الفصل العاشـر

الانحدار والارتباط

Regression and Correlation

(١٠-١) الانحدار Regression

يستخدم الانحدار لدراسة العلاقات بين متغيرات قابلة للقياس. تحليل الانحدار له تطبيقات كثيرة في معظم مجالات الدراسة التي تشتمل على العلوم الحيوية ، الفيزيائية ، العلوم الاجتماعية ، الاقتصادية ، الصناعية ... الخ. تختلف الطرق المستخدمة في تحليل الانحدار باختلاف التطبيقات ولكنها تشترك جميعها في الوصف والتنبأ بالقيم المستقبلية . عموما يتكون الانحدار من عدة خطوات. فلدراسة العلاقة بين عدد من المتغيرات ، تجمع البيانات على عينة من المفردات المعرضة لتلك المتغيرات. في نموذج الانحدار يكون هناك متغير واحد يسمي المتغير التابع depend variable أو متغير الاستجابة response variable ، بينما المتغيرات الأخرى تسمي متغيرات مستقلة independent variables . تستخدم البيانات للحصول على تقديرات لمعالم النموذج . في الحقيقة طرق تقدير معالم النموذج المجهولة كثيرة . سوف تقتصر دراستنا على طريقة المربعات الصغرى least squares. يشتمل البند التالي على الانحدار الخطى البسيط ، أي وجود متغير تابع ومتغير مستقل واحد والعلاقة بينهما خطية ، حيث ينحصر اهتمامنا على تعريف النموذج المناسب ومناقشة الفروض وإيجاد تقديرات المربعات الصغرى وعرض بعض طرق التقدير بفترة واختبارات الفروض .

(١٠-٢) الانحدار الخطى البسيط Simple Linear Regression

بفرض أننا اخترنا عينة عشوائية من الحجم n من المجتمع موضع الدراسة ممثلة بالفئة $\{(x_i, y_i)\ ; i = 1,2,...,n\}$. بتكرار المعاينة مع استخدام نفس قيم x ، فإننا نتوقع أن قيم y تختلف من عينه إلى أخرى. وعلى ذلك فإن قيمة y_i في الزوج المرتبة (x_i, y_i) هي قيمة لمتغير عشوائي Y_i. سوف نرمز للتوزيع الاحتمالي للمتغير $Y\,|\,x$ المقابل لقيمة ثابتة x بالرمز $f(y\,|\,x)$. من الواضح ، أنه إذا كانت $x = x_i$ ، فإن الصيغة $Y\,|\,x_i$ تمثل متغير عشوائي Y_i . اهتمامنا سوف يكون في توزيعات فئة المتغيرات العشوائية $\{Y_i, i = 1,2,...,n\}$. والتي يفترض أنها مستقلة. أيضا للحصول على فترات ثقة واختبارات فروض ، لابد أن تكون $Y_1, Y_2, ..., Y_n$ تتبع توزيعا طبيعيا .

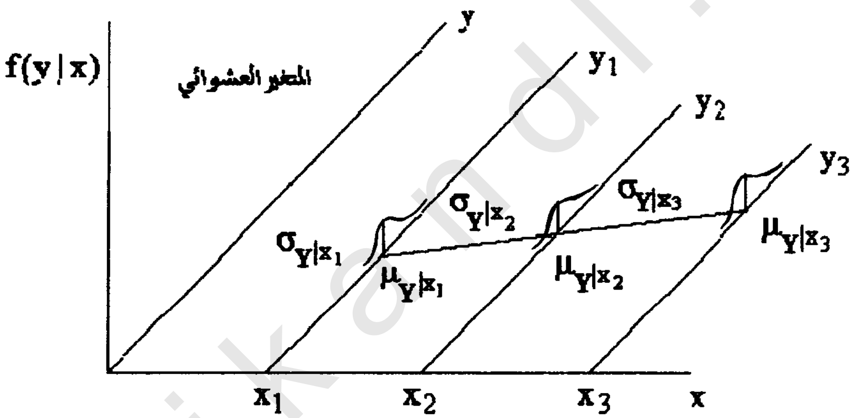
في مشكلة الانحدار سوف نعرف متوسط توزيع Y عند قيمة معطاة x بالرمز $\mu_{Y|x} = E(Y\,|\,x)$ وتباين توزيع Y بالرمز $\sigma^2_{Y|x}$ وذلك تحت فرض أن تباينات المتغيرات $Y_1, Y_2, ..., Y_n$ متساوية ، أي أن $\sigma^2_{Y|x} = \sigma^2$، لجميع قيم x و المعلمة $\mu_{Y|x}$ ثابتة لأي x ولكن قد تختلف باختلاف قيم x .

المنحنى الذي يربط متوسط التوزيعات للمتغيرات $Y_1, Y_2, ..., Y_n$ يسمى خط الانحـــدار **regression curve** . إذا وقعت جميع المتوسطات $\mu_{Y|x}$ على خط مستقيم كما هو موضح في شكل (١٠-١) فإن الانحدار يكون خطى ويمكن تمثيله بالمعادلة :

$$\mu_{Y|x} = \beta_0 + \beta_1\ x,$$

حيث β_1 , β_0 معالم النموذج المجهولة والتي يراد تقديرها من بيانات العينـــة. ســـوف نرمـــز لتقديرات هذه المعالم بالرموز b_1 , b_0 على التوالي . وعلى ذلك يمكن تقدير $\mu_{Y|x}$ بواســـطة $\hat{y}_x$ من خط الانحدار المقدر التالي :

$$\hat{y}_x = b_0 + b_1 x$$



شكل (١٠-١)

(١٠-٢-١) شكـــل الانتشار **Scatter Plot**

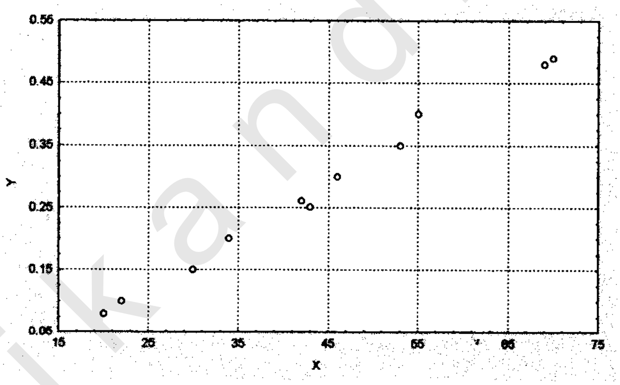
الأسلوب المفيد لبدء تحليل الانحدار البسيط هو تمثيل البيانات بيانياً وهو ما يعرف بشـــكل الانتشار **scatter plot** وذلك لمحاولة اكتشاف الصورة التقريبية للعلاقة. للحصول على شكل الانتشار يخصص محور **x** (المحور الأفقي) للمتغير المستقل بينما يخصص محور **y** (المحور الرأســـي) للمتغير التابع . لكل زوج **(x , y)** من أزواج المشاهدات التي عددها **n** نقوم بتوقيع نقطـــة على الرسم. كثير من البرامج الإحصائية مثل برنامج **SPSS** و**Statistica** يمكن اســـتخدامهم للحصول على أشكال الانتشار .

مثال (١٠-١) في إحدى التجارب وزن قرون عدد من الغزلان المختلفة الأعمار و كانت النتائج كما هي مدونة في الجدول (١٠-١) . المطلوب رسم شكل الانتشار وتحديد شكل العلاقة بين المتغيرين .

جدول (١٠-١)

العمر x	20	22	30	34	42	43	46	53	55	69	70
الوزن y	0.08	0.10	0.15	0.20	0.26	0.25	0.30	0.35	0.40	0.48	049

الحل . يتضح من شكل (١٠-٢) أن النقط عموماً ، ليس بالضبط ، تقع على خط مستقيم. هذا يجعلنا نقترح أن العلاقة بين المتغيرين يمكن وصفها (كتقريب أولي) بمعادلة خط مستقيم .



شكل (١٠-٢)

(١٠-٢-٢) بناء نموذج الانحدار البسيط Building a Simple Regression Line

بفرض أن كل المتوسطات ، $\mu_{Y|x}$ تقع على خط مستقيم وعلى ذلك يمكن كتابة خط انحدار المجتمع على الشكل :

$$\mu_{Y|x} = \beta_0 + \beta_1 x$$

كما يمكن كتابة المتغير العشوائي $Y_i = Y \mid x_i$ على الشكل :

$$Y_i = \beta_0 + \beta_1 x_i + E_i,$$

حيث E_i متغير عشوائي ، بناء على الفروض السابقة على Y_i ، من الضروري أن يكون له متوسط صفر وتباين $\sigma^2_{Y|x} = \sigma^2$. كل مشاهدة $(x_i \ , \ y_i)$ في العينة لا بد أن تحقق العلاقة :

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i .$$

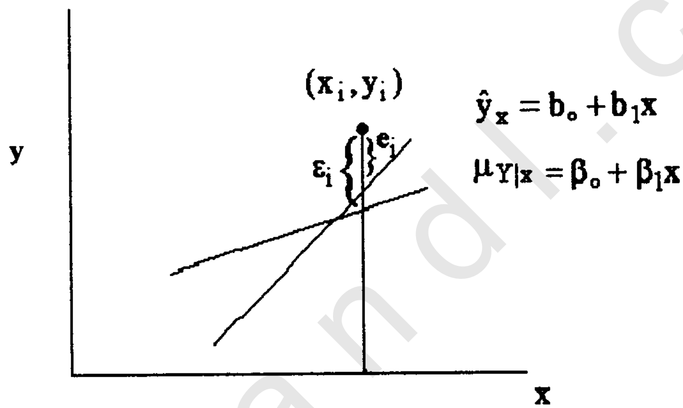
لعينة عشوائية من الحجم n من المشاهدات $\{(x_i, y_i), i = 1,2,...,n\}$ فإن خط الانحـــدار ، الذي يعتبر تقدير لــ $\mu_{Y|x}$، يمكن إيجاده بالمعادلة :

$$\hat{y}_x = b_0 + b_1 x,$$

وكل زوج من المشاهدات لا بد أن يحقق العلاقة :

$$y_i = b_0 + b_1 x_i + e_i$$

حيث e_i يسمى الباقي **residual**. الفرق بين ϵ_i و e_i موضح في شكل (١٠-٣).



شكل (١٠-٣)

يعرف مجموع مربعات البواقي كالآتي :

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left[y_i - b_0 - b_1 x_i\right]^2$$

ويرمز له بالرمز S S E والذي غالباً ما يسمى مجموع مربعـــات الأخطـــاء **sum of squares of the errors** حول خط الانحدار .

(١٠-٢-٣) طريقة المربعات الصغرى Method of Least Squares

سوف نوجد التقديرين b_0 , b_1 للمعالم β_1 , β_0 على التـــوالي بحيـــث أن مجمـــوع مربعات الأخطاء يكون أقل ما يمكن. هذه الطريقة لتقدير المعالم تسمى طريقة المربعات الصغرى.

باستخدام حساب التفاضل نجد أن قيمتي β_1 , β_0 اللتين تحققان النهاية الصغـــرى لمجمـــوع مربعات الأخطاء يمكن تقديرهما بقيمتي b_1 , b_0 اللتين تحققان المعادلتين :

$$b_0\ n + b_1 \Sigma\ x_i = \Sigma\ y_i ,$$

$$b_0\ \Sigma\ x_i + b_1 \Sigma\ x_i^2 = \Sigma\ x_i\ y_i$$

وتسميان بالمعادلتين الطبيعيتين وبحل هاتين المعادلتين آنيا نحصل على :

$$b_1 = SXY / SXX ,$$

$$SXX = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} ,$$

$$SXY = \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} ,$$

$$b_0 = \bar{y} - b_1\ \bar{x}.$$

مثال (١٠-٢) أجريت تجربة لدراسة العلاقة بين التسميد ومحصول الذرة . البيانات الـــتي تم الحصول عليها معطاة في جدول (١٠-٢)

جدول (١٠-٢)

السماد x	0.3	0.6	0.9	1.2	1.5	1.8	2.1	2.4
المحصول y	10	15	30	35	25	30	50	45

(أ) أرســـم شكل الانتشار

(ب) أوجد معادلة خط الانحدار المقدرة

الحل . (أ) يتضح من شكل الانتشار (١٠-٤) أن الخط المستقيم هو أفضل طريقة لتمثيل هذه البيانات :

أي أننا نفترض النموذج الخطى البسيط :

(ب) بما أن β_1 , β_0 مجهولتان فإننا نقدرهما من مشاهدات العينة حيث :

$$n = 8 \quad \Sigma\ x_i = 10.8 \quad \Sigma\ x_i^2 = 18.36$$

$$\Sigma\ x_i y_i = 385.5 \quad , \quad \bar{x} = 1.35 \quad , \quad \bar{y} = 30, \quad \Sigma\ y_i = 240.$$

$$b_1 = \frac{SXY}{SXX} = \frac{\Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}}{\Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}}$$

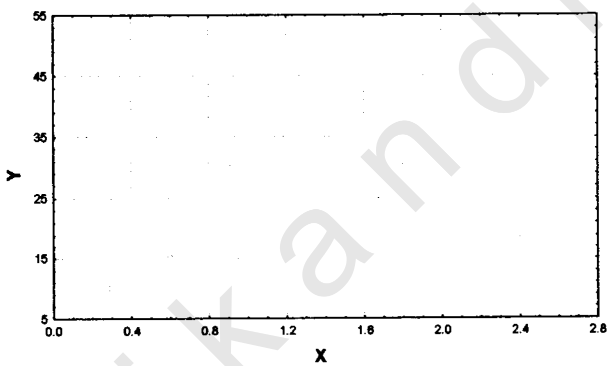
$$= \frac{385.5 - \frac{(10.8)(240)}{8}}{18.36 - \frac{(10.8)^2}{8}}$$

$$= \frac{61.5}{3.78} = 16.27,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 30 - (16.27)(1.35) = 8.036.$$

معادلة الانحدار المقدرة سوف تكون على الشكل :

$$\hat{y}_x = 8.036 + 16.27 \ x.$$



شكل (١٠-٤)

(١٠-٢-٤) تحليل الانحدار Analysis of Variance

لاختبار معنوية معامل الانحدار β_1 أي اختبار فرض العدم :

$$H_0 : \beta_1 = 0$$

ضد الفرض البديل :

$$H_1 : \beta_1 \neq 0$$

يجب دراسة مكونات مجموع المربعات وتجزئته إلى مكوناته الأساسية على النحو التالي :

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

وذلك بإضافة وطرح $\hat{y}_i$ ($\hat{y}_i = b_0 + b_1 x_i$) من الطرف الأيمن وتربيع طرفي المعادلة والتجميع مع ملاحظة أنه يمكن إثبات أن المقدار التالي مساوي الصفر :

$$\Sigma(\hat{y}_i - \bar{y})(y_i - \hat{y}_i)^2 = 0 .$$

وعلى ذلك :

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

أي أن مجموع المربعات حول المتوسط يساوى مجموع مربعات الانحدار مضاف إلى مجموع المربعات حول الانحدار ويمكن الرمز إليه بالمتساوية التالية :

$$\mathbf{SSTO = SSR + SSE} .$$

وهناك صيغ مبسطة للقيم **SSTO , SSR , SSE** حيث :

$$SSTO = SYY = \Sigma {y_i}^2 - \frac{\Sigma y_i^2}{n},$$

$$SSR = \frac{(SXY)^2}{SXX},$$

$$\mathbf{SSE = SSTO - SSR} .$$

من الناحية الإحصائية نجد أن لكل مجموع مربعات درجات حرية خاصة به ، فإذا كان لدينا n من المشاهدات فإن توزيع درجات الحرية يكون على الشكل الموضح في جدول (١٠-٣)

جدول (١٠-٣)

مجموع المربعات	درجات الحرية
مجموع مربعات الانحدار	1
مجموع مربعات الخطأ	n-2
مجموع المربعات الكلي	n – 1

بقسمة مجموع المربعات بدرجات الحرية الخاصة به نحصل على ما يسمي متوسط المربعات **mean squares** ويعتبر تباين العينة s^2 مثال لمتوسط المربعات. وعلى ذلك متوسط مجموع مربعات الانحدار نرمز له بالرمز **MSR** ، هو :

$$MSR = \frac{SSR}{1}.$$

ومتوسط مربعات الخطأ ، نرمز له بالرمز **MSE** ، هو :

$$MSE = \frac{SSE}{n-2}.$$

من النتائج السابقة يمكن اشتقاق جدول تحليل التباين ANALYSIS OF VARIANCE ، للاختصار جدول ANOVA ، والموضح في جدول (١٠-٤) . الآن :

لاختبار فرض العدم $H_0 : \beta_1 = 0$ ضد الفرض البديل $H_1 : \beta_1 \neq 0$ وباعتبار أن فرض العدم صحيح فإن :

$$f = \frac{MSR}{MSE},$$

قيمة لمتغير عشوائي F يتبع توزيع F بدرجات حرية $v_1 = 1, \quad v_2 = n - 2$. لمستوى معنوية α منطقة الرفض $F > f_{\alpha}(1, n-2)$ حيث $f_{\alpha}(1, n-2)$ تستخرج من جدول توزيع F في ملحق (٦) أو ملحق (٧) بدرجات حرية $v_1 = 1, \quad v_2 = n - 2$. إذا وقعت f في منطقة الرفض نرفض H_0.

