الفصل التاسع

تحليل الانحدار

Regression Analysis

(9-1) مقدمة :

إن نموذج الانحدار يعبر عن علاقة بين متغير تابع وبين واحد أو أكثر من المتغيرات المستقلة، فإذا احتوى النموذج على متغير مستقل واحد فيعرف بنموذج الانحدار البسيط Regression Model Simple أو ما يسمى Ordinary Least Squares، وإذا احتوى على أكثر من متغير مستقل فهو نموذج الانحدار المتعدد، كما أن النموذج قد يكون خطياً أو غير خطي.

(9-2) نموذج الانحدار الخطي البسيط Simple Regression

إن الغرض من استخدام أسلوب تحليل الانحدار الخطي البسيط، هو دراسة وتحليل أثر متغير كمي على متغير كمي آخر، ومن الأمثلة على ذلك ما يلي:

- دراسة أثر كمية السماد في إنتاجية القمح.
- دراسة أثر الإنتاج في التكلفة.
- دراسة أثر كمية البروتين التي يتناولها الأبقار في الزيادة في الوزن.
- أثر الدخل في الإنفاق الاستهلاكي.
- دراسة أثر الذكاء في التحصيل الدراسي.

وهكذا هناك أمثلة في كثير من النواحي الاقتصادية، والزراعية، والتجارية، والعلوم السلوكية، وغيرها من المجالات الأخرى.

في تحليل الانحدار البسيط، نجد أن الباحث يهتم بدراسة أثر أحد المتغيرين ويسمى بالمتغير المستقل أو المتنبأ منه، على المتغير الثاني ويسمى بالمتغير التابع أو المتنبأ به، ومن ثم يمكن عرض نموذج الانحدار الخطي في شكل معادلة خطية من الدرجة الأولى، تعكس المتغير التابع كدالة في المتغير المستقل كما يلي:

$$y = \beta_0 + \beta_1 \chi + e \qquad (1-9)$$

حيث إن:

- ✓ y : هو المتغير التابع (الذي يتأثر).
- ✓ x : هو المتغير المستقل (الذي يؤثر).

obeikandl.com

✓ β_0 : هو الجزء المقطوع من المحور الرأسي y، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل χ، أي في حالة $\chi = 0$.

✓ β_1 : ميل الخط المستقيم $(\beta_0 + \beta_1\chi)$، ويعكس مقدار التغير في y إذا تغيرت بوحدة واحدة.

✓ e : هو الخطأ العشوائي، الذي يعبر عن الفرق بين القيمة الفعلية، والقيمة المقدرة $\hat{y} = \beta_0 + \beta_1\chi$، أي إن $e = y-(\beta_0 + \beta_1\chi)$.

تقدير نموذج الانحدار الخطي البسيط

يمكن تقدير معاملات الانحدار $(\beta_0$ و $\beta_1)$ في النموذج (9-1) باستخدام طريقة المربعات الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات الأخطاء العشوائية $\sum e^2 = \sum (y-(\beta_0 + \beta_1\chi))^2$ أقل ما يمكن، وتكون القيمة المقدرة للمتغير التابع هو: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1\chi$ ، ويطلق على هذا التقدير" تقدير معادلة انحدار y على x.

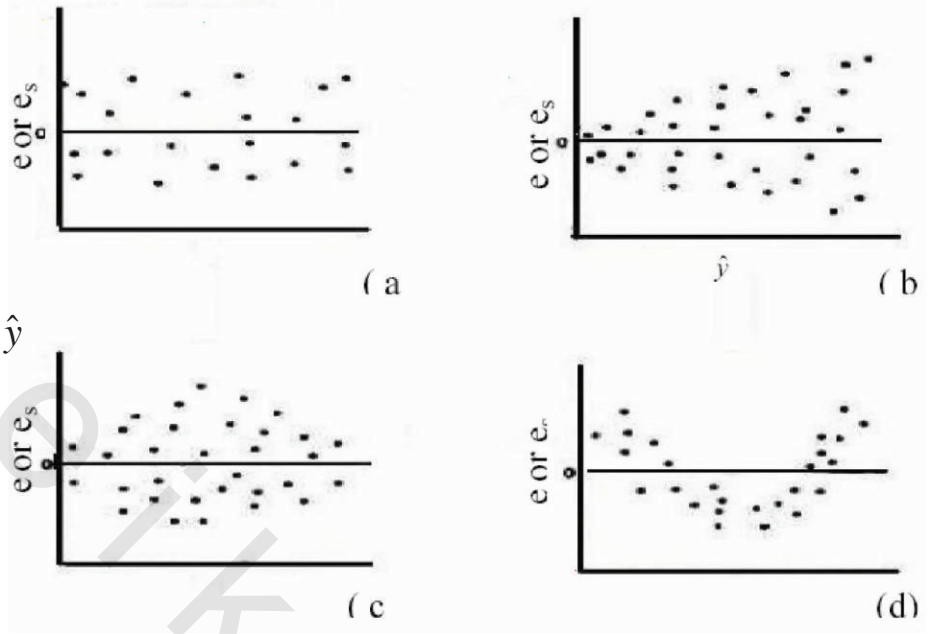
علماً بأن نموذج الانحدار الخطي البسيط يجب أن يحقق مجموعة الفروض الآتية:

- وجود علاقة خطية بين X وY.
- أن الأخطاء العشوائية تتوزع بمتوسط مساوٍ للصفر.
- أن الأخطاء العشوائية لها تباين ثابت يساوي σ^2 (فرضية تجانس تباين الخطأ العشوائية Ho-moscedasticity).
- الأخطاء تتوزع طبيعياً وهذا الشرط ليس ضرورياً لتقدير المعالم بطريقة OLS، ولكنه ضروري لاختبار الفرضيات المعلقة بمعاملات الانحدار β_0 و β_1.

عدم وجود ارتباط ذاتي Autocorrelation بين الأخطاء العشوائية.

ويمكن التحقق من توافر فرضيات النموذج الخطي البسيط من خلال تخطيط Scatter plots بتمثيل $\hat{y}$ على المحور الأفقي أو (X) يقابله الخطأ العشوائي e أو الأخطاء المعيارية Standardized Residual التي يرمز لها e_s على المحور الرأسي، وذلك كما هو موضح في الشكل الآتي:

obeikandl.com



(a) توافر فروض التحليل جميعها (عدم وجود مشكلة).

(b) زيادة تباين الخطأ العشوائي بزيادة $\hat{y}$.

(c) زيادة وتناقص في تباين الخطأ العشوائي (مشكلة عدم تجانس تباين الخطأ العشوائي).

(d) عدم ملاءمة العلاقة الخطية (يجب استعمال نماذج أخرى مثلا نموذج الدرجة الثانية).

وفيما يلي نبين كيف يمكننا اختبار هذه الشروط وكيفية تقدير معادلة الانحدار وتحديد الأخطاء باستعمال البرنامج:

تطبيقات:

مثال (9-1)

فيما يلي بيانات عن كمية البروتين اليومي بالجرام التي يحتاج إليها العجل الرضيع، ومقدار الزيادة في وزن العجل بالكليو جرام، وذلك لعينة من العجول الرضيعة حجمها 10.

جدول (9-1)

كمية البروتين X	10	11	14	15	20	25	46	50	59	70
الزيادة في الوزن Y	10	10	12	12	13	13	19	15	16	20

والمطلوب:

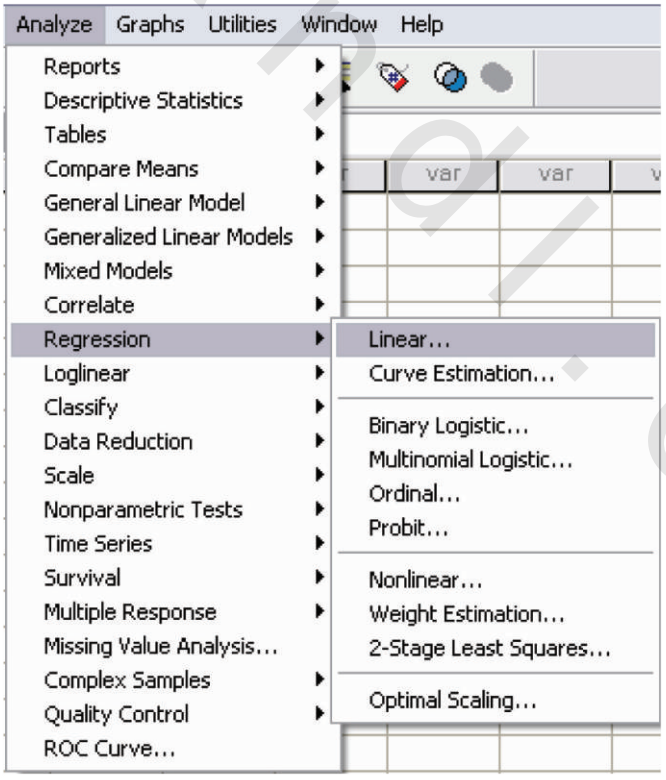
- **قدر معادلة انحدار الوزن** y على كمية البروتين x.

