\sum y_i e_i = \sum y_i (y_i - (\bar{y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 x_i)$$

بأخذ $\hat{\beta}_1$ عامل مشترك وفتح القوس نحصل على:

$$\sum y_i e_i = \sum y_i (y_i - (\bar{y} + \hat{\beta}_1 (x_i - \bar{X}_1))$$
$$\sum y_i e_i = \sum y_i (y_i - \bar{y} - \hat{\beta}_1 (x_i - \bar{X}))$$
$$\sum y_i e_i = \sum y_i ((y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{X}))$$

وبضرب $\sum y_i$ بالقوس نحصل على:

$$\sum y_i e_i = \sum y_i (y_i - \bar{y}) - \hat{\beta}_1 \sum y_i (x_i - \bar{X})$$
$$\because S_{yy} = \sum y_i (y_i - \bar{y})$$
$$\because S_{Xy} = \sum y_i (x_i - \bar{X})$$
$$\therefore \sum y_i e_i = S_{yy} - \hat{\beta}_1 S_{Xy}$$
$$\therefore SSe = S_{yy} - \hat{\beta}_1 S_{Xy}$$

بذلك مجموع مربعات الاخطاء (البواقي) هو:

SSe =Residual sum of squares

5- وزن الأخطاء بالقيم التنبؤية المناظرة والتي تساوي صفر.

أي ان

$$\sum \hat{y}_i e_i = 0$$

Proof:

$$\hat{y}_i = (\bar{y} + \hat{\beta}_1 (x_i - \bar{X}))$$
$$e_i = (y_i - (\bar{y} + \hat{\beta}_1 (x_i - \bar{X}))$$

بذلك فان:

$$\sum \hat{y}_i e_i = \sum (\bar{y} + \hat{\beta}_1 (x_i - \bar{X}))(y_i - (\bar{y} + \hat{\beta}_1 (x_i - \bar{X}))$$

من خلال القوسين يتم ترتيب القوس الثاني وكما يلي:

$$\sum \hat{y}_i e_i = \sum (\bar{y} + \hat{\beta}_1 (x_i - \bar{X}))((y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{X}))$$

يتم ضرب القوسين مع بعضهما وكما يلي:

$$\sum \hat{y}_i e_i = \sum (\bar{y}(y_i - \bar{y}) - \bar{y}\hat{\beta}_1 (x_i - \bar{X}) + \hat{\beta}_1 (x_i - \bar{X})(y_i - \bar{y}) - (\hat{\beta}_1 (x_i - \bar{X}))^2)$$

بإدخال المجموع الى القوس نحصل على:

$$\sum \hat{y}_i e_i = \bar{y}\sum (y_i - \bar{y}) - \bar{y}\hat{\beta}_1 \sum (x_i - \bar{X}) + \hat{\beta}_1 \sum (x_i - \bar{X})(y_i - \bar{y}) - \hat{\beta}_1^2 \sum (x_i - \bar{X})^2$$

وبذلك ستكون النتيجة:

$$\sum \hat{y}_i e_i = \hat{\beta}_1 \sum (x_i - \bar{X})(y_i - \bar{y}) - \hat{\beta}_1^2 \sum (x_i - \bar{X})^2$$

بما ان:

$$\because S_{Xy} = \sum y_i (x_i - \bar{X})$$

$$\because S_{XX} = \sum (x_i - \bar{X})^2$$

بذلك فان:

$$\sum \hat{y}_i e_i = \hat{\beta}_1 S_{Xy} - \hat{\beta}_1^2 S_{XX}$$

ويمكن اختصار المعادلة أعلاه كما يلي:

$$\sum \hat{y}_i e_i = \hat{\beta}_1 S_{Xy} - \hat{\beta}_1 \frac{S_{Xy}}{S_{XX}} S_{XX}$$

$$\sum \hat{y}_i e_i = \hat{\beta}_1 S_{Xy} - \hat{\beta}_1 S_{Xy}$$

$$\sum \hat{y}_i e_i = 0$$

6- ان المعلمة المقدرة $\hat{\beta}_1$ هو تقدير غير متحيز للمعلمة β_1

أي ان $E(\hat{\beta}_1) = \beta_1$

Proof:

$$E(\hat{\beta}_1) = E\left[\frac{S_{xy}}{S_{xx}}\right] = E\left[\frac{\sum(x_i - \bar{X})y_i}{\sum(x_i - \bar{X})^2}\right]$$

بإدخال التوقع الى الطرفين نحصل على:

$$= \frac{\sum(x_i - \bar{X})y_i}{\sum(x_i - \bar{X})^2} = \frac{\sum(x_i - \bar{X})(\beta_0 + \beta_1 x_i + e_i)}{\sum(x_i - \bar{X})^2}$$

$$= \frac{\beta_0\sum(x_i - \bar{X}) + \beta_1\sum(x_i - \bar{X})x_i + \sum(x_i - \bar{X})e_i)}{\sum(x_i - \bar{X})^2}$$

$$= \frac{\beta_1\sum(x_i - \bar{X})x_i + \sum(x_i - \bar{X})e_i)}{\sum(x_i - \bar{X})^2}$$

مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر

بما ان $S_{XX} = \sum(x_i - \bar{X})x_i$ ، وادخال المجموع وضرب الـ e_i على القوس في الحد الثاني نحصل على:

$$= \frac{\beta_1 S_{XX} + \sum x_i e_i - \bar{X}\sum e_i)}{S_{XX}}$$

بما ان $\sum x_i e_i = 0$ وان $\sum e_i = 0$ بذلك فان:

$$= \frac{\beta_1 S_{XX}}{S_{XX}}$$

$$\therefore E(\hat{\beta}_1) = \beta_1$$

7- ان المعلمة المقدرة $\hat{\beta}_0$ هو تقدير غير متحيز للمعلمة β_0

أي ان $E(\hat{\beta}_0) = \beta_0$

Proof:

$$E(\hat{\beta}_0) = E(\bar{y} - \hat{\beta}_1\bar{X})$$

$$= \bar{y} - E(\hat{\beta}_1)\bar{X}$$

$$= \left(\frac{\sum y_i}{n}\right) - \bar{X}\beta_1$$

$$= \left(\frac{\sum(\beta_0 + \beta_1 x_i + e_i)}{n}\right) - \bar{X}\beta_1$$

وبإدخال المجموع على القوس نحصل على:

$$= \left(\frac{n\beta_0 + \beta_1 \sum x_i + \sum e_i)}{n} \right) - \bar{X}\,\beta_1$$

$$= \frac{n\beta_0}{n} + \beta_1 \frac{\sum X_i}{n} + \frac{\sum e_i}{n} - \bar{X}\,\beta_1$$

$$= \beta_0 + \beta_1 \bar{X} + 0 - \beta_1 \bar{X}$$

$$\therefore E(\hat{\beta}_0) = \beta_0$$

8- تقدير تباين المجتمع σ^2

SSe =Error Sum of Square

=Residual Sum of squares

يمكن تقدير تباين المجتمع من خلال مجموع مربعات الخطأ وكما يلي:

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

بما ان: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ وان $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}$ بذلك فان:

$$SSe = \sum (y_i - ((\bar{y} - \hat{\beta}_1 \bar{X}) + \hat{\beta}_1 x_i))^2$$

بتبسيط المعادلة أعلاه نحصل على:

$$SSe = \sum ((y_i - \bar{y}) + \hat{\beta}_1 \bar{X} - \hat{\beta}_1 x_i))^2$$

باستخراج $\hat{\beta}_1$ عامل مشترك نحصل على:

$$SSe = \sum ((y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{X}))^2$$

بفتح القوس بالتربيع:

$$SSe = \sum ((y_i - \bar{y})^2 - 2\hat{\beta}_1 (x_i - \bar{X})(y_i - \bar{y}) + \hat{\beta}_1^2 (x_i - \bar{X})^2)$$

وبإدخال المجموع على القوس نحصل على:

$$SSe = \sum (y_i - \bar{y})^2 - 2\hat{\beta}_1 \sum (x_i - \bar{X})(y_i - \bar{y}) + \hat{\beta}_1^2 \sum (x_i - \bar{X})^2$$

وبما ان: $S_{XX} = \sum (x_i - \bar{X})^2, S_{yy} = \sum (y_i - \bar{y})^2, S_{Xy} = \sum (x_i - \bar{X})(y_i - \bar{y})$:

$$SSe = S_{yy} - 2\hat{\beta}_1 S_{Xy} + \hat{\beta}_1^2 S_{XX}$$

أي ان:

$$SSe = S_{yy} - 2\hat{\beta}_1 S_{Xy} + \hat{\beta}_1 \hat{\beta}_1 S_{XX}$$

وبما ان $\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}}$:

$$SSe = S_{yy} - 2\hat{\beta}_1 S_{Xy} + \hat{\beta}_1 \frac{S_{Xy}}{S_{XX}} S_{XX}$$

$$SSe = S_{yy} - 2\hat{\beta}_1 S_{Xy} + \hat{\beta}_1 S_{Xy}$$

أي ان:

$$SSe = S_{yy} - \hat{\beta}_1 S_{Xy}$$

ان مجموع مربعات الخطأ SSe لهُ (n−2) من درجات الحرية لأنه يحتاج الى درجتين لتقدير β_0, β_1 للحصول على $\hat{y}_i$ بذلك فان:

$$\hat{\sigma}^2 = S^2_{y/x} = MSe = V(y_i) = \frac{SSe}{n-2} = \frac{S_{yy} - \hat{\beta}_1 S_{Xy}}{n-2}$$

9- تقدير تباينات معامل الانحدار β_1

أي ان: $V(\hat{\beta}_1)$

Proof:

$$\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}} = \frac{\sum (x_i - \bar{X}) y_i}{\sum (x_i - \bar{X})^2} = \sum \left(\frac{(x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \right) y_i$$

If we have

$$a = a_1 y_1 + a_2 y_2 + ... + a_n y_n$$

بأخذ التباين للطرفين نحصل على:

$$V(a) = a_1^2 V(y_1) + a_2^2 V(y_2) + ... + a_n^2 V(y_n)$$

وبما ان $V(y_1) = V(y_2) = ... = V(y_n) = V(y_i)$ بذلك فأن:

$$V(a) = V(y_i)(a_1^2 + a_2^2 + ... + a_n^2)$$

$$V(a) = \sum_{i=1}^{n} a_i^2 V(y_i)$$

بذلك فان:

$$V(\hat{\beta}_1) = V\left(\sum \left(\frac{(x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \right) y_i \right)$$

$$V(\hat{\beta}_1) = \sum \left(\frac{(x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \right)^2 V(y_i)$$

$$V(\hat{\beta}_1) = \sum \left(\frac{(x_i - \bar{X})}{S_{XX}} \right)^2 V(y_i)$$

$$V(\hat{\beta}_1) = \frac{\sum (x_i - \bar{X})^2}{S_{XX}^2} V(y_i)$$

$$V(\hat{\beta}_1) = \frac{S_{XX}}{S_{XX}^2} \hat{\sigma}^2$$

$$\therefore V(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{XX}}$$

10- تقدير تباينات معامل الانحدار β_0

أي ان: $V(\hat{\beta}_0)$

Proof:

$$V(\hat{\beta}_0) = V(\bar{y} - \hat{\beta}_1 \bar{X})$$

$$= V(\bar{y} + (-\bar{X})\hat{\beta}_1)$$

بإدخال التباين للقوس في الطرف الأيمن نحصل على:

$$V(\hat{\beta}_0) = V(\bar{y}) + 2\operatorname{cov}(\bar{y}, (-\bar{X})\hat{\beta}_1) + \bar{X}^2 V(\hat{\beta}_1)$$

الحد الأول $V(\bar{y})$:

Proof:

$$\begin{aligned}
\bar{y} = \frac{\sum y_i}{n} &= \left(\frac{1}{n}y_1 + \frac{1}{n}y_2 + \ldots + \frac{1}{n}y_n\right) \\
&= \frac{1}{n^2}V(y_1) + \frac{1}{n^2}V(y_2) + \ldots + \frac{1}{n^2}V(y_n) \\
&= \frac{n}{n^2}V(y_i) \\
&= \frac{V(y_i)}{n} \\
&= \frac{\hat{\sigma}^2}{n}
\end{aligned}$$

الحد الثاني $\operatorname{cov}(\bar{y}, (-\bar{X})\hat{\beta}_1)$:

Proof:

$$\operatorname{cov}(\bar{y}, (\bar{X})\hat{\beta}_1) = (-\bar{X})\operatorname{cov}(\bar{y}, \hat{\beta}_1)$$

$$a = a_1y_1 + a_2y_2 + \ldots + a_ny_n$$

$$c = c_1y_1 + c_2y_2 + \ldots + c_ny_n$$

$$\begin{aligned}
\operatorname{cov}(a,c) &= a_1c_1\operatorname{cov}(y_1,y_1) + a_2c_2\operatorname{cov}(y_2,y_2) + \ldots + a_nc_n\operatorname{cov}(y_n,y_n) \\
&= a_1c_1V(y_1) + a_2c_2V(y_2) + \ldots + a_nc_nV(y_n)
\end{aligned}$$