جدول (١٠-٤)

الاختلاف	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات
الانحدار	1	SSR	$MSR = \frac{SSR}{1}$
الخطأ	n-2	SSE	$MSE = \frac{SSE}{n-2}$
الكلي	$n-1$	SSTO	

مثال (١٠-٣) لازواج القياسات المعطاة في جدول (١٠-٥) :

المطلوب : (أ) إيجاد معادلة الانحدار الخطى المقدرة .

(ب) اختبار فرض العدم $H_0 : \beta_1 = 0$ ضد الفرض البديل $H_1 : \beta_1 \neq 0$ عند مستوى معنوية $\alpha = 0.05$.

جدول (١٠-٥)

x	4	6	2	5	7	6	3	8	5	3	1	5
y	197	272	100	228	327	279	148	377	238	142	66	239

الحل .

$$n = 12 \quad \Sigma\, x_i = 55 \quad \Sigma\, x_i y_i = 14060$$

$$\Sigma\, y_i = 2613 \quad , \quad \bar{x} = 4.58333 \quad , \quad \bar{y} = 217.75,$$

$$\Sigma x_i^2 = 299\ , \ \Sigma y_i^2 = 661865,$$

$$SXY = \Sigma x_i y_i - \frac{\Sigma x_i \Sigma y_i}{n}$$

$$= 14060 - \frac{(55)(2613)}{12} = 2083.75,$$

$$SXX = \Sigma {x_i}^2 - \frac{(\Sigma x_i)^2}{n}$$

$$= 299 - \frac{(55)^2}{12} = 46.91667,$$

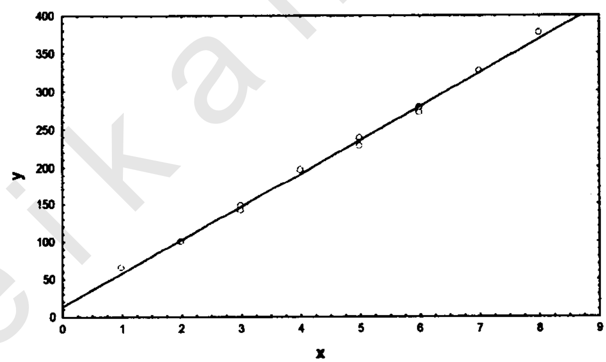
$$b_1 = \frac{SXY}{SXX} = \frac{2083.75}{46.91667} = 44.41385\,,$$

$$b_0 = \bar{y} - b_1\ \bar{x}$$
$$= 217.75 - (44.41385)(4.58333)x$$
$$= 14.187\,.$$

وعلى ذلك فإن معادلة خط الانحدار المقدرة هي :

$$\hat{y}_x = 14.187 + 44.41385x.$$

والموضحة في شكل (١٠-٥) .



شكل (١٠-٥)

الآن نحسب :

$$SSTO = SYY = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n}$$

$$= 661865 - \frac{(2613)^2}{12} = 92884.25,$$

$$SSR = \frac{(SXY)^2}{SXX} = \frac{(2083.75)^2}{46.91667}$$
$$= 92547.362.$$

وبطرح SSR من SSTO نحصل على :

$$SSE = SSTO - SSR$$
$$= 92884.25 - 92547.362$$
$$= 336.888,$$

$$MSR = \frac{SSR}{1} = \frac{92547.362}{1} = 92547.362.$$

$$MSE = \frac{SSE}{n-2} = \frac{336.888}{10} = 33.6888.$$

جدول تحليل التباين معطى في جدول (١٠-٦).

جدول (١٠-٦)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات
الانحدار	1	92547.362	92547.362
الخطأ	10	336.888	33.6888
الكلي	11	92884.25	

$$f = \frac{MSR}{MSE} = \frac{92547.362}{33.6888}$$
$$= 2747.126.$$

$f_{0.05}(1,10) = 4.96$ والمستخرجة من جدول توزيع F في ملحق (٦) بدرجات حرية $\nu_1 = 1, \nu_2 = 10$. منطقة الرفض $F > 4.96$. وبما ن f تقع في منطقة الرفض نرفض H_0 .

(١٠-٢-٥) تقدير σ^2 Estimating σ^2

في الحقيقة ، يفضل الحصول على تقدير للمعلمة σ^2 لا يعتمد على النموذج. عموماً، مثل هذا التقدير يمكن الحصول عليها فقط في فئات البيانات التي تحتوي على قيم متعددة لـ y عند كل قيمة من قيم x أو عند توفر بعض المعلومات القبلية. في حالة عدم توفر هذه الحالات فإن التقدير للمعلمة σ^2 يعتمد على النموذج ، أي يكون دالة في مجموع مربعات الأخطاء SSE ، ويحسب من المعادلة الآتية :

$$s^2 = \frac{SSE}{n-2} = MSE$$

أي أن s^2 يساوى متوسط مجموع مربعات الخطأ وهو تقدير غير متحيز للمعلمة σ^2. التقدير بنقطة للمعلمة σ هو $\sqrt{s^2}$ والذي يسمي الخطأ المعياري للانحدار **standard error of regression**. للمثال (١٠-٣) ومن جدول (١٠-٦) فإن :

$$s = \sqrt{MSE} = \sqrt{33.6888} = 5.804.$$

(١٠-٢-٦) معامل التحديد البسيط <u>Coefficient of Simple Determination</u>

يعرف معامل التحديد البسيط r^2 كالتالي :

$$r^2 = \frac{SSR}{SSTO} = \frac{SSTO - SSE}{SSTO} = 1 - \frac{SSE}{SSTO}$$

للمثال (١٠-٣) ومن جدول (١٠-٦) فإن :

$$r^2 = \frac{92547.362}{92884.25} = 0.996.$$

يأخذ r^2 الواحد الصحيح عندما تقع القيم $y_1, y_2, \ldots, y_n$ على خط الانحدار المقدر وعلى ذلك فإن $SSE = 0$. وذلك لأن :

$$SSR = SSTO - SSE$$

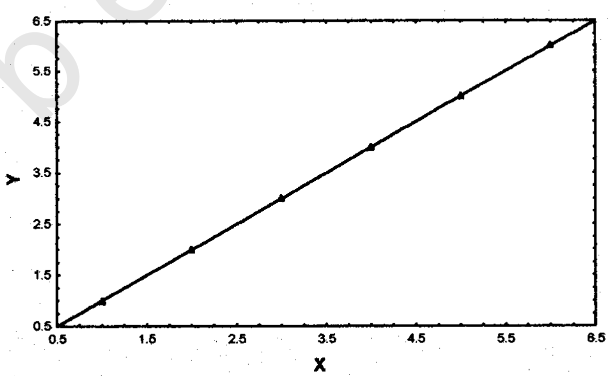
وفي هذه الحالة فإن :

$$SSR = SSTO - 0.0 = SSTO$$

وعلى ذلك فإن :

$$r^2 = \frac{SSR}{SSTO} = \frac{SSTO}{SSTO} = 1$$

هذه الحالة موضحة في شكل (١٠-٦) .

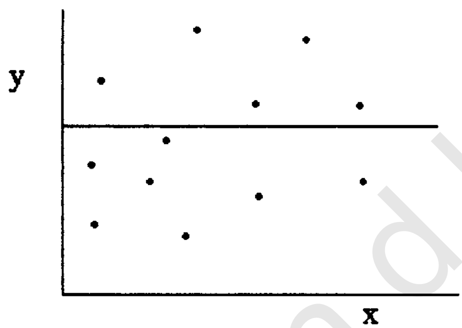


شكل (١٠-٦)

عندما $r^2 = 0$ فهذا يدل على عدم وجود علاقة خطيه بين المتغيرين وبالتالي فـإن $SSR = 0$ ومنها :

$$r^2 = \frac{SSR}{SSTO} = \frac{0.0}{SSTO} = 0.0$$

في هذه الحالة فإن معادلة الانحدار المقدرة سوف تكون موازية للمحور الأفقي ، أي أن $b_1 = 0$ كما هو موضح في شكل (١٠-٧)

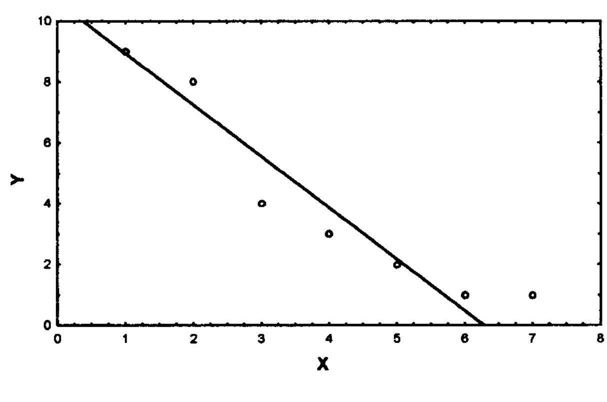


شكل (١٠-٧)

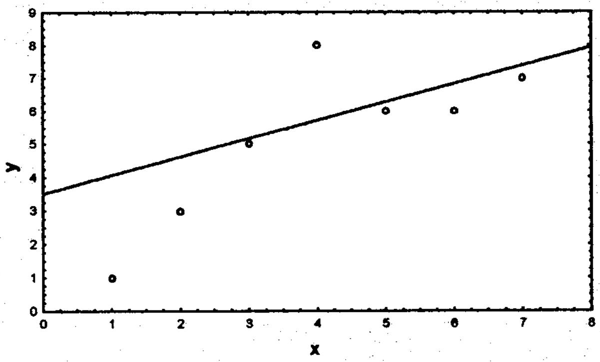
معامل التحديد دائما موجب وتتراوح قيمته بين الصفر والواحد الصحيح أي أن :

$$0 \le r^2 \le 1.$$

يتضح من شكل (١٠-٨) ، حيث $r^2 = 0.9025$ ، أن المشاهدات تقترب بدرجة كبـيرة من خط الانحـدار المقـدر وذلـك بالمقارنـة للمشـاهدات في شـكل (١٠-٩) حيـث $r^2 = 0.3249$.



شكل (١٠-٨)



شكل (١٠-٩)

(١٠-٢-٧) تقدير المعلمتين β_0, β_1 **Estimation of the parameters**

في دراستنا السابقة في البند (١٠-٢) عرفنا أن b_1 , b_0 تقديرين للمعلمتين الحقيقتين β_0, β_1 وذلك بالاعتماد على عينة عشوائية من الحجم n. بتكرار المعاينة من الحجم n وحساب b_1 , b_0 لكل عينة فإن القيمتين b_1 , b_0 سوف يختلفان من عينة إلى أخرى . وعلى ذلك فإن التقديرين b_1 , b_0 قيمتين لمتغيرين عشوائية B_0, B_1 .

وبما أن قيمة x لازالت ثابتة ، فإن قيم β_0, β_1 سوف تعتمد على قيم y أو بدقة أكثر ، على قيم المتغيرات العشوائية $Y_1, Y_2, \ldots, Y_n$ المستقلة والتي تتبع توزيعاً طبيعياً ، وباستخدام نظريات الإحصاء يمكن إثبات أن المتغير العشوائي B_1 أيضاً يتبع توزيعاً طبيعياً بمتوسط :

$$\mu_{B_1} = \beta_1$$

وتباين :

$$\sigma_{B_1}^2 = \frac{\sigma^2}{SXX}.$$

وتبعاً لنظرية (٨-١) فإن :

$$Z = \frac{B_1 - \beta_1}{\sqrt{\dfrac{\sigma^2}{SXX}}}$$

متغير عشوائي يتبع التوزيع الطبيعي القياسى .

عادة الانحراف المعياري ، σ، مجهول ويستبدل بالإحصاء S في صيغة Z لنحصل على

$$T = \frac{B_1 - \beta_1}{\sqrt{\dfrac{S^2}{SXX}}}$$

حيث T متغير عشوائي له توزيع t بدرجات حرية n-2 . سوف نستخدم المتغير T في إيجاد $(1-\alpha)100\%$ فترة ثقة للمعلمة β_1 حيث :

$$P(-t_{\alpha/2} < \frac{B_1 - \beta_1}{\sqrt{S^2/SXX}} < t_{\alpha/2}) = 1-\alpha.$$

وبإتباع الخطوات الجبرية التي استخدمناها في الفصل الثامن يمكن كتابة :

$$P(B_1 - t_{\alpha/2}\sqrt{\frac{S^2}{SXX}} < \beta_1 < B_1 + t_{\alpha/2}\sqrt{\frac{S^2}{SXX}}) = 1-\alpha.$$

لعينة عشوائية معطاة من الحجم n نحسب s و SXX وذلك للحصول على $(1-\alpha)100\%$ فترة ثقة كالتالي :

$$b_1 - t_{\alpha/2}\sqrt{\frac{s^2}{SXX}} < \beta_1 < b_1 + t_{\alpha/2}\sqrt{\frac{s^2}{SXX}}.$$

مثال (١٠-٤) أوجد 95% فترة ثقة للمعلمة β_1 في معادلة الانحدار $\mu_{y|x} = \beta_0 + \beta_1 x$ بالاعتماد على البيانات في جدول (١٠-٥) .

الحل . للمثال (١٠-٣) فإن :

$$SXX = 46.91667 \quad , \quad b_1 = 44.41385 \quad , \quad s^2 = 33.6888$$

باستخدام جدول توزيع t في ملحق (٤) فإن $t_{.025}=2.228$ بدرجات حرية $\nu = 10$. وعلى ذلك فإن95% فترة ثقة للمعلمة β_1 تحسب كالآتي :

$$44.41385 - 2.228\sqrt{\frac{33.6888}{46.91667}} < \beta_1 < 44.41385 + 2.228\sqrt{\frac{33.6888}{46.91667}}$$

أي أن :

$$44.41385 - 2.228(0.847382) < \beta_1 < 44.41385 + 2.228(0.847382)$$

والتي تختصر إلى :

$$42.5 < \beta_1 < 46.3.$$

لاختبار فرض العدم $H_0 : \beta_1 = \beta_1^*$ ضد فرض بديل مناسب يمكننا استخدام توزيع t بدرجات حرية n-2 للحصول على منطقة الرفض. قرارنا سوف يعتمد على القيمة :

$$t = \frac{b_1 - \beta_1^*}{\sqrt{s^2 / SXX}}.$$

مثال (١٠-٥) باستخدام القيمة المقدرة b_1=44.41385 في مثال (١٠-٣) ، أختبر فرض العدم أن $\beta_1 = 0$ عند مستوى معنوية $\alpha = 0.05$.

الحل .

$$H_0 : \beta_1 = 0,$$
$$H_1 : \beta_1 \neq 0.$$
$$\alpha = 0.05.$$

$t_{.025}$=2.228 ومنطقة الرفض $T > 2.228$ أو $T < -2.228$

$$t = \frac{b_1 - 0.0}{\sqrt{s^2 / SXX}}$$

$$= \frac{44.41385}{\sqrt{\frac{33.6888}{46.91667}}} = \frac{44.41385}{0.847382} = 52.41.$$

وبما أن t تقع في منطقة الرفض نرفض H_0 .

أيضا المتغير العشوائي B_0 له توزيع طبيعي بمتوسط :

$$\mu_{B_0} = \beta_0$$

وتباين :

$$\sigma_{B_0}^2 = \sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{SXX})$$

وتبعا لنظرية (٨-١) فإن :

$$Z = \frac{B_0 - \beta_0}{\sqrt{\sigma^2 (\frac{1}{n} + \frac{\bar{x}^2}{SXX})}}$$

متغير عشوائي يتبع التوزيع الطبيعي القياسي وباستبدال σ بالإحصاء S في صيغة Z نحصل على المتغير العشوائي :

$$T = \frac{B_0 - \beta_0}{\sqrt{S^2(\frac{1}{n} + \frac{\overline{x}^2}{SXX})}}$$

والذي يتبع توزيع t بدرجات حرية n −2 .