- فسر معادلة الانحدار.
- ما مقدار الزيادة في الوزن عند إعطاء العجل 50 جرامًا من البروتين؟
- وما مقدار الخطأ العشوائي؟
- استخراج فترة ثقة 95% لكل من معلمتي الانحدار β_0 و β_1 .
- استخراج جدول تحليل التباين ANOVA للمعاملات.
- اختبر جودة توفيق النموذج الخطي (باستعمال معامل التحديد R^2) مع تحليل الأخطاء العشوائية بالرسم البياني.
- اختبر التوزيع الطبيعي للأخطاء العشوائية بيانياً.

الحل:

لتنفيذ المطلوب السابق باستخدام البرنامج نتبع الخطوات الآتية:

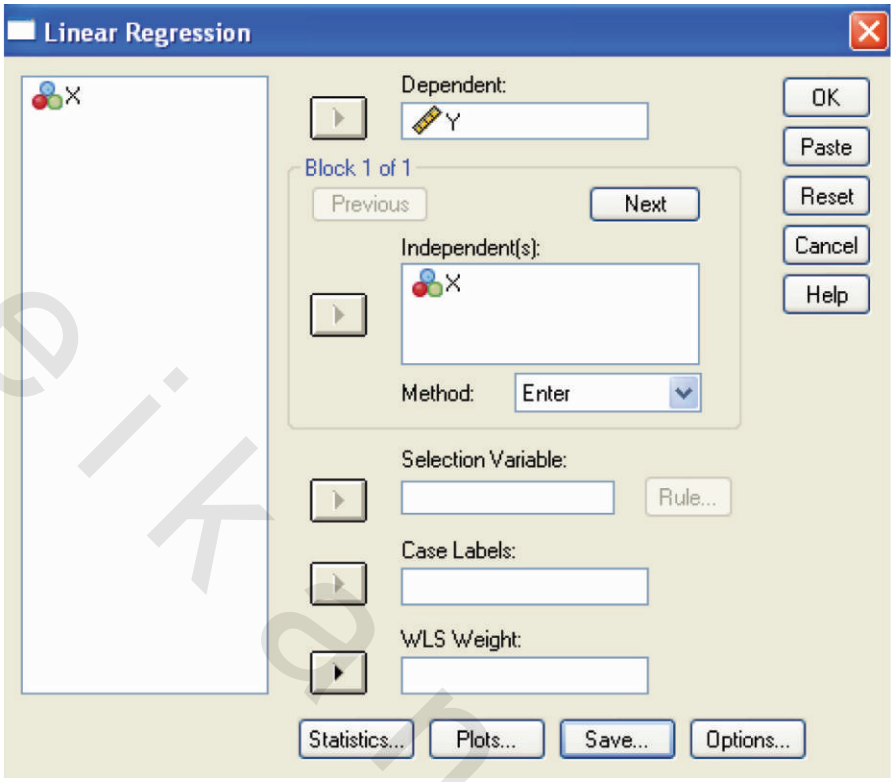
Analyze ثم ⇐ Regression ⇐ Linear كما في الشكل (9-1):



شكل (9-1)

obeikandl.com

فيظهر صندوق الحوار الآتي:



شكل (9-2)

ونجد أن الشكل (9-2) يتكون من:

Dependent: يمثل المتغير التابع.

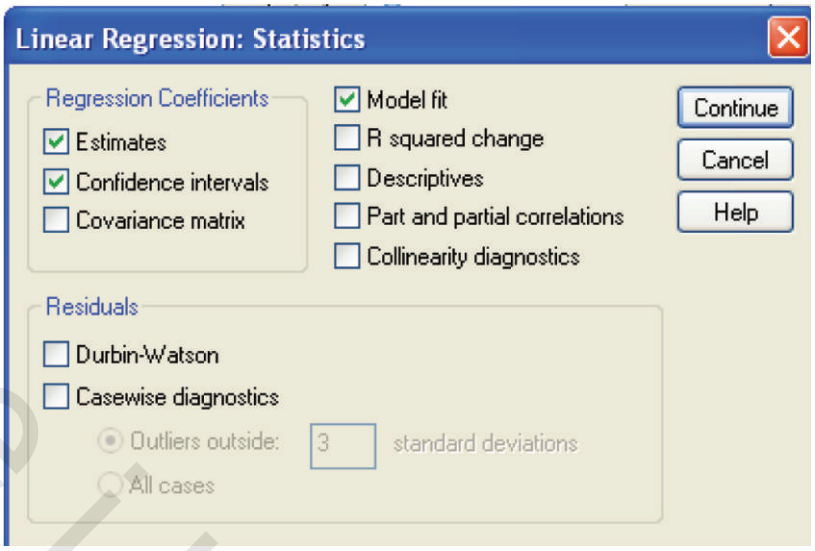
Independent: وهو يمثل المتغير المستقل ويمكننا إدخال مجموعة من المتغيرات المستقلة هنا كما سوف نرى في مبحث الانحدار المتعدد، ولأننا نريد معادلة انحدار لـ y على X فإننا نعني أن y متغير تابع وX متغير مستقل.

Method: نوع الطريقة المستخدمة في الانحدار (الطريقة الاعتيادية هي Enter).

Selection Variable: يستعمل في تحديد التحليل لمجموعة معينة من الحالات التي لها قيمة معينة لمتغير الاختيار (مثلا اقتصار نموذج الانحدار على الحالات التي تكون فيها قيمة المتغير Observat أكبر من 5) يتم التحديد بواسطة الزر Rule.

Case Lebels: متغير يستخدم قيمة كعناوين لنقاط شكل الانتشار Scatterplots.

Statistics: عند النقر على هذا الزر يظهر مربع الحوار الآتي:



شكل (9-3)

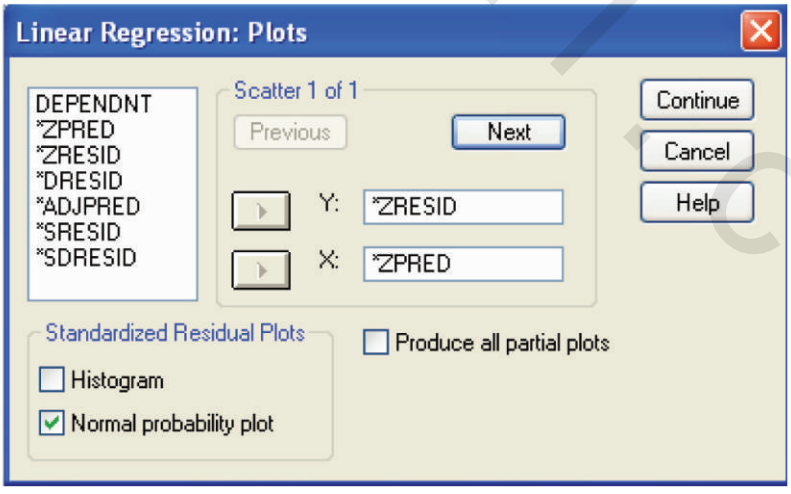
وقد تم تأشير الخيارات الآتية:

✓ Estimate: لتقدير معالم نموذج الانحدار واختبارات t المرافقة.

✓ Confidence Interval: لتقدير فترة ثقة 95% لكل من معلمتي الانحدار.

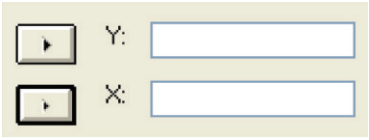
✓ Model Fit: لعرض R^2 و ANOVA.

وبعـد الانتهاء النقر على زر Continue. وهناك زر آخر في شكل (9-2) بعنوان Plots وعند النقر عليه يظهر مربع الحوار الآتي:



شكل (9-4)

ونلاحظ أننا اخترنا Normal probability plot: لاختبار التوزيع الطبيعي للأخطاء العشوائية المعيارية (المطلوب ن). كما أننا نجد أعلى شكل (9-4) الشكل الآتي:



في هـذا الشـكل يتم إدخـال المتغيرات التي نـود ان تظهر في شـكل تحليل الأخطاء في المحـور السيني والمحور الصادي والمتغيرات موجودة في الجانب الأيمن من الشكل السابق وتوجد عدة خيارات:

1- (*ZPRED) Standardized predicted values: القيمة المعيارية للقيمة **المتوقعة** $\hat{y}$.

2- (*ZRESID) Standardized residuals: القيمة المعيارية للأخطاء.

3- (*DRESID) Deleted residuals: حذف الأخطاء.