وبما ان $V(y_1) = V(y_2) = \ldots = V(y_n) = V(y_i)$ بذلك فأن:

$$\begin{aligned}
\operatorname{cov}(a,c) &= (a_1c_1 + a_2c_2 + \ldots + a_nc_n)V(y_i) \\
&= (a_1c_1 + a_2c_2 + \ldots + a_nc_n)\hat{\sigma}^2 \\
&= \sum_{i=1}^{n}(a_ic_i)\hat{\sigma}^2
\end{aligned}$$

نفرض ان $a = \bar{y}$, $c = \hat{\beta}_1$ أي ان:

$$\begin{aligned}
\bar{y} = \frac{\sum y_i}{n} &= \left(\frac{1}{n}y_1 + \frac{1}{n}y_2 + \ldots + \frac{1}{n}y_n\right) \\
&= \left(\frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n}\right)y_i
\end{aligned}$$

$$\therefore a_i = \frac{1}{n}$$

وان:

$$\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}} = \frac{\sum (x_i - \bar{X})y_i}{\sum (x_i - \bar{X})^2} = \frac{(x_1 - \bar{X})y_1}{\sum (x_i - \bar{X})^2} + \frac{(x_2 - \bar{X})y_2}{\sum (x_i - \bar{X})^2} + \ldots + \frac{(x_n - \bar{X})y_n}{\sum (x_i - \bar{X})^2}$$

$$\hat{\beta}_1 = \left(\frac{(x_1 - \bar{X})}{\sum (x_i - \bar{X})^2} + \frac{(x_2 - \bar{X})}{\sum (x_i - \bar{X})^2} + \ldots + \frac{(x_n - \bar{X})}{\sum (x_i - \bar{X})^2} \right) y_i$$

بذلك فان: $c = \frac{(x_i - \bar{X})}{\sum (x_i - \bar{X})^2}$

عليهِ فان:

$$\operatorname{cov}(\bar{y}, \hat{\beta}_1) = \operatorname{cov}(a, c) = \sum \left(\frac{1}{n} \frac{(x_i - \bar{X})}{\sum (x_i - \bar{X})^2} \right) \hat{\sigma}^2$$

$$= \left(\frac{1}{n} \frac{\sum (x_i - \bar{X})}{S_{XX}} \right) \hat{\sigma}^2$$

مجموع انحرافات القيم عن وسطها الحسابي يساوي صفر

$$\operatorname{cov}(\bar{y}, \hat{\beta}_1) = 0$$

الحد الثالث $\bar{X}^2 V(\hat{\beta}_1)$:

Proof:

$$\bar{X}^2 V(\hat{\beta}_1) = \bar{X}^2 \frac{\hat{\sigma}^2}{S_{XX}}$$

بذلك فان $V(\hat{\beta}_0)$:

$$V(\hat{\beta}_0) = \frac{\hat{\sigma}^2}{n} + \bar{X}^2 \frac{\hat{\sigma}^2}{S_{XX}}$$

$$V(\hat{\beta}_0) = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}} \right)$$

$$V(\hat{\beta}_0) = \hat{\sigma}^2 \left(\frac{S_{XX} + n\bar{X}^2}{nS_{XX}} \right)$$

$$V(\hat{\beta}_0) = \hat{\sigma}^2 \left(\frac{\sum X_i^2 - n\bar{X}^2 + n\bar{X}^2}{nS_{XX}} \right)$$

$$V(\hat{\beta}_0) = \hat{\sigma}^2 \frac{\sum X_i^2}{nS_{XX}}$$

11- تقدير تباين متوسط الاستجابة

أي ان $V(\hat{y}_0)$:

Proof:

$$\begin{aligned}\hat{y}_0 &= \hat{\beta}_0 + \hat{\beta}_1 X_0 \\ &= (\bar{y} - \hat{\beta}_1 \bar{X}) + \hat{\beta}_1 X_0 \\ &= \bar{y} + \hat{\beta}_1 (X_0 - \bar{X})\end{aligned}$$

باخذ التباين للطرفين نحصل على:

$$\begin{aligned}V(\hat{y}_0) &= V(\bar{y} + \hat{\beta}_1 (X_0 - \bar{X})) \\ &= V(\bar{y}) + \operatorname{cov}(\bar{y}, \hat{\beta}_1 (X_0 - \bar{X})) + (X_0 - \bar{X})^2 V(\hat{\beta}_1)\end{aligned}$$

وقد تم برهان ان $\operatorname{cov}(\bar{y}, \hat{\beta}_1) = 0$ بذلك فان:

$$V(\hat{y}_0) = V(\bar{y}) + 0 + (X_0 - \bar{X})^2 V(\hat{\beta}_1)$$

وبما ان $V(\hat{\beta}_1) = \dfrac{\hat{\sigma}^2}{S_{XX}}$ وان $V(\bar{y}) = \dfrac{\hat{\sigma}^2}{n}$ بذلك فان:

$$V(\hat{y}_0) = \frac{\hat{\sigma}^2}{n} + (X_0 - \bar{X})^2 \frac{\hat{\sigma}^2}{S_{XX}}$$

12- التباين المشترك $\operatorname{cov}(\bar{y}, \hat{\beta}_0)$

$$\operatorname{cov}(\bar{y}, \hat{\beta}_0)$$

Proof:

من خلال المعادلة الاتية:

$$\operatorname{cov}(A, B - \alpha D) = \operatorname{cov}(A, B) - \alpha \operatorname{cov}(A, D)$$

بذلك فان:

$$\begin{aligned}\operatorname{cov}(\bar{y}, \hat{\beta}_0) &= \operatorname{cov}(\bar{y}, \bar{y} - \bar{X}\hat{\beta}_1) \\ &= \operatorname{cov}(\bar{y}, \bar{y}) - \bar{X} \operatorname{cov}(\bar{y}, \hat{\beta}_1)\end{aligned}$$

وبما ان $\operatorname{cov}(\bar{y}, \hat{\beta}_1) = 0$ بذلك فان:

$$\operatorname{cov}(\bar{y}, \hat{\beta}_0) = \operatorname{cov}(\bar{y}, \bar{y}) - 0$$

$$\operatorname{cov}(\bar{y},\hat{\beta}_0)=\operatorname{cov}(\bar{y},\bar{y})$$

$$\therefore \operatorname{cov}(\bar{y},\hat{\beta}_0)=V(\bar{y})$$

$$\therefore \operatorname{cov}(\bar{y},\hat{\beta}_0)=\frac{\hat{\sigma}^2}{n}$$

13- التباين المشترك $\operatorname{cov}(\hat{\beta}_0,\hat{\beta}_1)$

$$\operatorname{cov}(\hat{\beta}_0,\hat{\beta}_1)$$

Proof:

$$\operatorname{cov}(\hat{\beta}_0,\hat{\beta}_1)=\operatorname{cov}(\bar{y}-\bar{X}\hat{\beta}_1,\hat{\beta}_1)$$

$$\operatorname{cov}(\hat{\beta}_0,\hat{\beta}_1)=\operatorname{cov}(\bar{y},\hat{\beta}_1)-\bar{X}\operatorname{cov}(\hat{\beta}_1,\hat{\beta}_1)$$

$$\operatorname{cov}(\hat{\beta}_0,\hat{\beta}_1)=0-\bar{X}V(\hat{\beta}_1)$$

$$\operatorname{cov}(\hat{\beta}_0,\hat{\beta}_1)=-\bar{X}\frac{\hat{\sigma}^2}{S_{XX}}$$

مثال:

البيانات الاتية تمثل المتغير المستقل (X) والذي يمثل متغير العمر بالسنوات، والمتغير المعتمد او متغير الاستجابة (y) والذي يمثل ضغط الدم، لعشرة اشخاص تم اختيارهم عشوائياً. والبيانات موضحة في الجدول ادناه:

ت	y_i	X_i	$X_i y_i$	X_i^2	y_i^2
1	112	35	3920	1225	12544
2	128	40	5120	1600	16384
3	130	38	4940	1444	16900
4	138	44	6072	1936	19044
5	158	67	10586	4487	24964
6	162	64	10368	4096	26244
7	140	59	8260	3481	19600
8	175	69	12075	4761	30625
9	125	25	3125	625	15625
10	142	50	7100	2500	20164
المجموع	**1410**	**1491**	**71566**	**26157**	**202094**

المطلوب:

1- إيجاد تقدير معلمات الانحدار β_1, β_0 .

2- إيجاد معادلة خط الانحدار البسيط.

3- ما مقدار ما تتوقعهُ (القيمة التنبؤية) من ضغط الدم لشخص عمرهُ 80 سنة.

4- إيجاد تقدير تباين المجتمع σ^2 .

5- إيجاد تقدير تباين كلُ من β_1, β_0 .

الحل:

1-نجد تقدير معلمات الانحدار β_1, β_0 كما يلي:

$$\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}} = \frac{\sum (x_i - \bar{X})(y_i - \bar{y})}{\sum (x_i - \bar{X})^2} = \frac{\sum X_i y_i - \frac{(\sum X_i)(\sum y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

$$\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}} = \frac{\sum X_i y_i - \frac{(\sum X_i)(\sum y_i)}{n}}{\sum X_i^2 - \frac{(\sum X_i)^2}{n}}$$

$$\hat{\beta}_1 = \frac{71566 - \frac{(491)(1410)}{10}}{26157 - \frac{(491)^2}{10}} = 1.1396$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_0 = 141 - (1.1396)(49.1)$$

$$\hat{\beta}_0 = 85.043$$

2-نجد معادلة خط الانحدار البسيط.

$$\hat{y}_i = 85.043 + 1.1396 x_i$$

3-نجد مقدار ما تتوقعهُ (القيمة التنبؤية) من ضغط الدم لشخص عمرهُ 80 سنة.

$$\hat{y}_i = 85.043 + 1.1396(80)$$

$$\hat{y}_i = 176.2$$

4-نجد تقدير تباين المجتمع σ^2.

$$\hat{\sigma}^2 = S^2_{y/x} = MSe = V(y_i) = \frac{SSe}{n-2} = \frac{S_{yy} - \hat{\beta}_1 S_{Xy}}{n-2}$$

او هو:

$$\hat{\sigma}^2 = S^2_{y/x} = MSe = V(y_i) = \frac{SSe}{n-2} = \frac{S_{yy} - \hat{\beta}_1^2 S_{XX}}{n-2}$$

نستطيع استخدام احد الصيغتين.

باستخدام الصيغة الاولى:

نجد اولاً: S_{yy}

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$
$$= 202094 - \frac{(1410)^2}{10}$$
$$= 3284$$

ثم نجد S_{Xy}

$$S_{Xy} = \sum X_i y_i - \frac{(\sum y_i)(\sum X_i)}{n}$$
$$S_{Xy} = 71566 - \frac{(1410)(491)}{10}$$
$$S_{Xy} = 2335$$

بذلك فان:

$$\hat{\sigma}^2 \frac{SSe}{n-2} = \frac{3284 - (1.1396)(2335)}{10-2}$$
$$\hat{\sigma}^2 = 77.87$$

باستخدام الصيغة الثانية:

ثم نجد S_{XX}

$$S_{XX} = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$$
$$= 26157 - \frac{(491)^2}{10}$$
$$= 2048.9$$

بذلك فان:

$$\hat{\sigma}^2 = \frac{SSe}{n-2} = \frac{3284 - (1.1396)^2(2048.9)}{10-2}$$
$$\hat{\sigma}^2 = 77.87$$

5-نجد تقدير تباين كلُ من β_0, β_1.

نجد تقدير تباين β_0 أولا:

$$V(\hat{\beta}_0) = \hat{\sigma}^2 \frac{\sum X_i^2}{nS_{XX}}$$

$$V(\hat{\beta}_0) = (77.9)\frac{(26157)}{(10)(2048.9)}$$

$$V(\hat{\beta}_0) = 99.4$$

ثم نجد تقدير تباين β_1

$$V(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{S_{XX}}$$

$$V(\hat{\beta}_1) = \frac{77.9}{2048.9}$$

$$V(\hat{\beta}_1) = 0.038$$

اختبار الفرضيات Test Hypothesis

اختبار الفرضيات هو اختبار عودة عينة الى مجتمع معين، أي هل تعود الى المجتمع نفسهُ ام الى مجتمع اخر.

او اختبار هل العينة تعود الى مجتمع ما لهُ قيمة معينة.

هناك نوعين من الفرضيات:

1-فرضية العدم: Null Hypothesis

ويرمز لها بالرمز H_0 وهي تتعلق بالمعلمة موضوع الدراسة وهي تحدد القيم التي يعتقد بأنها تعبر عن القيمة الحقيقية للمعلمة.

2-الفرضية البديلة: Alternative Hypothesis

ويرمز لها بالرمز H_1 وهي التي تحدد قيم المعلمة التي يعتقد بأنها صحيحة، والتي نأمل ان تكون بيانات العينة تؤدي الى قبول الفرضية البديلة H_1 على أساس انها فرضية صحيحة.