سوف يستخدم المتغير T للحصول على $(1-\alpha)100\%$ فترة ثقة للمعلمة β_0 كالتالي :

$$b_0 - t_{\alpha/2}\sqrt{s^2(\frac{1}{n} + \frac{\overline{x}^2}{SXX})} < \beta_0 < b_0 + t_{\alpha/2}\sqrt{s^2(\frac{1}{n} + \frac{\overline{x}^2}{SXX})}.$$

مثال (١٠-٦) أوجد 95% فترة ثقة للمعلمة β_0 في خط الانحـــدار $\mu_{y|x} = \beta_0 + \beta_1 x$ بالاعتماد على البيانات في جدول (١٠-٣) .

الحل . من المثال (١٠-٣) فإن :

$$s^2 = 33.6888 \;,\; SXX = 46.91667 \;,\; \overline{x} = 4.58333 \;,\; b_0 = 14.187$$

وعلى ذلك 95% فترة ثقة للمعلمة β_0 تعطى على الشكل :

$$14.187 - 2.228\sqrt{33.6888\left[\frac{1}{12} + \frac{4.58333^2}{46.91667}\right]} < \beta_0 <$$

$$14.187 + 2.228\sqrt{33.6888\left[\frac{1}{12} + \frac{4.58333^2}{46.91667}\right]}.$$

أي أن :

$$14.187 - 2.228(4.2298) < \beta_0 < 14.187 + 2.228(4.2298)$$

والتي تختزل إلى :

$$4.8 < \beta_0 < 23.6.$$

لاختبار فرض العدم $H_0 : \beta_0 = \beta_0^*$ ضد أي فرض مناسب فإننا مرة أخرى ســـوف نستخدم توزيع t بدرجات حرية n-2 للحصول على منطقة الرفض وبالتالي فـــإن قرارنـــا سوف يعتمد على القيمة :

$$t = \frac{b_0 - \beta_0^*}{\sqrt{s^2(\frac{1}{n} + \frac{\overline{x}^2}{SXX})}}$$

الطريقة المتبعة موضحة في المثال التالي :

مثال (١٠-٧) باستخدام القيمة المقدرة $b_0=14.187$ في مثال (١٠-٣) ، أختبر فرض العدم $H_0 : \beta_0 = 0$ ضد الفرض البديل $H_1 : B_1 \neq 0$ عند مستوى معنوية $\alpha = 0.05$.

الحل .

$$H_0 : \beta_0 = 0$$
$$H_1 : \beta_0 \neq 0$$
مستوى معنوية $\alpha = 0.05$

$$t = \frac{b_0 - 0.0}{\sqrt{s^2(\frac{1}{n} + \frac{\bar{x}^2}{SXX})}}$$

$$= \frac{14.187}{\sqrt{33.6888\left[\frac{1}{12} + \frac{4.58333^2}{46.91667}\right]}} = 3.354.$$

$t_{.025}=2.228$ ومنطقة الرفض $T > 2.228$ أو $T < -2.228$ وحيث أن t وقعت في منطقة الرفض نرفض H_0 .

(١٠-٢-٨) التنبؤ Prediction

يمكن استخدام المعادلة $\hat{y}_x = b_0 + b_1 x$ للتنبأ بقيمة $\mu_{Y|x'}$ ، حيث x' ليس من الضرورة أن تكون واحدة من $x_1, x_2, ..., x_n$ في العينة العشوائية من الحجم n للمشاهدات $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. أيضا يمكن استخدام المعادلة $\hat{y}_x = b_0 + b_1 x$ للتنبأ بقيمة واحدة $y_{x'}$ للمتغير $Y | x'$. سوف نتوقع أن خطأ التنبأ سوف يكون أعلى في حالة قيمة واحدة متنبأ بها عنه في حالة التنبأ بالمتوسط وهذا سوف يؤثر على طول فترة الثقة للمعالم المراد تقديرها.

عند الرغبة في الحصول على فترة ثقة للمعلمة $\mu_{Y|x'}$ يكون من الضروري إيجاد التوزيع العيني للفروق بين القيم $\hat{y}_{x'}$ والتي نحصل عليها من خط الانحدار المقدر بتكرار المعاينة والقيمة الحقيقية المقابلة $\mu_{Y|x'}$ من خط الانحدار الحقيقي . باستخدام نظريات الإحصاء يمكن إثبات أن التوزيع العيني للإحصاء $\hat{Y}_{x'} - \mu_{Y|x'}$ يتبع توزيعاً طبيعياً بمتوسط :

$$\mu_{\hat{Y}_{x'} - \mu_{Y|x'}} = E\left[\hat{Y}_{x'} - \mu_{Y|x'}\right] = 0$$

وتباين :

$$\sigma^2_{\hat{Y}_{x'}-\mu_{Y|x'}} = \sigma^2\left[\frac{1}{n}+\frac{(x'-\bar{x})^2}{SXX}\right].$$

في التطبيق يستبدل σ^2 بالقيمة s^2 والتي تمثل قيمة للإحصاء S^2 وعلى ذلك ، فإن الإحصاء :

$$T = \frac{\hat{Y}_{x'}-\mu_{Y|x'}}{\sqrt{S^2\left[\frac{1}{n}+\frac{(x'-\bar{x})^2}{SXX}\right]}}$$

له توزيع t بدرجات حرية n-2. يمكن الحصول على $(1-\alpha)100\%$ فترة ثقة للمعلمة $\mu_{Y|x'}$ من الصيغة التالية :

$$\hat{y}_{x'} - t_{\alpha/2}\sqrt{s^2\left[\left(\frac{1}{n}+\frac{(x'-\bar{x})^2}{SXX}\right)\right]} < \mu_{Y|x'} < \hat{y}_{x'} + t_{\alpha/2}\sqrt{s^2\left[\left(\frac{1}{n}+\frac{(x'-\bar{x})^2}{SXX}\right)\right]}.$$

مثال (١٠-٨) باستخدام البيانات في جدول (١٠-٥) أوجد 95% فترة ثقة للمعلمة $\mu_{Y|4}$

الحل . من معادلة الانحدار المقدرة فإن :

$$\hat{y}_4 = 14.187 + (44.41385)(4)$$
$$= 191.84.$$

عرفنا مما سبق أن :

$$SXX = 46.91667,\ \bar{x} = 4.58333,\ s^2 = 33.6888,$$

$t_{.025} = 2.228$ بدرجات حرية 10 . وعلى ذلك 95% فترة ثقة للمعلمة $\mu_{Y|4}$ هي :

$$191.84 - 2.228\sqrt{33.6888\left[\frac{1}{12}+\frac{(4-4.58333)^2}{46.91667}\right]} < \mu_{Y|4} <$$
$$191.84 + 2.228\sqrt{33.6888\left[\frac{1}{12}+\frac{(4-4.58333)^2}{46.91667}\right]}\ .$$

أي أن :

$$191.84 - (2.228)(1.7469) < \mu_{Y|4} < 191.84 + (2.228)(1.7469)$$

والتي تختصر إلى :

$$187.94791 < \mu_{Y|4} < 195.73209.$$

للحصول على $(1-\alpha)100\%$ فترة ثقة لأي قيمة مفردة $y_{x'}$ للمتغير $Y \mid x'$ ، يكون من الضروري تقدير التباين للفروق بين القيم $\hat{y}_{x'}$ المحسوبة من خط الانحدار المقدر بتكرار المعاينة والقيمة المقابلة الحقيقية $y_{x'}$. باستخدام نظريات الإحصاء يمكن إثبات أن التوزيع العيني للإحصاء $\hat{Y}_{x'} - y_{x'}$ يتبع توزيعا طبيعيا بمتوسط $\mu_{\hat{Y}_{x'}-y_{x'}} = 0$ وتباين :

$$\sigma^2_{\hat{Y}_{x'}-y_{x'}} = \sigma^2\left[1+\frac{1}{n}+\frac{(x'-\bar{x})^2}{SXX}\right]$$

في التطبيق يستبدل σ^2 بالقيمة s^2 والتي تمثل قيمة للإحصاء S^2 وعلـــى ذلـــك ، فـــإن الإحصاء :

$$T = \frac{\hat{Y}_{x'} - y_{x'}}{\sqrt{S^2\left[1+\frac{1}{n}+\frac{(x'-\bar{x})^2}{SXX}\right]}}.$$

يتبع توزيع t بدرجات حرية $n-2$.

يمكن الحصول على $(1-\alpha)100\%$ فترة ثقة لقيمة مفردة $y_{x'}$ من الصيغة التالية :

$$\hat{y}_{x'} - t_{\alpha/2}\sqrt{s^2\left[1+\frac{1}{n}+\frac{(x'-\bar{x})^2}{SXX}\right]} < y_{x'} < \hat{y}_{x'} + t_{\alpha/2}\sqrt{s^2\left[1+\frac{1}{n}+\frac{(x'-\bar{x})^2}{SXX}\right]}.$$

مثال (١٠-٩) للبيانات في جدول (١٠-٥) أوجد (ا) 95% فترة ثقة لــ y_4 (ب) وضـح بيانيا 95% فترة ثقة لــ $y_{x'}$.

الحل . (ا) $\bar{x} = 4.58333$, $s^2 = 33.6888$, $n = 12$ وعلى ذلك 95% فترة ثقـة لــ y_4 هي :

$$191.84 - 2.228\sqrt{33.6888\left[1+\frac{1}{12}+\frac{(4-4.58333)^2}{46.91667}\right]} < y_4 <$$

$$191.84 + 2.228\sqrt{33.6888\left[1+\frac{1}{12}+\frac{(4-4.58333)^2}{46.91667}\right]}$$

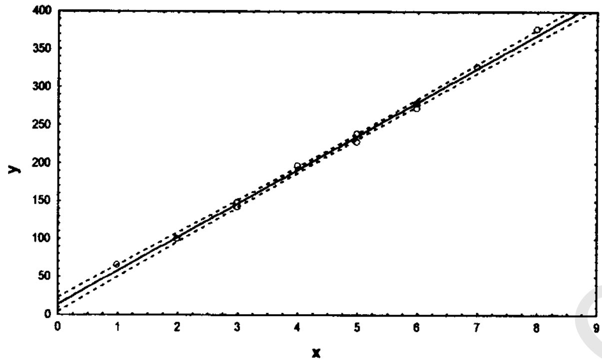
أي أن :

$$191.84-2.228(6.061)<y_4 < 191.84 + 2.228\ (6.06)$$

والتي تختصر إلى :

$$178.3 < y_4 < 205.3$$

(ب) يوضح شكل (١٠-١٠) 95% فترة ثقة لـ y_x، .



شكل (١٠-١٠)

(١٠-٢-٩) اختبار خطية الانحدار **Test for Linearity of Regression**

لمشكلة معطاة فإننا نفترض إما أن الانحدار خطي ونتبع خطوات التقدير التي تناولناها في البند (١٠-٢-٣) أو نستنتج أن الانحدار غير خطي ونبحث عن نمو زج آخـــر. في الجــزء التالي سوف نتناول اختبار الخطية في حالتين .

الحالة الأولي عندما تكون σ^2 معلومة (التباين معلوم)

لاختبار فرض العدم H_0 : النموذج خطي ضد الفرض البديل H_1 : النموذج غير خطي فإن:

$$\chi^2 = \frac{SSE}{\sigma^2}$$

قيمة لمتغير عشوائيX^2 له توزيع χ^2 بدرجات حرية $v = n - 2$ وذلك بافتراض صحة فرض العدم. لمستوى معنوية α فإن منطقة الرفض $X^2 > \chi^2_\alpha$ حيث χ^2_α تستخرج من جدول توزيع χ^2 في ملحق (٥) بدرجات حرية $v = n - 2$. إذا وقعت χ^2 في منطقة الرفض نرفض H_0.

مثال (١٠-١٠) لأزواج القياسات في جدول (١٠-٧) أختبر فرض العدم H_0 : النمــوذج خطي ضد الفرض البديل H_1 : النموذج غير خطي وذلك عند مســـتوى معنويـــة $\alpha = 0.01$ وتحت فرض أن $\sigma^2 = 1$.

جدول (١٠-٧)

x	0.345	0.287	0.251	0.225	0.207	0.186	0.161	0.132	0.084	0.060
y	367	311	295	268	253	239	220	213	193	192

الحل . لاختبار فرض العدم H_0 : النموذج خطي ضد الفرض البديل H_1 : النموذج غير خطي نحسب جدول تحليل التباين والمعطى في جدول (١٠-٨) .ونحسب القيمة :

$$\chi^2 = \frac{SSE}{\sigma^2} = \frac{21.95}{1} = 21.95.$$

بمستوي معنوية $\alpha = 0.01$ فإن $\chi^2_{0.01} = 20.09$ والمستخرجة من جدول توزيع χ^2 في ملحق (٥) بدرجات حرية $v = n - 2 = 10 - 2 = 8$. منطقة الرفض $X^2 > 20.09$. وبما أن χ^2 تقع في منطقة الرفض نرفض H_0 ، أي أن العلاقة بين المتغيرين لا تتبع النموذج الخطي ويجب أن نبحث عن نموذج آخر.

جدول (١٠-٨)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات
الانحدار	1	342.0	342.0
الخطأ	8	21.95	2.744
الكلي	9	363.95	

الحالة الثانية عندما تكون σ^2 مجهولة :

لاختبار فرض العدم H_0 : النموذج خطي ضد الفرض البديل H_1: النموذج غير خطي نختار عينة عشوائية من الحجم n من المشاهدات بحيث أن لكل قيمة من x يوجد أكثر من قيمة لـ y . أي أن العينة التي حجمها n تحتوي على قيم مختلفة من x عددها k بحيث تحتوي العينة على n_1 من القيم المشاهدة $y_{11}, y_{12}, \ldots, y_{1n_1}$ المقابلة للقيمة x_1 و n_2 من القيم المشاهدة $y_{21}, y_{22}, \ldots, y_{2n_2}$ المقابلة للقيمة x_2 و... و n_k من القيم المشاهدة $y_{k1}, y_{k2}, \ldots, y_{kn_k}$ المقابلة للقيمة x_k حيث $n = \sum_{i=1}^{k} n_i$.

لإجراء الاختبار نتبع الخطوات التالية :

(أ) نحسب مجموع مربعات الخطأ الخالص من x_1 pure error of sum squares من الصيغة التالية :

$$\sum_{u=1}^{n_1} (y_{1u} - \bar{y}_1)^2$$

، حيث $\bar{y}_1 = \sum_{u=1}^{n_1} y_{1u} / n_1$ ، بدرجات حرية $n_1 - 1$. بنفس الشكل يمكن حساب مجموع مربعات الخطأ الخالص من x_2 و ... و x_k . مجموع مربعات الخطأ الخالص الكلي يحسب من الصيغة التالية :

$$SSP = \sum_{i=1}^{k} \sum_{u=1}^{n_i} (y_{iu} - \bar{y}_i)^2$$

بدرجات حرية :

$$n_e = \sum_{i=1}^{k} (n_i - 1) = \sum_{i=1}^{k} n_i - k$$

(ب) نحسب متوسط مجموع مربعات الخطأ الخالص من الصيغة التالية :

$$s_e^2 = \frac{SSP}{n_e}$$

والذي يعتبر تقدير للمعلمة σ^2

(ج) نحسب مجموع مربعات قصور التوفيق sum squares lake of fit كالتالي :

$$SSL = SSE - SSP$$

بدرجات حرية $n_L = (n-2) - n_e$.

من الحسابات السابقة فإن جدول تحليل التباين يكون كما هو موضح في جدول (١٠-٩)

جدول (١٠-٩)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات
الانحدار الخطـــأ	1 n-2	SSR SSE	SSR/1 SSE/n-2
الكلي	n-1		
قصور التوفيق الخطأ الخالص	n_L n_e	SSL SSP	$MSL = SSL/n_L$ $s_e^2 = SSP / n_e$

نحسب قيمة f' كالتالي :

$$f' = \frac{MSL}{s_e^2}.$$

بافتراض صحة فرض العدم فإن f' تمثل قيمة لمتغير عشوائي F له توزيع F بدرجات حرية n_L , n_e . عند مستوي معنوية α نحسب منطقة الرفض كالتالي : $F > f_\alpha(n_L, n_e)$ حيـــث:

$f_\alpha(n_L, n_e)$ تستخرج من جدول توزيع F في ملحق (٦) أو ملحق (٧) بدرجات حرية $\nu_1 = n_L, \nu_2 = n_e$. إذا وقعت f' في منطقة الرفض نرفض H_0 .