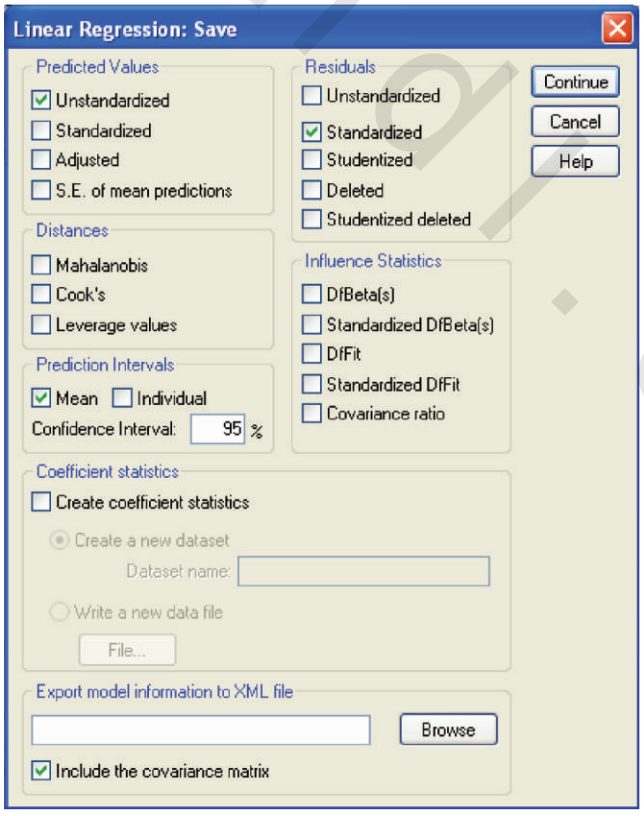
4- (*ADJPRED) Adjusted predicted values: القيمة المعيارية المعدلة لقيمة متوقعة $\hat{y}$.

5- (*SRESID) Studentized residuals: جعل الأخطاء تتبع توزيع t.

6- (*SDRESID) Studentized deleted residuals: الأخطاء المحذوفة التي تتبع توزيع t.

وفي تلك الحالة يتم تخصص كل من: ZPRED لمحور السينات، ZRESID لمحور الصادات وبعد الانتهاء يتم النقر على زر Continues.

وهناك رز آخر في شكل (9-2) بعنوان Save وعند النقر عليه يظهر مربع الحوار الآتي:



شكل (9-5)

ونلاحظ أننا اخترنا Unstandardized predicted Values أي $\hat{y}$ وكذلك Standardized Residual أي e_s سيتم استعمال هذين المتغيرين لرسم Scatterplots الجزء الثاني من المطلوب م، علماً بأنه سيتم إضافة هذين المتغيرين إلى ورقة Data Editor إلى جانب متغيرات X و Y و Observat حيث يضاف المتغير Unstandardized predicted Values باسم Pre_1 ويضاف Standardized Residual باسم Zre_1. وبعد الانتهاء ينقر زر Continues.

و في النهاية ننقر على Ok في الشكل (10-2) لإظهار النتائج كالآتي:

جدول (9-2)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	9.436	.864		10.926	.000	7.444	11.427
	X	.143	.023	.913	6.325	.000	.091	.195

a. Dependent Variable: Y

من خلال الجدول أعلاه يمكن كتابة النموذج كما يلي:

$$\hat{y} = 9.436 + 0.143\,\chi$$

وتشير المعادلة السابقة إلى أن زيادة البروتين مقدار جرام تعمل على زيادة الوزن بمقدار 0.143 جرام، بينما توجد زيادة ليست متعلقة بالبرويتن مقدارها 09.436 لكل جرام في الوزن، وعليه فإن زيادة البروتين بمقدار 50 جرامًا تعمل على أن يكون

$$\hat{y} = 9.436 + 0.143\,(50)$$
$$= 16.436$$

ومن ثم مقدار الخطأ العشوائي يساوي:

$$\hat{e}_{\chi=50} = y_{\chi=50} - \hat{y}_{\chi=50}$$
$$= 15 - 16.436$$
$$= -1.43$$

ويستعمل اختبار t لاختبار الفرضية الآتية لمعلمة الميل β_1:

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

ويستعمل اختبار t لاختبار الفرضية الآتية لمعلمة الميل β_0:

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

نستخدم قيمة *P-value* المرافقة لإحصائية t للمعلمة في الاختبار كما يلي:

إذا كانت *P-value<0.05* نرفض فرضية العدم بمستوى 5%.

إذا كانت *P-value<0.01* نقبل فرضية العدم بمستوى 1%.

عكس هذا نقبل فرضية العدم.

ونلاحظ أن *P-value* لمعلمة الميل تساوي صفرًا، وهي أقل من 0.01، إن *P-value* لمعلمة الحد الثابت تساوي صفرًا، وهي أقل من 0.01، ولهذا نرفض فرضي العدم لكل من المعلمتين، أي إن كلا من المعلمتين تختلف جوهرياً عن الصفر. إن ظهور معلمة الميل معنوية يعكس أهمية متغير البروتين في التأثير على الوزن في النموذج.

ويمكن كتابة فترة الثقة للحد الثابت، وذلك باستعمال نتائج الشكل (10-6) كما يلي:

$$\Pr[7.444 \leq \beta_0 \leq = 11.427] = 95\%$$

حيث إن Pr تمثل الاحتمال كما يمكن كتابة فترة ثقة 95% لمعلمة الميل كما يلي:

$$\Pr[0.091 \leq \beta_0 \leq = 0.195] = 95\%$$

الجدول الآتي يعرف بجدول تحليل التباين ANOVA، ويشتمل على إحصائية F لاختبار الفرضية الخاصة بمعلمة الميل β_1 نفسها، وهذا الاختبار مكافئ تماماً لاختبار t لمعلمة الميل (لاحظ أن قيمة P-value متساوية لكلا الاختبارين) علما بأن $F = t^2$.

جدول (9-3)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	90.000	1	90.000	40.001	.000[a]
	Residual	18.000	8	2.250		
	Total	108.000	9			

a. Predictors: (Constant), X

b. Dependent Variable: Y

obeikandl.com

الجدول الآتي يتضمن أهم مؤشر لنموذج الانحدار وهو معامل التحديد Coefficient of Determination ويرمز له R^2 ويعد مقياساً لجودة توفيق النموذج:

جدول (4-9)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.913[a]	.833	.813	1.49999

a. Predictors: (Constant), X

b. Dependent Variable: Y

ويحتسب من جدول تحليل التباين كما يلي:

$$R^2 = \frac{\text{Explained Variations}}{\text{Total Variations}}$$

$$= \frac{SSR}{SST}$$

$$= \frac{90}{108.00}$$

$$= 0.833333 \quad , 0 \le R^2 \le 1$$

وتفسير ذلك أن 83.333% من التباينات (الانحرافات الكلية من قيم المتغير y) تفسرها العلاقة الخطية أي نموذج الانحدار، وأن 15.76% من التباينات ترجع إلى عوامل عشوائية كأن تكون هناك متغيرات مهمة لم يتضمنها النموذج. وعلى العموم كلما اقتربت قيمة R^2 من 100% دل ذلك على جودة توفيق النموذج. هذا، أن $r = \sqrt{R^2} = 0.91287$، حيث إن r معامل الارتباط الخطي البسيط لبيرسون، وإن إشارة r هي إشارة معلمة الميل نفسها، أي إن الإشارة موجبة، وهذا يعني أن العلاقة بين مقدار البروتين والوزن علاقة طردية وقوية.

لاحظ أن معامل التحديد لو أضيف متغير مستقل للنموذج فإن قيمته سترتفع حتى لو لم تكن هناك أهمية للمتغير المستقل في النموذج، حيث إن إضافة متغير مستقل إلى نموذج الانحدار تؤدي إلى زيادة R^2 بسبب زيادة مجموع المربعات العائدة للانحدار SSR مع ثبات مجموع المربعات الكلية SST، ولهذا يتم احتساب معامل التحديد المصحح Adjusted R Square الذي يأخذ في الحسبان النقصان الحاصل في درجات الحرية، وقيمته دائما أقل من قيمة معامل التحديد (غير المصحح) وفي هذا المثال 81.3% ولهذا يمكن القول إن النموذج جيد التوفيق.

أمـا الخطـأ المعيـاري للتقديـر Standard Error of Estimate فيقيس تشتت القيـم المشاهدة عن خط الانحدار، وأن الحصول على قيمة صغيرة لهذا المؤشر يعني صغر الأخطاء العشوائية، ومن ثم جودة تمثيل خط الانحدار لنقاط شكل الانتشار.

لتحليل الأخطاء العشوائية بيانياً نكون شكل الانتشار Scatterplots بتمثيل القيم التقديرية $\hat{y}$ على المحور الأفقي والأخطاء المعيارية على المحور العمودي أو الرأسي كما يلي:

ففي شاشة Data Editor نجد أنه كما سبق أن أوضحنا أنه سوف تتم إضافة المتغير Unstandardized predicted Values باسم Pre_1 ويضاف Standardized Residual باسم Zre_1 كما يلي:

	X	Y	PRE_1	ZRE_1
1	10	10.00	10.86212	-.57475
2	11	10.00	11.00475	-.66984
3	14	12.00	11.43264	.37824
4	15	12.00	11.57527	.28315
5	20	13.00	12.28843	.47439
6	25	13.00	13.00158	-.00105
7	46	19.00	15.99684	2.00212
8	50	15.00	16.56736	-1.04491
9	59	16.00	17.85104	-1.23403
10	70	20.00	19.41998	.38668

شكل (9-6)