أ- اختبار الفرضيات المتعلقة بـ β_0

1- لاختبار الفرضية القائلة بأن β_0 تساوي صفر هي:

$$H_0 : \beta_0 = 0$$

$$H_1 : \beta_0 \neq 0$$

ولاختبار هذه الفرضية نستخدم المختبر الاحصائيt والذي يأخذ الصيغة الاتية:

$$t = \frac{\hat{\beta}_0}{S_{\hat{\beta}_0}}$$

ان المختبر الاحصائي t يتوزع توزيع t بدرجة حرية (n−2).
فإذا كانت القيمة المحسوبة اصغر من القيمة الجدولية، أي ان:

$$|t_{cal.}| < t(\frac{\alpha}{2}, n-2)$$

عندئذٍ تقبل فرضية العدم H_0 وترفض الفرضية البديلة.
واذا كانت القيمة المحسوبة اكبر من القيمة الجدولية، أي ان:

$$|t_{cal.}| > t(\frac{\alpha}{2}, n-2)$$

عندئذٍ ترفض فرضية العدم H_0 وتقبل الفرضية البديلة H_1.

مثال:

اختبر β_0 تساوي صفر للمثال السابق عندما $\alpha = 0.05$ (مستوى المعنوية).

الحل:

يتم كتابة الفرضية:

$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 \neq 0$$

يتم حساب درجة الحرية لاستخراج قيمة t الجدولية من الجداول الخاصة من توزيع t، وكما يلي:

$$n-2$$
$$10-2=8$$

ومستوى المعنوية يكون:

$$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$$

ومن جداول توزيع t تكون القيمة الجدولية كالاتي:

$$t(0.025, 8) = 2.306$$

يتم الان حساب المختبر الاحصائي t:

$$t = \frac{\hat{\beta}_0}{S_{\hat{\beta}_0}} = \frac{85.043}{S_{\hat{\beta}_0}}$$

حيث ان $S_{\hat{\beta}_0}$ هو الجذر التربيعي لتباين المعلمة المقدرة $\hat{\beta}_0$ والتي تكون الصيغة الحسابية لهُ كما يلي:

$$S^2_{\hat{\beta}_0} = V(\hat{\beta}_0) = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}} \right)$$

$$S^2_{\hat{\beta}_0} = 77.9\left(\frac{1}{10}+\frac{(49.1)^2}{2048.9}\right)$$

$$S^2_{\hat{\beta}_0} = 99.45$$

$$S_{\hat{\beta}_0} = 9.97$$

بذلك تكون قيمة المختبر الاحصائي t كما يلي:

$$t = \frac{\hat{\beta}_0}{S_{\hat{\beta}_0}} = \frac{85.043}{9.97} = 8.53$$

القرار: بما ان قيمة t المحسوبة (8.53) اكبر من القيمة الجدولية (2.31) أي ان:

$$|t_{cal.}| = 8.53 > t(0.025,8) = 2.31$$

لذا نرفض فرضية العدم ونقبل الفرضية البديلة، أي ان $H_1 : \beta_0 \neq 0$.

2- اما الفرضية القائلة بان β_0 تساوي قيمة معينة مثلاً β_{00} ، فان الفرضية سوف تكون كما يلي:

$$H_0 : \beta_0 = \beta_{00}$$
$$H_1 : \beta_0 \neq \beta_{00}$$

وسيكون المختبر الاحصائي t موضح بالصيغة الاتية:

$$t = \frac{\hat{\beta}_0 - \beta_{00}}{S_{\hat{\beta}_0}}$$

فإذا كانت $|t_{cal.}| > t(\frac{\alpha}{2}, n-2)$ ترفض فرضية العدم H_0 وتقبل الفرضية البديلة H_1.

مثال:

اختبر الفرضية الاتية باستخدام بيانات المثال السابق:

$$H_0 : \beta_0 = 100$$
$$H_1 : \beta_0 \neq 100$$

الحل:

تحديد القيمة الجدولية وكما يلي:

$$t(0.025,8) = 2.306$$

حساب المختبر الاحصائي t وكما يلي:

$$t = \frac{\hat{\beta}_0 - \beta_{00}}{S_{\hat{\beta}_0}}$$
$$= \frac{85.043 - 100}{9.97}$$
$$t = -1.5$$

يتم مقارنة قيمة t المحسوبة مع القيمة الجدولية وكما يلي:

$$|t_{cal.}| = |-1.5| < t(0.025, 8) = 2.306$$

وبما ان قيمة t المحسوبة اصغر من القيمة الجدولية ، بذلك تقبل فرضية العدم وترفض الفرضية البديلة أي ان $\beta_0 = 100$.

أ- اختبار الفرضيات المتعلقة بمعامل الانحدار β_1

وهناك طريقتين:

اولاً: باستخدام المختبر الاحصائي t :

1.1- اختبار الفرضية الاتية:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$
$$t_{cal.} = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$$

حيث ان اذا كانت t المحسوبة اكبر من t الجدولية $t(\frac{\alpha}{2}, n-2)$ فأننا نرفض فرضية العدم ونقبل الفرضية البديلة.

2.1- اما الفرضية القائلة بان β_1 تساوي قيمة معينة مثلاً β_{00} ، فان الفرضية سوف تكون كما يلي:

$$H_0 : \beta_1 = \beta_{00}$$
$$H_1 : \beta_1 \neq \beta_{00}$$

وسيكون المختبر الاحصائي t موضح بالصيغة الاتية:

$$t_{cal.} = \frac{\hat{\beta}_1 - \beta_{00}}{S_{\hat{\beta}_1}}$$

فإذا كانت $|t_{cal.}| > t(\frac{\alpha}{2}, n-2)$ ترفض فرضية العدم H_0 وتقبل الفرضية البديلة H_1.

<u>ملاحظة:</u>

اذا قبلنا فرضية العدم التي تنص على:

$$H_0 : \beta_1 = 0$$

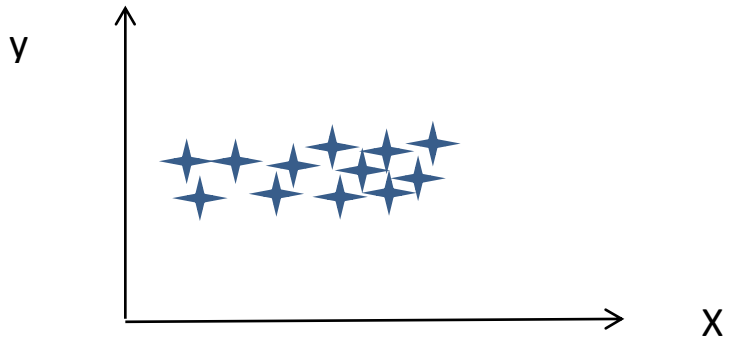
فأن ذلك يدل على انهُ لا توجد علاقة خطية بين x و y وهذا قد ينتج ذلك لسببين:

1- اما ان x_i لا يفسر التغيرات الحاصلة في y لأي قيمة من قيم x وكما هو موضح في الشكل (2).

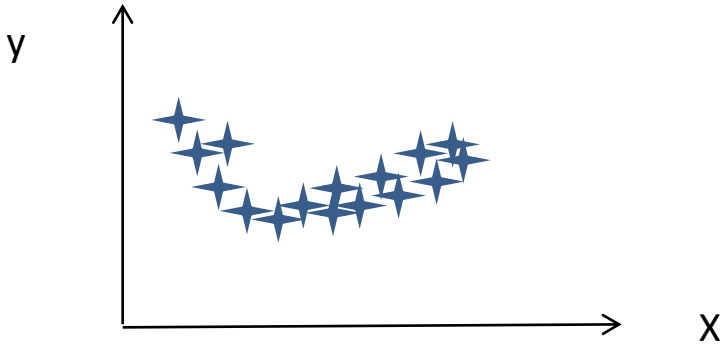
2- واما ان العلاقة بين x و y ليست علاقة خطية أي هي علاقة غير خطية وبذلك يكون الرسم بالشكل (3).

اما اذا رفضنا فرضية العدم $H_0 : \beta_1 = 0$.

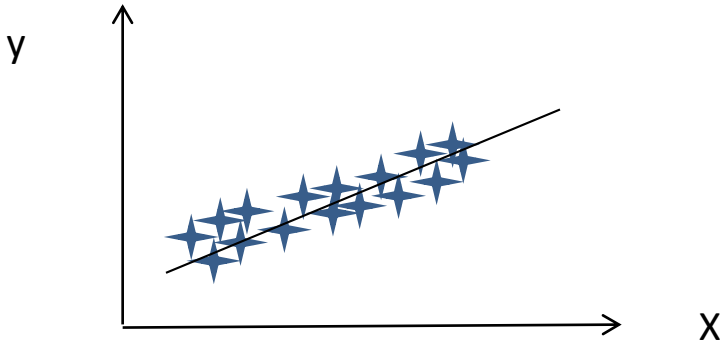
فان ذلك يدل على ان X لهُ أهمية في تفسير الاختلافات مع y وهذا يعني ان العلاقة بين X و y تتمثل بخط مستقيم أي ان النموذج الخطي يمثل البيانات تمثيلاً جيداً كما في الشكل (4).



الشكل (2) يوضح ان المتغير X لا يفسر التغيرات الحاصلة في y



الشكل (3) يوضح العلاقة بين x و y هي علاقة غير خطية



الشكل (4) يوضح العلاقة بين x و y تتمثل بخط مستقيم

ثانياً: باستخدام المختبر الاحصائي أي جدول تحليل التباين وكما يلي:

<u>تحليل التباين Analysis of variance</u>

يتكون انحراف القيم الاصلية y_i عن الوسط الحسابي $\bar{y}$ من عنصرين:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

بأخذ التربيع للطرفين نحصل على:

$$(y_i-\bar{y})^2=[(\hat{y}_i-\bar{y})+(y_i-\hat{y}_i)]^2$$

بأخذ المجموع للطرفين نحصل على:

$$\sum(y_i-\bar{y})^2=\sum[(\hat{y}_i-\bar{y})+(y_i-\hat{y}_i)]^2$$

بفتح التربيع للقوس الكبير في للطرف الأيمن:

$$\sum(y_i-\bar{y})^2=\sum[(\hat{y}_i-\bar{y})^2+2(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)+(y_i-\hat{y}_i)^2]$$

وبإدخال المجموع على الاقواس للطرف الأيمن:

$$\sum(y_i-\bar{y})^2=\sum(\hat{y}_i-\bar{y})^2+2\sum(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)+\sum(y_i-\hat{y}_i)^2 \quad \dots \ (*)$$

يمكن برهان ان الحد الثاني من الطرف الأيمن يساوي الصفر وكما يلي:

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

وبإدخال المجموع على الاقواس:

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

$$\begin{aligned}\sum\hat{y}_i y_i &= \sum\hat{y}_i y_i \\ &= \sum\hat{y}_i(\beta_0+\beta_1 X_i+e_i) \\ &= \sum\hat{y}_i(\hat{y}_i+e_i) \\ &= \sum\hat{y}_i\hat{y}_i+\sum\hat{y}_i e_i \\ &= \sum\hat{y}_i^2+\sum\hat{y}_i e_i\end{aligned}$$

بذلك فان:

$$\sum(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)=\sum\hat{y}_i^2+\sum\hat{y}_i e_i-\bar{y}\sum y_i-\sum\hat{y}_i^2+\bar{y}\sum\hat{y}_i$$

بالاختصار نحصل على:

$$\sum(\hat{y}_i-\bar{y})(y_i-\hat{y}_i)=\sum\hat{y}_i^2+\sum\hat{y}_i e_i-\bar{y}\sum y_i-\sum\hat{y}_i^2+\bar{y}\sum\hat{y}_i$$

$$\sum\hat{y}_i e_i=0$$

بذلك فان الناتج سيكون مساوي للصفر أي ان:

$$\sum(\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$$

بذلك فان المعادلة (*) ستكون بالشكل الاتي:

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

بما ان $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ بذلك فان:

$$\sum(y_i - \bar{y})^2 = \sum(\hat{\beta}_0 + \hat{\beta}_1 X_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

بما ان $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}$ بذلك فان:

$$\sum(y_i - \bar{y})^2 = \sum(\bar{y} - \hat{\beta}_1 \bar{X} + \hat{\beta}_1 X_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

وبأخذ $\hat{\beta}_1$ عامل مشترك نحصل على:

$$\sum(y_i - \bar{y})^2 = \sum(\bar{y} + \hat{\beta}_1(X_i - \bar{X}) - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2$$

وبإدخال المجموع على القوس الأول:

$$\sum(y_i - \bar{y})^2 = \hat{\beta}_1^2 \sum(X_i - \bar{X})^2 + \sum(y_i - \hat{y}_i)^2$$

وحسب الرموز التي لدينا نحصل على:

$$S_{yy} = \hat{\beta}_1^2 S_{XX} + SSe$$

$$\therefore SSe = S_{yy} - \hat{\beta}_1^2 S_{XX}$$

بذلك يمكن تكوين جدول تحليل التباين كما يلي:

Source of variation مصادر التباين S.O.V	Degree of Fredom درجات الحرية	Sum of squares مجموع المربعات	Mean of squares متوسط المربعات	Fcal. قيمة F المحسوبة
Due to Regression مصدر يعود الى الانحدار	1 عدد المتغيرات المستقلة في النموذج	$SSR(X_1) = \hat{\beta}_1^2 S_{XX}$ $Or\ SSR(X_1) = \hat{\beta}_1 S_{Xy}$	$M.S.R(X_1) = \frac{SSR(X_1)}{df.R}$	$F_{cal.} = \frac{M.S.R}{M.S.e}$
About Regression Error مصدر يعود للخطأ	n-2 n مطروح منها عدد المعلمات في النموذج	$SSe = SST - SSR(X_1)$ $SSe = S_{yy} - \hat{\beta}_1^2 S_{XX}$ $or\ SSe = S_{yy} - \hat{\beta}_1 S_{Xy}$	$M.S.e = \frac{SSe}{df.e}$	
Total المجموع	n-1	$SST = S_{yy}$		