مثال (١٠-١١) لازواج القياسات في جدول (١٠-١٠) اختبر فرض العدم H_0 : النموذج خطي ضد الفرض البديل H_0 : النموذج غير خطي عند مستوى معنوية $\alpha = 0.05$.

جدول (١٠-١٠)

المشاهدات	x	y	المشاهدات	x	y	المشاهدات	x	y
1	1.3	2.3	9	3.7	1.7	17	5.3	3.5
2	1.3	1.8	10	4.0	2.8	18	5.3	2.8
3	2.0	2.8	11	4.0	2.8	19	5.3	2.1
4	2.0	1.5	12	4.0	2.2	20	5.7	3.4
5	2.7	2.2	13	4.7	5.4	21	6.0	3.2
6	3.3	3.8	14	4.7	3.2	22	6.0	3.0
7	3.3	1.8	15	4.7	1.9	23	6.3	3.0
8	3.7	3.7	16	5.0	1.8	24	6.7	5.9

الحل .

H_0 : النموذج الخطى :

H_1 : النموذج غير خطى :

$\alpha = 0.05$.

سوف نوجد مجموع المربعات الخالص ثم مجموع مربعات قصور التوفيق كالتالي :

مجموع مربعات الخطأ الخالص عند $x = 1.3$ هو :

$$(2.3)^2 + (1.8)^2 - \{ (2.3 + 1.8)^2 / 2 \} = 0.125$$

بدرجة حرية واحدة ($n_1 = 2\text{-}1 = 1$) .

مجموع مربعات الخطأ الخالص عند $x = 2.0$ هو :

$$(2.8)^2 + (1.5)^2 - \{ (2.8 + 1.5)^2 / 2 \} = 0.845$$

بدرجة حرية واحدة ($n_2 = 2\text{-}1 = 1$) . بنفس الطريقة يتم حساب مجموع مربعات الخطأ الخالص للقيم الباقية من x كما في جدول (١٠ - ١١) .

جدول (١٠-١١)

مستوى x	$\Sigma(y_{iu} - \bar{y}_i)^2$	درجات حرية
1.3	0.125	1
2.0	0.845	1
3.3	2.000	1
3.7	2.000	1

←

4.0 4.7 5.3 6.0	0.240 6.260 0.980 0.020	2 2 2 1
المجموع	12.470	11

جدول تحليل التباين معطى في جدول (١٠-١٢).

جدول (١٠-١٢)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات	f المحسوبة
الانحدار	1	6.326	6.326	$f=\dfrac{6.326}{0.963}$ $=6.569$ عند مستوى معنوية $\alpha=0.05$
الخطأ	22	21.192	$0.963=s^2$	
قصور التوفيق الخطأ الخالص	11 11	8.722 12.470	$0.793=MSL$ $1.134=s_e^2$	$f'=\dfrac{0.793}{1.134}$ $=0.699$

من جدول (١٠-١٢) فإن القيمة $f'=0.699$ غير معنوية لأنها أقل من الواحد.

(١٠-٣) بعض نماذج الانحدار الغير خطية

Some Nonlinear Regression Models

يوجد العديد من النماذج الرياضية الغير خطية المستخدمة في تمثيل العلاقات الاقتصادية والاجتماعية ... الخ. سوف تقتصر دراستنا في هذا الجزء على النموذج الأسي ونموذج القوى.

(١٠-٣-١) النموذج الأسي The Exponential Model

معادلة النموذج الأسي تكون على الصورة التالية :

$$\mu_{Y|x}=\gamma\,\delta^{x}$$

حيث γ,δ ثابتان والمطلوب تقديرهما من البيانات بالتقديرين c , d على التوالي . يمكن تقدير $\mu_{Y|x}$ بالقيمة $\hat{y}_x$ من منحنى الانحدار المقدر التالي :

$$\hat{y}_x=c\;d^{x}.$$

بأخذ لوغاريتمات الطرفين (للأساس e) في المعادلة السابقة فإن منحنى الانحدار المقدر يمكن كتابته على الشكل :

$$\ln\hat{y}_x=\ln c+(\ln d)x,$$

وكل زوج من المشاهدات في العينة يحقق العلاقة :

$$\ln y_i = \ln c + (\ln d)\, x_i + e_i = b_0 + b_1 x_i + e_i,$$

حيث أن : $b_1 = (\ln d), b_0 = \ln c$. وعلى ذلك يمكن إيجاد b_0, b_1 بالصيغ المستخدمة لنموذج الانحدار الخطى ، التي سبق أن تناولناها ، باستخدام النقاط $(x_i , \ln y_i)$ ثم إيجاد c , d بأخذ القيم المقابلة للوغاريتمات لـــ b_0, b_1 على التوالي، أي أن :

$$d = \exp(b_1), c = \exp(b_0) .$$

طريقة المربعات الصغرى لتوفيق منحنى أسى لفئة من المشاهدات موضحه في المثال التالي .

مثال (١٠-١٢) لازواج القياسات في جدول (١٠-١٣) أوجد معادلة الانحدار المقدرة تحت فرض النموذج الأسى .

جدول (١٠ - ١٣)

x	1	2	3	4	5	6	7
y	304	341	393	457	548	670	882

الحل . بوضع $y_i' = \ln y_i$ فإن : $\Sigma y_i' = 43.243148$

و $\Sigma x_i^2 = 140 \quad , \quad \Sigma x_i = 28 \quad , \quad n = 7$

$$\Sigma x_i y_i' = 177.85134 \quad , \quad \bar{x} = 4 , \frac{\Sigma y_i'}{n} = 6.1775926$$

$$b_1 = \frac{177.85134 - \dfrac{(28)(43.243148)}{7}}{140 - \dfrac{(28)^2}{7}}$$

$$= 0.174241 .$$

$$b_0 = 6.1775926 - (0.174241)\ (4)$$

$$= 5.4806286 .$$

معادلة الانحدار المقدرة هي :

$$\hat{y}_x = 5.4806286 + 0.174241x.$$

وعلى ذلك :

$$\ln d = b_1 = 0.174241 \ , \ \ln c = b_0 = 5.4806286,$$

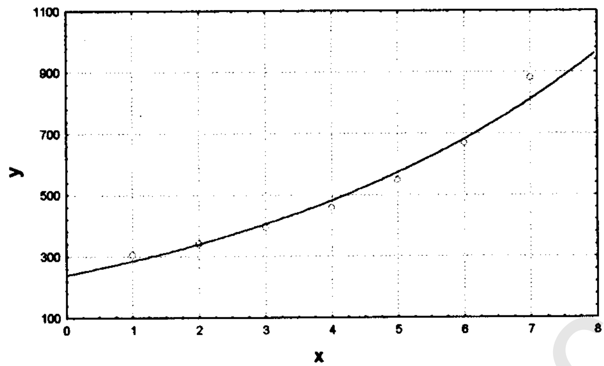
$$d = \exp(b_1) = 1.1903424 \ , c = \exp(b_0) = 239.99752.$$

وبالتالي فإن منحنى الانحدار المقدر بالمربعات الصغرى هو :

$$\hat{y}_x = c \ \ d^x$$

$$= (239.99752)(1.1903424)^x$$

والتمثيل البياني لها موضح في شكل (١٠– ١١) .



شكل (١٠– ١١)

(١٠–٣–٢) نموذج القوى **Power Model**

معادلة نموذج القوى تكون على الصورة التالية :

$$\mu_{Y|x} = a_0 \ x^{\ a_0'}$$

حيث a_0 , a_0' ثابتان والمطلوب تقديرهما من البيانات بالتقديرين c_o , d_o على التوالي. يمكن تقدير $\mu_{Y|x}$ بالقيمة $\hat{y}_x$ من منحنى الانحدار المقدر التالي :

$$\hat{y}_x = c_o \ x^{d_o} .$$

بأخذ لوغاريتمات الطرفين (للأساس e) فإن منحنى الانحدار يمكن كتابته على الشكل :

$$\ln \hat{y}_x = \ln \ c_o + d_o(\ln \ x)$$

كل زوج من المشاهدات في العينة يحقق العلاقة :

$$\ln y_i = \ln \ c_o + d_o(\ln \ x_i) + e_i$$
$$= b_o + b_1(\ln \ x_i) + e_i$$

حيث $b_1 = d_o, b_o = \ln \ c_o$. وعلى ذلك يمكن إيجاد b_0 , b_1 بالصيغ المستخدمة لنموذج الانحدار الخطى، التي سبق أن تناولتها ، باستخدام النقاط $(\ln \ x_i, \ln \ y_i)$ ثم إيجاد c_o , d_o حيث $b_1 = d_o, \ln c_o = b_o$.

مثال (١٠–١٣) لازواج القياسات في جدول (١٠–١٤) أوجد معادلة الانحدار المقدرة تحت فرض نموذج القوى .

جدول (١٠ – ١٤)

x	600	600	600	600	500	500	500	500	400	400	400	400
y	2.35	2.65	3.0	3.6	6.4	7.8	9.8	16.5	21.5	24.5	26.0	33.0

الحل .

$$n = 12 \ , \ \Sigma \ln x_i = 74.412 \ , \ \Sigma \ln y_i = 26.22601 \ ,$$

$$\Sigma \ln x_i^2 = 461.75874 \ , \ \Sigma (\ln x_i)(\ln y_i) = 160.84601 \ ,$$

$$\Sigma \ln y_i^2 = 67.74609.$$

الحل.

$$b_1 = \frac{160.84601 - \dfrac{(74.412)(26.22601)}{12}}{461.75874 - \dfrac{(74.412)^2}{12}}$$

$$= -5.3996,$$

$$b_0 = \frac{26.22061 - (-5.3996)(74.412)}{12}$$

$$= 35.6684.$$

معادلة الانحدار المقدرة هي :

$$\hat{y}_x = 35.6684 - 5.3996x.$$

وعلى ذلك :

$$\ln c_o = b_0 = 35.6684$$

أي أن :

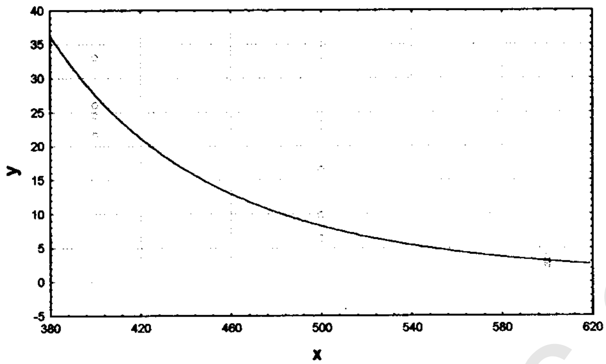
$$c_0 = \exp(b_0) = 3.094491530.10^{15}$$

$$d_0 = b_1 = -5.3996$$

والمعادلة الأساسية المقدرة هي :

$$\hat{y}_x = c_0 \ x^{d_0} = (3.094491530 \cdot 10^{15}) \cdot x^{-5.3996}.$$

والتمثيل البياني لها موضح في شكل (١٠– ١٢).



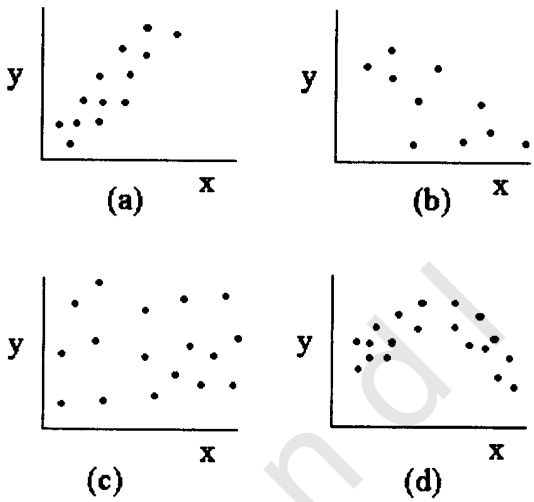
شكل (١٠- ١٢)

(١٠-٤) معامل الارتباط الخطى البسيط

The Simple Linear Correlation Coefficient

في مشكلة الانحدار كان اهتمامنا بالتنبأ بمتغير وذلك من المعلومات عن المتغيرات المستقلة ، بينما في مشكلة الارتباط فإن اهتمامنا سوف يكون في قياس العلاقة بين متغيرين أو أكثر. مرة أخرى فإن قيم المتغيرات المستقلة كانت ثابتة في مشكلة الانحدار .الآن سوف يختلف الوضع. سوف نعرف معامل الارتباط الخطى بأنه مقياس للعلاقة بين متغيرين عشوائيين X,Y. وسوف نرمز له بالرمز r . سوف نفترض أن المتغيران X,Y لهما توزيع احتمالي ثنائي. لحساب معامل الارتباط الخطى نختار عينة عشوائية من أزواج المشاهدات (x , y) . إذا كانت نقاط شكل الانتشار تتركز فوق وحول خط انحدار له ميل موجب ، فهذا يدل على ارتباط موجب قوى بين المتغيرين (ارتباط طردي) كما هو موضح في شكل (١٠-١٣) (a) . ومن ناحية أخرى ، إذا كانت نقاط شكل الانتشار تتركز فوق وحول خط انحدار له ميل سالب فهذا يدل على ارتباط قوى سالب بين المتغيرين (ارتباط عكسي) كما هو موضح في شكل (١٠-١٣) (b) كلما زاد انتشار نقاط شكل الانتشار حول وفوق خط الانحدار فإن الارتباط يقل عدديا بين المتغيرين ، إذا كانت نقاط شكل الانتشار تنتشر بطريقة عشوائية كما في شكل (١٠-١٣) (c) فهذا يعنى أن $r=0$ ونستنتج عدم وجود علاقة بين X,Y. ولما كان معامل الارتباط بين متغيرين يعتبر مقياس للعلاقة الخطيه بينهما وعلى ذلك فإن $r=0$ تعنى قصور في الخطية وليست قصور في الارتباط . على سبيل المثال قد تكون هناك علاقة ولكنها علاقة غير خطيه . فعلى سبيل المثال إذا

وجدت علاقة قوية من الدرجة الثانية بين المتغيرين X,Y كما هو موضح في شكل (١٠-١٣) (d) فهذا يعنى أن r=0 . يعتبر معامل الارتباط الخطى (معامل بيرسون للارتباط) أو اختصارا معامل الارتباط أكثر مقاييس الارتباط الخطى انتشارا.



شكل (١٠-١٣)

يتم حساب معامل الارتباط من المعادلة التالية :

$$r = \frac{\Sigma x_i y_i - \dfrac{\Sigma x_i \Sigma y_i}{n}}{\sqrt{\left[\Sigma x_i^2 - \dfrac{(\Sigma x_i)^2}{n}\right]\left[\Sigma y_i^2 - \dfrac{(\Sigma y_i)^2}{n}\right]}}$$

$$= \frac{SXY}{\sqrt{SXX.SYY}}$$

من الجزء (١٠-٢-٦) يمكن حساب معامل الارتباط كالتالي :

$$r = \pm\sqrt{r^2}$$

حيث r^2 هو معامل التحديد البسيط والإشارة تخص التقدير b_1 . وبما أن $0 \le r^2 \le 1$ فـإن r يمكن أن تأخذ الإشارة الموجبة أو السالبة ، أي أن :

$$-1 \le r \le 1$$

عادة يفضل حساب معامل الارتباط من معادلته وليس من r^2 لصعوبة الحساب.

مثال (١٠-١٤) لدراسة العلاقة بين تركيز الأوزون Ozone (X) (مقاس PPM) وتركيز الكربون (Y) (مقاس $\mu g / m^3$) تم الحصول على البيانات المعطاة في جدول (١٠-١٥).

جدول (١٠-١٥)

x	0.066	0.088	0.120	0.050	0.162	0.186	0.057	0.100
y	4.6	11.6	9.5	6.3	13.8	15.4	2.5	11.8
x	0.112	0.055	0.154	0.074	0.111	0.140	0.071	0.110
y	8.0	7.0	20.6	16.6	9.2	17.9	2.8	13.0

أوجد معامل الارتباط البسيط .

