الانحدار الخطي البســـيط
(النموذج الخطي لمتغيرين)

1 مقدمه:

تحليل الانحدار من أكثر الأدوات المستعملة في التحليل القياسي لذا سوف نبدأ بتحديد الخطوط العريضة لتحليل الانحدار. نبدأ بالسؤال الأساسي: **ما هو تحليل الانحدار؟**

تحليل الانحدار يهتم بوصف وتقييم العلاقة بين متغير (عادة يسمى المتغير التابع) وواحد آو اكثر لمتغيرات أخرى (تسمى عادة المتغيرات المفسرة آو المتغيرات المستقلة) ويرمز للمتغير المفسر بـ y والمتغيرات المفسرة بـ $x_1 \ x_2 \ x_3 \ldots . x_n$.

التفسير الحرفي لكلمة انحدار تعني" ارتداد أو انكفاء أو رجوع" في الحقيقة تحليل الانحدار لا يربطه بهذا المعنى أي رابط.

كلمة انحدار استخدمت من قبل سير فرنسيس جالتون Sir Francis Galton (1911-1982) من إنجلترا. والذي كان يدرس العلاقة بين طول الأبناء وطول الآباء والذي لاحظ جالتون أن الطول يميل إلى المعدل مع أن الآباء الطوال يكون أبنائهم طوال

والآباء القصار يميل أبنائهم لان يكونوا قصار أي أن هناك ميل عند ألأبناء للمعدل أي أن هناك انحدار نحو المعدل.في دراسات أخرى مشابهه تحصل على نفس النتيجة التي تحصل عليها جالتون.

بالعودة إلى الرموز التي استخدمناها حيث رمزنا للمتغير المفسَر بـ y والمتغيرات المفسرة بـ x_3 x_1 x_2 .... x_n . إذا كانت $k=1$ ، أي إن هناك متغير مستقل واحد فقط من المتغيرات المفسرة. أي ان هناك x واحدة فقط. يعرف هذا بالانحدار البسيط. وهو ما سوف يتم مناقشته في هذا الفصل. إذا كانت $k>2$ ، آي أن هناك اكثر من x واحد و متغير مستقل. نحصل على ما يعرف بالانحدار المتعدد. والذي سوف نناقشه في الفصل القادم.

مثال 1 : الانحدار البسيط.

y = المبيعات

x = النفقات الاعلانيه.

حيث يتم تحديد العلاقة بين المبيعات والنفقات الأعلانيه.

مثال 2: الانحدار المتعدد.

Y = استهلاك ألا سره.

X_1 = دخل ألا سره.

X_2 = الأصول المالية للاسره

X_3 = حجم ألا سره

تحديد العلاقة بين نفقات استهلاك ألا سره من جهة والدخل، والأصول المالية و حجم لأ سره من جهة اخرى.

هناك عدة أسباب لدراسة هذه العلاقات. يمكن استخدام ذلك في :

1- تحليل تأثير بعض السياسات التي تتضمن تغير قيم لفرد معين. في المثال الأول نستطيع أن نحلل تأثير النفقات الاعلانيه على كمية المبيعات.

2- التنبؤ بقيم Y من قيم X.

3- اختبار مدى معنوية العلاقة بين آي من X و Y.

في مناقشتنا نفرق بين المتغيرY و المتغيرات X. افترضنا أن المتغيرات X هي المتغير الذي يؤثر على المتغير Y. هناك العديد من المصطلحات التي نطلقها على X , Y توجد في الجدول 1.3 .

جدول 3: مصطلحات المتغير التابع و المتغير المستقل.

X	Y
1-مًتنبأ	متَنبأ به
2-مفسر	مفَسر
3-مستقل	تابع
4-مسبب	متأثر
5-خارجي	داخلي
6-المتغير المتحكم	المتغير الهدف

كل من هذه المصطلحات يستخدم حسب الغرض من تحليل الانحدار فالمصطلح الأول يستخدم في عملية التنبؤ بينما المصطلحات الأخرى تستخدم في مناقشة الانحدار. اما المصطلح خارجي وداخلي تستخدم فقط من قبل القياسيين. بينما المصطلح الأخير يستخدم في التجارب الخاصة بدراسة تأثير مسببات معينه على متغير مستهدف.

2 تحديد العلاقة:

العلاقة بين Y و X تمثل بالتالي:

$$Y = f(X) \qquad \ldots\ldots \qquad \ldots. \quad 1.2$$

حيث ترمز لـ Y كدالة لـ X. نستطيع إن نقسم العلاقة إلى نوعين:-

1- علاقة رياضية محدده.Deterministic

2- علاقة إحصائيه لا تعطي قيمه فريدة لـ Y من قيمه محدده من X. ولكن يمكن أن توصف بصيغة احتماليه.

سوف نتحدث هنا في تحليل الانحدار عن العلاقة من النوع الأول على سبيل المثال العلاقة بين المبيعات و النفقات الاعلانيه يمكن أن توصف بما يلي:-

$$Y = 2500 + 100X \qquad \ldots\ldots\ldots \quad 2.2$$

هذه علاقة محدده deterministic . حيث يمكن تحديد المبيعات لكل مستوى من النفقات الاعلانيه، كما يلي:-

Y	X
2500	0
4500	20
7500	50
12500	100

في الجانب الأخر، نفترض أن العلاقة بين المبيعات والنفقات الاعلانيه كما يلي:-

$$Y = 2500 + 100X + u \qquad 3.2$$

قيمة u تتراوح بين قيم معينه حسب جدول للاحتمالات يعطي لكل قيمه احتمال معين على سبيل المثال:- الاحتمالات أن قيمة u اكبر من 500 يساوي ½ واقل من 500 يساوي ½ . لذا لا نستطيع تحديد قيمة Y من قيمة X نقول إن هناك العديد من قيمY المقابلة لقيمة واحدة من X . إذا كان هناك توزيع طبيعي لقيم Y المقابلة لقيمه واحدة من X .

بالرجوع إلى المعادلة التي تمثل الدالة.نستطيع القول إن الدالة تمثل الخط $f(x) = \alpha + \beta X$

بينما العلاقة العشوائية هي $f(x) = \alpha + \beta X + u$

حيث تمثل u الخطأ العشوائي. وتمثل α β معاملات الانحدار.

لماذا يتم إضافة الخطأ العشوائي للمعادلة؟

1- يمثل عنصر العشوائية في استجابة الإنسان مثل اختلاف النفقات الاستهلاكية من فرد لأخر مع العلم انهم قد يتساووا في الدخل.