ان قيمة F الجدولية يتم استخراجها من جداول توزيع F وذلك بمعرفة درجة حرية المصدر الذي يعود للانحدار ودرجة حرية الخطأ ومستوى المعنوية. ويمكن التعبير عن ذلك بصيغة رموز أي ان:

$$F_{tab.} = F(\alpha, v_1, v_2)$$

وفي حالة الانحدار الخطي البسيط والذي يتكون من متغير مستقل واحد تكون قيمة F الجدولية كما يلي:

$$F_{tab.} = F(\alpha, 1, n-2)$$

ان جدول تحليل التباين يختبر الفرضية نفسها التي يختبرها اختبار t حول المعلمة β_1 والتي هي:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

يمكن تطبيق جدول تحليل التباين على المثال السابق وكما يلي:

مثال:

البيانات الاتية تمثل المتغير المستقل (X) والذي يمثل متغير العمر بالسنوات، والمتغير المعتمد او متغير الاستجابة (y) والذي يمثل ضغط الدم، لعشرة اشخاص تم اختيارهم عشوائياً. والبيانات موضحة في الجدول ادناه:

ت	y_i	X_i	$X_i y_i$	X_i^2	y_i^2
1	112	35	3920	1225	12544
2	128	40	5120	1600	16384
3	130	38	4940	1444	16900
4	138	44	6072	1936	19044
5	158	67	10586	4487	24964
6	162	64	10368	4096	26244
7	140	59	8260	3481	19600
8	175	69	12075	4761	30625
9	125	25	3125	625	15625
10	142	50	7100	2500	20164
المجموع	**1410**	**1491**	**71566**	**26157**	**202094**

المطلوب كون جدول تحليل التباين لاختبار الفرضية الاتية:

$H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$

الحل: يتم كتابة الفرضية أولا وكما يلي:

$H_0 : \beta_1 = 0$
$H_1 : \beta_1 \neq 0$

إيجاد مجموع مربعات العائدة للانحدار، ومن خلال القيم التي تم الحصول عليها في حل المثال في المحاضرة 6 يمكن ان نعوض في الصيغ الاتية:

$SSR(X_1) = \hat{\beta}_1^2 S_{XX}$
$SSR(X_1) = (1.1396)^2 (2048.9)$
$SSR(X_1) = 2660.9$

إيجاد مجموع المربعات الكلية وكما يلي:

$SST = S_{yy}$
$SST = 3284$

حساب مجموع مربعات الخطأ:

$SSe = SST - SSR(X_1)$
$SSe = S_{yy} - \hat{\beta}_1^2 S_{XX}$
$SSe = 3284 - 2660.9$
$SSe = 623.1$

S.O.V	d.f	S.S	M.S	Fcal.
R(X1)	1	$SSR(X_1) = 2660.9$	$M.S.R(X_1) = \frac{2660.9}{1} = 2660.9$	$F_{cal.} = \frac{2660.9}{77.9} = 34.16$
Error	8	$SSe = 623.1$	$M.S.e = \frac{623.1}{8} = 77.9$	
Total	9	$SST = 3284$		

اما قيمة F الجدولية ومن خلال جداول توزيع F فسوف تكون:

$$F_{tab.} = F(0.05,1,8)$$
$$F_{tab.} = 5.32$$

ومن خلال المقارنة بين القيمة المحسوبة والقيمة الجدولية:

$$F_{cal.} = 34.16 > F_{tab.} = 5.32$$

القرار: بما ان القيمة المحسوبة اكبر من القيمة الجدولية اذاً نرفض فرضية العدم ونقبل الفرضية البديلة أي ان $H_1 : \beta_1 \neq 0$.

ملاحظة:

ان استخدام المختبر الاحصائي t لهُ افضلية على استخدام المختبر الاحصائي F للأسباب الاتية:

1- يمكن استخدام المختبر الاحصائي t لاختبار فرضية ذات طرف واحد أي $t(\alpha, n-1)$.

2- يمكن استخدام المختبر الاحصائي t لاختبار فرضية العدم القائلة بأن β_1 تساوي قيمة معينة مثل:

$H_0 : \beta_1 = 10$

$H_1 : \beta_1 \neq 10$

3- يمكن استخدام المختبر الاحصائي t في تكوين حدود الثقة لـ β_1 .

حدود الثقة $(\alpha-1)\%$ Confidence intervals

حدود الثقة توضع للمعلمة وليس للإحصائية أي مثلاً توضع للـ μ وليس للـ $\bar{X}$.

ان تقدير فترة او حدود الثقة هو عملية إيجاد القيمة الحقيقية للمعلمة بين حدي ثقة اعلى وادنى بحيث تكون على ثقة مقدارها $(\alpha-1)\%$. ان القيمة الحقيقية للمعلمة تقع بين حدي ثقة اللذان يأخذان الصيغة الاتية:

$$\hat{\theta} \mp t(\frac{\alpha}{2}, n-2) S(\hat{\theta})$$

فاذا كانت الحدود المراد ايجادها للمقدر $\bar{X}$ عندئذٍ تكون حدود الثقة كما يلي:

$$Lower\ bound = \bar{X} - t(\frac{\alpha}{2}, n-2) S(\bar{X})$$

$$Upper\ bound = \bar{X} + t(\frac{\alpha}{2}, n-2) S(\bar{X})$$

ملاحظة: ان حدود الثقة هي عبارة عن الحدود التي تقبل فيها فرضية العدم والتي تقع بينها أي الحد الأعلى والحد الأدنى.

1- تقدير فترة (حدود) الثقة للمعلمة β_0 .

تكون الصيغة المستخدمة لإيجاد حدود الثقة للمعلمة β_0 هي كالاتي:

$$\hat{\beta}_0 - t(\frac{\alpha}{2}, n-2) S(\hat{\beta}_0) \leq \beta_0 \leq \hat{\beta}_0 + t(\frac{\alpha}{2}, n-2) S(\hat{\beta}_0)$$

2- تقدير فترة (حدود) الثقة للمعلمة β_1 .

تكون الصيغة المستخدمة لإيجاد حدود الثقة للمعلمة β_1 هي كالاتي:

$$\hat{\beta}_1 - t(\frac{\alpha}{2}, n-2) S(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t(\frac{\alpha}{2}, n-2) S(\hat{\beta}_1)$$

3- تقدير فترة (حدود) الثقة للـ σ^2 (تباين المجتمع).

يمكن تكوين حدود الثقة $(n-\alpha)\%$ لتباين المجتمع σ^2 من خلال الصيغة الاتية:

$$\frac{(n-1)MSe}{\chi^2(\frac{\alpha}{2},n-2)} \le \sigma^2 \le \frac{(n-1)MSe}{\chi^2(1-\frac{\alpha}{2},n-2)}$$

حيث ان $\frac{(n-1)MSe}{\sigma^2}$ يتوزع توزيع χ^2 بدرجة حرية (n–2).

4– حدود الثقة لخط الانحدار نفسهُ لمتوسط الاستجابة $\hat{y}_i$ أي $\hat{y}_0$.

نفرض بان X_0 هي قيمة من القيم الاصلية للمتغير X والتي نرغب بعمل تقدير فترة ثقة لمتوسط الاستجابة أي ان:

$$\hat{y}_0 - t(\frac{\alpha}{2},n-2)S(\hat{y}_0) \le E(y/X_0) \le \hat{y}_0 + t(\frac{\alpha}{2},n-2)S(\hat{y}_0)$$

يمكن تطبيق حدود الثقة من كل ما ذكر أعلاه باستخدام المثال الاتي.

مثال:(المثال الأول حيث تم استخدام بياناتهُ)

البيانات الاتية تمثل المتغير المستقل (X) والذي يمثل متغير العمر بالسنوات، والمتغير المعتمد او متغير الاستجابة (y) والذي يمثل ضغط الدم، لعشرة اشخاص تم اختيارهم عشوائياً. والبيانات موضحة في الجدول ادناه:

ت	y_i	X_i	$X_i y_i$	X_i^2	y_i^2
1	112	35	3920	1225	12544
2	128	40	5120	1600	16384
3	130	38	4940	1444	16900
4	138	44	6072	1936	19044
5	158	67	10586	4487	24964
6	162	64	10368	4096	26244
7	140	59	8260	3481	19600
8	175	69	12075	4761	30625
9	125	25	3125	625	15625
10	142	50	7100	2500	20164
المجموع	**1410**	**1491**	**71566**	**26157**	**202094**

المطلوب:

1- تكوين حدود الثقة للمعلمة β_0 .

2- تكوين حدود الثقة للمعلمة β_1 .

3- تكوين حدود الثقة لتباين المجتمع σ^2 .

4- تكوين حدود الثقة لمتوسط الاستجابة عند $X_0 = 40$

الحل:

1- تكوين حدود الثقة للمعلمة β_0 .

من خلال الحل للمثال أعلاه تكون قيمة β_0 وقيمة $S_{\hat{\beta}_0}$ كالاتي:

$$\hat{\beta}_0 = 85.043, S_{\hat{\beta}_0} = 9.97$$

بذلك تكون حدود الثقة كما يلي:

$$85.043 - t(0.025,8)(9.79) \le \beta_0 \le 85.043 + t(0.025,8)(9.79)$$
$$85.043 - (2.31)(9.79) \le \beta_0 \le 85.043 + (2.31)(9.79)$$
$$62.05 \le \beta_0 \le 108.03$$

2- تكوين حدود الثقة للمعلمة β_1 .

من خلال الحل للمثال أعلاه تكون قيمة β_1 وقيمة $S_{\hat{\beta}_1}$ كالاتي:

$$\hat{\beta}_1 = 1.1396, S_{\hat{\beta}_1} = 0.194$$

بذلك تكون حدود الثقة كما يلي:

$$1.139 - t(0.025,8)(0.194) \le \beta_1 \le 1.139 + t(0.025,8)(0.194)$$
$$1.139 - (2.31)(0.194) \le \beta_1 \le 1.139 + (2.31)(0.194)$$
$$0.69 \le \beta_1 \le 1.587$$

3- تكوين حدود الثقة لتباين المجتمع σ^2 .

من خلال الحل للمثال أعلاه تكون القيم المطلوبة لحدود الثقة لتباين المجتمع كالاتي:

$$MSe = 77.9$$

اما القيم الجدولية فقد تم ايجادها من جداول توزيع كاي سكوير χ^2 وهي كالاتي:

$$\chi^2(\frac{\alpha}{2}, n-2) = \chi^2(0.025,8) = 17.53$$

$$\chi^2(1-\frac{\alpha}{2}, n-2) = \chi^2(0.975,8) = 2.18$$

بذلك تكون حدود الثقة كالاتي:

$$\frac{(n-1)MSe}{\chi^2(\frac{\alpha}{2},n-2)} \le \sigma^2 \le \frac{(n-1)MSe}{\chi^2(1-\frac{\alpha}{2},n-2)}$$

$$\frac{(8)(77.9)}{\chi^2(0.025,8)} \le \sigma^2 \le \frac{(8)(77.9)}{\chi^2(0.975,8)}$$

$$35.55 \le \sigma^2 \le 285.87$$

4- تكوين حدود الثقة لمتوسط الاستجابة عند $X_0 = 40$

من خلال الحل للمثال أعلاه تكون القيم المطلوبة لحدود الثقة لمتوسط الاستجابة كالاتي:

في بادئ الامر نعوض قيمة $X_0 = 40$ في معادلة خط الانحدار المقدرة للحصول على $\hat{y}_0$ وكما يلي:

$$\hat{y}_{X=40} = 85.043 + 1.1396(40)$$

$$\hat{y}_{X=40} = 130.627$$

وتباين متوسط الاستجابة فأنهُ سيكون:

$$V(\hat{y}_0) = \frac{\hat{\sigma}^2}{n} + (X_0 - \bar{X})^2 \frac{\hat{\sigma}^2}{S_{XX}}$$

$$V(\hat{y}_0) = \hat{\sigma}^2 \left(\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{XX}} \right)$$

$$V(\hat{y}_0) = (77.9) \left(\frac{1}{10} + \frac{(40-49.1)^2}{2048.9} \right)$$

$$V(\hat{y}_0) = 10.938$$

اما الانحراف القياسي لتباين متوسط الاستجابة فأنهُ سيكون:

$$\sqrt{V(\hat{y}_0)} = \sqrt{10.938} = 3.31$$

بذلك تكون حدود الثقة لمتوسط الاستجابة عن النقطة $X_0 = 40$ هي كالاتي:

$$\hat{y}_0 - t(\frac{\alpha}{2},n-2)S(\hat{y}_0) \le E(y/X_0) \le \hat{y}_0 + t(\frac{\alpha}{2},n-2)S(\hat{y}_0)$$

$$130.6 - (2.306)(3.31) \le E(y/X_0) \le 130.6 + (2.306)(3.31)$$

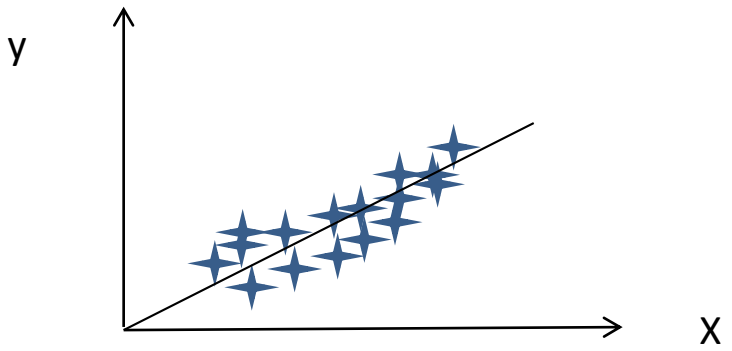
$$123.0 \le E(y/X_0) \le 138.2$$

الانحدار او المرور من نقطة الأصل Regression through the origin

ان بعض حالات الانحدار قد يمر خط الانحدار فيها من نقطة الأصل. وفي هذه الحالة فان النموذج الخطي يكون بدون المعلمة β_0 (No intercept).