الحل .

$$n = 16 \ , \ \Sigma x_i = 1.656 \ , \ \Sigma y_i = 170.6 \ ,$$

$$\Sigma x_i^2 = 0.196912 \ , \ \Sigma x_i \ y_i = 20.0397 \ ,$$

$$\Sigma y_i^2 = 2253.56.$$

$$SXY = \Sigma x_i \ y_i - \frac{\Sigma x_i \ \Sigma y_i}{n}$$

$$= 20.0397 - \frac{(1.656)(170.6)}{16}$$

$$= 2.3826,$$

$$SXX = \Sigma x_i^2 - \frac{(\Sigma x_i)^2}{n}$$

$$= 0.196912 - \frac{(1.656)^2}{16} = 0.025516,$$

$$SYY = \Sigma y_i^2 - \frac{(\Sigma y_i)^2}{n} = 2253.56 - \frac{(170.6)^2}{16}$$

$$= 434.5375.$$

وعلى ذلك :

$$r = \frac{SXY}{\sqrt{SXX.SYY}} = \frac{2.3826}{\sqrt{(0.025516)(434.5375)}} = 0.716.$$

في المناقشة السابقة لم نضع فروض قوية على توزيع المجتمع الذي اختبرت منه العينة وذلك للحصول على تقدير بنقطة للمعلمة ρ والتي ترمز إلى معامل ارتباط المجتمع. للحصول على

$(1-\alpha)100\%$ فترة ثقة للمعلمة ρ أو اختبارات فروض تخص ρ فإننا نفترض أن العينة مأخوذة من مجتمع يتبع التوزيع الطبيعي الثنائي **the bivarate normal distribution** أي أن X , Y متغيرين عشوائيين حيث دالة التوزيع الهامشية لكل من X , Y تتبع التوزيع الطبيعي . يمكن اختبار الاعتدال لقيم X وقيم Y على حدة بطريقة رياضية سوف نتناولها في البند (١٢-٥) من الفصل الثاني عشر .

اختبارات فروض وفترات ثقة تخص ρ

Tests Hypotheses and Confidence Intervals Concerning ρ

لاختبار فرض العدم $H_0: \rho = 0$ ضد الفرض البديل $H_1: \rho \neq 0$ أو الفرض البديل $H_1: \rho > 0$ أو الفرض البديل $H_1: \rho < 0$ وبافتراض صحة فرض العدم فإن :

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

هي قيمة لمتغير عشوائي T له توزيع t بدرجات حرية $\nu = n-2$. وعلى ذلك لمستوى معنوية α وللفرض البديل $H_1: \rho \neq 0$ (اختبار ذي جانبين) فإن منطقة الرفض سوف تكون $T < -t_{\alpha/2}$ or $T > t_{\alpha/2}$ حيث $t_{\alpha/2}$ هي قيمة t المستخرجة من جدول توزيع t بدرجات حرية $\nu = n-2$. للبديل $H_1: \rho > 0$ فإن منطقة الرفض $T > t_\alpha$ وللبديل $H_1: \rho < 0$ فإن منطقة الرفض $T < -t_\alpha$.

مثال (١٠-١٥) بفرض أن البيانات في مثال (١٠-١٤) مأخوذة من مجتمع يتبع التوزيع الثنائي الطبيعي . المطلوب اختبار فرض العدم $H_0: \rho = 0$ ضد الفرض البديل $H_1: \rho > 0$ وذلك عند مستوى معنوية $\alpha=0.01$.

الحل .

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.716\sqrt{14}}{\sqrt{1-(0.716)^2}} = 3.84.$$

$t_{0.01} = 2.624$ والمستخرجة من جدول توزيع t في ملحق (٤) بدرجات حرية $\nu = n-2 = 16-2 = 14$. منطقة الرفض $T > 2.624$. وبما أن t تقع في منطقة الرفض ، نرفض H_0 .

وبما أن ρ يقيس قوة الارتباط الخطي بين متغيرين X,Y في المجتمع فإن فرض العدم $H_0: \rho=0$ يدل على عدم وجود ارتباط بين المتغيرين في المجتمع. في البند (١٠-٢-٧) استخدمنا

القيمة $t = \dfrac{b_1}{\sqrt{s^2/SXX}}$ لاختبار فرض العـــدم $H_0:\ \beta_1 = 0$ هنـا يمكـن إثبـات أن $t = r\sqrt{n-2}/\sqrt{1-r^2} = b_1/\sqrt{s^2/SXX}$ وهذا يعنى أن الاختبارين متكافئين . وعلى ذلك إذا كان الاهتمام فقط بقياس قوة العلاقة بين متغيرين X , Y وليس الحصول على معادلـة الانحدار الخطى فإن اختبار $H_0:\ \rho = 0$ يكون اسهل من اختبار t لأنه يتطلب كمية قليلة مـن الحسابات.

الأسلوب المستخدم لاختبار $H_0:\ \rho = \rho_0$ عندما $\rho_0 \neq 0$ لا يكافئ أي طريقة مستخدمة في تحليل الانحدار. بفرض أن أزواج المشاهدات $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ تمثل عينة عشوائية مأخوذة من مجتمع يتبع التوزيع الثنائي الطبيعي وإذا كانت n كبيرة وبـافتراض صحـة فرض العدم فإن :

$$v = \frac{1}{2}\ \ln\left(\frac{1+r}{1-r}\right)$$

هي قيمة لمتغير عشوائي V تقريباً يتبع التوزيع الطبيعي بمتوسط $\mu_V = \dfrac{1}{2}\ \ln\left(\dfrac{1+\rho_0}{1-\rho_0}\right)$ وتباين $\sigma_V^2 = \dfrac{1}{n-3}$ ، حيث التباين لا يعتمد على ρ، وعلى ذلك فإن :

$$z = \frac{v - \frac{1}{2}\ \ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{1/\sqrt{n-3}}.$$

هي قيمة لمتغير عشوائي Z تقريبا يتبع التوزيع الطبيعي القياسـي. الجـدول (١٠-١٦) يعطـى الفروض البديلة ومنطقة الرفض لكل فرض بديل عند مستوى معنوية α.

جدول (١٠-١٦)

منطقة الرفض	الفروض البديلة
$Z < -z_{\alpha/2}$ or $Z > z_{\alpha/2}$.	$H_1: \rho \neq \rho_0$
$Z > z_\alpha$	$H_1: \rho > \rho_0$
$Z < -z_\alpha$	$H_1: \rho < \rho_0$

مثال (١٠-١٦) إذا كان لديك البيانات التالية :

$$n = 20 \ , \ \Sigma y_i = 690.30 \ , \ \Sigma y_i^2 = 29040.29 \ ,$$

$$\Sigma x_i y_i = 10818.56 \ , \ \Sigma x_i = 285.90 \ \ \Sigma x_i^2 = 4409.55,$$

أختبر الفرض $0.5 < \rho < 0.8$ عند مستوى معنوية $\alpha = 0.05$

الحل . أي أننا نرغب في اختبار :

$$H_0 : \rho = 0.5 \ ,$$

$$H_1 : \rho > 0.5$$

$$\alpha = 0.05.$$

وحيث أن $r = .733$ **فإن :**

$$v = \frac{1}{2}\ln\left(\frac{1+.733}{1-.733}\right) = .935,$$

$$\mu_V = \frac{1}{2}\ln\left(\frac{1+.5}{1-.5}\right) = .549.$$

وعلى ذلك فإن :

$$z = \frac{v - \frac{1}{2}\ln[(1+\rho_0)/(1-\rho_0)]}{1/\sqrt{n-3}}$$

$$= (.935 - .549)\sqrt{17} = 1.59.$$

$z_{0.05} = 1.645$ والمستخرجة من جدول التوزيع الطبيعي القياسي في ملحق (٣) . منطقة الرفض $Z > 1.645$. وبما أن z تقع في منطقة القبول نقبل H_0 .

يمكن الحصول على $(1-\alpha)100\%$ فترة ثقة للمعلمة ρ من الصيغة التالية :

$$\frac{e^{2c_1} - 1}{e^{2c_1} + 1} \le \rho \le \frac{e^{2c_2} - 1}{e^{2c_2} + 1}$$

حيث أن $c_1 = v - \dfrac{z_{\alpha/2}}{\sqrt{n-3}}$, $c_2 = v + \dfrac{z_{\alpha/2}}{\sqrt{n-3}}$

للبيانات في مثال (١٠-١٦) فإن :

$$r = 0.733 \quad , \quad v = 0.935 \quad , \quad n = 20,$$

$$c_1 = .935 - 1.96/\sqrt{17} = .460,$$

$$c_2 = .935 + 1.96/\sqrt{17} = 1.410$$

بالتعويض في الصيغة التالية يمكن الحصول على 95% فترة ثقة للمعلمة ρ كالتالي :

$$\frac{e^{2(.460)}-1}{e^{2(.460)}+1} \le \rho \le \frac{e^{2(1.410)}-1}{e^{2(1.410)}+1}$$

والتي تختزل إلى :

$$0.43 \le \rho \le 0.89$$

(١٠-٥) الانحدار الخطي المتعدد Linear Multiple Regression

في الغالب تكون العلاقات الفعلية سواء الاقتصادية أو الاجتماعية أو السياسية معقدة يمثل فيها متغير واحد تابع وعدد من المتغيرات الأساسية المستقلة ومن الأمثلة العديدة على ذلك في مجال الاقتصاد نجد أن الكمية المستهلكة من سلعة ما تتأثر بسعر السلعة ذاتها علاوة على أسعار السلع البديلة وأيضا بالإضافة إلى ذوق المستهلك. كذلك كمية الإنتاج تتأثر بالعمل ورأسي المال والموارد الوسيطية وغيرها من عناصر العملية الإنتاجية. وفي مجال التأمين يتوقف القسط التأميني على عمر المؤمن ودخله وقيمة الوثيقة وطول فترات التأمين .

(١٠-٥-١) طريقة المربعات الصغرى Least Square Method

الآن سوف نتناول مشكلة التقدير والتنبأ بقيمة متغير تابع بالاعتماد على فئة من المشاهدات المأخوذة من عدة متغيرات مستقلة $X_1, X_2, \ldots, X_p$. كما في حالة الانحدار الخطي البسيط ، القيمة لكل متغير مستقل والمختارة بواسطة الباحث سوف تظل ثابتة . إذا تم اختيار عينة عشوائية من الحجم n من المجتمع فإن بيانات العينة سوف تكون على الشكل :

$$\{x_{1i}, x_{2i}, \ldots, x_{pi}; y_i); i = 1,2,\ldots,n$$

مرة أخرى القيمة y_i تمثل قيمة لمتغير عشوائي Y_i . نموذج الانحدار الخطي المتعدد النظري سوف يكون على الشكل :

$$\mu_{Y|x_1,x_2,\ldots,x_p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p ,$$

حيث أن $\beta_0, \beta_1, \ldots, \beta_p$ تمثل المعالم المطلوب تقديرها. نموذج الانحدار الخطي المتعدد المقدر سوف يكون على الشكل :

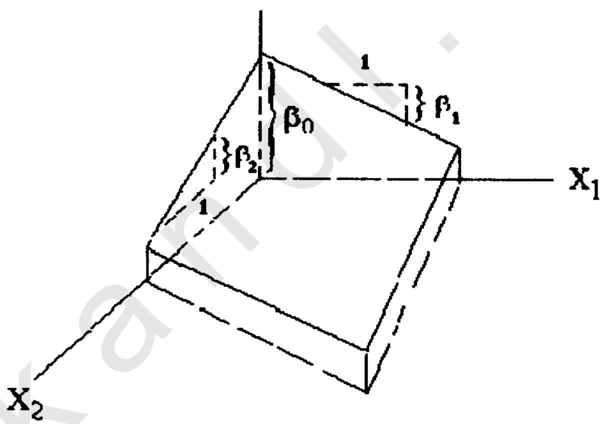
$$\hat{y}_{x_1,x_2,\ldots,x_p} = b_0 + b_1 x_1 + \ldots. + b_p x_p ,$$

حيث أن $b_0, b_1, ...b_p$ التقديرات المطلوب الحصول عليها للمعالم $\beta_0, \beta_1, ..., \beta_p$. سوف تقتصر دراستنا في هذا البند على حالة وجود متغيرين مستقلين X_1 , X_2 $(p=2)$. النتائج يمكن تعميمها إلى عدة متغيرات مستقلة . في حالة وجود أكثر من متغير مستقل فإن معرفتنا لنظرية المصفوفات سوف يساعدنا في عملية الحساب .

في حالة وجود متغيرين مستقلين فإن نموذج الانحدار الخطي المتعدد سوف يكون على الشكل :

$$\mu_{Y|x_1,x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2,$$

الشكل البياني للمعادلة السابقة موضح في شكل (١٠-١٤) و الذي يسمى المستوى .



شكل (١٠-١٤)

ونموذج الانحدار الخطي المقدر سوف يكون على الشكل :

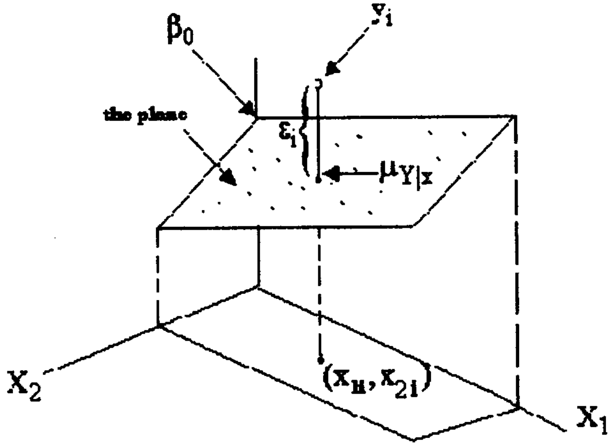
$$\hat{y}_{x_1,x_2} = b_0 + b_1 x_1 + b_2 x_2,$$

وكل فئة من المشاهدات تحقق العلاقة :

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + e_i, i = 1,2,...,n,$$

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i, i = 1,2,...,n.$$

الشكل البياني للمعادلة السابقة موضح في شكل (١٠-١٥) .



شكل (١٠-١٥)

تقديرات المربعات الصغرى b_0, b_1, b_2 يمكن الحصول عليها بحل المعادلات الخطية التالية آنيا :

$$b_0 n + b_1 \Sigma x_{1i} + b_2 \Sigma x_{2i} = \Sigma y_i$$
$$b_0 \Sigma x_{1i} + b_1 \Sigma x_{1i}^2 + b_2 \Sigma x_{1i} x_{2i} = \Sigma x_{1i} y_i$$
$$b_0 \Sigma x_{2i} + b_1 \Sigma x_{1i} x_{2i} + b_2 \Sigma x_{2i}^2 = \Sigma x_{2i} y_i$$

مثال (١٠-١٧) يتأثر الاستهلاك السنوي للطعام على كل من الدخل السنوي للأسرة x_1 وحجم الأسرة x_2 . أوجد معادلة الانحدار الخطى المقدرة وذلك بفرض توفر البيانات المعطاة في جدول (١٠-١٧) .