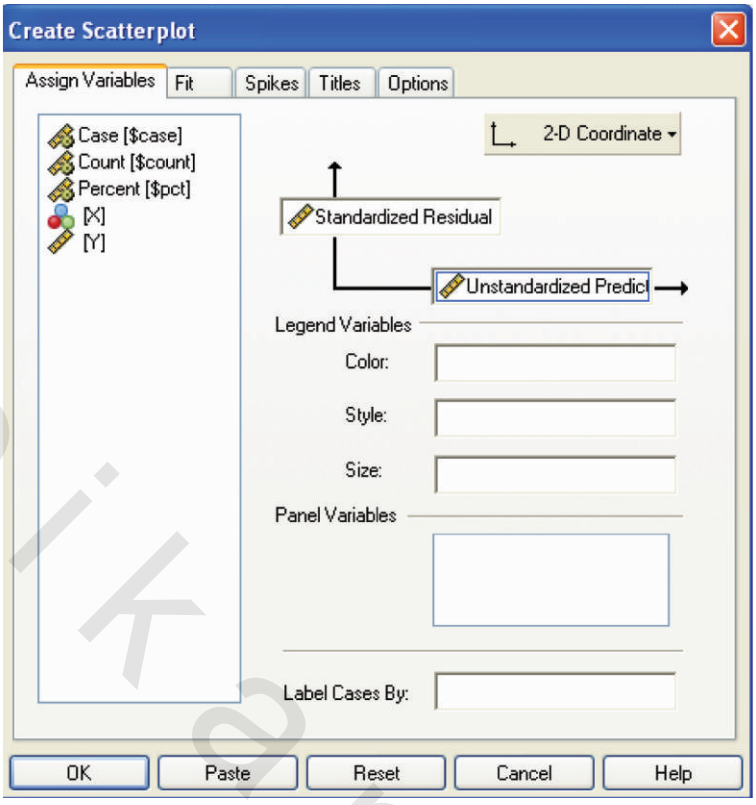
ثـم بعد ذلـك من شريـط القوائـم نختـار Simple ⇐ Scatter ⇐ Interactive ⇐ Graphs فيظهر صندوق حوار بعنوان Create Scatterplots:

فنقوم بالنقر على Zre_1 لإدخاله إلى الخانة Y في صندوق الحوار وبالمثل

فنقوم بالنقر على Pre_1 لإدخاله إلى الخانة X في صندوق الحوار وبالمثل

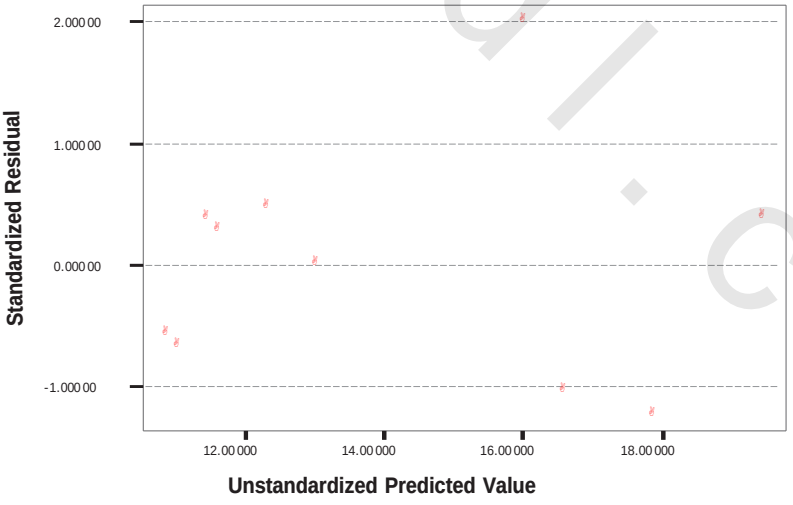
فيظهر مربع الحوار Create Scatterplots كما يلي:

obeikandl.com



شكل (9-7)

فننقر على Ok فيظهر المخطط الآتي:



شكل (9-8)

نلاحظ أن النقاط تتوزع بشكل شريط أفقي متساو حول الصفر، ما يدل علي توافر فرضيات التحليل بصورة عامة، حيث لا يعاني النموذج مشكلة عدم تجانس تباين الخطأ العشوائي، ولا توجد حاجة لاستخدام علاقة من درجات أعلى.

obeikandl.com

مثال (9-2)

لدينـا شركة تقوم بصناعة قطع غيار معينة على دفعات شهريـة، وقد أعطتنا الشركة البيانات لعشر دفعات في الجدول الآتي:

جدول (9-5)

دورة الإنتاج	حجم الدفعة	ساعات العمل
1	30	73
2	20	50
3	60	128
4	80	170
5	40	87
6	50	108
7	60	135
8	30	69
9	70	148
10	60	132

تريـد الشركة معرفة نموذج الانحدار الخطي الخاص لهـا لكي تتنبأ بأحجام الدفعات المستقبلية بناء على ساعات العمل عند مستوى معنوية 0.05.

الحـل:

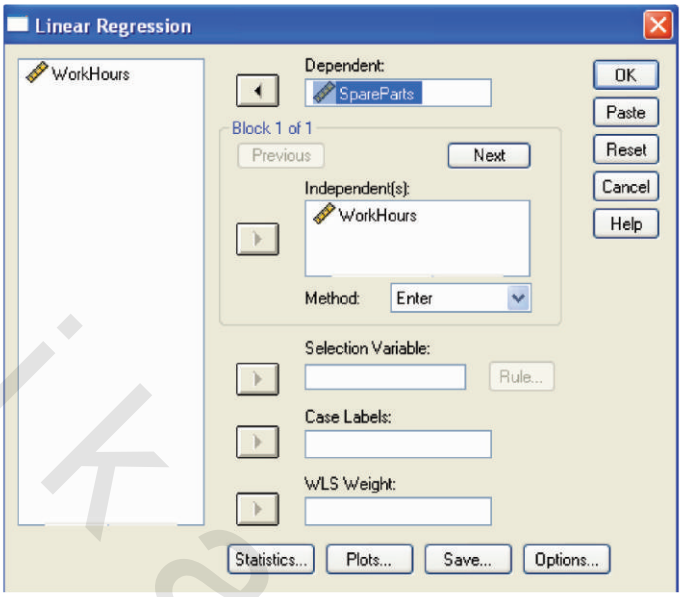
ندخل البيانات في عمودين أحدهما لحجم الدفعة من قطع الغيار Spare Parts والآخر لساعات العمل على كل دفعة Work Hours كما يأتي:

	SpareParts	WorkHours
1	30.00	73.00
2	20.00	50.00
3	60.00	128.00
4	80.00	170.00
5	40.00	87.00
6	50.00	108.00
7	60.00	135.00
8	30.00	69.00
9	70.00	148.00
10	60.00	132.00

شكل (9-9)

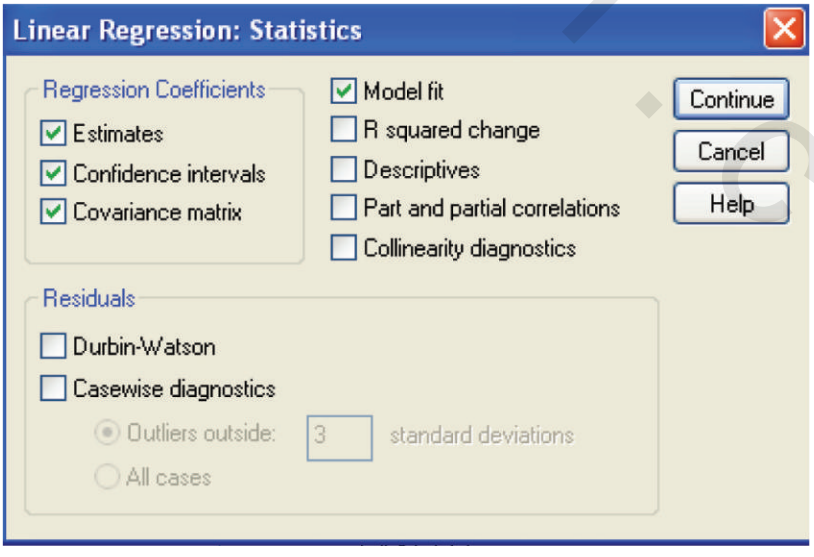
obeikandl.com

نقـوم بتفعيـل نافذة الانحدار ونختار المتغير المستقل هو Work Hours، أما التابع فهو Spare Parts كما في الشكل الآتي:



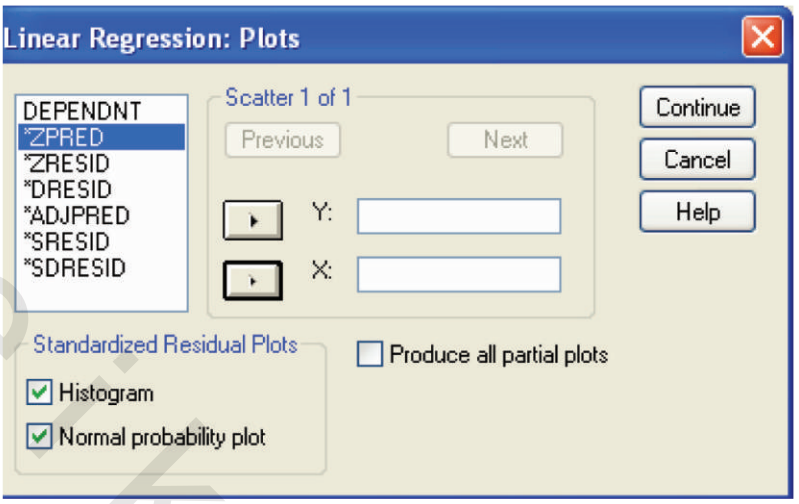
شكل (9-10)