$$y_i = \beta_1 x_{i1} + e_i \quad , \quad i = 1,2,...,n$$

أي ان معادلة خط الانحدار تكون كما موضح بالشكل الاتي:



الشكل (4) يوضح العلاقة بين x و y تتمثل بخط مستقيم

فلإيجاد تقدير المعلمة β_1 بطريقة المربعات الصغرى نتبع ما يلي:

من المعادلة أعلاه نجد ان:

$$e_i = y_i - \beta_1 x_i$$

وبتربيع الطرفين واخذ المجموع لهما نحصل على:

$$\begin{aligned} Q = \sum e_i^2 &= \sum (y_i - \beta_1 x_i)^2 \\ &= \sum (y_i^2 - 2\beta_1 x_i y_i + \beta_1^2 X_i^2) \\ &= \sum y_i^2 - 2\beta_1 \sum x_i y + \beta_1^2 \sum X_i^2 \end{aligned}$$

نشتق بالنسبة الى β_1 لنحصل على:

$$\frac{\partial Q}{\partial \beta_1} = 0 - 2\sum x_i y + 2\beta_1 \sum X_i^2$$

أي ان:

$$2\beta_1 \sum X_i^2 = 2\sum x_i y$$

بذلك المعلمة المقدرة $\hat{\beta}_1$ ستكون كما يلي:

$$\hat{\beta}_1 = \frac{\sum x_i y}{\sum X_i^2}$$

بذلك فان:

$$\hat{y}_i = \hat{\beta}_1 x_i$$

اما القيمة المقدرة لتباين المجتمع σ^2 فيمكن الحصول عليهِ كما يلي:

$$\begin{aligned} SSe &= \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 \\ &= \sum (y_i - \hat{\beta}_1 x_i)^2 \\ &= \sum (y_i^2 - 2\beta_1 x_i y_i + \beta_1^2 X_i^2) \\ &= \sum y_i^2 - 2\beta_1 \sum x_i y + \beta_1^2 \sum X_i^2 \\ &= \sum y_i^2 - 2\beta_1 \sum x_i y + \beta_1 \frac{\sum x_i y}{\sum X_i^2} \sum X_i^2 \\ &= \sum y_i^2 - 2\beta_1 \sum x_i y + \beta_1 \sum x_i y \end{aligned}$$

$$\therefore SSe = \sum y_i^2 - \beta_1 \sum x_i y$$

بذلك فان MSe في هذه الحالة سيكون:

$$\therefore MSe = \frac{SSe}{n-1} = \frac{\sum y_i^2 - \beta_1 \sum x_i y}{n-1}$$

اما تباين المعلمة β_1 فسيكون كالاتي:

$$S_{\hat{\beta}_1}^2 = \frac{MSe}{\sum X_i^2}$$

والسؤال الان أي النموذجين نستخدم، هل بوجود β_0 ام عدم وجودها ؟

ونقترح في هذه الحالة اتباع ما يلي:

1- رسم الشكل الانتشاري والذي يعطي بعض الدلالة عن خط الانحدار من حيث انهُ قد يمر من نقطة الأصل ام لا.

2- استخدام النموذج الذي يحوي على β_0 ثم اختبر هل ان $H_0 : \beta_0 = 0$ ام لا، فاذا قبلت فرضية العدم فاستخدم النموذج الذي لا يحتوي على β_0 فقد يكون اكثر مطابقة للبيانات.

3- استخدم كلا النموذجين وقارن بين متوسط مربعات البواقي MSe، فالنموذج الذي يمتلك MSe اقل فهو الاحسن مطابقة للبيانات.

مثال:

البيانات التالية تمثل عدد وحدات الإنتاج المصنعة X وكلفة العمل الكلية المختلفة لها y خلال فترة معينة والمأخوذة من 12 مصنعاً متشابهاً.

ت	y_i	X_i	$X_i y_i$	X_i^2	y_i^2
1	114	20	2280	400	12996
2	921	196	180516	38416	848241
3	560	115	64400	13225	313600
4	245	50	12250	2500	60025
5	575	122	70150	14884	330625
6	475	100	47500	10000	225625
7	138	33	4554	1089	19044
8	727	154	111958	23716	528529
9	375	80	30000	6400	140625
10	670	147	98490	21609	448900
11	828	182	150696	33124	685584
12	762	160	121920	25600	580644
المجموع	**6390**	**1359**	**894714**	**190963**	**4194438**

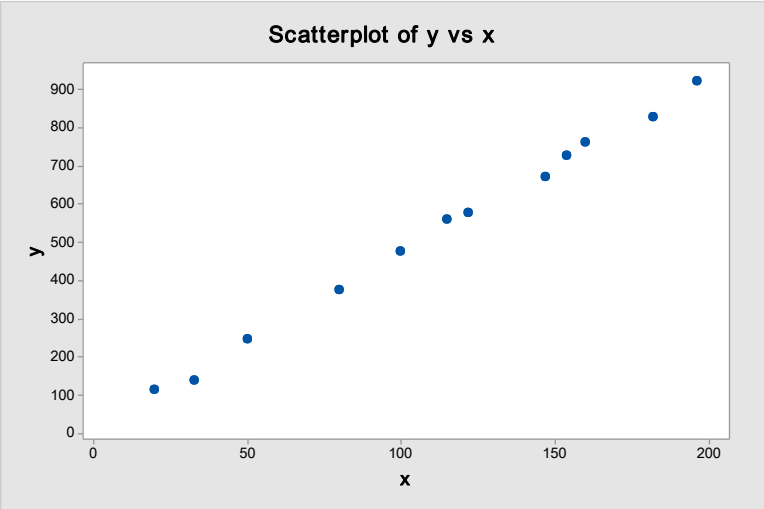
لاحظ هنا بانهُ عندما لا يكون هناك وحدات منتجة فان الكلفة تساوي صفراً أي عنما يكون X=0 فان y=0 ولذلك فان النموذج بدون المعلمة β_0 قد يكون اكثر مطابقة للبيانات. فضلاً عن اننا لو لاحظنا رسم الشكل الانتشاري فانهُ يدل على ان خط الانحدار قد يمر من نقطة الأصل (0,0).

الحل:

نجد الاعمدة لكل من $X_i y_i$ و X_i^2 و y_i^2 .

نرسم الشكل الانتشاري بين X و yوكما يلي:

من خلال الشكل يلاحظ ان معظم نمط النقاط يدل على مرورها من نقطة الأصل. بذلك سنقدر النموذج بدون المعلمة $\hat{\beta}_0$.



بذلك تكون المعلمة $\hat{\beta}_1$ كالاتي:

$$\hat{\beta}_1 = \frac{\sum x_i y}{\sum X_i^2} = \frac{894714}{190963} = 4.685$$

بذلك تكون معادلة خط الانحدار كما يلي:

$$\hat{y}_i = 4.685 x_i$$

اما تقدير تباين المجتمع σ^2 فسيكون كالاتي:

$$\therefore MSe = \frac{SSe}{n-1} = \frac{\sum y_i^2 - \beta_1 \sum x_i y}{n-1}$$
$$= \frac{4194438 - (4.685)(894714)}{12-1}$$
$$= 223.399$$

اما تباين المعلمة $\hat{\beta}_1$ فيكون كالاتي:

$$S^2_{\hat{\beta}_1} = \frac{MSe}{\sum X_i^2}$$
$$S^2_{\hat{\beta}_1} = \frac{223.399}{190963}$$
$$S^2_{\hat{\beta}_1} = 0.00117$$

اما حدود الثقة فستكون كالاتي:

$$\hat{\beta}_1 - t(\frac{\alpha}{2}, n-2)S(\hat{\beta}_1) \le \beta_1 \le \hat{\beta}_1 + t(\frac{\alpha}{2}, n-2)S(\hat{\beta}_1)$$
$$4.685 - (2.201)(0.0342) \le \beta_1 \le 4.685 + (2.201)(0.0342)$$
$$4.61 \le \beta_1 \le 4.76$$

لو استخدمنا النموذج الخطي والذي يحتوي على المعلمة β_0 نجد ان:

$S_{XX} = 37056.25$

$S_{yy} = 791763$

$\hat{\beta}_0 = 9.75$

$\hat{\beta}_1 = 4.615$

$\hat{y}_i = 9.75 + 4.62X_i$

$SSe = 2236.13$

$MSe = \frac{2236.13}{10} = 223.612$

$S_{\hat{\beta}_0} = 9.799$

ولو أجرينا اختبار t حول المعلمة β_0 وحسب الفرضية الاتية:

$H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

فسينتج لنا الاتي:

$$t = \frac{\hat{\beta}_0}{S_{\hat{\beta}_0}} = \frac{9.75}{9.799} = 0.994$$

القرار: بما ان قيمة t المحسوبة (0.994) اصغر من القيمة الجدولية (2.22) أي ان:

$$|t_{cal.}| = 0.994 < t(0.025,10) = 2.22$$

لذا تقبل فرضية العدم وترفض الفرضية البديلة، أي ان $H_1 : \beta_0 = 0$. أي ان النموذج الخطي الذي لا يحتوي على $\hat{\beta}_0$ هو اكثر تطابقاً.

كما اننا اذا قارنا قيمة MSe لكل من النموذجين سنلاحظ النموذج الذي لا يحتوي على المعلمة β_0 يمتلك MSe اقل من النموذج الثاني ولو حتى بمقدار قليل.

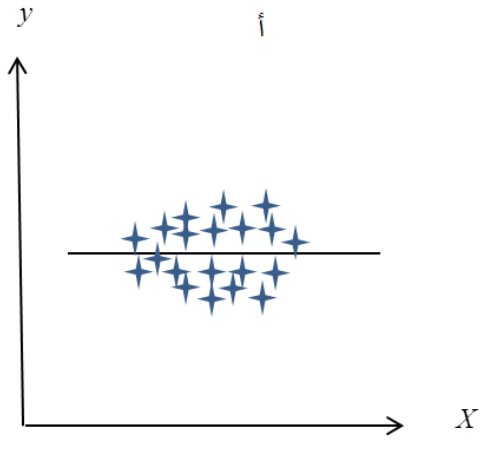
معامل التحديد R^2 Coefficient of Determination

تعريف: يعرف معامل التحديد هو ما يفسرهُ أهمية النموذج الرياضي في وصف العلاقة بين X و y ليعطي نسبة ما تفسرهُ هذه العلاقة من خلال النموذج للتغيرات الحاصلة في y.

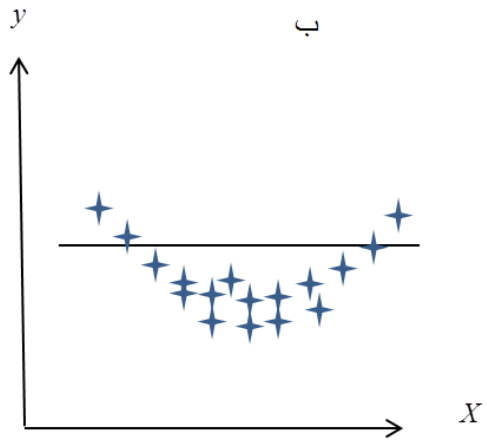
وتتراوح قيمة R^2 بين الصفر والواحد أي $0 \le R^2 \le 1$. وصيغة معامل التحديد تكون كما يأتي:

$$R^2 = \frac{Sum\ of\ Squares\ of\ due\ to\ Regression}{Sum\ of\ Squares\ of\ Total}$$

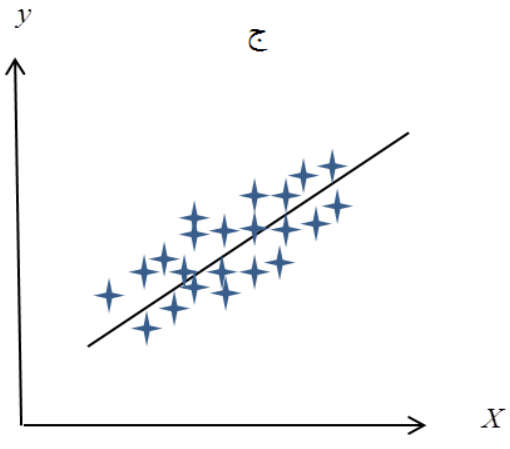
لاحظ ان R^2 قد يكون مقياساً غير جيد للدلالة على ملائمة النموذج للبيانات، لاحظ الرسوم الاتية:



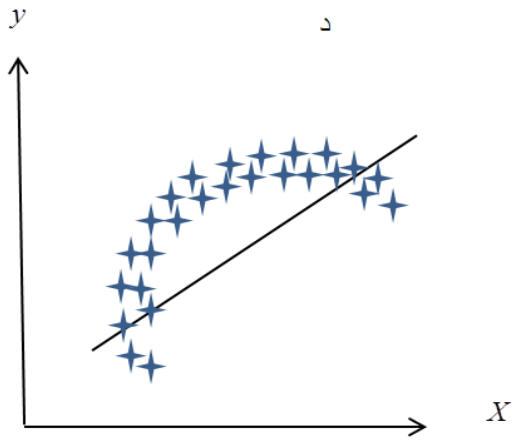
$R^2 = 0$ بينما لا يوجد علاقة بين X و y



$R^2 = 0$ بينما يوجد علاقة بين X و y



قيمة R^2 عالية والنموذج الخطي يلائم البيانات



قيمة R^2 عالية بينما النموذج الغير خطي يلائم البيانات

معامل الارتباط وعلاقتهُ بالانحدار الخطي البسيط

معامل الارتباط

ان معامل الارتباط البسيط Correlation Coefficient ويرمز لهُ بالرمز r وهو مقياس لدرجة الترابط والالتزام بين متغيرين مستقلين. ويمكن ايجادهُ من خلال الصيغة الاتية:

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}\, S_{yy}}}$$

ان العلاقة بين معامل الانحدار $\hat{\beta}_1$ ومعامل الارتباط هي:

$$r_{Xy} = \hat{\beta}_1 \sqrt{\frac{S_{XX}}{S_{yy}}}$$

وكذلك يمكن

$$\hat{\beta}_1 = r_{Xy} \sqrt{\frac{S_{yy}}{S_{XX}}}$$

ويمكن برهنة الصيغتين أعلاه كما يأتي:

1- برهن ان $r_{Xy} = \hat{\beta}_1 \sqrt{\frac{S_{XX}}{S_{yy}}}$.