الحل . من البيانات في جدول (١٠-١٧) فإن :

$$n = 10 \quad , \quad \Sigma x_{1i} = 60 \quad , \quad \Sigma x_{2i} = 40 \quad ,$$
$$\Sigma x_{1i}^2 = 406 \quad , \quad \Sigma x_{1i} x_{2i} = 269 \quad \Sigma x_{2i}^2 = 182,$$
$$\Sigma y_i = 180 \quad , \quad \Sigma x_{1i} y_i = 1159 \quad \Sigma x_{2i} y_i = 766, \Sigma y_i^2 = 3396.$$

بالتعويض بالقيم السابقة في المعادلات الثلاثة نحصل على :

$$10\, b_0 + 60\, b_1 + 40\, b_2 = 180$$
$$60\, b_0 + 406\, b_1 + 269\, b_2 = 1159$$
$$40\, b_0 + 269\, b_1 + 182\, b_2 = 766.$$

الحل . لهذه الفئة من المعادلات نحصل على :

$$b_0 = 7.918 \ , \ b_1 = 2.363 \ , \ b_2 = -1.024$$

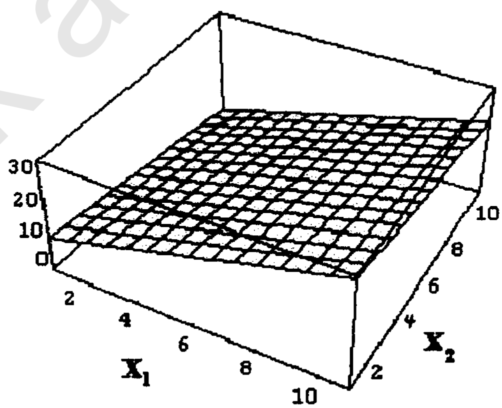
جدول (١٠-١٧)

الأســـرة	الاستهلاك السنوي للطعام (بمئات الدولارات)	الدخل السنوي الصافي (بمئات الدولارات)	حجم الأسرة (عدد الأفراد في الأسرة الواحدة)
1	22	8	6
2	23	10	7
3	18	7	5
4	9	2	2
5	14	4	3
6	20	6	4
7	21	7	4
8	18	6	3
9	16	4	3
10	19	6	3

وعلى ذلك فإن نموذج الانحدار الخطى المقدر يمكن كتابته على الشكل :

$$\hat{y}_{x_1,x_2} = 7.918 + 2.363x_1 - 1.024x_2$$

والتمثيل البياني لها موضح في شكل (١٠-١٦) .



شكل (١٠-١٦)

(١٠-٥-٢) تحليل الانحدار **Analysis of Variance**

عادة ، الخطوة الأولى بعد الحصول على معادلة الانحدار المتعدد هو اختبار ما إذا كانت هناك علاقة بين المتغير التابع وفئة المتغيرات المستقلة. يعتبر اختبار F في الانحدار المتعدد تعميم لاختبار F في حالة الانحدار الخطى البسيط في حالة متغيرين مستقلين . فرض العدم سوف يكون :

$$H_0 : \beta_1 = 0, \beta_2 = 0$$

ضد الفرض البديل : على الأقل واحد من $\beta_i \neq 0$ و $H_1 : i = 1,2$

يعطى جدول (١٠-١٨) مجاميع المربعات ودرجات الحرية المقابلة لها .

جدول (١٠ – ١٨)

درجات الحرية	مجموع المربعات
n-1	$SSTO = \Sigma(y_i - \bar{y})^2$
n-3	$SSE = \Sigma(y_i - \hat{y}_i)^2$
2	$SSR = \Sigma(\hat{y}_i - \bar{y})^2$

حيث : $\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i}$

جدول تحليل التباين معطى في جدول (١٠-١٩).

جدول (١٠-١٩)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات
الانحدار	2	SSR	$MSR = \frac{SSR}{2}$
الخطأ	n-3	SSE	$MSE = \frac{SSE}{n-3}$
الكلى	n-1	SSTO	

بافتراض صحة فرض العدم فإن :

$$f = \frac{MSR}{MSE}$$

تمثل قيمة لمتغير عشوائي F يتبع توزيع F بدرجات حرية $v_2 = n - 3$, $v_1 = 2$. لمستوى معنوية α فإن منطقة الرفض $F > f_\alpha(v_1, v_2)$ حيث $f_\alpha(v_1, v_2)$ تستخرج من الجدول في ملحق (٦) بدرجات حرية v_1, v_2 . إذا وقعت f في منطقة الرفض نرفض H_0 .

للبيانات في مثال (١٠-١٧) فرض العدم سوف يكون :

$$H_0 : \beta_1 = 0, \beta_2 = 0$$

ضد الفرض البديل : على الأقل واحد من $0 \neq \beta_i$ و $H_1 : i = 1,2$

يمكن وضع فرض العدم والفرض البديل على الشكل :

H_0 : الانحدار غير معنوي .

H_1 : الانحدار معنوي .

الحل . من البيانات في جدول (١٠-١٧) فإن جدول تحليل التباين معطى في جدول (١٠-٢٠) .

جدول (١٠-٢٠)

مصدر الاختلاف	درجات الحرية	مجموع المربعات	متوسط مجموع المربعات
الانحدار الخطأ	2 7	139.56725 16.43275	69.78363 2.34754
الكلي	9	156	

$$f=\frac{69.78363}{2.34754}=29.7263.$$

$f_{0.05}(2,7)=4.74$ والمستخرجة من جدول توزيع F في ملحق (٦) بدرجات حرية $v_1 = 2$, $v_2 = 7$. وبما أن f تقع في منطقة الرفض نرفض H_0 .

يعتبر MSE تقدير للتباين σ^2. وعلى ذلك فإن التقدير بنقط للمعلمة σ^2 هو $s^2=2.34754$. وعلى ذلك فإن التقدير بنقطة للانحراف المعياري σ هو :

$$s = \sqrt{2.34754} = 1.5321684.$$

(١٠-٥-٣) معامل التحديد المتعدد Coefficient of Multiple Determination

معامل التحديد المتعدد ، يرمز له بالرمز R^2 هو :

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}.$$

في حالة وجود متغير مستقل واحد فإن R^2 يصبح معامل التحديد البسيط r^2 . يتراوح قيمة R^2 من الصفر إلى الواحد الصحيح ، أي أن :

$$0 \leq R^2 \leq 1$$

عندما $R^2 = 0$ فهذا يعنى أن $b_1 = b_2 = 0$ وعندما $R^2 = 1$ فهذا يعنى أن جميع القيم المشاهدة y_i تقع على المستوى المقدر.

(١٠-٥-٤) الارتباط المتعدد والجزئي Multiple and Partial Correlation

الجذر التربيعي الموجب لمعامل التحديد المتعدد R^2 يسمى الارتباط المتعدد ويرمز له بالرمز R ، حيث :

$$R = +\sqrt{R^2}$$

للمثال (١٠-١٧) فإن :

$$R = +\sqrt{0.89466} = 0.94586$$

عموما يقيس معامل التحديد المتعدد R العلاقة بين المتغير التابع Y والمتغيرات المستقلة كلها مجتمعة. بفرض أننا نرغب في إيجاد العلاقة بين المتغير التابع Y وأحد العوامل فقط (بفرض أن العوامل الأخرى ثابتة (أي بحذف تأثير المتغيرات الأخرى) ، وهنا نستخدم معامل الارتباط الجزئي **coefficient of partial correlation** .

فإذا كنا نرغب في إيجاد معامل الارتباط الجزئي بين المتغير Y والمتغير X_1 مع استبعاد أثر المتغير X_2 فإن معامل الارتباط الجزئي ، يرمز له بالرمز $r_{y1.2}$ هو :

$$r_{y1.2} = \frac{r_{y1} - r_{y2}\ r_{12}}{\sqrt{1-r_{y2}^2}\sqrt{1-r_{12}^2}}$$

حيث r_{y1} هو معامل الارتباط البسيط بين المتغير Y والمتغير X_1 وبالمثل يكون y_{y2} و r_{12}.

أيضا معامل الارتباط الجزئي بين المتغير Y والمتغير X_2 مع استبعاد أثر المتغير X_1 ، يرمز له بالرمز $r_{y2.1}$ ، هو :

$$r_{y2.1} = \frac{r_{y2} - r_{y1}\ r_{12}}{\sqrt{1-r_{y1}^2}\sqrt{1-r_{12}^2}}$$

معامل الارتباط الجزئي يمكن أن يكون موجباً أو سالباً حيث تقع قيمته في الفترة [1 , 1-] .

لحساب معاملات الارتباط الجزئية من مثال (١٠-١٧) نقوم أولاً بحساب معاملات الارتباط البسيطة التالية :

$$r_{y1} = \frac{\Sigma\ x_{1i}\ y_i - \dfrac{\Sigma\ x_{1i}\ \Sigma\ y_i}{n}}{\sqrt{\left[\Sigma\ x_{1i}{}^2 - \dfrac{(\Sigma\ x_{1i})^2}{n}\right]\left[\Sigma\ y_i^2 - \dfrac{(\Sigma\ y_i)^2}{n}\right]}}$$

$$= \frac{1159 - \frac{(60)(180)}{10}}{\sqrt{\left[406 - \frac{(60)^2}{10}\right]\left[3396 - \frac{(180)^2}{10}\right]}} = 0.9325795,$$

$$r_{y2} = \frac{\Sigma\ x_{2i}\ y_i - \frac{\Sigma\ x_{2i}\ \Sigma\ y_i}{n}}{\sqrt{\left[\Sigma\ x_{2i}{}^2 - \frac{(\Sigma\ x_{2i})^2}{n}\right]\left[\Sigma\ y_i^2 - \frac{(\Sigma\ y_i)^2}{n}\right]}}$$

$$= \frac{766 - \frac{(40)(180)}{10}}{\sqrt{\left[182 - \frac{(40)^2}{10}\right]\left[3396 - \frac{(180)^2}{10}\right]}} = 0.785207,$$

$$r_{12} = \frac{\Sigma\ x_{1i}\ x_{2i} - \frac{\Sigma\ x_{1i}\ \Sigma\ x_{2i}}{n}}{\sqrt{\left[\Sigma\ x_{1i}{}^2 - \frac{(\Sigma\ x_{1i})^2}{n}\right]\left[\Sigma\ x_{2i}^2 - \frac{(\Sigma\ x_{2i})^2}{n}\right]}}$$

$$= \frac{269 - \frac{(60)(40)}{10}}{\sqrt{\left[406 - \frac{(60)^2}{10}\right]\left[182 - \frac{(40)^2}{10}\right]}} = 0.9116072.$$

وعلى ذلك فإن :

$$r_{y1.2} = \frac{r_{y1} - r_{y2}\ r_{12}}{\sqrt{1 - r_{y2}^2}\sqrt{1 - r_{12}^2}}$$

$$= \frac{0.9325795 - (0.785207)(0.9116072)}{\sqrt{(1-(0.7895207)^2}\sqrt{1-(0.9116072)^2}} = 0.859283884.5,$$

$$r_{y2.1} = \frac{r_{y2} - r_{y1}\ r_{12}}{\sqrt{1-r_{y1}^2}\sqrt{1-r_{21}^2}}$$

$$= \frac{0.785207 - (0.9325795)(0.9116072)}{\sqrt{(1-(0.9325795)^2}\sqrt{1-(0.9116072)^2}} = -0.4376576.$$

(١٠-٦) الانحدار من الدرجة الثانية Quadratic Regression

في بعض الأحيان تكون العلاقة بين متغيرين على شكل منحنى من الدرجة الثانية فعلى سبيل المثال بفرض أننا نرغب في تقدير معالم النموذج :

$$\mu_{Y|x} = \beta_0 + \beta_1 x + \beta_2 x^2$$

في الحقيقة نرغب في تقدير معالم النموذج انحدار خطى متعدد على الشكل :

$$\mu_{Y|x} = b_0 + b_1 x_1 + b_2 x_2$$

وذلك بوضع $x_1 = x$, $x_2 = x^2$ في المعادلة السابقة .

مثال (١٠-١٨) لازواج القياسات المعطاة في جدول (١٠ – ٢١) أوجد الانحدار المقدرة لنموذج انحدار من الدرجة الثانية .

جدول (١٠-٢١)

x	1	2	3	4	5	6	7	8	9	10
y	5.0	10.20	15.35	20.50	25.95	32.20	38.50	46.00	53.80	62.00

الحل . من البيانات في جدول (١٠ -٢١) فإن :

$$n = 10\ ,\ \Sigma x_i = 55\ ,\ \Sigma y_i = 309.5\ ,$$

$$\Sigma x_i y_i = 2218.1\ ,\ \Sigma x_i^2 = 385,$$

$$\Sigma x_i^2 y_i = 17708.2\ ,\ \Sigma x_i^3 = 3025\ ,\ \bar{y} = 30.95,$$

$$\Sigma x_i^4 = 25333\ ,\ \Sigma y_i^2 = 12831.845\ ,$$

تستخدم القيم السابقة في الحصول على المعادلات التالية :

$$10\,b_0 + 55\,b_1 + 385\,b_2 = 309.5$$
$$55\,b_0 + 385\,b_1 + 3025\,b_2 = 2218.1$$
$$385\,b_0 + 3025\,b_1 + 25333\,b_2 = 17708.2$$

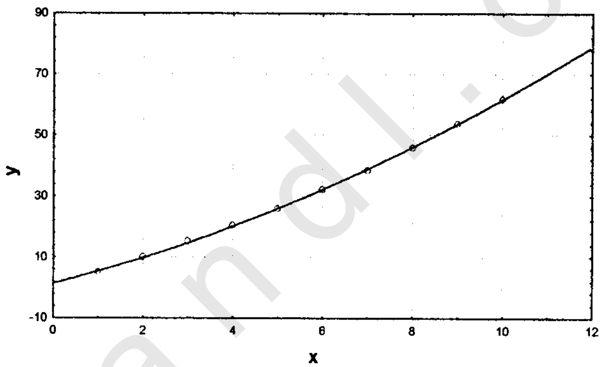
بحل المعادلات السابقة آنيا يمكن إيجاد b_0, b_1, b_2 .

الحل لهذه الفئة من المعادلات هو $b_0 = 1.48083$ و $b_1 = 3.792313$ و $b_2 = 0.223674$

معادلة الانحدار المقدرة هي :

$$\hat{y}_x = 1.48083 + 3.792313x + 0.223674x^2.$$

والتمثيل البياني لها موضح في شكل (١٠- ١٧) .



شكل (١٠- ١٧)

معامل الارتباط من الدرجة الثانية **Second – Degree Correlation Coefficient**

عندما تكون العلاقة بين متغيرين على شكل معادلة من الدرجة الثانية ، بمعنى أن خط الانحدار يكون على شكل منحني (أي غير مستقيم) ، يقال في هذه الحالة أن الارتباط غير مستقيم وفي هذه الحالة لا يصلح قياس الارتباط بمعامل الارتباط الخطي البسيط r وذلك لعدم استقامة الارتباط بل يستخدم المقياس التالي :

$$r = \sqrt{\frac{b_0\Sigma y_i + b_1\Sigma x_i y_i + b_2\Sigma x_i^2 y_i - n\bar{y}^2}{\Sigma y_i^2 - n\bar{y}^2}}$$

$$= \sqrt{\frac{1.4808(309.5) + 3.7923(2218.1) + 0.22367(17708.2) - 10(30.95)^2}{12831.845 - 10(30.95)^2}}$$

$$= \sqrt{\frac{3251.7763}{3252.82}} \quad = \quad \sqrt{0.9997}$$

$= 0.99984.$

أيضا معامل التحديد مازال r^2 يساوى $r^2 = 0.9997$:

تمــــارين :

-١- أجريت تجربة لدراسة تأثير درجة الحرارة X على نتائج إحدى العمليـــات الكيميائيـــة وتم الحصول على البيانات التالية (في شكل شفرة Coded) .

x	-5	-4	-3	-1	-1	0	1	2	3	4	5
y	1	5	4	7	10	8	9	13	14	13	18

أوجد نموذج الانحدار الخطى المقدر .

-٢- تعتبر كمية الرطوبة في منتج ما لها تأثير على كثافة المنتج النهائية. تم مراقبة المنتج وقيـــاس كثافته وتسجيل البيانات التالية (في شكل شفرة Coded) .

رطوبـــة المنتج	x	4.7	5.0	5.2	5.2	5.9	4.7	5.9	5.2	5.9	5.6	5.0
كثافـــة المنتج.	y	3	3	4	10	2	9	3	7	6	6	4

(أ) قدر معالم نموذج الانحدار الخطى البسيط .

(ب) أوجد 95% فترة ثقة للمعلمة β_0 .

(ج) أوجد 95% فترة ثقة للمعلمة β_1.

(د) هل تعتقد أن المعادلة الخطية مناسبة لتوفيق هذه البيانات ؟

-٣- إذا كانت تكاليف صيانة سيارات الشحن تزيد مع عمر السيارة . اســـتخدام البيانـــات التالية في (أ) تقدير معالم نموذج الانحدار الخطي البسيط (ب) هل يعتبر النمـــوذج الخطـــي هـــو الأفضل لتوفيق العلاقة في هذا المثال ؟

العمـــر بالسنوات	x	4.5	4.5	4.5	4.0	4.0	5.0	5.0	5.5	5.0	5.0	0.5	6.0	6.0
التكـاليف خـــلال ٦ شهور	y	619	1049	1033	495	723	681	890	1522	987	1194	163	182	764

-٤- في دراسة عن تأثير درجة الحرارة ، في عملية decolorizing على لون المنتج النــهائي تم الحصول على البيانات التالية :

درجـــة الحرارة	x	460	450	440	430	420	410	450	440	430

←

اللون	y	0.3	0.3	0.4	0.6	0.5	0.6	0.6	0.5	0.5

(أ) أوجد معادلة نموذج الانحدار الخطى المقدرة.