2- تأثير عوامل أخرى محذوفة مثل العادات حجم ألا سره وغيرها من العوامل.

3- خطأ في قياس المتغير التابع.

الهدف هو الحصول على تقدير للمعاملات الغير معروفه α β للقيام بعملية التقدير يجب افتراض بعض الافتراضات الخاصة بالخطأ العشوائي:

1- الوسط الصفري. $E(u)=0$

-أن وسط التوزيع الاحتمالي الخاص بالمتغير العشوائي = الصفر. إي أن قيم u تتمركز حول الصفر.

2- تساوي التباين $V(u)=\sigma^2$.

تباين التوزيع الاحتمالي الخاص بالعناصر العشوائية u يساوي قيمه ثابتة وموجبة.

3- استقلالية الخطأ العشوائي: أي أن التغاير، درجة الارتباط بين قيم العشوائي = الصفر

أي انها مستثقله عن بعضها. $COV(u_i, u_j)=0$

4- التوزيع الطبيعي للخطأ العشوائي. $u_i \sim N$

تمثل هذه الافتراضات بالتالي $u_i \sim N(0, \sigma^2)$

3 تقدير نموذج الانحدار باستخدام طريقة المربعات الصغرى:-

هناك عدة طرق لتقدير معاملات معادلة الانحدار أهمها 1)طريقة المربعات الصغرى. 2) طريقة الإمكانية العظمى.

في المرحلة الأولى نفترض وجود الفروض الاساسيه لمعالجة النموذج الخطي. وفي المراحل اللاحقة نتعرض للحالات التي تكون فيها هذه الفروض غير صحيحه.

نموذج الانحدار بالافتراضات الأساسية كما يلي:

$$Y_i = \alpha + \beta X_i + u$$

هي المعادلة الأساسية التي تصور العلاقة بين التابع والمستقل حيث i تعتمد على العينة التي يبلغ حجمها n . بالأضافه إلى المعادلة ألأساسيه نقول أن النموذج يحتوي افتراضات عن المتغير العشوائي.

تقدير النموذج يتم بغرض الحصول على مقدرات معالم نموذج الانحدار البسـيط. نموذج الانحدار البسيط يتضمن ثلاث معالم هي, α , معلمة القاطع، β , معلمة الميل، σ^2 معلمة التباين. المراد هو استخدام إحصائيات المتغيرات التابعة والمتغيرات المستقلة حسب الطرق الإحصائيه الملائمة للحصول على مقدرات لهذه المعالم.

طريقة المربعات الصغرى

تعتمد طريقة المربعات الصغرى العادية على الحصول على مقدرات , الانحدار حيث تمثل α معلمة القاطع، , β معلمة الميل. بحيث يتم تصغير مجموع مربعات البوا قي إلى آدني قيمه

لها. بحيث يجري تعريف مكون يطلق علية مجموع المربعات البوا قي وبعد ذلك يشرع في الحصول على α ، β , بحيث يتم تصغير هذا المكون إلى أدنى قيمه له.

طريقة المربعات الصغرى تعطينا مقدرات الانحدار α ، β , ولكن لا تعطينا مقدرة التباين وهذا يعتبر من نقاط ضعف طريقة المربعات الصغرى.

المعيار الخاص في المربعات الصغرى العادية: النموذج المقدر هو كما يلي

$$Y_i = \hat{\alpha} + \hat{\beta}X + u_i$$

u هي البوا قي والتي تساوي من النموذج $u_i = Y_i - (\hat{\alpha} + \hat{\beta}X)$ نموذج الانحدار ممكن أن يمر من خلال انتشـار البيانات الخاصة بـ Y, X، الخط المقدر هنا هو الذي يعطي Y المقدرة $\hat{Y} = \hat{\alpha} + \hat{\beta}X$

إذا أخذنا إحداثيات القيم Y,X إحداثيات النقطة الأولى تنقسم إلى قسمين، قسم من المحور الأفقي في النموذج المقدر، هذا عبارة عن $\hat{Y} = \hat{\alpha} + \hat{\beta}X$ الجزء الثاني عبارة عن قيمة البوا قي. فالمشاهدة Y هي حصيلة جمع $u + \hat{Y} = \hat{\alpha} + \hat{\beta}X$ أي أن أي مشاهده مكونه من جانبين، جانب الخط المقدر والبواقي. البواقي بحكم أنها مقدرة العنصر العشوائي يمكن أن تكون موجبة وممكن أن تكون سالبه وكذلك من الناحية النظرية يمكن أن تساوي الصفر.

للحصول على مقدرات المربعات الصغرى العادية يجب أن نحصل أولا على البواقي:

$$u_i^2 = (\hat{Y} - (\hat{\alpha} + \hat{\beta}X))^2$$

مجموع مربعات البواقي $=\sum u^2$

$$\sum u_i^2 = \sum (Y_i - \hat{\alpha} - \hat{\beta}X)^2$$

يتم التوصل إلى الخط الذي تكون فيه مجموع مربعات البواقي اصغر ما يمكن [اختيار الخط الذي يدني مجموع مربعات البواقي إلى أصغر ما يمكن]. باستخدام الرياضيات فأن شرط الدرجة الأولى يتطلب أجراء التفاضل بالنسبة للمجاهيل α β نستخدم التفاضل الجزئي وبعد ذلك نساوي المعادلات التي تم أل تحصل عليها بالصفر ثم نطبق المعادلات الآنية للحصول على قيم المقدرات.

$$\frac{\partial\left(\sum u_i^2\right)}{\partial\hat{\alpha}} = (2)\sum(Y_i - \hat{\alpha} - \hat{\beta}X)(-1) = 0$$

$$= (-2)\sum(Y_i - \hat{\alpha} - \hat{\beta}X) = 0$$

نساوي بالصفر $\sum(Y_i - \hat{\alpha} - \hat{\beta}X) = 0$

بادخال المجموع Σ وحيث ان αعدد ثابت فأن $n\ \alpha = \Sigma\ \alpha$ ثم بقسمة المعادلة على n نحصل على مايلي:

$$\sum Y_i - \sum_i \hat{\alpha} - \hat{\beta}\sum_i X = 0$$
$$\sum_i \hat{\alpha} = \sum Y_i - \hat{\beta}\sum_i X$$
$$n\hat{\alpha} = \sum Y_i - \hat{\beta}\sum_i X$$
$$\hat{\alpha} = \frac{\sum Y_i}{n} - \hat{\beta}\frac{\sum_i X}{n}$$
$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$

$$\frac{\partial\left(\sum u_i^2\right)}{\partial\hat{\beta}} = (2)(\sum(Y_i - \hat{\alpha} - \hat{\beta}X)(-X) = 0$$

$$\frac{\partial\left(\sum u_i^2\right)}{\partial\hat{\beta}} = -2\sum X(Y_i - \hat{\alpha} - \hat{\beta}X) = 0$$

$$(\sum X(Y_i - \hat{\alpha} - \hat{\beta}X) = 0$$

$$-\sum XY + \sum X\alpha + \sum \beta X^2 = 0$$

نساوي بالصفر

$$\sum XY = \alpha\sum X + \beta\sum X^2$$

$$\sum_{i=1}^{n} X_i Y_i = \left(\sum_{i=1}^{n} X_i\right)\hat{\alpha} + \left(\sum_{i=1}^{n} X^2\right)\hat{\beta} \qquad 2.5$$