Proof:

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}\, S_{yy}}}$$

نضرب ونقسم الطرف الأيمن على $\sqrt{S_{XX}}$ نحصل على:

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}\, S_{yy}}} \frac{\sqrt{S_{XX}}}{\sqrt{S_{XX}}}$$

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}}\sqrt{S_{yy}}} \frac{\sqrt{S_{XX}}}{\sqrt{S_{XX}}}$$

وبضرب $\sqrt{S_{XX}}$ الموجودين في المقام نحصل على:

$$r_{Xy} = \frac{S_{Xy}}{S_{XX}} \frac{\sqrt{S_{XX}}}{\sqrt{S_{yy}}}$$

وبما ان $\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}}$ نحصل على:

$$\therefore\ r_{Xy} = \hat{\beta}_1 \frac{\sqrt{S_{XX}}}{\sqrt{S_{yy}}}$$

2- برهن ان $\hat{\beta}_1 = r_{Xy}\sqrt{\frac{S_{yy}}{S_{XX}}}$.

Proof:

بما ان :

$$\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}}$$

نضرب ونقسم $\sqrt{S_{yy}}$

$$\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}} \frac{\sqrt{S_{yy}}}{\sqrt{S_{yy}}}$$

نفتح S_{XX} الى جذرين $\sqrt{S_{XX}}$ بذلك نحصل على:

$$\hat{\beta}_1 = \frac{S_{Xy}}{\sqrt{S_{XX}}\sqrt{S_{XX}}} \frac{\sqrt{S_{yy}}}{\sqrt{S_{yy}}}$$

وبترتيب الجذور من اجل الحصول على صيغة r_{Xy} نحصل على:

$$\hat{\beta}_1 = \frac{S_{Xy}}{\sqrt{S_{XX}}\sqrt{S_{yy}}} \frac{\sqrt{S_{yy}}}{\sqrt{S_{XX}}}$$

بذلك نصل الى:

$$\therefore \quad \hat{\beta}_1 = r_{Xy} \frac{\sqrt{S_{yy}}}{\sqrt{S_{XX}}}$$

إيجاد جدول تحليل التباين للانحدار باستخدام معامل الارتباط

يمكن تنظيم جدول تحليل التباين للانحدار باستخدام معامل الارتباط وكما يأتي:

S.O.V	d.f	S.S
R(X1)	1	$r^2_{Xy}S_{yy}$
Error	n-2	$(1-r^2_{Xy})S_{yy}$
Total	n-1	

اختبار الفرضيات حول معامل الارتباط

من اجل اجراء اختبار حول معامل الارتباط فانهُ يجب اختبار الفرضية الاتية:

$$H_0 : \rho = r_0$$
$$H_1 : \rho \neq r_0$$

حيث ان r_0 هي أي قيمة.

ويتم ذلك باستخدام اختبار t والذي يأخذ الصيغة الاتية:

$$t = \frac{r - r_0}{\sqrt{\frac{1-r^2}{n-1}}}$$

وقيمة t الجدولية تكون $t_{tab.} = t(\frac{\alpha}{2}, n-2)$ فاذا كانت $|t_{cal.}| > t_{tab.}$ فهذا يعني رفض فرضية العدم وقبول الفرضية البديلة اي ان معامل الارتباط لا يساوي r_0.

مثال: من بيانات المثال المتعلق بضغط الدم y والعمر X ، جد ما يأتي:

y	112	128	130	138	158	162	140	175	125	142
X	35	40	38	44	67	64	59	69	25	50

1- معامل الارتباط بين y و X.

2- كون جدول تحليل التباين باستخدام معامل الارتباط.

3- جد معامل التحديد.

4- اختب الفرضية الاتية: $H_0 : \rho = 0$, $H_1 : \rho \neq 0$

الحل: 1- معامل الارتباط بين y و X.

في بادى الامر نجد معامل الارتباط وكما يلي:

حيث ان

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}\, S_{yy}}}$$

$$S_{Xy} = \sum X_i y_i - \frac{(\sum X_i)(\sum y_i)}{n}$$

$$S_{XX} = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

نكون جدول من اجل إيجاد مجاميع المربعات، وكما يلي:

y_i^2	$X_{i1}y_i$	X_{i1}^2	X_{i1}	y_i
12544	3920	1225	35	112
16384	5120	1600	40	128
16900	4940	1444	38	130
19044	6072	1936	44	138
24964	10586	4489	67	158
26244	10368	4096	64	162
19600	8260	3481	59	140
30625	12075	4761	69	175
15625	3125	625	25	125
20164	7100	2500	50	142
202094	1410	491	26157	71566

$$S_{Xy} = \sum X_i y_i - \frac{(\sum X_i)(\sum y_i)}{n} \Rightarrow = 71566 - \frac{(491)(1410)}{10} = 2335$$

$$S_{XX} = \sum X_i^2 - \frac{(\sum X_i)^2}{n} \Rightarrow = 26157 - \frac{(491)^2}{10} = 2048.9$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} \Rightarrow = 202094 - \frac{(1410)^2}{10} = 3284$$

بذلك فان معامل الارتباط يساوي:

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}\, S_{yy}}} = \frac{2335}{\sqrt{(2048.5)(3284)}} = 0.9$$

1-كون جدول تحليل التباين باستخدام معامل الارتباط.

S.O.V	d.f	S.S	M.S	Fcal.
R(X1)	1	$r_{Xy}^2 S_{yy} = (0.9)^2(3248) = 2660.04$	2662.04	34.1
Error	n-2 10-2 8	$(1-r_{Xy}^2)S_{yy} = (1-(0.9)^2)(3248) = 623.96$	77.99	
Total	n-1 9			

وبمقارنة $F_{cal.}$ مع $F_{tab.}$:

$$F_{cal.} = 34.1 > F_{tab.} = F(0.05,1,8) = 5.32$$

اذاً ترفض فرضية العدم وتقبل الفرضية البديلة، أي ان هناك تأثير معنوي للمتغير المستقل المتمثل بالعمر على المتغير التابع المتمثل بضغط الدم.

ملاحظة: في هذا الاختبار لجدول تحليل التباين باستخدام معامل الارتباط فان الاختبار هو لاختبار الفرضية ادناه وهي نفس الفرضية التي نستخدمها عادتاً عند تكوين جدول تحليل التباين.

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

3- إيجاد معامل التحديد.

$$R^2 = \frac{Sum\,of\ Squares\,of\ due\,to\,Regression}{Sum\,of\ Squares\,of\ Total} = \frac{2661.04}{3284} = 0.81$$

أي ان المتغير X يفسر 0.81 من التغيرات الحاصلة في y من خلال النموذج الخطي الذي يصف العلاقة بينهما.

4- اختب الفرضية الاتية: $H_0 : \rho = 0$ / $H_1 : \rho \neq 0$

من اجل اجراء اختبار هذه الفرضية نستخدم اختبار t وكما يأتي:

$$t = \frac{r - r_0}{\sqrt{\frac{1-r^2}{n-1}}} = \frac{0.9 - 0}{\sqrt{\frac{1-(0.9)^2}{10-2}}} = 5.84$$

وهي نفس القيمة لاختبار t عند اختبار $H_0 : \beta_1 = 0$. وباستخدام جداول توزيع tنحصل على قيمة t الجدولية وكما يأتي:

$$t_{(\frac{\alpha}{2},n-2)} = t_{(0.025,8)} = 2.306$$

ومن خلال مقارنة القيمة المحسوبة مع القيمة الجدولية يلاحظ ان:

$$t_{cal.} = 5.84 > t_{(0.025,8)} = 2.306$$

وبذلك ومن خلال المقارنة ترفض فرضية العدم وتقبل الفرضية البديلة أي ان معامل الارتباط لا يساوي صفر وبالتالي هناك ارتباط معنوي بين المتغيرين X, y .

المخالفات او الخلل في فروض التحليل لنموذج الانحدار البسيط
كيفية اكتشافها وتصحيحها

لقد تم في محاضرات سابقة توضيح فروض التحليل لنموذج الانحدار البسيط من خلال أربعة نقاط، ومن اجل جعل هذه الفروض ملائمة بشكل اكثر مع نموذج الانحدار الخطي البسيط سنضيف فرض خامس وهو:

5- وجود علاقة خطية بين $\hat{y}, X$ تتمثل بمعادلة الخط المستقيم الاتية:

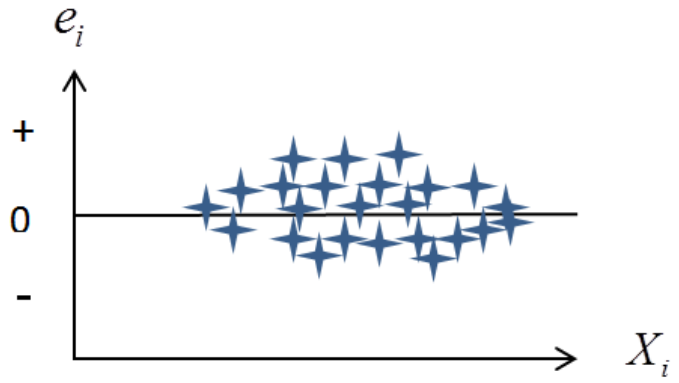
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 X_{i1}$$

ان اجراء اختبار لهذه الفروض واكتشاف ان كانت هناك مخالفات او خلل في احداها او اكثر وبيان كيفية تصحيحها والطريقة التي تستخدم لهذا الغرض هو ما يسمى بتحليل الأخطاء او المتبقيات (Residual Analysis) .

يمكن اختبار توفر فروض التحليل من خلال الرسم وكما يأتي:

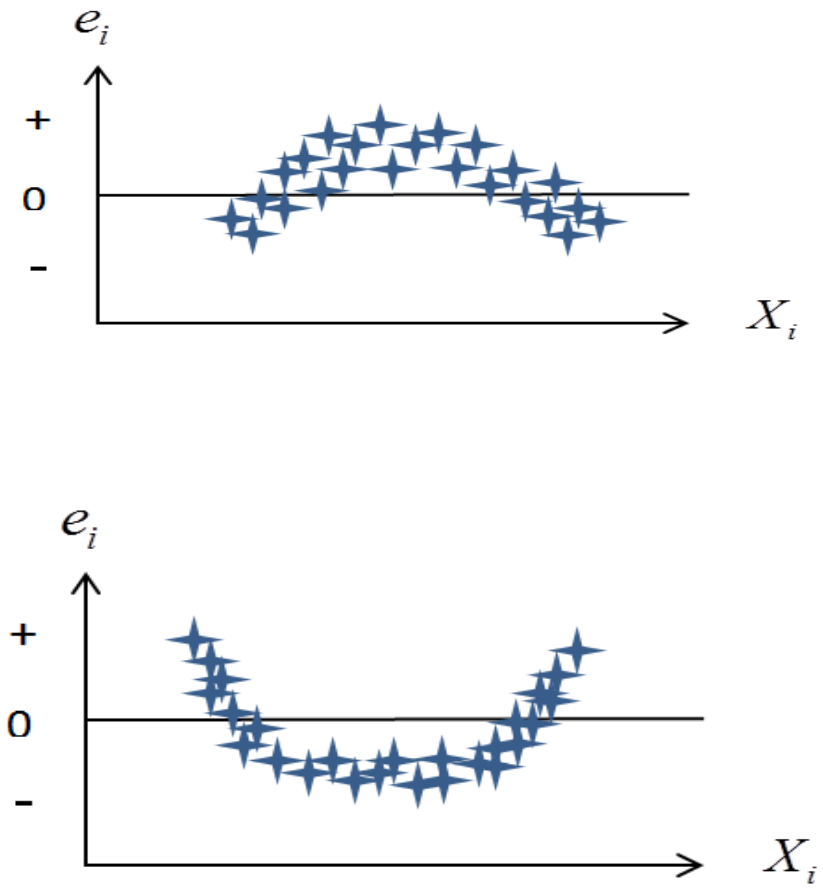
1- باستخدام الرسم البياني لقيم e_i ضد $\hat{y}_i$ او X_i والذي يفسر لنا بصورة عامة ما اذا كانت فروض التحليل متوفرة ام لا.