نختـار الطريقـة المعيارية وهـي Enter في قائمة الخيـارات Method؛ لأنها الطريقـة الخاصة بتقدير معاملات الانحدار المجهولة للمتغيرات المستقلة، بعد ذلك نقوم باختيار الأمر الفرعي Statistics ونتأكد من اختيار كل من Model fit، Covariance matrix، Confidence intervals، Estimates، كما في الشكل الآتي:



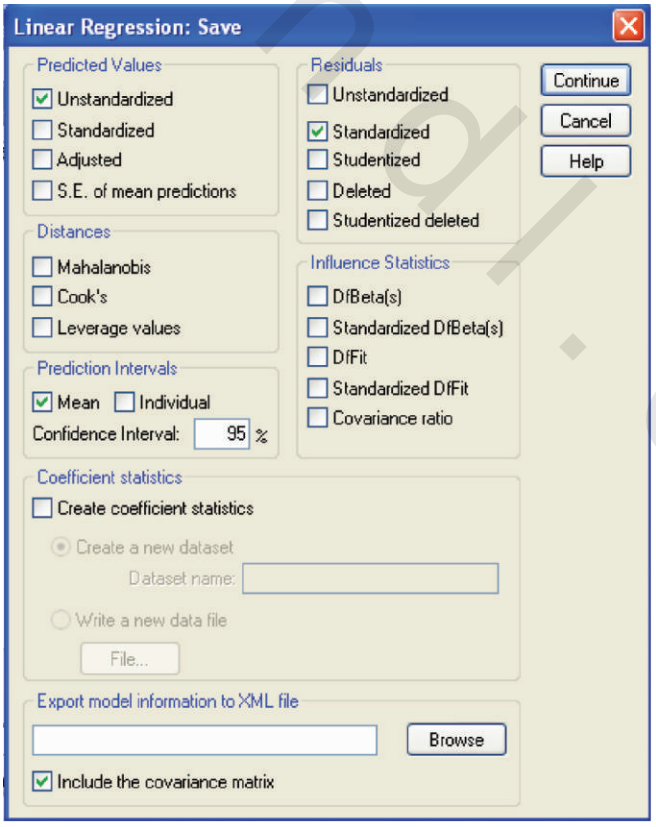
شكل (9-11)

ثـم نعود إلى مربع الحوار السابق بالنقر علـى Continue، ونذهب إلى الأمر الفرعي Plots ونتأكد من اختيار Histogram وNormal probability plot كما يأتي:



شكل (9-12)

ثـم نعود إلى مربع الحـوار السابق بالنقر على Continue ونذهب إلى الأمـر الفرعي Save ونتأكد من الاختيارات السابقة في المثال السابق كما يلي:



شكل (9-13)

obeikandl.com

ثم نعود إلى مربع الحوار السابق بالنقر على Continue ثم ننقر OK فتظهر لدينا النتائج الآتية:

يظهـر لدينا جدول يبـين الطريقة المستخدمة في الانحـدار والمتغير التابع وهـو Spare Parts والمتغيرات المستقلة الداخلة في نموذج الانحدار الخطي، وفي حالتنا يوجد متغير واحد فقط وهو Work Hours كما يلي:

جدول (9-6)

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Age, Weight[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: BloodPresure

الجدول الآتي يمثل بعض المقاييس التي تهمنا لمعرفة هل المتغيرات المستقلة لها تأثير كبير في النموذج أم لا هو:

جدول (9-7)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.998[a]	.996	.995	1.36630

a. Predictors: (Constant), WorkHours

b. Dependent Variable: SpareParts

يظهر فيه معامل التحديد R وله قيمة عالية جدًا ولدينا أيضًا R Square المعدلة، و قيمته 0,996 أي إن المتغير المستقل يقوم بتفسير ما يقارب 99% من النموذج. ولدينا خطأ التقدير، وهو هنا خطأ صغير قيمته 1,36، وكلما اقتربت قيمته من الصفر كان أفضل.

جدول تحليل التباين:

جدول (9-8)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3385.066	1	3385.066	1813.333	.000[a]
	Residual	14.934	8	1.867		
	Total	3400.000	9			

a. Predictors: (Constant), WorkHours

b. Dependent Variable: SpareParts

obeikandl.com

من هذا الجدول نستطيع اختبار معنوية الانحدار وذلك باستخدام الاختبار الآتي:

الفروض الإحصائية:

H_0: الانحدار غير معنوي.

H_1: الانحدار معنوي.

خطوات الاختبار:

نقارن P-value = .000 بمستوى المعنوية 0.05 نلاحظ أن القيمة P-value أقل من مستوى المعنوية.

القرار الإحصائي:

نقوم برفض الفرض العدمي، وهذا معناه أن الانحدار معنوي (دال).

كما يمكننا معرفة مجموع المربعات الآتية ودرجات حرياتها:

- مجموع مربعات البواقي SSE = 14.934 بدرجة حرية 8
- مجموع مربعات الانحدار SSR = 3358.066 بدرجة حرية 1
- مجموع المربعات الكلي SST = 4300.00 بدرجة حرية 9

وأيضًا هناك متوسطات المربعات الآتية:

- MSE = 1.867 متوسط مربع الخطأ.
- MSR =3385.066 متوسط مربع الانحدار

جدول المعالم المقدرة:

جدول (9-9)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-4.758	1.357		-3.508	.008	-7.887	-1.630
	WorkHours	.498	.012	.998	42.583	.000	.471	.525

a. Dependent Variable: SpareParts

من هذا الجـدول نستطيع الحصول على تقديرات لقيم معالم الانحدار المجهولة، وبذلك نحصل على معادلة الانحدار المقدرة الآتية:

Spare Parts = -4.758 + 0.498Workhours

obeikandl.com

حيث:

$$\hat{\beta}_0 = -4.758$$
$$\hat{\beta}_1 = 0.498$$

مـن معادلة الانحدار المقدرة بإمكاننا التنبؤ بقطع الغيار التي سنحصل عليها بناء على ساعات العمل فمثلا إذا كانت ساعات العمل هي 70 ساعة فيكون الإنتاج المتوقع لقطع الغيار كالآتي:

Spare Parts = -4.758 + .498 (70)

=30.102

أي تقريبًـا سيكـون الإنتاج 30 قطعة غيار، كما يمكن من هذا الجـدول معرفة الخطأ المعياري لكل من التقديـرات و فترات الثقة لكل تقدير. وأيضًا يمكن إجـراء اختبارات المعنوية لكل من هذه التقديرات على حدة.

فلنختبر معنوية β_1 عند مستوى معنوية 0.05:

الفروض الإحصائية:

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

خطوات الاختبار:

نقارن P-value = .008 بمستوى المعنوية 0.05 نلاحظ أن القيمة P-value أقل من مستوى المعنوية.

القرار الإحصائي:

نقوم برفض الفرض العدمي وهذا معناه أن β_0 معنوي.

openkandil.com

جدول البواقي:

جدول (9-10)

Residuals Statistics(a)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	20.1318	79.8682	50.0000	19.39377	10
Std. Predicted Value	1.540 -	1.540	000.	1.000	10
Standard Error of Predicted Value	433.	824.	597.	137.	10
Adjusted Predicted Value	20.2070	79.7930	50.0073	19.35863	10
Residual	-2.44510	1.44949	00000.	1.28816	10
Std. Residual	-1.790	1.061	000.	943.	10
Stud. Residual	-1.936	1.143	003.-	1.026	10
Deleted Residual	-2.86228	1.68296	00726.-	1.52759	10
Stud. Deleted Residual	-2.485	1.169	067.-	1.153	10
Mahal. Distance	003.	2.372	900.	852.	10
Cook's Distance	004.	320.	089.	103.	10
Centered Leverage Value	000.	264.	100.	095.	10

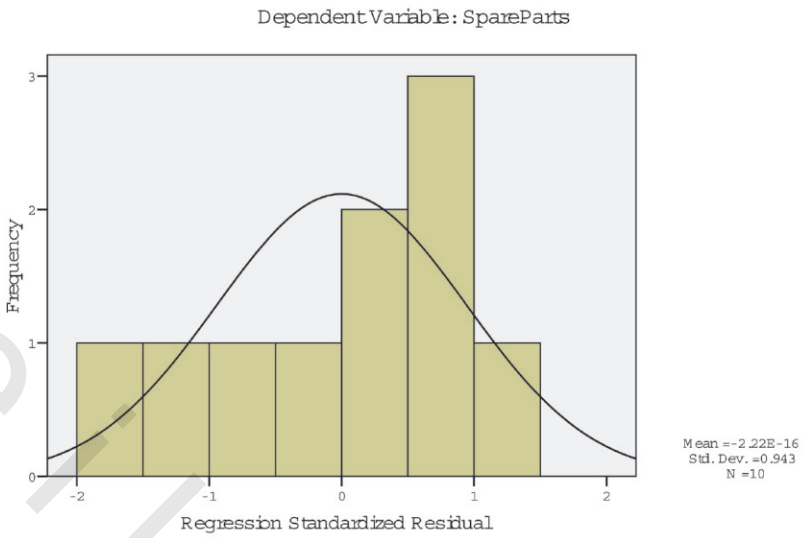
a Dependent Variable: Spare Parts

الأشكال البيانية:

- المدرج التكراري: ويستخدم لمعرفة إذا كانت البيانات تتوزع توزيعًا طبيعيًا أم لا، والمدرج التكراري الخاص بالنموذج هو:

obeikandl.com

Histogram



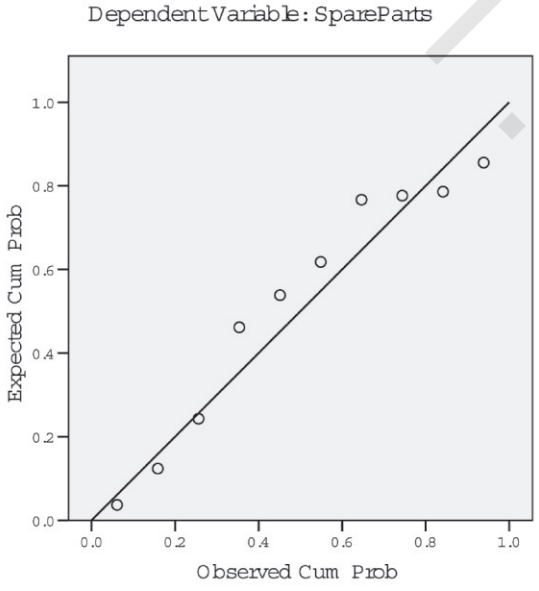
شكل (10-14)

نلاحظ أن شكله قريب من شكل التوزيع الطبيعي، وهذا معناه أن البيانات تتبع التوزيع الطبيعي ونستطيع تطبيق تحليل الانحدار عليها.

رسم منحى الانحدار:

نشاهد أن النقاط تتوزع بشكل متساو، حيث إن النقاط الأعلى تساوي النقاط الأدنى من الخط، ولذلك فإن فروض النموذج محققة فيه.

Normal P-Pot of Regression Standardized Residual



شكل (9-15)

(9-3) الانحدار الخطي المتعدد Multiple Linear Regrassion

تحليل الانحدار المتعدد هو من بين الأدوات الإحصائية كافة، الأداة المستخدمة على أوسع نطاق، الفرق الأساسي بينه وبين الانحدار الخطي البسيط هو أنه هنا لدينا متغير تابع واحد ومتغيرات مستقلة متعددة لها تأثيرات مختلفة في هذا المتغير التابع، ويكون شكل النموذج الخطي العام الذي يحتوي على عدد p-1 متغيرات مستقلة هو:

$$Y_1 = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

حيث p هو عدد المعالم المجهولة في النموذج، ونلاحظ هنا أنه إذا وضعنا p=2 نحصل على النموذج الخطي البسيط.

تدعى المعالم $\beta_1, \beta_2, \ldots, \beta_{p-1}$ بمعالم انحدار جزئية Partial regression Coefficients أو الميول الجزئية؛ لأن كل واحد منها يعكس تأثيرًا جزئيًا على المتغير التابع للمتغير المستقل الخاص به عندما يكون المتغير المستقل الآخر ثابتًا.

إن فروض النموذج الخطي المتعدد هي فروض النموذج البسيط نفسها يضاف إلى ذلك فرض عدم وجود ارتباط خطي متعدد بين المتغيرات المستقلة (Multicollinearity).

مثال (9-3)

نريد دراسة العلاقة بين ضغط دم الأطفال الرضع وعمرهم بالأيام ووزنهم عند الولادة مقاسًا بالأونصة عن طريق نموذج انحدار خطي متعدد لستة عشر رضيعًا باستخدام البيانات الآتية:

جدول (9-11)

عدد	الوزن عند الولادة	العمر بالأيام	ضغط الدم
1	135	3	89
2	120	4	90
3	100	3	83
4	105	2	77
5	130	4	92
6	125	5	98
7	125	2	82
8	105	3	85
9	120	5	96
10	90	4	95
11	120	2	80
12	95	3	79
13	120	3	86
14	150	4	97
15	160	3	92
16	125	3	88

obeikandl.com

حيث نريد تقدير معاملات الانحدار لهذا النموذج، أي إيجاد نموذج الانحدار المقدر عند مستوى معنوية 0.05.

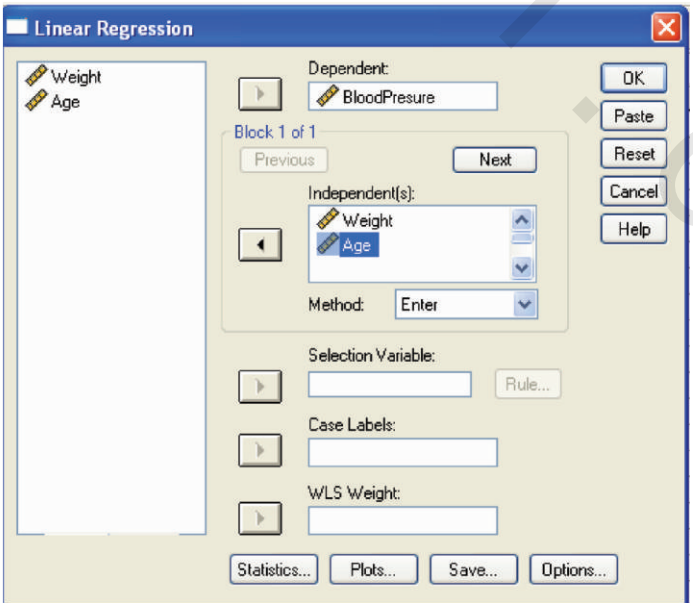
الحل:

نقوم بإدخال البيانات في ثلاثة أعمدة أحدها الوزن Weight، العمر Age ، وضغط الدم Blood Pressure كما يلي:

	Weight	Age	BloodPressure
1	135.00	3.00	89.00
2	120.00	4.00	90.00
3	100.00	3.00	83.00
4	105.00	2.00	77.00
5	130.00	4.00	92.00
6	125.00	5.00	98.00
7	125.00	2.00	82.00
8	105.00	3.00	85.00
9	120.00	5.00	96.00
10	90.00	4.00	95.00
11	120.00	2.00	80.00
12	95.00	3.00	79.00
13	120.00	3.00	86.00
14	150.00	4.00	97.00
15	160.00	3.00	92.00
16	125.00	3.00	88.00

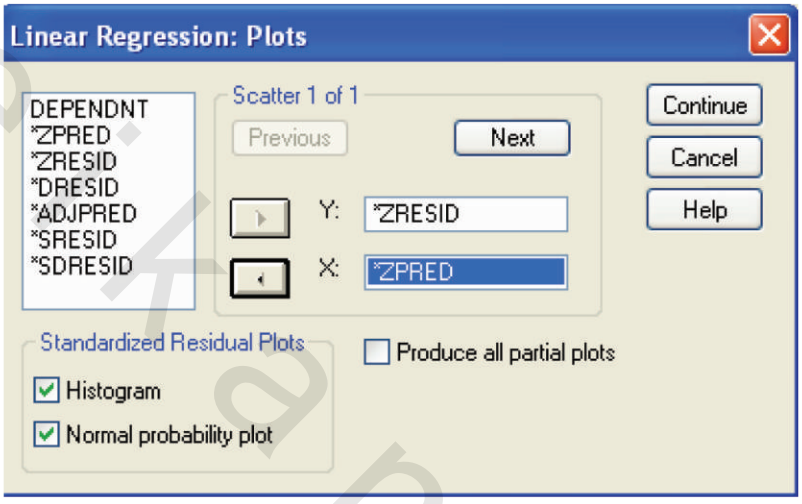
شكل (9-16)

نقوم بتفعيل نافذة الانحدار شكل (9-17) ونختار المتغيرات المستقلة هي Weight، Age، أما المتغير التابع فهو Blood Pressure كما يلي:



شكل (9-17)

بعـد ذلك نقـوم باختيار الأمر الفرعي Statistics و نتأكد من اختيـار كل من Estimates، Confidence intervals، Covariance matrix، Model fit، Casewise diagnostics كمـا في الانحـدار الخطي البسيط، ثم نعـود إلى مربـع الحوار السابق بالنقر علـى Continue ونذهب إلى الأمر الفرعـي Plots ونتأكد من اختيار Histogram وNormal probability plot وأيضًـا نضـع *ZRESID في المستطيـل Y، ونضـع *ZPRED في المستطيل X كما يلي:



شكل (9-18)

ثم نعود إلى مربع الحوار السابق بالنقر على Continue ثم ننقر OK فتظهر لدينا النتائج الآتية: يظهر لدينا جدول يبـين الطريقة المستخدمة في الانحدار والمتغير التابع وهوBlood Pressure والمتغيرات المستقلة الداخلة في نموذج الانحدار الخطي وفي حالتنا يوجد متغيران وهما Age، Weight كما يلي:

جدول (9-12)

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Age, Weight[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: BloodPresure

بعض المقاييس التي تم حسابها للنموذج:

جدول (9-13)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.939[a]	.881	.863	2.47917

a. Predictors: (Constant), Age, Weight

b. Dependent Variable: BloodPresure

لدينا المقياس R وكما نلاحظ فقيمته عالية وأيضا لدينا المقياس R Square الذي قيمته تقريبًا .88 وقيمته المعدلة 0.863، أي إن المتغيرين المستقلين في هذا النموذج لهما تأثير بمقدار 86,3% في المتغير التابع، كما أنه لدينا الخطأ المعياري للتقدير في النموذج وهو 2.48 تقريبًا.