(ب) أختبر معنوية كل من β_0, β_1 .

(ج) أوجد 95% فترة ثقة لكل من β_0, β_1.

-٥- في أحد أماكن بيع السيارات كانت المبيعات كالتالي :

عمر السيارة	x	3	2	1	1	5	6	1	4
ثمن البيع	y	31	10	59	68	16	15	69	28

(أ) أوجد معادلة الانحدار الخطى المقدرة .

(ب) أوجد 95% فترة ثقة لكل من β_0, β_1.

-٦- إذا كان معروف أن هناك علاقة بين فترة الإصابة بمرض معين وعدد البكتريا في العضو المصاب . تم اختيار 10 مصابا بهذا المرض وسجلت أطوال فترات إصابتهم بالمرض عند بدا دخولهم المستشفى فتم الحصول على البيانات التالية :

عدد البكتريا (بالألف)	x	9	10	5	7	10	6	7	4	8	6
فترة الإصابة (باليوم)	y	12	11	8	9	13	10	14	8	11	7

أوجد معادلة الانحدار المقدرة ؟

-٧- فيما يلي أوزان وأطوال مجموعة من الذكور البالغين والمطلوب إيجاد معادلة خط الانحدار المقدرة .

الطول	x	159	180	175	150	170	171	165	176
الوزن	y	68	88	79	65	70	73	63	74

-٨- الجدول التالي يمثل الدخل والأنفاق لعينة من الأسر .

الدخل بمئات الجنيهات	x	42	65	41	43	37	26	38	39
الإنفاق بمئات الجنيهات	y	25	37	25	21	18	24	19	26

(أ) أوجد معادلة الانحدار المقدرة.

(ت) أختبر معنوية β_0 عند مستوى معنوية 0.05.

(ج) أختبر معنوية β_1 عند مستوى معنوية 0.05.

(د) أوجد 95% فترة ثقة لكل من β_0, β_1.

-٩- الجدول التالي يوضح السن وضغط الدم لعشرة من الإناث .

السن	x	41	35	62	52	41	58	48	67	66	68
ضغط الدم	y	124	115	138	149	145	144	145	152	150	150

(أ) أوجد معادلة الانحدار المقدرة.

(ب) أوجد 99% فترة ثقة لكل من $\beta_0, \beta_1, y_{69}, \mu_{Y|170}$

-١٠- أوضحت الدراسة أن عدد العلب الصفيح التي تتعرض للتلف في عربة الشحن دالـــة في سرعة السيارة. استخدم البيانات التالية في إيجاد :

x ســرعة السيارة	4	3	5	8	4	3	3	4	3	5	7	3	8
y عــدد العلب التالفة	63	54	86	36	65	69	28	75	53	33	168	47	52

(أ) معادلة الانحدار الخطى المقدرة.

(ب) 95% فترة ثقة لكل من β_0, β_1

-١١- قام مركز تجاري بدراسة العلاقة بين تكاليف الإعلانات الأسبوعية و المبيعـــات وقـــد تم تسجيل البيانات والجدول التالي :

x التكاليف	3	5	8	4	3	3	4	3	5	7	3	8
y المبيعات $	63	54	86	36	65	109	28	75	53	168	47	52

(أ) أوجد معادلة الانحدار الخطي المقدرة؟

(ب) أوجد تقدير للتباين σ^2؟

ج أوجد 95% فترة ثقة للمعلمة β_1 ؟

(د) أختبر فرض العدم $H_0 : \beta_1 = 0$ ضد الفرض البديل $H_1 : \beta_1 \neq 0$ وذلـــك عنـــد مستوى معنوية 0.05.

-١٢- أجريت دراسة في نهر ما على عينة الثلج في الماء و تم الحصول على البيانات التالية :

x كمية الثلج في أبريل	23.1	31.8	31.7	31.0	23.0	34.1	24.1	51.1
y كمية المياه من أبريـــل إلى مايو	10.2	16.2	18.1	16.9	16.2	10.4	23.0	24.9

←

x كمية الثلج في أبريل	36.9	30.4	25.1	12.3	35.0	31.4	21.0	27.5
y كمية المياه من أبريل إلى مايو	22.7	14.0	12.8	8.7	17.3	14.8	10.4	15.1

(أ) أوجد معادلة الانحدار الخطي المقدرة ؟

(ب) أوجد تقدير للتباين σ^2 و 95% فترة ثقة للمعلمة β_0, β_1 ؟

-١٣- البيانات التالية تمثل عدد السيارات الخاصة والتي لديها رخصة سارية خلال التسع سنوات الأخيرة في بلد ما. وقد تم تسجيل البيانات من قبل شركة لبيع السيارات وذلك للتعرف على مبيعاتها من السيارات .

x السنة	1	2	3	4	5	6	7	8	9
y عدد السيارات	8247	8919	4513	10303	10816	11228	11515	112059	12717

(أ) أوجد معادلة خط الانحدار المقدرة ؟

(ب) أوجد عدد السيارات الخاصة برخصة سارية في العام ؟

-١٤- يحتوى الجدول التالي على كمية مركب كيميائي والذي يذوب في 100 جرام من الماء عند درجات حرارة مختلفة .

C^o x	x جرام		
0	7	5	7
15	11	9	13
30	24	20	23
45	30	32	25
60	43	38	41
75	47	50	43

(أ) أوجد معادلة الانحدار الخطي المقدرة ؟

(ب) أوجد كمية المركب الكيميائي والذي يذوب في 100 جرام من الماء عند $50c^o$ ؟

(ج) أختبر فرض العدم H_0 : العلاقة خطية ضد الفرض البديل H_1 : العلاقة غير خطية ؟

-١٥- قام باحث بدراسة العلاقة بين الضغط (المتغير المستقل ، Kg/mm^3)، والزمن اللازم لقطع ألواح من الصلب القاتل للصدأ (المتغير التابع ، بالساعات) في وسط ما والبيانات في الجدول التالي :

i	1	2	3	4	5	6	7	8	9	10
الضغط x_i	2.5	5	10	15	175	20	25	30	35	40
زمن القطع y_i	63	88	55	61	62	37	38	45	46	19

(أ) أوجد شكل الانتشار ؟

(ب) أوجد معادلة خط الانحدار المقدرة ؟

(ج) أختبر معنوية كل من β_0, β_1.

- ١٦- من البيانات التالية أرسم شكل الانتشار وأوجد معادلة الانحدار الخطي المقدرة .

x حجم المبيعات (بالألف دولار) لسلعة ما	5	6	7	8	9	10
y السعر (بالألف دولار)	74	77	82	86	92	95

وما توقعك عن السعر عندما يكون حجم المبيعات 7500؟

-١٧- البيانات التالية تم تسجيلها خلال 8 فترات .

الفتـــرة	الوحدات المصدرة x	التكاليف الكلية y
1	10000	32000
2	20000	39000
3	30000	58000
4	40000	52000
5	50000	61000
6	60000	70000
7	7000	64000
8	90000	66000

(أ) أرسم شكل الانتشار ؟

(ب) أوجد خط الانحدار المقدر ؟

(ج) أوجد التكاليف الكلية عندمـــا يكـــون عـــدد الوحـــدات المصـــدرة , 48750 12600وحده.

-١٨- إذا كان لديك البيانات التالية :

$$n = 12 \ , \ \Sigma x_i y_i = 2000 \ , \ \bar{y} = 19 \ , \ \Sigma y_i^2 = 4612,$$
$$\bar{x} = 8 \ , \ \Sigma x_i^2 = 1502$$

(أ) أوجد معادلة الانحدار المقدرة ؟

(ب) أختبر الانحدار عند α=0.05 ؟

-١٩- إذا كان لديك البيانات التالية :

$$n = 25 \ , \ \Sigma(y - \bar{y})^2 = 500 \ , \ \Sigma(y_i - \hat{y}) = 125,$$
$$\Sigma(\hat{y}_i - \bar{y})^2 = 375$$

أوجد تقدير للتباين σ^2 ومعامل التحديد؟

-٢٠- إذا كان لديك البيانات التالية :

$$n = 100 \quad , \quad \Sigma(y_i - \bar{y})^2 = 8000 \quad , \quad \Sigma(\hat{y}_i - \bar{y})^2 = 7000,$$

قدر معامل الارتباط مع العلم أن إشارة β_1 موجبة ؟

-٢١- أخذت عينة عشوائية من 10 موظفين وتم الحصول على البيانات التالية :

x عدد سنوات الدراسة	12.5	14	12	16	11.0	16.0	8.0	11.2	15.0	12.7
y الأجر بالآلف دولار	20.2	13.7	19.8	12.5	21.8	35.7	39.2	22.6	25.9	17.3

(أ) ارسم شكل الانتشار ؟

(ب) قدر معادلة الانحدار الخطى

(ج) أختبر فرض العدم $H_0: \beta_1 = 0$ ضد الفرض البديل $H_1: \beta_1 \neq 0$ عند مستوى معنوية $\alpha=0.05$.

-٢٢- قامت شركة بإجراء اختبار لـ 500 موظف جديد فإذا كانت x تمثل الدرجة التي حصل عليها الموظف الجديد. وبعد ثلاثة سنوات أعطيت درجات على الأداء في العمل y الدرجة التي حصل عليها . فإذا كان لديك البيانات التالية :

$$\bar{x} = 100 \quad , \quad s_x = 10 \quad , \quad \bar{y} = 130 \quad , \quad s_y = 20 \quad , \quad r = 0.7$$

حيث S_x , S_y تمثل الانحراف المعياري من العينة لقيم x , y على التوالي :

(أ) أوجد معادلة الانحدار المقدرة ؟

(ب) أوجد قيمة y عندما $x = 90$ و $x = 125$ ؟

-٢٣- قام باحث بالحصول على بيانات من 1000 مريض نفسي وذلك لمدة 5 سنوات ، فإذا كان x يمثل الدرجة التي حصل عليها الشخص في بداية العلاج ، y الدرجة التي حصل عليها بعد تلقى العلاج باستخدام البيانات التالية :

$$\Sigma x_i = 3000 \quad , \quad \Sigma y_i = 5000 \quad , \quad \Sigma x_i y_i = 3000$$

$$\Sigma x_i^2 = 14000.$$

(أ) أوجد معادلة الانحدار المقدرة؟

(ب) أوجد قيمة y عندما $x=4$؟

(ت) إذا كانت قيمة $s_y=10$ أوجد معامل الارتباط r.

-٢٤- للبيانات التالية :

$$\bar{x} = 200 \;,\; \bar{y} = 90 \;,\; r = -0.9 \;,\; s_x = 9 \;,\; s_y = 5$$

أوجد معادلة الانحدار المقدرة وتنبأ بقيمة y عندما x = 164 .

-٢٥- في مكتب للشرطة تم إجراء دراسة للعلاقة بين عدد الجرائم في اليوم وأعلى درجة حرارة في اليوم . اختيرت عينة عشوائية في 10 أيام وتم الحصول على البيانات التالية .

x أعلى درجة حرارة	12	89	40	52	75	60	50	20	32	90
y عدد الجرائم	2	15	6	8	14	12	7	3	5	16

(أ) أوجد معادلة الانحدار المقدرة؟

(ب) أختبر فرض العدم $H_0: \beta_1 = 0$ ضد الفرض البديل $H_1 : \beta_1 \neq 0$ عند مستوى معنوية α=0.05 .

-٢٦- الجدول التالي يعطى العمر لأحد النباتات (بالأسابيع) وطوله بالسنتيمتر .

العمر بالأسبوع x	1	2	3	4	5	6	7
الطول بالسنتيمتر y	5	13	16	33	23	38	40

أوجد معادلة خط الانحدار المقدرة ؟

-٢٧-أجريت تجربة لتقدير العلاقة بين درجة الحرارة ومعدل دقات القلب في الضفدعة المسماة Rana Pipiens والبيانات معطاة في الجدول التالي :

الحيوان	1	2	3	4	5	6	7	8	9
x درجة الحرارة	2	4	6	8	10	12	14	16	18
y دقات القلب بالدقيقة	5	11	11	14	22	23	32	29	32

(أ)أوجد معادلة الانحدار المقدرة؟

(ب) أختبر فرض العدم $H_0: \beta_1 = 0$ ضد الفرض البديل $H_1 : \beta_1 \neq 0$ عند مستوى

معنوية $\alpha=0.05$.

-٢٨- أجريت تجربة لتقدير العلاقة بين العمر ودقات القلب (في الدقيقة) في الإنـــاث الذيـــن أعمارهم تتراوح من واحد إلى 15سنة . استخدم البيانات المعطاة في الجدول التالي في إيجاد :

(أ) معادلة الانحدار المقدرة؟

(ب) أختبر معنوية معادلة الانحدار؟

(ج) أوجد 95% فترة ثقة لكل من B_1 , B_0 .

الأنثى	1	2	3	4	5	6	7	8	9	10
العمر x	1	2	3	4	5	6	7	8	9	10
دقـــات القلب y	111	108	108	102	99	92	93	88	90	90
الأنثى	11	12	13	14	15					
العمر x	11	12	13	14	15					
دقـــات القلب y	88	84	83	83	82					

-٢٩- يعطى الجدول التالي عدد سنوات الخبرة للأستاذ في جامعة ما والدخل السنوي بـــالدولار (بالآلاف):

السنوات x	5	10	15	20	25	30
الدخل y	59.0	69.0	78.0	88.0	97.5	107.0

(أ) أختبر العلاقة بين سنوات الخبرة والدخل السنوي بيانيا؟

(ب) اقترح شكل العلاقة بين المتغيرين وقدر معالم النموذج؟

-٣٠- يعطى الجدول التالي أعمار الزوج والزوجة بالسنوات لعينة من 6 أزواج:

عمر x الزوجة	35	25	51	25	53	42
عمر y الزوج	38	25	49	31	55	44

أوجد معامل الارتباط .

-٣١- قام باحث بدراسة العلاقة بين ضغط الغاز (مقـــاس milliliters) ودرجـــة الحـــرارة (مقاسه k^0) وتم الحصول على البيانات التالية :

x درجة الحرارة	200	250	300	350	200	250	300	350

←

الضغط y	251	315	374	440	241	302	362	423

ارسم شكل الانتشار وهل تعتقد أن النموذج الخطى مناسب لتوفيق البيانات السابقة ؟

-٣٢- الجدول التالي يمثل درجات امتحان أعمال السنة ودرجات الامتحان النهائي لعينة عشوائية من 8 طلبة .

x درجة أعمال السنة	75	49	70	71	80	93	95	98
y درجة الامتحان النهائي	80	65	77	33	40	84	98	98

(أ) أوجد معادلة الانحدار المقدرة؟

(ب) أوجد الدرجة النهائية لطالب إذا كانت درجته في أعمال السنة 83.

-٣٣- تعمل آلة عند سر عات مختلفة ولكن السرعة العالية تؤثر على عمر يد المثقاب البيانات التالية تمثل أعمار يد المثقاب عند دورات مختلفة من الآلة.

xعدد الدورات في الدقيقة	18	20	43	20	23	26	26	28	31	32	32	40	41	42
y عمر يد المثقاب	162	54	69	171	162	138	140	129	125	106	97	95	105	109

(أ) ارسم شكل الانتشار ؟

(ب) قدر عمر يد المثقاب عندما تكون عدد دورات الآلة 30 دورة في الدقيقة ؟

-٣٤- أعطى اختبار في المعلومات العامة يتكون من 100 سؤال إلى 20 طالب في أعمار مختلفة وكانت النتائج كالتالي :

الطالب	العمر		عدد الإجابات الصحيحة
	سنوات	شهر	
A	16	8	40
B	16	2	45
C	17	9	47
D	17	1	46
E	18	6	67
F	18	7	45
G	19	10	53
H	20	4	54

المطلوب (أ) رسم شكل الانتشار ؟

(ب)إيجاد معادلة الانحدار الخطى المقدرة ؟

(ج) أوجد باستخدام خط الانحدار المقدر عدد الإجابات الصحيحة إذا كان عمر الطالب 17 سنة .