بالتعويض بقيمة α نحصل على

$$\sum XY = \sum X\left(\frac{\sum Y}{n} - \beta\frac{\sum X}{n}\right) + \beta\sum X^2 \qquad 2.6$$

بالضرب في n

$$n\sum XY = \sum X\sum Y - \beta(\sum X)^2 + \beta\ n\sum X^2$$

$$\begin{aligned} n\sum XY - \sum X\sum Y &= -\beta(\sum X)^2 + \beta\ n\sum X^2 \\ &= \beta\ n\sum X^2 - \beta(\sum X)^2 \qquad 2.7 \\ &= \beta\left(n\sum X^2 - (\sum X)^2\right) \end{aligned}$$

معادلة 2.7 تسمى المعادلات الطبيعية ونستطيع استخراج قيم β α منها

بالتعويض نحصل على

$$\boxed{\hat{\beta} = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum^{n} X_i \sum Y_t}{n\sum X_i^2 - \left(\sum X\right)^2} =}$$

من الممكن الحصول على المقدرات باستخدام الانحرافات كما يلي:

$$\sum y_i = \sum (Y - \bar{Y})^2 \qquad = \sum Y_i^2 - n\bar{Y}^2$$

$$\sum xy = \sum (X - \bar{X})(Y - \bar{Y}) = \sum XY - n\overline{XY}$$

$$\sum x^2 = \sum (X - \bar{X})^2 \qquad = \sum X^2 - n\bar{X}$$

ولكن

$$\sum Y_i X_i = \sum X_i (\bar{Y}_i - \hat{\beta}\bar{X}) + \hat{\beta}\sum X_i^2$$

$$\sum Y_i X_i = n\bar{X}(\bar{Y} - \hat{\beta}\bar{X}) + \hat{\beta}\sum X_i^2$$

$$\sum Y_i X_i = n\bar{X}\bar{Y} - \hat{\beta}n\bar{X}\bar{X}) + \hat{\beta}\sum X_i^2$$

$$\sum Y_i X_i - n\bar{X}\bar{Y} = -\hat{\beta}n\bar{X}\bar{X} + \hat{\beta}\sum X_i^2$$

$$\sum XY - n\overline{XY} = -\hat{\beta}n\bar{X}^2 + \hat{\beta}\sum X_i^2$$

$$\sum xy = \beta\left[nX^2 - \sum X^2\right]$$

$$\sum xy = \beta\sum x^2$$

$$\beta = \frac{\sum xy}{\sum x^2}$$

مثال (1)

X	Y	X^2	x	y	XY	xy	x^2
2	4	4	−2	−4	8	8	4
3	7	9	−1	−1	21	1	1
1	3	1	−3	−5	3	15	9
5	9	25	1	1	45	1	1
9	17	81	5	9	153	45	25
∑X= 20	**∑Y =40**	**$\sum X^2$= 120**			**230 ∑XY=**	**∑xy=70**	**$\sum x^2$= 40**

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum^{n} X_i \sum Y_t}{n\sum X_i^2 - \left(\sum X\right)^2} =$$

$$\hat{\beta} = \frac{(5)(230)-(20)(40)}{5(120)-(20)^2} = 1.75$$

$$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X} = 8 - 1.75(4) = 1$$

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X =$$

$$\hat{Y} = 1 + 1.75X$$

باستخدام الانحرافات

$$\hat{\beta} = \frac{\sum xy}{\sum x^2} = \frac{70}{40} = 1.75$$

	X	Y	x	y	xy	x^2	XY	X^2

1	**100**	**55**	30	-45	1650	900	5500	10000
2	**90**	**70**	20	-30	1400	400	6300	8100
3	**80**	**90**	10	-10	900	100	7200	6400
4	**70**	**100**	0	0	0	0	7000	4900
5	**70**	**90**	0	-10	0	0	6300	4900
6	**70**	**105**	0	5	0	0	7350	4900
7	**70**	**80**	0	-20	0	0	5600	4900
8	**65**	**110**	-5	10	-550	25	7150	4225
9	**60**	**125**	-10	25	-1250	100	7500	3600
10	**60**	**115**	-10	15	-1150	100	6900	3600
11	**55**	**130**	-15	30	-1950	225	7150	3025
12	**50**	**130**	-20	30	-2600	400	6500	2500
المجموع	**840**	**1200**			**-3550**	**2250**	**80450**	**61050**
	X=70	**Y=100**						
					β=-3550/2250=-1.6			

4 خصائص مقدرات المربعات الصغرى العادية (م ص ع)

الخصائص الإحصائي التي تتميز فيها مقدرات المربعات الصغرى العادية.

تتميز المقدرات β α بثلاث خواص أساسيه:

1) الخطية 2)عدم التحيز 3) الكفاءة

1) الخطية: $\hat{\alpha}$تعتبر داله خطية للعنصر العشوائي التابع Y . أهمية هذه الخاصة أنها تعطينا درجه من البساطة في أجراء الحسابات حيث انه لحساب α β نستعمل المتغير التابع في صوره خطيه فقط هذه لتبسيط الحسابات.

2) عدم التحيز: مقدرات (م ص ع) $\hat{\alpha}$ مقدرة غير متحيزة للمعلمة α . عدم التحيز يتطلب بأن القيمه المتوقعة لـ $\hat{\alpha}$ و التي هي قيمة المعلومة الحقيقية بمعنى آخر متوسط $\hat{\alpha}$

= α . إذا جمعت عينات كثيرة وفي كل عينه نحسب $\hat{\alpha}$ يتم أخذ المتوسط. ذلك المتوسط نظريا يجب أن يتساوى مع المعلمة الحقيقية. $E(\hat{\alpha}) = \alpha$ مقدرات (م ص ع) $\hat{\beta}$ مقدرة غير متحيزة للمعلمة β حيث أن $E(\hat{\beta}) = \beta$. أي أن توقع $\hat{\beta}$ يجب أن يساوي المعلم هـ الحقيقية بمعنى آخر متوسط قيم $\hat{\beta}$ أو في المتوسط $\hat{\beta}$ تساوي القيمة الحقيقية للمعلمه β.