جدول تحليل التباين:

جدول (9-14)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	591.036	2	295.518	48.081	.000[a]
	Residual	79.902	13	6.146		
	Total	670.938	15			

a. Predictors: (Constant), Age, Weight

b. Dependent Variable: BloodPresure

وهو يظهر لنا قيم مجاميع المربعات ودرجات الحرية الخاصة بكل مجموع مربعات، بالإضافة إلى متوسطي مربعات الخطأ والبواقي، كما نلاحظ أنه عند إجراء اختبار المعنوية عند مستوى 0,05 نجد أن نموذج الانحدار المتعدد معنوي من آخر قيمة في الجدول.

جدول المعالم المقدرة:

جدول (9-15)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	53.450	4.532		11.794	.000	43.660	63.241
	Weight	.126	.034	.352	3.657	.003	.051	.200
	Age	5.888	.680	.833	8.656	.000	4.418	7.357

a. Dependent Variable: BloodPresure

obeikandl.com

يختلف هذا الجدول عن جدول المعالم المقدرة الخاص بالانحدار الخطي البسيط بوجود أكثر من معلمتين فهنا لدينا ثلاثة معالم لنموذج الانحدار المتعدد وتكون معادلة الانحدار المقدرة:

Blood Pressure = 53.45 +0 .126Weight + 5.888Age

حيث:

$$\hat{\beta}_0 = 53.45$$

$$\hat{\beta}_1 = 0.126$$

$$\hat{\beta}_2 = 5.888$$

يمكننا أيضًا القيام باختبارات المعنوية لهذه المعالم كل على حده بمقارنة آخر قيمة في الجدول بمستوى المعنوية فنجد أن جميع المعالم معنوية.

جدول الارتباطات والتغايرات:

Coefficient Correlations **(a)**

جدول (9-16)

Model			Age	Weight
1	Correlations	Age	1.000	-.107
		Weight	107.-	1.000
	Covariances	Age	463.	-.002
		Weight	002.-	001.

a Dependent Variable: BloodPressure

يظهر لنا هذا الجدول قيم الارتباطات والتغايرات بين المتغيرات Age, Weight.

obeikandi.com

جدول البواقي:

Residuals Statistics **(a)**

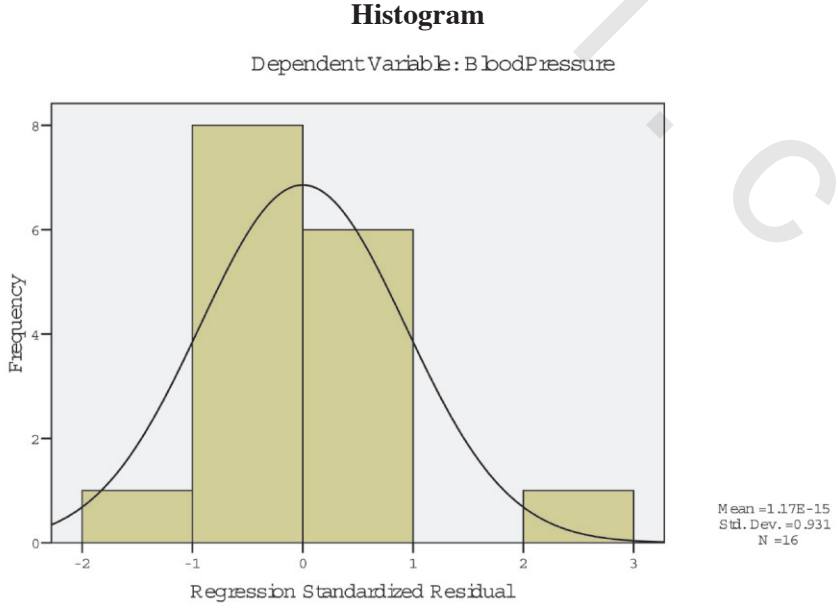
جدول (9-17)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	78.4119	98.5867	88.0625	6.27713	16
Std. Predicted Value	-1.537	1.677	000.	1.000	16
Standard Error of Predicted Value	655.	1.532	1.043	261.	16
Adjusted Predicted Value	78.8120	98.8089	87.9800	6.31927	16
Residual	-4.04376	6.69644	00000.	2.30798	16
Std. Residual	-1.631	2.701	000.	931.	16
Stud. Residual	-1.808	3.208	014.	1.074	16
Deleted Residual	-4.96979	9.44826	08249.	3.08420	16
Stud. Deleted Residual	-2.008	6.756	223.	1.871	16
Mahal. Distance	109.	4.793	1.875	1.358	16
Cook's Distance	000.	1.410	123.	349.	16
Centered Leverage Value	007.	320.	125.	091.	16

a Dependent Variable: BloodPressure

الأشكال البيانية:

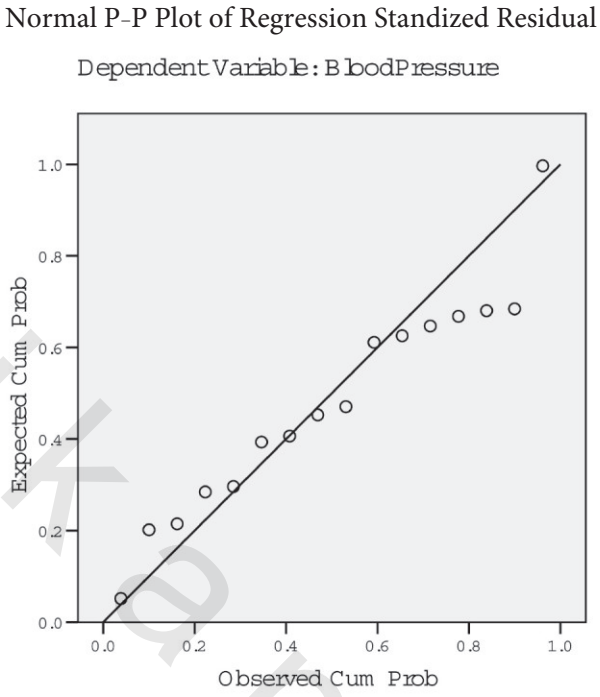
رسم المدرج التكراري:



شكل (9-19)

نستطيع منه القول إن البيانات تتوزع وفق التوزيع الطبيعي.

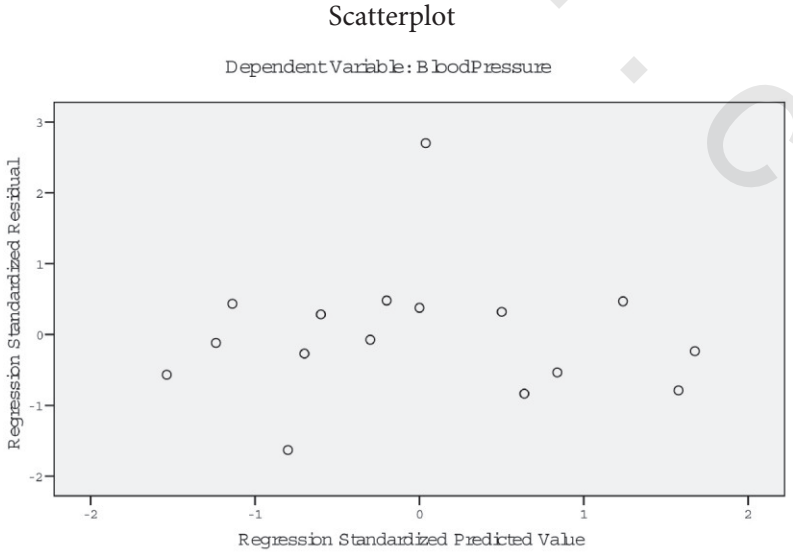
رسم منحنى الانحدار:



شكل (9-20)

هنـا أيضاً البيانات لها تقريبًا نمط انحـدار واضح على الرغم من وجود بعض القيم الشاذة عن خط الانحدار شذوذًا كبيرًا.

رسم انتشار البواقي:



شكل (9-21)

obeikandi.com

مـن هـذا الشكل نستطيع أن نقول إنـه لا يوجد نمط معين لانتشار البواقي ضد القيـم المتوقعة، أي إن الانتشار عشوائي، وهذا يتفق مع شرط الخطية المطلوب في الانحدار سواء كان بسيطًا أم متعددًا.