-٣٥- أجرى ثلاثة باحثين في الجيولوجيا في القطب الشمالي بحث أستمر خمسة فصول شتوية وذلك لدراسة العلاقة بين عدد الأيام التي تنخفض فيها درجة الحرارة عن 50^oF وطول القرون لحيوان الموط (حيوان ضخم من الأيائل). أوجد معامل الارتباط الذي يعتمد على هذه البيانات .

x عدد الأيام التي تنخفض درجة الحرارة عن 50^oF	30	20	10	30	10
متوسط الطول لقروق الحيوان (بالمتر)	0.9	0.8	0.5	1.0	0.8

-٣٦- أحسب معامل الارتباط الخطى بين الفئات التالية من الأرقام

درجات اختبار ذكاء	97	121	84	105	93	126	109	97	112	116	86	103
الدخل الأسبوعي عند العمر 23	20	16	19	18	18	17	22	13	14	9	20	21

-٣٧- الجدول الآتي يبين طول الجمجمة X وعرضها Y بالمليمتر والمطلوب إيجاد معامل الارتباط الخطي .

x الطول	63	80	70	76	66	79	73	72	58	71
y العرض	40	42	45	38	39	46	42	37	39	35

-٣٨- الجدول التالي يبين المبيعات اليوم بالجنية X لعشرة عمال في متجر ومدة خدمتهم Yبالسنين أوجد معامل الارتباط الخطى .

x	9	10	12	9	10	9	6	5	4	5
y	7	10	11	10	4	8	9	4	2	5

-٣٩- الجدول التالي يعرض عدد العدسات التي تنتجها إحدى المصانع وتكلفة العدسة الواحدة بالدولار .

xعدد العدسات	1	3	5	10	12
yتكلفة العدسة	20	15	10	7	5

أوجـــد معامل الارتباط؟

٤٠- اختيرت عينة عشوائية من 12 ورقة من شجرة ما وتم قياس الطول والعرض لكل ورقة إلى أقرب ملليمتر . البيانات معطاة في الجدول التالي :

الورقة	1	2	3	4	5	6	7	8	9	10	11	12
العرض	35	21	25	35	26	40	35	40	35	42	23	25
الطول	55	44	46	60	55	57	64	68	5	61	46	44

(أ) أوجد معامل الارتباط ؟

(ب) أختبر معنوية معامل الارتباط المجتمع ρ؟

أوجـــد معامل الارتباط؟

(ت) أختبر معنوية معامل الارتباط المجتمع ρ؟

٤١- الجدول التالي يبين درجات مجموعة مكونة من 10 طلاب في كل من مـادتي الإحصاء والرياضيات في أحد الامتحانات لأعمال الفصليين .

x الإحصاء	17	13	8	17	14	10	7	17	8	9
y الرياضيات	18	14	6	16	14	9	10	13	7	8

(أ) أوجد معامل الارتباط ؟

(ب) أوجد 95% فترة ثقة للمعلمة ρ ؟

٤٢- الجدول التالي يعطى أطوال الأب X وأطوال ابنه الأصغر عند بلوغه سن معين Y (البيانات مقاسه لاقرب بوصة).

x	68	64	70	72	69	74
y	67	68	69	73	66	70

أوجد معامل الارتباط الخطي البسيط ؟

٤٣- لدراسة العلاقة بين عدد الكيلوجرامات التي يفقدها شخص في برنامج لإنقاص الـــوزن وعدد الأسابيع التي يقضيها لإنقاص الوزن ، اختيرت عينة عشوائية من 5 أشخاص ممن يتبعـون هذا البرنامج الغذائي وتم الحصول علي البيانات التالية :

x	3	2	1	4	5
y	6	5	4	9	11

أوجد معامل الارتباط الخطي ؟

٤٤- لدراسة العلاقة بين المدة اللازمة لتدريب العامل في مصنع ما وعمر العامل تم الحصـول علي البيانات التالية :

x العمر (بالسنوات)	44	38	42	24	34	23	27	21	20	18
y المدة اللازمة للتدريب	8	6	10	8	11	7	8	6	5	8

أحسب معامل الارتباط الخطي البسيط ؟

٤٥- البيانات التالية تعطي متوسط درجة الحرارة اليومية وكمية الكهرباء المستهلكة وذلك خلال ثمانية أيام اختيرت في سنة ما :

x الحرارة (F^0)	48	42	44	40	40	35	32	37
y استهلاك الكهرباء Megawatt Hours	3.3	3.4	3.4	3.7	3.6	3.7	3.8	3.7

أحسب معامل الارتباط الخطي البسيط ؟

٤٦- لدراسة العلاقة بين عدد الموظفين في شركة للخدمات وعدد الطلبات المتدفقة على الشركة تم الحصول على البيانات التالية :

عدد الموظفين	15	12	8	5	3	2	1
عدد الطلبات	40	35	30	25	20	15	10

(أ) أوجد نموذج الانحدار الأسى المقدر ؟

(ب) أوجد عدد الموظفين عندما تكون عدد الطلبات 45 ؟

٤٧-لدراسة العلاقة بين النسبة المئوية للعاطلين عن العمل والتغير في الأجور خلال عشر سنوات تم الحصول على البيانات التالية :

	10	9	8	7	6	5	4	3	2	1
النسبة % x العاطلين عن العمل	11.6	11.5	11.7	12.6	12.1	11.6	11.7	12.5	12.2	11.6
y المتغير في الأجور	4.4	3.5	4.6	2.9	2.7	4.1	2.1	2.7	3.2	5.0

أوجد معامل الارتباط الخطي البسيط ؟

٤٨- أجريت تجربة لاختبار تأثير نوع معين من العقاقير على سريان الادرينا لين في حيوان مـا. إذا كان Y يمثل عدد المليجرامات من العقار اللازمة لإحداث سريان من الأدرينالين مقـداره X وإذا كانت العلاقة التي تربط بين X,Y يمثلها معادلة على الشكل $\mu_{Y|x} = b\,a^{x}$ حيث a , b ثابتان. أوجد تقديرات لمعلمتين a , b إذا كان لديك البيانات التالية :

x	1	2	3	4	5	6
y	3.1	6.2	11.3	22.0	48.0	92.0

٤٩- أجريت تجربة على أحد الأنواع الجديدة من السيارات لتحديد مسافة التوقف للسـيارة عند سر عات مختلفة ، وكانت نتائج التجربة كما يلي :

السرعة x بالميل/ساعة	20	30	40	50	60	70
مسافة y التوقف بالقدم	50	85	130	200	280	290

إذا كانت العلاقة بين السرعة ومسافة التوقف هي على الشكل :

$$\mu_{Y|x} = a_o x^{b_o}$$

استخدم البيانات السابقة في تقدير معالم النموذج ثم قدر مسافة التوقف لسيارة سرعتها 53 ميلا في الساعة .

٥٠- البيانات التالية تمثل أعداد السكان بالمليون في مدينة ما خلال الفترة من 1980-1972.

فإذا كانت العلاقة بين السنوات وأعداد السكان على الشكل :

$$\mu_{Y|x} = a_o x^{b_o}$$

أوجد معادلة الانحدار المقدرة ؟

العام x	1970	1971	1972	1973	1974	1975
عدد y السكان بالمليون	1.010	1.050	1.060	1.080	1.111	1.156
العام x	1976	1977	1978	1979	1980	
عدد y السكان بالمليون	1.155	1.170	1.201	1.230	1.330	

-٥١- في دراسة أجريت على إحدى أنواع الثدييات وجد أن حجم المخ يتغير مع وزن الجسم من فرد لآخر وأن العلاقة بين حجم المخ ووزن الجسم على الشكل :

$$\mu_{Y|x} = a_o x^{b_o}$$

استخدم البيانات التالية لإيجاد معادلة الانحدار المقدرة ؟

x وزن الجسم	30	35	37	40	41	44	46	47	49	52	54
y حجم المخ	360	379	380	390	409	408	412	419	425	435	439

-٥٢- الجدول التالي يوضح متوسط طول الجمجمة مقاساً بالمليمترات ومتوسط طول الوجه y مقاساً بالمليمترات لنوع معين من القردة. فإذا كانت العلاقة بين طول الجمجمة وطول الوجه على الشكل

$$\mu_{Y|x} = a_o x^{b_o}$$

استخدم البيانات التالية للحصول على تقديرات لكل من a_o , b_o .

x طول الجمجمة	75.5	99.1	107.8	113.7	115.25	120.0
y طول الوجه	30	63.5	93.7	131.0	141.2	143.22

-٥٣- البيانات التالية تمثل سعر بيع السيارة وعمر السيارة لماركة خاصة بالسنوات :

x العمر بالسنوات	1	2	2	5	5	5
y سعر البيع	2350	1690	1749	1390	981	890

إذا كانت العلاقة بين عمر السيارة وسعر البيع على الشكل : $\mu_{Y|x} = \gamma\ \delta^x$

(أ) أوجد تقديرات لكل من δ , γ.

(ب) أوجد سعر البيع عندما يكون عمر السيارة 4 سنوات .

-٥٤- يعطي الجدول التالي الضغط لغاز ما عند قيم مختلفة من الحجم .

x $in.^3$ الحجم	50	60	70	90	100

الضغط ib/in^3 y	63.6	50.2	40.1	24.8	7.7

فإذا كان قانون الغاز معطي بالعلاقة $\mu_{Y|x} \quad x^a = c$ حيث a , c ثوابت أوجد :

(أ) تقديرات المربعات الصغرى للمعالم a , c

(ب) أوجد y عندما x = 70 .

-٥٥- في دراسة عن العلاقة بين عدد الحوادث في مدينة ما خلال 10 سنوات وبين عدد الإشارات تم الحصول علي البيانات التالية (من 1969 إلي 1978).

السنوات	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978
x عدد الإشارات	452	425	500	517	350	480	603	611	650	718
y عدد الحوادث	42	1	6	5	1	4	7	10	15	18

استخدم النموذج الأسى في حساب معادلة الانحدار المقدرة ؟

-٥٦-البيانات التالية تعطى عدد العمال الزراعيين في الفترة من 1968 إلي 1975 .

السنوات	1965	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975
عدد العمال	1050	1000	900	800	750	675	500	450	425	350	300

(أ) أوجد معادلة الانحدار المقدرة تحت فرض النموذج الأسى ؟

(ب) قدر معادلة الانحدار من الدرجة الثانية ؟

(ج) قدر معادلة الانحدار الخطية ؟

(د) أى المعادلات منطقية ولماذا ؟

-٥٧- إذا كان لديك البيانات التالية :

x	0	1	2	3	4	5	6	7	8	9
y	9.0	7.2	3.1	4.5	4.7	2.8	5.6	7.0	8.7	10.1

(أ) المطلوب توفيق هذه البيانات باستخدام النموذج :

$$\mu_{Y|x_1,x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

(ب) أوجد y عند x = 2 .

-٥٨- الجدول التالي يعطي أعداد البيتزا المباعة ومتوسط التكلفة في محل لبيع البيتزا .

x عدد الوحدات المباعة	100	110	120	130	140	150	160	170	180
y متوسط التكلفة لوحده البيتزا	0.90	0.88	0.85	0.72	0.65	0.70	0.73	0.79	0.74

(أ) ارسم شكل الانتشار ؟

(ب) أوجد معادلة الانحدار المقدرة تحت فرض معادلة من الدرجة الثانية ؟

-٥٩- يعطى الجدول التالي عدد القوارب المباعة في مدينة ما خلال السنوات من 1970 إلي 1975 .

x السنوات	1970	1971	1972	1973	1974	1975
y عدد القوارب المباعة	100	90	85	90	75	80
x السنوات	1976	1977	1975			
y عدد القوارب المباعة	86	110	115			

أستخدم معادلة من الدرجة الثانية لتوفيق هذه البيانات وأوجد عدد القوارب المباعة سنة 1981

-٦٠- استخدم البيانات التالية في إيجاد نموذج الانحدار الخطي المتعدد المقدر ؟

x_1 السنوات	1	2	3	4	5	6	7	8	9	10
x_2 عدد الإشارات	452	425	500	517	350	480	603	611	850	718
y عدد الحوادث	42	1	6	5	1	4	7	10	15	18

-٦١- يتأثر محصول الفراولة بكمية الإمطار وكمية السماد المستخدم.أستخدم البيانات في الجدول التالي لتوفيق معادلة انحدار خطي متعدد باستخدام كمية الإفطار وكمية السماد كمتغيرات مستقلة ؟

x_1 كمية الامطار	16	22	23	13	17	25	18	20	21	19	22
x_2 السماد بالطن	510	450	500	425	450	475	515	500	490	510	525

←

المحصول y	1000	450	1200	700	800	1100	1050	1150	1000	950	1300

-٦٢- تبيع شركة لإنتاج السجق منتجاها من خلال بعض المراكز التجارية البيانات التالية تعطي مبيعات الشركة (بالألف دولار) في 10 مراكز وعدد الوحدات المباعة (بالمائة) وعدد المراكز التجارية أوجد معادلة الانحدار الخطي المتعدد المقدرة ؟

الموقع	1	2	3	4	5	6	7	8	9	10
المبيعات y	75	38	25	98	93	54	78	85	65	88
عدد x_1 الوحدات المباعة	15	10	7	21	14	8	14	24	9	23
عدد x_2 المراكز التجارية	25	29	15	25	11	13	22	14	13	11

-٦٣- للبيانات التالية :

x_1	1	4	9	11	3	8	5	10	2	7	6
x_2	8	2	-8	-10	6	06	0	-12	4	-2	-4
y	6	8	1	0	5	3	2	-4	10	-3	5

استخدم النموذج التالي :

$$\mu_{Y|x_1,x_2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

في إيجاد

(أ) معادلة الانحدار الخطي المتعدد المقدرة ؟

(ب) جدول تحليل التباين واستخدام $\alpha = 0.05$ في اختبار معنوية b_0, b_1, b_2 .

(ج) معامل التحديد ؟

-٦٤- البيانات التالية تعطي بيانات لثلاثة متغيرات X, Y, Z حيث X , Y متغيرات مستقلة :

$$\sum_{i=1}^{20} x_i = 23,\quad \sum_{i=1}^{20} y_i = 40, \sum_{i=1}^{20} z_i = 67,\quad \sum_{i=1}^{20} x_i^2 = 105,\quad \sum_{i=1}^{20} y_i^2 = 294,$$

$$\sum_{i=1}^{20} x_i y_i = \sum_{i=1}^{20} x_i z_i = 290,\quad \sum_{i=1}^{20} y_i z_i = 4888, \sum_{i=1}^{20} z_i^2 = 815$$

استخدم البيانات السابقة في إيجاد معادلة الانحدار الخطي المقدرة ؟

-٦٥- تتوقف قيمة المبيعات داخل أي جمعية استهلاكية على المربع السكني الذي توجد به الجمعية وذلك من حيث عدد السكان ومتوسط الدخل الشهري لقاطن هذا الحي بفرض أن كل جمعية يشتري منها سكان الحي الذي تتواجد به وفيما يلي بيانات لعشر جمعيات استهلاكية .

الجمعية	1	2	3	4	5	6	7	8	9	10	11	12
x_1 عدد السكان	275	175	375	200	175	260	90	300	190	50	400	270
x_2 متوسط الدخل	240	320	370	282	320	375	300	240	200	210	250	400
المبيعات بالآلاف جنية	160	115	220	130	119	165	80	190	115	50	250	144

أوجد معادلة الانحدار الخطي المتعدد المقدرة ؟

-٦٦- في دراسة عن العلاقة بين امتصاص الماء في دقيق القمح والخواص المختلفة للدقيق وتحت فرض نموذج انحدار خطي متعدد تم الحصول على البيانات في الجدول التالي حيث (%)Y تمثل كمية امتصاص الماء و (%)x_1 كمية البروتين و (%)x_2 كمية النشا الذي يتعرض للفقد التحطم مقاس بوحدات **Farrand** أوجد معادلة الانحدار المقدرة ؟

x_1	8.5	8.9	10.6	10.2	9.8	10.8	11.6	12.0	12.5	10.4
x_2	2	3	3	20	22	20	31	32	31	28
y	30.9	32.7	36.7	41.9	40.9	42.9	46.3	47.2	44.0	47.7
x_1	12.2	11.9	11.3	13.0	12.9	12.0	12.9	13.1	11.4	13.2
x_2	36	28	30	27	24	25	28	28	32	28
y	43.9	46.8	46.2	47.0	46.8	45.9	48.8	46.2	47.8	49.2