هذه الأوضاع كلها نظريه بحتة في الواقع لا يكون عندنا عدد من العينات، بكون في الواقع عينه واحدة فقط وتعطينا قيمه واحدة $\hat{\alpha}$، قيمه واحدة $\hat{\beta}$ يعتمد عليها في التحليل، من الناحية النظرية نقول أن هذه المقدرات يتوقع أنها تساوي القيمة الحقيقية من الناحية الأخرى القيمة الحقيقة لا نعرفها وبالتالي هذه الخصائص خصائص نظريه بحتة.

تباين المقدرات: تباين اي قيمة تتوزع حول وسط معين هو معدل تشتت هذه القيم عن الوسط

ويكون القانون الخاص بتباين مقدرة القاطع:

$$V(\hat{\alpha}) = E\{\hat{\alpha} - E(\hat{\alpha})\}^2$$

بإجراء بعض الخطوات يمكن إن نبرهن إن تباين $\hat{\alpha}$ يساوي

$$V(\hat{\alpha}) = \frac{\sigma^2 \sum X^2}{n \sum x^2}$$

من المعادلة نلاحظ إن تباين $\hat{\alpha}$ تعتمد على تباين u فإذا زاد تباين u توقع زيادة تباين $\hat{\alpha}$ لان هناك علاقة طرديه بين تباين $\hat{\alpha}$ وتباين u . وتوجد صيغه أخرى لتباين $\hat{\alpha}$ على انه يساوي :

$$V(\hat{\alpha}) = \sigma^2 \{\frac{1}{n} + \frac{\overline{X}^2}{\sum x^2}\}$$

اما القانون الخاص بتباين $\hat{\beta}$:

$$V(\hat{\beta}) = \{\hat{\beta} - E(\hat{\beta})\}^2$$

يمكن إن نثبت إن التباين الخاص بـ$\hat{\beta}$ يساوي $V(\hat{\beta}) = \frac{\sigma^2}{\sum x^2}$

ومن المعادلة نلاحظ إن تباين $\hat{\beta}$ يعتمد طرديا على تباين u وعكسيا على مجموع مربعات انحرافات المتغير المستقل، فكلما ازدادت درجة انتشار المتغير المستقل (آي بيانات X مختلفة كثيرا عن بعضها) نتوقع إن يزيد المكون الموجود في المقام وبالتالي ينخفض تباين $\hat{\beta}$ مما يشعر إلى دقة التقديرات.

3- **ادنى (أصغر) تباين**:

الخاصية الثالثة لمقدرات م ص ع تمتلك أدنى تباين هذه الخاصية لها أهمية بالغة في الاقتصاد القياسي لان أدنى تباين يعتبر مؤشر إلى دقة القياسات، أدنى تباين يعتبر مؤشر إلى دقة القياسات، أدنى تباين يعني أعلى دقة من ناحية القياسات.

هناك علاقة عكسية بين التباين ودقة القياسات كلما زاد التباين كلما انخفضت دقة القياسات وكلما قل ارتفعت دقة القياسات. لأن مقدرات م ص ع $\hat{\alpha}$ $\hat{\beta}$ تلك

المقدرات تمتلك آدني تباين نعني مقارنه بمقدرات أخرى تقاس بطريقه مختلفة عن م ص ع فان مقدرات م ص ع تمتلك آدني تباين إي تتحلى بأعلى دقه. نفترض إن هناك مقدرات لـ α β تحصل عليها بطريقه مختلفة ونفترض إن المقدرات الأخرى α'', β'' ا ذا افترضنا أن تلك المقدرتين خطيه وغير متحيزة سيكون الاختلاف في خاصية أن مقدرات م ص ع $\hat{\alpha}$ $\hat{\beta}$ تمتلك أعلى دقة.

من النتائج التي توصلنا أليها عن مقدرات م ص ع يمكن أن نقول أن شـــكل التوزيع الاحتمالي الخاص بالمقدرات $\hat{\alpha}$ $\hat{\beta}$

$$\hat{\alpha} \sim N\left[\alpha, \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x^2}\right)\right], \quad \hat{\beta} \sim N\left[\beta, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}\right]$$

من المعادلتين يتبين انه :

1- كلما زاد التباين σ^2 كلما زاد تباين المقدرات $\hat{\alpha}$ $\hat{\beta}$.

2-كلما كان انتشار قيم X اكبر كلما قل تباين $\hat{\alpha}$ $\hat{\beta}$

مثال (2):

البيانات التالية عن السعر وكميه البرتقال الذي تم بيعه في أحد أسواق الخضار في مدة 12 يوم إذا رمزنا للسعر بـ X والكميه بـY

باستخدام المعادلة التالية:

$$Y = \alpha + \beta X + u$$

$$\bar{X} = 70, \sum xy = -35550$$

$$\bar{Y} = 100, \quad \sum x^2 = 2250$$

$$\hat{\beta} = \frac{-3550}{2250} = -1.578$$

$$\hat{\alpha} = 100 - (-1.578)70 = 210.460X + u$$

$$\hat{Y} = 210.46 - 1.578X.$$

$$V(\hat{\alpha}) = \sigma^2 (\frac{1}{n} + \frac{\overline{X}^2}{\sum x^2}) = \sigma^2 (\frac{1}{12} + \frac{(70)^2}{2250}) = 2.2611\sigma^2$$

$$V(\hat{\beta}) = \beta, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2} = \frac{\sigma^2}{2250} = 0.00044\sigma^2$$

تقدير التباينσ^2 : حيث أن σ^2 مجهولة والتي نحتاجها لنتمكن من حساب تباين $\hat{\alpha}$ $\hat{\beta}$. نستخدم مقدرة σ^2 = مجموع مربعات البواقي/درجة الحرية

$$\sigma^2 = \frac{\sum u^2}{n-2} \qquad 7.2$$

بحساب مربعات مجموع البواقي من قيمة $\sum u^2 = \sum (Y_i - \hat{Y}_i)^2 = 698$

إذا مقدرة التباين $\hat{\sigma}^2 = \frac{\sum u^2}{n-2} = \frac{698}{12-2} = 69.8$