فاذا كانت النقاط متمركزة بصورة منظمة حول الصفر فان ذلك يدل على ان فروض التحليل متوفرة كما في الرسم الاتي:



اما اذا كان الرسم البياني يختلف عن هذا فان ذلك يدل على ان هناك خلل في فروض التحليل.

كما يمكن ان يكون الرسم بالشكلين ادناه:



ان الرسمين السابقين يختبران علاقة X بـ y حيث ان الرسمين يدلان على ان هناك معادلة ذات درجة اعلى من معادلة خط المستقيم وهي الملائمة اكثر للبيانات.

2-باستخدام مختبر احصائي

وهناك حالتين:

أ- عند عدم وجود تكرار لقيم المتغير المستقل X_i فأننا في هذه الحالة نستمر بإضافة حدود أخرى للمعادلة لتصبح من درجة ثانية او ثالثة ... الخ، الى ان نجد المعادلة الملائمة.

ب-اما عند وجود تكرار لقيم X_i فأننا نختبر نقص المطابقة Lack of fit، والذي يمكن وصفهُ كما يأتي:

اختبار نقص المطابقة A test of Lack of fit

يقصد باختبار نقص المطابقة هو مدى ملائمة النموذج الخطي للبيانات على شرط ان يكون هناك تكرار في قيم X_i .

فاذا كان الخطأ الناتج من عدم مطابقة النموذج معنوي فنقول ان النموذج لا يلائم البيانات. بذلك سيكون هنالك لدينا نوعين من الأخطاء.

1- الخطأ الصافي Pure Error

2- خطأ نقص المطابقة Lack of fit error

ويمكن توضيح الهدف من اختبار نقص المطابقة وهو اختبار هل ان النموذج الخطي ملائم للبيانات ام لا. ومن اجل اجراء هذا الاختبار يكون هناك فرضيتين هما:

H_0 : عدم وجود نقص في المطابقة أي ان النموذج الخطي يلائم البيانات.

H_1 : يوجد نقص في المطابقة أي مطابقة النموذج الخطي البيانات أي ان النموذج الخطي لا يلائم البيانات.

في حالة تم اختبار نقص المطابقة وظهر ان النموذج الخطي يلائم البيانات (قبول فرضية العدم) فانهُ نجري اختبار معنوية β_1 ، أي نجري اختبار الفرضية الاتية:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

ونرى فيما اذا كان التغيرات الحاصلة في y هو سببها المتغير X . اما ذا تم رفض H_0 في اختبار نقص المطابقة أي ان النموذج الخطي لا يلائم البيانات فلا داعي لاختبار معنوية β_1 لأنه أصلا النموذج الخطي لا يلائم البيانات حيث انهُ يوجد ربما نموذج من درجة اعلى مثلاً يكون اكثر ملائمة من النموذج الخطي.

كيفية اجراء اختبار نقص المطابقة

على فرض انهُ لدينا جدول مكون من قيم X_i (تتضمن تكرارات في قيم X_i) المقابلة لقيم y_i وكما يأتي:

X_i	y_i
$x_{11}=a$	y_1
$x_{21}=b$	y_2
$x_{31}=a$	y_3
$x_{41}=c$	y_4
$x_{51}=b$	y_5
$x_{61}=a$	y_6
⋮	⋮
$x_{n1}=F$	y_n

يتم أولاً تريب قيم الجدول اعلاه بالشكل الاتي:

قيم X_i المتكررة	قيم y_i المقابلة لقيم X_i المتكررة	d.f	S.S
X_i	$y_{11}, y_{12}, y_{13}, \cdots, y_{1k}$	r_1-1	SS_{y_1}
X_i	$y_{21}, y_{22}, y_{23}, \cdots, y_{2k}$	r_2-1	SS_{y_2}
⋮	⋮ ⋮ ⋮ ⋮ ⋮	⋮	⋮
X_i	$y_{k1}, y_{k2}, y_{k3}, \cdots, y_{kk}$	r_k-1	SS_{y_k}
		$\sum_{i=1}^{k}(r_i-1)$	

$$\sum_{i=1}^{k} SS(pure\ Error)$$

S.O.V	d.f	S.S	M.S	$F_{cal.}$
$R(X_i)$	1	$SSR(X_i)$	$M.S.R(X_1)$	
$Error$	$n-2$	SSe	$M.S.e$	
$L.o.f$	$d.f(L.o.f)=$ $=d.f.e-d.f(p.e)$	$SS(L.o.f)$	$M.S(L.o.f)$	$F_{cal.}=\frac{M.S(L.o.f)}{M.S(p.e)}$
$P.e$	$d.f(p.e)$	$SS(p.e)$	$M.S(p.e)$	
$Total$	$n-1$			

حيث ان $SS(L.o.f) = SSe - S.S(p.e)$

اما F الجدولية فتكون كما يأتي:

$$F_{(\alpha, v1, v2)}$$

$L.o.f$ درجة حرية $P.e$ درجة حرية

مثال: من خلال البيانات التالية اختبر هل النموذج الخطي يلائم البيانات.

X_i	y_i
35	112
25	125
40	128
35	115
64	162
25	130
50	142
67	158
69	175
50	140
70	170
50	145

الحل: نرتب القيم تصاعدياً حسب X_i .

X_i	y_i
25	125
25	130
35	112
35	115
40	128
50	142
50	140
50	145
64	162
67	158
69	175
70	170

يتم تكوين جدول لحساب مجموع مربعات الخطأ الصافي (p.e) وكما يلي:

قيم X_i المتكررة	قيم y_i المقابلة لقيم X_i المتكررة	d.f	S.S
$X=25$	125 130	1	12.5
$X=35$	112 115	1	4.5
$X=50$	142 140 145	2	12.67
Total		4	29.67=SS(p.e)

$$\bar{X}=\frac{125+130}{2}=127.5,$$

$$SS=\sum(X_i-\bar{X})^2$$

$$SS=(125-127.5)^2+(130-127.5)^2$$

$$=6.25+6.25$$

$$=12.5$$

نكون الان جدول تحليل التباين لاجراء الاختبار.

S.O.V	d.f	S.S	M.S	$F_{cal.}$
$R(X_i)$	1	3963.65	3963.69	
Error	10	836.02	83.602	
L.o.f	10−4=6	806.35	134.39	18.12
P.e	4	29.67	7.4175	
Total	11			

حيث ان

$$\begin{aligned} S.S(L.o.f) &= SSe - SS(p.e) \\ &= 836.02 - 29.67 \\ &= 806.35 \\ d.f(L.o.f) &= d.f.e - d.f(p.e) \\ &= 10 - 4 \\ &= 6 \end{aligned}$$

وقيمة F الجدولية هي:

$$\begin{aligned} F_{tab.} &= F(\alpha,\ d.f(L.o.f),\ d.f(p.e)) \\ &= F(0.05, 6, 4) \\ &= 4.16 \end{aligned}$$

وبمقارنة قيمة F المحسوبة مع قيمة F الجدولية نحصل على:

$$F_{cal.} = 18.12 > F_{tab.} = 4.16$$

ومن خلال المقارنة وبما ان قيمة F المحسوبة اكبر من قيمة F الجدولية ترفض فرضية العدم أي ان هناك نقص في المطابقة أي ان النموذج الخطي لا يلائم البيانات بل ان هناك معادلة من درجة ثانية او اكثر قد تلائم البيانات.

في حالة ضهور معنوية اختبار نقص المطابقة أي ان النموذج الخطي لا يلائم البيانات فاننا نكون اما حلين:

1- اما ام نجد نموذج اخر غير النموذج الخطي قد يلائم البيانات ملائمة اكثر ويكون هذا النموذج هو نموذجاً من الدرجة الثانية والثالثة أي ان:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_3^3 + ... + e_i$$

او قد يكون نموذجاً تضاعفي مثل:

$$y = \beta_0 + X_1^{\beta_1} + e_i$$

استخدام تحويلات اما لقيم y او لقيم X او لكليهما من اجل تقريب النموذج الى النموذج الخطي.

وهناك عدة طرق لتحويل البينات أهمها:

1- التحويل اللوغارتمي The Log transformation

ويشترط في هذا التحويل ان تكون جميع قيم y_i موجبة، ويمكن إضافة الواحد الصحيح او 0.5 الى القيم ثم اجراء التحويل. ان التحويل اللوغارتمي لا يحقق فقط العلاقة الخطية وانما ايضاً يجعل تباينات الأخطاء متجانسة.

فمثلاً :

$$Log\, y_i = \beta_0 + \beta_1 X_i + e_i$$

او يمكن اخذ الـ Log للـ X_i :

$$y_i = \beta_0 + \beta_1\, Log X_i + e_i$$

او كليهما وحسب طبيعة البيانات

$$Log\, y_i = \beta_0 + \beta_1 Log\, X_i + e_i$$

2- التحويل الجذري Square Root Transformation

هو اخذ الجذر التربيعي اما لقيم y_i او لقيم X_i او كليهما أي ان:

$$y_i' = \sqrt{y_i} \quad \text{او} \quad X_i' = \sqrt{X_i}$$

وهذا التحويل يلائم البيانات التي تتبع توزيع بواسون.

3- التحويل الزاوي The Aresine Transformation

وصيغتهُ هي كالاتي:

$$y = \sin^{-1}\sqrt{y_i}$$

ويستخدم هذا التحويل لتثبيت التباين وخاصة عندما تكون قيم y_i على شكل نسب مئوية.

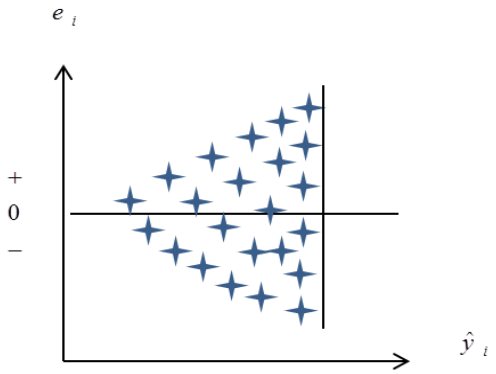
وبصورة عامة اذا كانت القيم بين 30% و 70% فلا حاجة للتحويل اما اذا كانت النسب بين 0 و 30% او بين 70% و 100% وليس كليهما نستخدم التحويل الجذري، اما اذا كانت قيم y_i ذات نسب غير ما ذكر أعلاه، فنستخدم التحويل الزاوي.

اختبار هل ان تباين الأخطاء او المتبقي ثابت ومتجانس

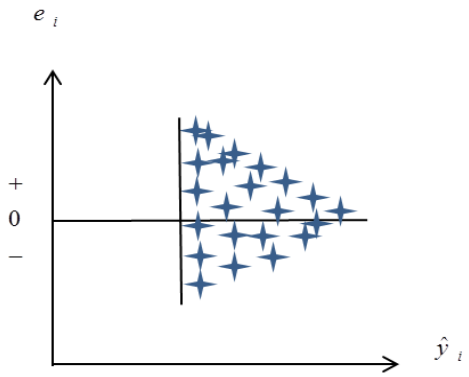
يمكن اجراء اختبار تباين الأخطاء من حيث كونهُ ثابت ومتجانس من خلال طريقتين:

1- باستخدام الرسم البياني

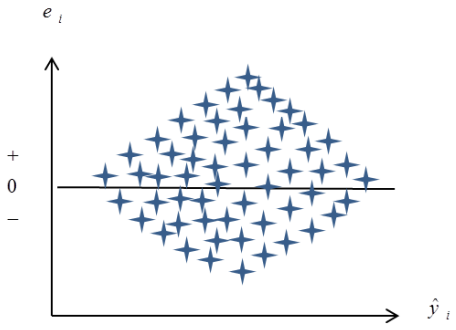
عند رسم قيم e_i مقابل قيم $\hat{y}_i$ او X_i وكان الشكل البياني كأحد الاشكال الاتية فان:



زيادة في تباين الخطأ



نقصان في تباين الخطأ



زيادة ونقصان في تباين الخطأ

فان ذلك يدل على عدم تجانس تباين الخطأ أي ان تباين الأخطاء غير ثابت Heteroscedastic.

2- باستخدام مختبر احصائي

ترتب البيانات تصاعدياً حسب قيم المتغير المستقل X_i ثم بعد ذلك تقسم البيانات المرتبة الى قسمين (مع حذف بعض البيانات الوسطى) ثم اجراء تحليل الانحدار لكل قسم لإيجاد $MSe_i = S_i^2$ لكل قسم ثم استخدام المختبر الاحصائي :

$$F = \frac{Max\ S_i^2}{Min\ S_i^2}$$

فاذا كانت قيمة F المحسوبة اكبر من F الجدولية دل ذلك على وجود اختلاف في التباين.