ومنها يمكن الحصول على مقدرات تباين $\hat{\alpha}$ $\hat{\beta}$.

$$V(\hat{\alpha}) = \sigma^2 (\frac{1}{n} + \frac{\overline{X}^2}{\sum x^2}) = 69.8(\frac{1}{12} + \frac{(70)^2}{2250}) = 157.82$$

$$V(\hat{\beta}) = \beta, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2} = \frac{69.8}{2250} = 0.0310$$

مثال 3

X	Y	X^2	x	y	XY	xy	x^2	Y^	u	u^2
2	4	4	-2	-4	8	8	4	4.5	0.50	0.2500
3	7	9	-1	-1	21	1	1	6.25	0.75	0.5625
1	3	1	-3	-5	3	15	9	2.75	0.25	0.0625
5	9	25	1	1	45	1	1	9.75	0.75	0.5625
9	17	81	5	9	153	45	25	16.75	0.25	0.0625
∑X=20	**∑Y=40**	**∑X²=120**			**230 ∑XY=**	**∑xy=70**	**∑x²=40**			**∑u²=1.5**

$$u = Y - \hat{Y} = Y - (\hat{\alpha} + \hat{\beta}X) = Y - 1 - 1.75(X)$$
$$X = 2,3,1,5,9$$

للحصول على مقدرة التباين نستخدم المعادلة التالية $\hat{\alpha}$ $\hat{\beta}$.

$$\sigma^2 = \frac{\sum u^2}{n-2} = \frac{1.5}{5-2} = 0.5$$

وبعد ذلك نستطيع أن نتحصل على تباين المقدرات

$$V(\hat{\alpha}) = \sigma^2 (\frac{1}{n} + \frac{\overline{X}^2}{\sum x^2}) = 0.5(\frac{1}{5} + \frac{(120)^2}{40}) = 0.3$$

$$V(\hat{\beta}) = \beta, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2} = \frac{0.5}{40} = .0125$$

وللحصول على الانحراف المعياري نتحصل على الجذر التربيعي للتباين:

$$Se(\hat{\beta}) = \sqrt{V(\hat{\beta}) =} \sqrt{0.0125} = 0.112$$

$$Se(\hat{\alpha}) = \sqrt{V(\hat{\alpha}} = \sqrt{0.3} = 0.548$$

5 فترات الثقة Confidence Interval

المقدرات مؤشرات مهمه يمكن إن تستخدم لاستخلاص نتائج عن المجتمع التي استخلصت منه هذه المقدرات.لبناء فترات الثقة وأجراء اختبارات الفروض نستخدم التوزيع الطبيعي:

فترة الثقة:

$$\hat{\beta} \sim N\left[\beta, \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}\right], \qquad \hat{\beta} \sim N\left[\beta, \sigma_{\hat{\beta}}^{\ 2}\right]$$

إذا كانت β تتوزع طبيعيا فستكون قيمة Z كما يلي:

$$Z = \frac{\hat{\beta} - \beta}{\sigma_{\hat{\beta}}}$$

إذا أخذت أي عشوائي يتوزع توزيع طبيعي وطرحت منه الوسط الخاص به وقسمته على الانحراف المعياري فان القيمة المتحصل عليها هي قيمة Z التي تتوزع طبيعيا بوسط صفر

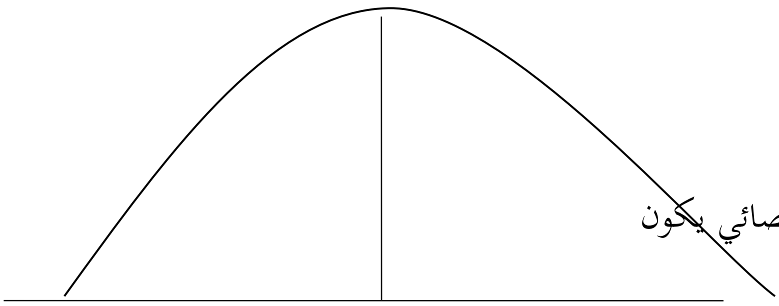
وتباين وانحراف معياري يساوي الواحد الصحيح. توزيع Z كما هو معروف يمكن استخلاص الاحتمالات الخاصة به من جدول التوزيع الطبيعي. وبذلك يمكن تحديد الاحتمال الخاص بحدوث أي قيمه من Z بالنظر إلي الجدول حيث يشير العمود الرأسي إلي اليسار إلي قيم Z والقيم بداخل الجدول تشير إلي الاحتمالات.

لاختبار الفرضية فإننا نختبر هل $\hat{\beta}$ تساوي β أم تختلف عنها وإذا كانت تختلف هل هذا الاختلاف قليل يمكن التعايش معه أي إن الاختلاف راجع إلي العشوائية فقط وليس بالاختلاف الكبير الذي يشير إلي انه لا ينتمي إلي نفس المجتمع. ونرفض الفرضية انهما متساويان.

في القانون أعلاه هناك معلمه غير متوفرة وهي معلمة تباين المجتمع فاستخدمنا مقدرة التباين.

مقدرة التباين لا تمتلك التوزيع الطبيعي ولكن تتبع توزيع t والذي يتحدد تبعا لدرجات الحرية المستعملة أي في هذه الحالة إلي n-2. توزيع t هو توزيع احتمالي مشابه للتوزيع الطبيعي وتوزيع t يتمركز حول الصفر ويأخذ شكل مماثل لتوزيع Z ،ويستخدم توزيعZ فقط عندما تكون حجم العينه كبيره $n>30$

توزيع t



الاختبار الإحصائي يكون

$$t = \frac{\hat{\beta} - \beta^*}{Se_{\hat{\beta}}}$$

يمكن حساب فترات الثقة كما يلي:

$$\hat{\beta} \pm t_{n-2.\frac{\lambda}{2}} Se(\hat{\beta})$$
$$\hat{\alpha} \pm t_{n-2.\frac{\lambda}{2}} Se(\hat{\alpha})$$

قيمة t تمثل القيمة اختبار t عند درجة حرية n-2 عند مساحة $\lambda/2$ من توزيع t من

المثال (3)

$$\hat{\beta} \pm t_{n-2.\frac{\lambda}{2}} Se(\hat{\beta})$$
$$1.7 \pm 2.228 \times 0.112$$
$$\pm 0.2495$$
$$1.45 \_\_\_\_ 1.94$$

$$\hat{\alpha} \pm t_{n-2.\frac{\lambda}{2}} Se(\hat{\alpha})$$
$$1 \pm 2.228 \times 0.548$$
$$1 \pm 1.2209$$
$$-0.22 \_\_\_\_ 2.22$$

إن شرح فترة الثقة يعني إن إن الاحتمال أن فترة الثقة المحددة تعطي المعلمة الحقيقية يساوي

$(1-\lambda)$. ويستخدم عادة مستوى الثقة 95% أو 99%.

6 اختبار الفرضيات

:يتعلق اختبار الفرضيات بإيجاد ألا جابه على هذا السؤال ما اذا كانت القيمة المحسوبة من العينة متوافقة مع الفرضية أم لا؟ الكلمة متوافقة هنا تعني أن القيمة المحسوبة قريبه من

القيمة المفترضة بحيث أننا لا نستطيع إن نرفض القيمة المفترضة. إي إذا كان هناك نظريه سابقه أو اعتقاد إن الميل الحقيقي لدالة الاستهلاك والدخل يساوي على سبيل المثال 1 هل القيمة المحسوبة أو المشاهدة والتي تساوي β=0.509 و تحصل عليها من العينة متفقه مع القيمة التي افترضناها سابقا؟ إذا كان الجواب بنعم فاننا لا نرفض الفرضية.في القياسي نسمي القيمة المفترضة بفرضية العدم لفرضية البديلة

فرضية العدم $H_0:\beta=\beta*$ الفرضية البديلة $H_A:\beta\neq\beta*$

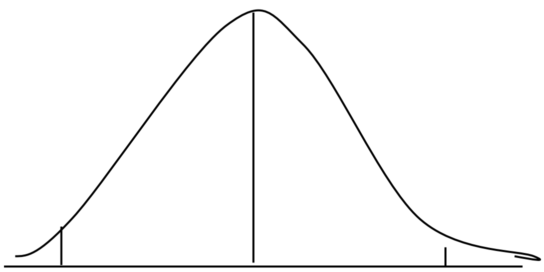
من المثال (1) نفترض إننا سوف نقوم باختبار الفرضية انه ليس هناك علاقة بين , Y, X

فرضية العدم $H_0:\beta=0$ الفرضية البديلة $H_A:\beta\neq 0$

$$t=\frac{\hat{\beta}-\beta_0}{Se_\beta}\quad t=\frac{1.75-0}{0.112}=15.65,\qquad t_{(n-k),(1-\alpha/2)}=t_{3,,(0.975)}=3.182$$

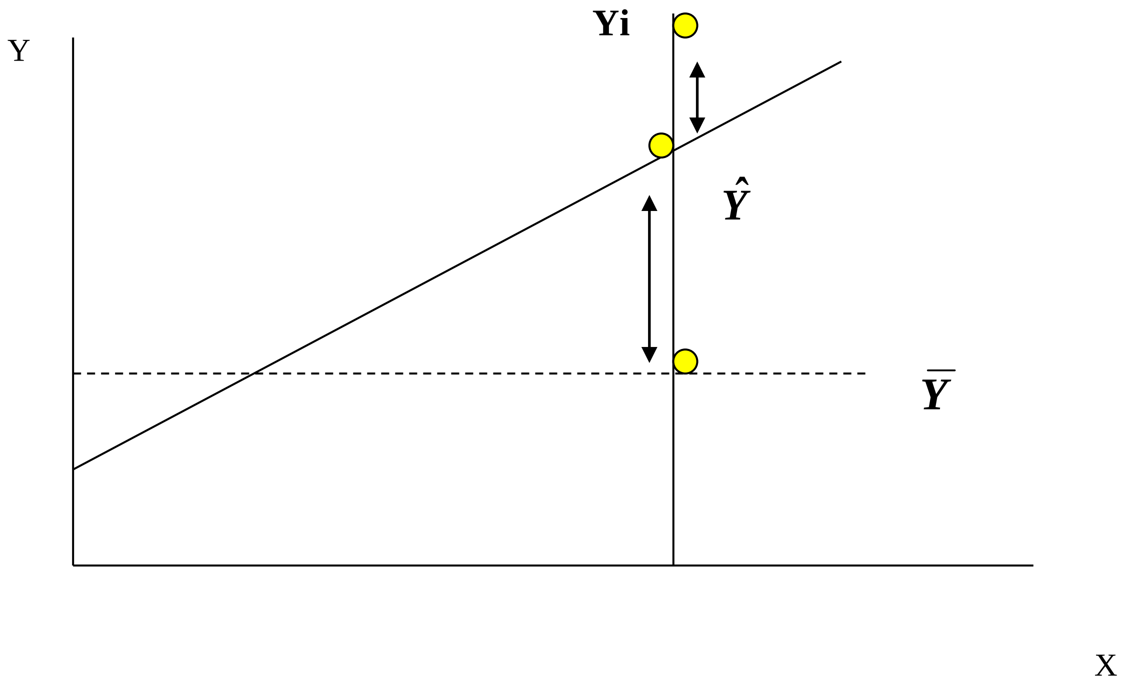
الاختبار يقارن بين ما تقوله الفرضية وما تقوله العينة إذا كان الفرق كبير إي اكثر من القيمة الجد وليه التي نحصلنا عليها من جدول t فأننا نرفض الفرض. إذا كان الفرق قليل فان هذا يعني إن العينة تؤيد ما يقوله الفرض وبالتالي نقبل الفرض.

توزيع t



t الجدوليه =3.182

يعني إننا يجب إن نوجد القيمة الفرضية من اجل اتخاذ القرار آما بقبول أو برفض . إن قرار القبول أو الرفض يتعلق بفرضية العدم وليس بالفرضيه البديلة.



$$\hat{Y} = \hat{\alpha} + \hat{\beta}X, \qquad Y_i = \hat{\alpha} + \hat{\beta}X + u \quad Y_i - \hat{Y} = u$$

7 اختبار جودة النموذج وتحليل التباين.

$$SST = \sum y_i = \sum (Y - \bar{Y})^2$$
$$SSR = \sum \hat{y}^2 = \beta \sum x_i^2 = \sum (\hat{Y} - \bar{Y})^2$$
$$SSE = \sum u_i^2 = \sum (\hat{Y} - Y)^2$$
$$SST = SSR + SSE$$