مثال:

الجدول التالي يمثل بيانات مكونة من متغيرين، الأول يمثل مقدار المتوسط الشهري لمبيعات سلعة ما المتمثل بـ (y) في كل من 30 متجراً ومقدار نفقات الإعلانات الشهرية عليها المتمثل بـ (X). اختبر هل ان تباين الخطأ ثابت ومتجانس.

رقم المتجر	X_i	y_i	رقم المتجر	X_i	y_i
1	3000	81464	16	12525	130963
2	3085	72344	17	13700	147041
3	3150	72661	18	14995	180732
4	3225	90743	19	15000	172021
5	5350	98588	20	15050	178187
6	6090	96507	21	15150	155931
7	8885	115814	22	15175	166200
8	8925	126574	23	15200	185304
9	8950	123181	24	16500	188851
10	9000	131434	25	16800	172577
11	9015	114133	26	17830	192424
12	11345	140564	27	19000	218715
13	12275	151352	28	19200	192482
14	12310	144630	29	19350	214317
15	12400	146926	30	19500	203112

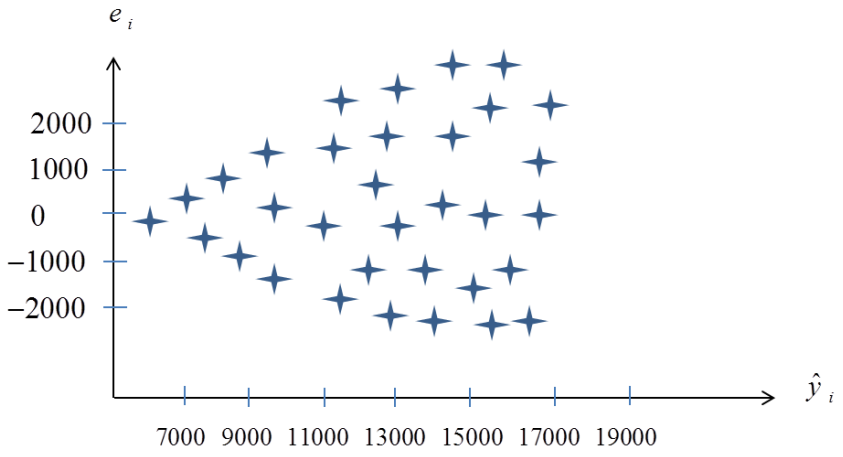
لقد تم تكوين معادلة خط الانحدار وجدول تحليل التباين وكما يلي:

$$\hat{y}_i = 49443.3838 + 8.0484\, X_{i1}$$

S.O.V	d.f	S.S	M.S	$F_{cal.}$
$R(X_i)$	1	4920001188	4920001188	607.46
Error	28	226780353	8099298.3	
Total	29			

ولمعرفة مدى توفر شروط فروض التحليل (تجانس تباين الخطأ) نتبع ما يلي:

1- استخدام الرسم البياني: عند رسم e_i مقابل قيم $\hat{y}_i$ كانت النتائج كالاتي:



من خلال الرسم البياني يتضح ان هناك وجود اختلاف في تباين الخطأ.

2- باستخدام مختبر احصائي

نرتب البيانات ترتيباً تصاعدياً تبعاً للمتغير المستقل X (والتي هي اصلاً مرتبة بهذا الترتيب) ثم يتم تقسيم البيانات الى قسمين بعد حذف الستة مشاهدات الوسطى (المرقمة من 13-18) وكما يلي:

رقم المتجر	X_i	y_i	رقم المتجر	X_i	y_i
1	3000	81464	16	12525	130963
2	3085	72344	17	13700	147041
3	3150	72661	18	14995	180732
4	3225	90743	19	15000	172021
5	5350	98588	20	15050	178187
6	6090	96507	21	15150	155931
7	8885	115814	22	15175	166200
8	8925	126574	23	15200	185304
9	8950	123181	24	16500	188851
10	9000	131434	25	16800	172577
11	9015	114133	26	17830	192424
12	11345	140564	27	19000	218715
13	12275	151352	28	19200	192482
14	12310	144630	29	19350	214317
15	12400	146926	30	19500	203112

وقد تم حساب كل مما يأتي:

متوسط مربعات الخطأ لبيانات القسم الأول فكان:

$$Mse_1 = 31066726.3$$

متوسط مربعات الخطأ لبيانات القسم الثاني فكان:

$$Mse_1 = 140450776.4$$

وباستخدام اختبار F نحصل على:

$$F = \frac{Max\ S_i^2}{Min\ S_i^2} = \frac{140450776.4}{31066726.3} = 4.52$$

وبمستوى معنوية $\alpha = 0.05$ ودرجات حرية المحسوبة بالصيغة الاتية:

$$v_1 = v_2 = \frac{n-d-2k}{2} = 10$$

حيث ان :

N: حجم العينة والذي تساوي 30 في مثالنا هذا.

K: عدد المعلمات المقدرة والذي تساوي 2 في مثالنا هذا.

d: عدد المشاهدات الوسطية المحذوفة والذي تساوي 6 في مثالنا هذا.

ومن جداول F نجد القيمة الجدولية والتي تساوي: $F(0.05,10,10) = 3.02$

فبما ان F المحسوبة اكبر من F الجدولية لذا فان S_1^2, S_2^2 مختلفات أي لا يوجد هناك تجانس في تباين الخطأ.

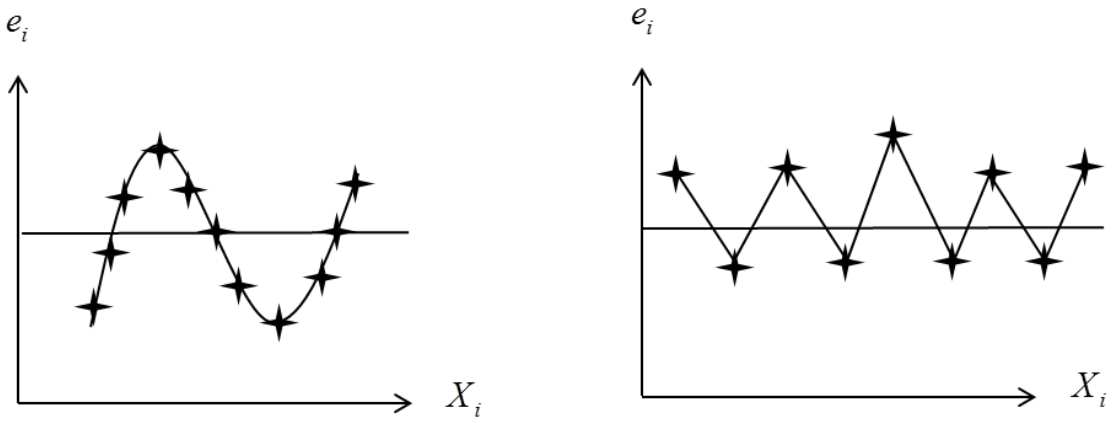
اختبار هل ان الأخطاء مستقلة ام ان هناك ارتباطاً ذاتياً فيما بينها

في كثير من المسائل وخاصة الاقتصادية منها تكون على شكل سلاسل زمنية مما يؤدي الى ان الخطأ في فترة زمنية معينة (e_i) يكون مرتبطاً مع الخطأ في الفترة الزمنية التي قبلهُ (e_{i-1}). والاخطاء المرتبطة على مدار الزمن تسمى أخطاء مرتبطة ذاتياً Autocorrelated .

وللكشف عن وجود ارتباط ذاتي بين الأخطاء نستخدم ما يلي:

1- باستخدام الرسم البياني

عند رسم قيم (e_i) مقابل قيم X_i فانهُ في حالة وجود ارتباط ذاتي بين الأخطاء يكون الرسم البياني كأحد الشكلين التاليين:



ارتباط ذاتي موجب ارتباط ذاتي سالب

في حالة ارتباط ذاتي موجب فان هناك عدة نقاط تكون موجبة تليها عدة نقاط سالبة ثم تليها عدة نقاط موجبة وهكذا. اما في حالة وجود ارتباط ذاتي سالب فان نقاط الأخطاء تتعاقب بالإشارة. فالاولى موجبة مثلاً والثانية سالبة والثالثة موجبة والرابعة سالبة وهكذا.

2- باستخدام مختبر احصائي

يمكن اجراء اختبار دربن واتسن Durbin-Watson في اختبار فيما اذا كان هناك ارتباط ذاتي بين الأخطاء ام لا.

ولإجراء هذا الاختبار يتم اختبار الفرضية الاتية:

$$H_0 : r = 0$$
$$H_1 : r \neq 0$$

تفسير فرضية العدم: لا يوجد ارتباط ذاتي بين الأخطاء.

تفسير الفرضية البديلة: يوجد ارتباط ذاتي بين الأخطاء.

ان صيغة اختبار دربن واتسن تأخذ الشكل الاتي:

$$D = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

فاذا كانت $D < dl$ نرفض العدم

واذا كانت $D > du$ نقبل العدم

اما اذا كانت $dl \le D \le du$ فيعد الاختبار غير حاسم.

حيث ان du هو الحد الأعلى وان dl فيمثل الحد الأدنى.

حيث ان قيم du و dl تستخرج من جداول دبن واتسن بدرجة حرية n ومستوى معنوية α ، وعدد المعلمات في النموذج.

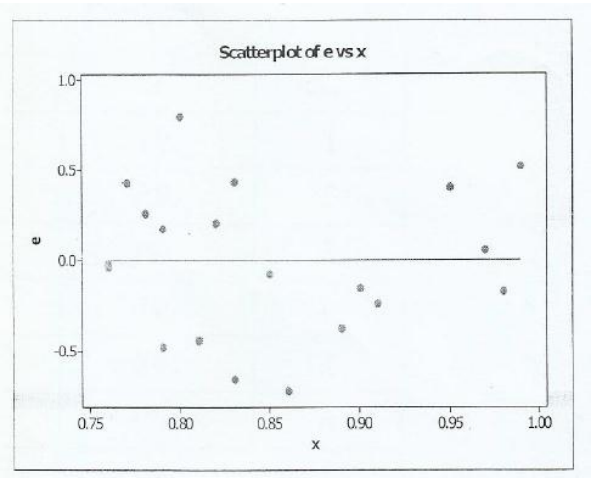
مثال

البيانات الاتية تمثل نسبة حصة متجر من معجون اسنان معين (y) ومقدار سعر البيع (X) لكل باوند منها ولعشرين شهراً متتالياً.

اختبر هل ان الأخطاء مستقلة ام انها مرتبطة ذاتياً.

الشهر	X	y	الشهر	X	y
1	0.97	3.36	11	0.79	7.25
2	0.95	4.2	12	0.83	6.09
3	0.99	3.33	13	0.81	6.8
4	0.91	4.54	14	0.77	8.65
5	0.98	2.89	15	0.76	8.43
6	0.90	4.87	16	0.80	8.29
7	0.89	4.9	17	0.83	7.18
8	0.86	5.29	18	0.79	7.9
9	0.85	6.18	19	0.76	8.45
10	0.82	7.2	20	0.78	8.23

1- نستخدم أسلوب الرسم البياني لقيم e_i مقابل قيم X_i وكما يأتي:



ان الرسم البياني يدل على ان هناك ارتباطاً ذاتياً موجباً بين الأخطاء حيث ان هناك مجموعة من الأخطاء موجبة تليها أخرى سالبة ثم تليها موجبة.

2- باستخدام المختبر الاحصائي لدربن واتسن وكما يلي:

حيث ان الصيغة الحسابية هي:

$$D = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

نجد قيم e_i وقيم e_i^2 أي الأخطاء وكما يلي:

الشهر	e_i	e_i^2	$(e_i - e_{i-1})^2$	الشهر	e_i	e_i^2	$(e_i - e_{i-1})^2$
1	0.2812	0.0791		11	-0.471	0.2218	0.4606
2	0.3654	0.1335	0.00708	12	-0.659	0.4348	0.0354
3	0.4670	0.2181	0.01032	13	-0.4352	0.1894	0.0502
4	-0.2662	0.0709	0.5375	14	0.4432	0.1964	0.7715
5	-0.2159	0.0466	0.00863	15	-0.0197	0.0004	0.214
6	-0.1791	0.0321	0.00135	16	0.8119	0.6592	0.6915
7	-0.3920	0.1537	0.453	17	0.4306	0.1854	0.1453
8	-0.7307	0.5339	0.1147	18	0.179	0.32	0.0633
9	-0.0836	0.007	0.4187	19	0.0003	0.0001	0.0319
10	0.2077	0.0431	0.0848	20	0.2661	0.0708	0.0706
				المجموع		3.3082	3.763

وبتطبيق صيغة دربن واتسن نحصل على:

$$D=\frac{\sum_{i=2}^{n}(e_i-e_{i-1})^2}{\sum_{i=1}^{n}e_i^2}=\frac{3.7634}{3.3082}=1.14$$

ومن جداول درن واتسن ولحجم عينة n=20 ومستوى معنوية 0.05 نحصل على قيم الحد الأدنى والحد الأعلى لدربن واتسن وكما يلي:

$dl=1.2$

$du=1.41$

وبما ان $D=1.14<dl=1.2$

بذلك ترفض العدم أي ان هناك ارتباط ذاتي بين الأخطاء.