SST Total Sum of Squares	مجموع المربعات الإجمالي للتغيرات التي تحدث في المتغير التابعY
SSR Regression Sum of Squares	يسمى بمجموع مربعات الانحدار يعني جزء من تباين Y الذي تم تفسيره بواسطة الانحدار. إي الجزء من المتغيرات التي تحدث في المتغير التابع والذي تم تفسيرها بواسطة النموذج المقدر
SSE Error	مجموع مربعات البواقي، $\sum u^2$, وهذا مؤشر للجزء الذي لم يفسر بواسطة نموذج الانحدار، إي الجزء الذي فشل النموذج في تفسيره

ويمثل نسبة مجموع مربعات الانحدار إلي مجموع المربعات الإجمالي ما يسمى بـ معامل التحديد

$$\boxed{R^2_{xy} = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = \frac{\hat{\beta}\sum xy}{\sum y_i^2}}$$

قيمة R^2 تتراوح بين صفر وواحد. إذا كانت مرتفعه أي قريبه من الواحد تعتبر X جيده في تفسير التغيرات في Y. إذا كانت قريبه من الصفر فان المتغير لا يشرح إلا القليل من التغير في Y.

$$R^2 = \frac{SSR}{SST} = \frac{\hat{\beta}\sum xy}{\sum y^2} = \frac{1.75(70)}{124} = \frac{122.5}{124} = 0.9879$$

جدول تحليل التباين لمعادلة الانحدار **ANOVA** **Analysis Of Variance)**:

وهو إن تحليل مجموع المربعات الصغرى إلي مجموع مربعات البواقي ومجموع مربعات الانحدار. الغرض من هذا التحليل لاختبار معنوية مجموع مربعات الانحدار وهذا أيضا يدخل في اختبار معنوية المعامل β. ونمثل هذا التحليل في جدول تحليل التباين:

جدول تحليل التباين ANOVA

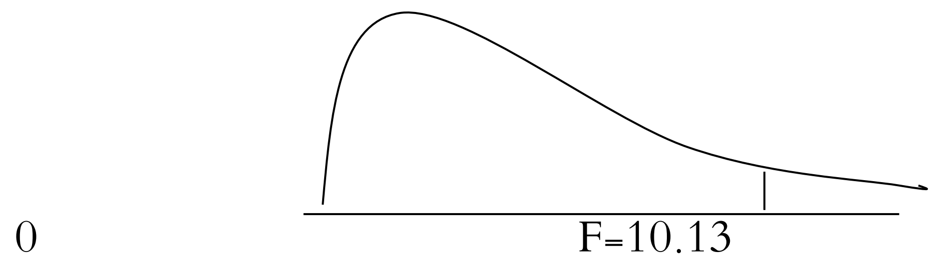
التباين	مجموع المربعات	درجة الحرية	متوسط المربعات
مجموع مربعات الانحدار	$SSR = \sum \hat{y}^2 = \hat{\beta}\sum xy = \hat{\beta}^2\sum x^2 =$	k-1=2-1	SSR/1
مجموع مربعات البواقي	$SSE = \sum u^2 = \sum y^2 - \sum \hat{y}^2$	n-k=n-2	SSE/(n-2)
مجموع مربعات الإجمالي	$SST = \sum y^2 = \sum \hat{y}^2 + \sum u^2$	n-2=3	$F = \frac{SSR}{SSE/n-k}$

جدول تحليل التباين للمثال (1)

التباين	مجموع المربعات	درجة الحرية	متوسط المربعات
مجموع مربعات الانحدار	SSR=122.5	K-1=1	122.5/1
مجموع مربعات البواقي	SSE=1.500	n-k=3	1.5/3
مجموع مربعات الإجمالي	SST=124	n-1=4	$F=\frac{122.5/1}{1.5/3}=245.00$

اختبار F هو اختبار لجودة النموذج. يحاول أن يجيب على السؤال هل افلح النموذج في تفسير التغيرات التي تحدث في المتغير التابع. ويختبر الفرضية إن معاملات المتغيرات المفسرة تساوي الصفر. أي أن فرضية العدم تقول انه لا يوجد علاقة بين المتغيرات المفسرة والمتغير التابع. وتقارن قيمة المحسوبة من الجدول مع الجد وليه بدرجة حرية للبسط تساوي k-1 ودرجة حرية المقام n-k. قيمة الجد وليه عند مستوى معنوية 5%تساوي 10.13.

توزيع F



8 التنبؤ باستخدام معادلة الانحدار:

معادلة الانحدار المقدرة $\hat{Y} = \hat{\alpha} + \hat{\beta}X + u_t$ تستخدم في عملية التنبؤ لقيم Y لقيم محدده من X . إذا كانت X_0 تمثل القيمة المحددة من X تستخدم في التنبؤ بقيمة Y_0 من قيم Y.

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta}X_0 + u_0$$

حيث u تمثل حد الخطأ.

حيث يمثل خطأ التنبؤ $\hat{Y}_0 - Y_0 = (\hat{\alpha} - \alpha) + (\hat{\beta} - \beta)X_0 + u_0$

حيث إن $E(\hat{\alpha} - \alpha) = 0, E(\hat{\beta} - \beta) = 0, E(u_0) = 0$

إذا تكون $E(\hat{y}_0 - y_0) = 0$

هذه تعني إن قيمة Y هي قيمه غير متحيزة ويكون تباين يساوي:

$$V(\hat{y}_0 - y_0) = V(\hat{\alpha} - \alpha) + X_0^2 V(\beta - \beta) + 2X_0 COV(\hat{\alpha} - \alpha, \hat{\beta} - \beta) + V(u)$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum x_i^2} \right] + \sigma^2 \frac{X_0^2}{\sum x_i^2} - 2X_0 \sigma^2 \frac{\overline{X}}{\sum x_i^2} + \sigma^2$$

$$\boxed{= \sigma^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum x^2} \right]}$$

أي إن التباين يرتفع بارتفاع تباين X أي باختلاف قيم X عن قيم متوسطها.

باستخدام البيانات أعلاه نحصل على $Y=10.0+0.90\ X$

$\hat{\sigma}^2 = 0.01, \overline{X} = 200, \sum x_i^2\ 4000,$ إذا استخدمنا $X_0 = 250$ ستكون قيمة Y_0 المتنبأ بها يساوي:

$$Y_0 = 10.0+0.9(250)=235$$

$$SE(\hat{Y}_0) = \sqrt{0.01\left(1+\frac{1}{12}+\frac{2500}{400}\right)} = 0.131$$

حيث أن t=2.228 من جدول مع 10 درجات حريه، و فترة الثقة 95% تكون

$$235 \pm 2.228\ (0.131) = 235 \pm\ 0.29$$

آي أن فترة الثقة تساوي $(234.71 \quad - \quad 235.29)$

التنبؤ للقيمة المتوقعة:-

أحيانا يرغب الباحث في التنبؤ بالقيمة المتوقعة لـY بدلا من Y_0 آي قيمة $E(Y_0)$ آي القيمة المتوسطة لـ $E(Y_0$ وليس Y_0 . عند التنبؤ بالقيمة المتوقعة فان $E(Y_0) = Y_0$ حيث أن $E(Y_0) = \alpha + \beta X_0 + u_0$ فان القيمة المتنبأ بها تساوي $\hat{E}(Y_0) = \hat{\alpha} + \hat{\beta} X_0 + u_0$ آي يساوي Y_0 ولكن الخطأ المعياري والتباين سيكون مختلفا. سيكون اصغر قيمه

$$\hat{E}(y_0) - E(y_0) = (\hat{\alpha} - \alpha) + (\hat{\beta} - \beta) X_0$$

التباين يساوي:

$$\boxed{Var[E(y_0) - E(y_0)] = \sigma^2 \left[\frac{1}{n} + \frac{(X_0 - \overline{X})^2}{\sum x^2}\right]}$$

والخطأ المعياري يساوي الجذر التربيعي للتباين وفترة الثقة تساوي

$$\boxed{\hat{E}(y_0) \pm t\ SE}$$

مثال:

	المبيعات Y	النفقات الأعلانيه X
1	3	1
2	4	2
3	2	3
4	6	4
5	8	5

للحصول على مقدرات النموذج:

مشاهده	المبيعات Y	النفقات الأعلانيه X	X^2	XY	u
1	3	1	1	3	0.80
2	4	2	4	8	0.60
3	2	3	9	6	2.60
4	6	4	16	24	0.20
5	8	5	25	40	1.00
	15	23	55		

$Y = 1.0 + 1.2\,X$ $\quad$ $X=3.0$ $\quad$ $SSE=8.8$

نفترض إن مدير المبيعات يرغب في التنبؤ بدخل المبيعات عندما تكون النفقات الاعلانيه تساوي 600 ريال. ويريد أيضا بناء 95% فترة ثقة لتنبؤه. $X_0 = 6$ إذا

$y=1.0 + 1.2(6)=8.2$

والتباين يساوي $= \sigma^2\left[1+\frac{1}{5}+\frac{(6-3)^2}{10}\right] = 2.1\sigma^2$

$$\hat{\sigma}^2 = \frac{SSE}{d.f} = \frac{8.8}{3} = 3.93$$

الخطأ المعياري $\sqrt{2.1(2.93)} = \sqrt{6.153} = 2.48$

عند مستوى المعنوية 5% ، ودرجة حرية df=2.353. و 90% فتره

تطبيق

دالة الاستهلاك الخاصة باقتصاد دولة ما:

$$C = \hat{\alpha} + \hat{\beta}Y$$

حيث ترمز X الى الاستهلاك الخاص و Y الى الدخل.

البيانات:

GDP الناتج الداخلي الخام	X الاستهلاك الخاص	Year
164.53	23.90	1975
205.06	34.75	1976
225.40	54.61	1977
249.54	68.61	1978
385.81	102.39	1979
520.59	114.91	1980
524.72	126..39	1981
415.23	151.29	1982
372.07	157.37	1983
351.40	159.35	1984
313.94	158.59	1985
271.09	140.15	1986
275.45	135.54	1987

1988	139.40	285.15
1989	145.03	310.82
1990	155.87	391.99
1991	168.75	442.04
1992	183.92	461.40
1993	193.91	443.84
1994	185.83	450.03
1955	191.10	470.70
1996	206.21	511.33
1997	207.35	547.41

$$\ln C = -3.33 + 1.38 \ln Y$$

$$n = 23, \quad K = 2$$

$$SSE = 0.341$$

$$SSR = 2.44$$

$$\overline{Y} = 4.82$$

$$SE(\hat{\alpha}) = 1.29, \quad SE(\hat{\beta}) = 0.219$$

$$t_{\alpha} = \quad = \frac{-3.33 - 0}{1.29} = -2.57$$

$$t_{\beta} = \quad = \frac{1.38 - 0}{0.219} = 6.31$$

$$t_{(n-k)(1-0.025)} = t_{(21, 0.975)} = 2.080$$

$$F = 39.88 \quad F_{(k-1)(n-k)} = F_{1,21} = 4.32$$

$$R^2 = 0.65$$

تمارين

1-عرف المصطلحات التالية:

1-الخطأ المعياري.

2-التباين.

2-عدد الافتراضات اللازمة لتطبيق طريقة المربعات الصغرى العادية.

3- للمشاهدات التالية

$X=\{3,2,1,,-1,0\}$ و $Y=\{5,2,3,2,-2\}$

أ- أوجدي القيم التاليه

$\sum_{i=1}^{5} x_i^2, \sum_{i=1}^{5} x_i y_i, \sum_{i=1}^{5} x_i, \quad \sum_{i=1}^{5} y_i, \quad \bar{X}, \bar{Y},$

ب-أوجد الخط الذي يمثل العلاقة $Y=a+bX$

ت-مالمقصود بـ a,b

4- حدد دالة الانتاج التي تعبر عن العلاقة بين كمية الانتاج وعنصر الانتاج العمل حسب البيانات المعطاة في الجدول التالي:

المشاهدات	Q	L
1	0.58	1
2	1.10	2
3	1.20	3
4	1.30	4
5	1.95	5
6	2.55	6
7	2.60	7
8	2.90	8

9	3.45	9
10	3.50	10
11	3.60	11
12	4.10	12
13	4.35	13
14	4.40	14
15	4.50	15

اذا افترضنا ان البيانات يمكن وصفها بعلاقة خطية تحت الفروض الخمسة. حدد تلك العلاقة . واشرح العلاقة الاقتصادية التي تربط المتغيرات حسب المعادلة.

تمرين :

حدد نتائج الانحدار للعلاقة بين الكمية المطلوبة لسلعه والسعر كما يلي

$$Y = \alpha + \beta X + u$$

للبيانات التالية

السعر X	الكميه Y						
100	55						
90	70						
80	90						
70	100						
70	90						
70	105						
70	80						
65	110						

60	125						
60	115						
55	130						
50	130						

1- حددي معاملات النموذج.

2- بناء مجال الثقة لمعامل الميل وحدد فرضية العدم و نتائج اختبار t,F. عند مستوى المعنوية 5%

3- حدد معامل التحديد R^2