(4-9) اختيار المتغيرات المستقلة الداخلة في النموذج الخطي:

المعلومـات النظريـة التي ستستخدم لاختيـار المتغيرات المستقلة قد يمكن تعيينهـا بسهولة في بعض الميادين، وغالبًا ما يمكن القيـام بتجارب تحت سيطرة المجرب تزوده ببيانات يمكـن على أساسها تقدير معـالم الانحـدار واختبار الشكل النظري لدالة الانحـدار. لكن في العديد من الميادين الأخرى مثل العلوم الاجتماعيـة، السلوكية، الصحية، والإدارية، يندر نسبيًا وجود نمـاذج قابلة للاستخدام، ولمزيد من تعقيد الأمـور قد تنطوي النماذج النظرية المتوافرة على متغيرات مستقلة غير قابلة للقياس مباشرة مثل الدخول المستقبلية للأسـرة في السنوات العشر القادمة، وتحت هذه الظروف يضطـر الباحثون إلى توقع متغيرات مستقلة يتصورون أنها يمكن أن تكون على صلة بالمتغير التابع المراد دراسته، ومن الواضح أن هذه المتغيرات المستقلة قد تكون مجموعة كبيرة وليس بالضرورة أن يكون لها تأثير في المتغير المراد دراسته فيلجأ الباحثون إلى عدة طرق يمكن أن يقوموا من خلالها بتحديد المتغيرات التي لها تأثير في المتغير المدروس حتى يتمكنوا مـن التوصل إلى نمـوذج انحدار نهائي يحتوي علـى أفضل المتغيرات التي لها تأثـير في النموذج، وسنقوم بدراسة إحدى هذه الطرق.

- الانحدار باستخدام الانحدار التدريجي.

تقوم طريقة الانحدار التدريجي التي تسمى أيضًا انحدار الخطوة فخطوة إلى الأمام، بالإضافة إلى تقدير معالم نموذج الانحدار المقدر باختيار أفضل نماذج انحدار يمكن التوصل إليها من متغيرات مستقلة عـدة مع متغـير تابع، حيث إنه ليـس بالضرورة أن تكون لجميـع المتغيرات المستقلة التـي اختارها الباحث التأثـير نفسـه في المتغير التابع، وهذا يـؤدي إلى أنه يمكـن ترشيح عدة نماذج انحـدار بمتغيرات مستقلة مختلفة وللمتغير التابع نفسه على أنها تكون نماذج جيدة.

يقـوم روتـين الانحـدار التدريجـي أولا بتوفيق نمـوذج انحدار خطـي بسيط لكل مـن المتغيرات X المرشحة وعددها p-1، حيث نقوم في كل نموذج انحدار من النماذج التي حصلنا عليها باختبار ما إذا كان الميل مساويًا للصفر أم لا، حيث نستخدم الإحصاءة الآتية:

$$F_k^* = \frac{MSR\ (X_k)}{MSE\ (X_k)}\ ,\ k = 1,2,...,\ P\text{-}1$$

حيـث إن $MSR\ (X_k)$ يقيس الانخفاض في التغير الكلي للمتغير التابع المصاحب لاستخدام المتغير Xk، والمتغـير Xi الـذي له أعلى قيمة في الإحصـاءة يكون هو المتغير المرشح لأول إضافـة في النموذج، حيث إذا تجاوزت قيمة *F مستوى محددًا سلفا يضاف المتغير X، وفيما عدا ذلك ينتهي البرنامج معتبرا أنه لا يوجد

obeikandl.com

أي متغير X مفيد بما يكفي لدخول نموذج الانحدار. لفرض أن المتغير X7 هو المتغير الذي تم اختياره في أول محاولة. سيقوم البرنامج بعدها بتوفيق جميع نماذج الانحدار المتضمنة لمتغيرين مستقلين أحدهما X7، ولكل نموذج نقوم باختبار ما إذا كان الميل مساويًا للصفر مستخدمين الإحصاءة:

$$F_k^* = \frac{MSR\ (X_k\ /\ X_7)}{MSR\ (X_k\ /\ X_7)}$$

فإذا تجاوزت قيمة F^* مستوى $F_{(1-a,1,n-p)}$ الذي نحدده سلفا يضاف المتغير الثاني، وفيما عدا ذلك ينتهي البرنامج باختيار المتغير X_7 فقط.

لنفرض أن X3 تمت إضافته في الخطوة السابقة عندئذ سيقوم روتين الانحدار التدريجي باختبار ما إذا كان ينبغي حذف أي من المتغيرات المستقلة الموجودة في النموذج، وبناءً عليه إما أن تحذف متغيرات أو يظل النموذج بالمتغيرات التي قد اختيرت سابقًا، ثم نرجع مرة أخرى إلى إضافة متغيرات مستقلة في النموذج ثم إلى حذف المتغيرات كما في الخطوات السابقة إلى أن نصل إلى مرحلة لا يمكننا فيها إضافة أو حذف متغير مستقل من النموذج، وعندها تنتهي عملية البحث عن النموذج، ويكون النموذج الموجود لدينا هو نموذج الانحدار المقدر بأفضل المتغيرات المستقلة.

مثال (9-4)

نريد معرفة المتغيرات التي لها تأثير في المصروفات المعيشية للأسرة قمنا باختيار 30 أسرة وسجلنا قيم المتغيرات الآتية لها:

جدول (9-18)

X4i	X3i	X2i	X1i	Yi	i
5	7	2	6	6,5	1
6	13	3	16	11,2	2
8	19	4	15	11,2	3
8	13	4	14	10,5	4
9	9,6	2	10	9,3	5
5	8	3	10	7,2	6
7	15	4	9	13,4	7
7	5	3	6	5	8
6	12	4	7	11,6	9
7	14	4	12	11,2	10
5	7	3	4	6,2	11
11	19	6	0	12	12

obeikandl.com

9	16	6	14	11,3	13
8	11	4	5	9,2	14
9	6	4	2	5,5	15
7	7,6	3	4	6	16
8	25	1	16	12,5	17
6	10	2	4	9,8	18
5	6,5	3	5	6,1	19
11	15,1	6	15	14,3	20
10	18	6	15	45,5	21
6	11	1	2	10,8	22
5	5,6	1	4	4,5	23
8	8,5	2	3	6,7	24
5	6,3	1	2	4,5	25
5	14	3	5	9,8	26
7	4,6	1	5	4	27
7	7,5	1	3	5,5	28
7	11,2	2	8	10	29
5	9,5	0	7	8,5	30

حيث:

Yi: المصروفات المعيشية (ألف جنيه مصري).

X1i: مستوى تعليم رب الأسرة (عدد سنوات دراسته).

X2i: عدد الأطفال.

X3i: دخل الأسرة (ألف جنيه مصري).

X4i: عدد أفراد الأسرة.

نريد إيجاد نموذج الانحدار الخطي الخاص بالمصروفات المعيشية للأسر عند مستوى معنوية 0.05.

الحل:

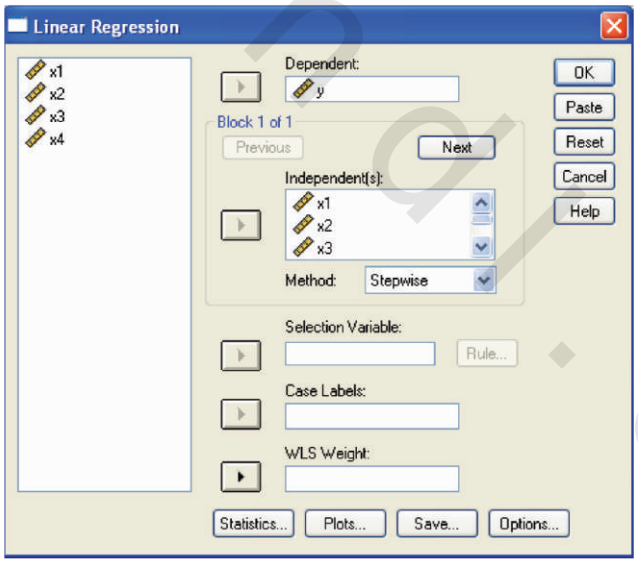
ندخل البيانات في SPSS البرنامج:

obeikandi.com

	y	x1	x2	x3	x4
1	6.50	6.00	2.00	7.00	5.00
2	11.20	16.00	3.00	13.00	6.00
3	11.20	15.00	4.00	19.00	8.00
4	10.50	14.00	4.00	13.00	8.00
5	9.30	10.00	2.00	9.60	9.00
6	7.20	10.00	3.00	8.00	5.00
7	13.40	9.00	4.00	15.00	7.00
8	5.00	6.00	3.00	5.00	7.00
9	11.60	7.00	4.00	12.00	6.00
10	11.20	12.00	4.00	14.00	7.00
11	6.20	4.00	3.00	7.00	5.00
12	12.00	.00	6.00	19.00	11.00
13	11.30	14.00	6.00	16.00	9.00
14	9.20	5.00	4.00	11.00	8.00

شكل (9-22)

نقـوم بتفعيـل نافذة الانحـدار ونختار المتغـير التابع y والمتغـيرات المستقلة هـي x1 x2 x3 x4 ونقوم باختيار الطريقة Stepwise كما يلي:



شكل (9-23)

بالنسبـة للخيـارات Statistics، Plots، Save نقـوم بما قمنا به في الانحدار المتعـدد ثم ننقر ok فينتج لدينا:

جدول المتغيرات الداخلة في النموذج:

جدول (9-19)

Model	Variables Entered	Variables Removed	Method
1	x3	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050، Probability-of-F-to-remove >= .100).

Variables Entered/Removed (a)

a Dependent Variable :y

يتضح لنا من هذا الجدول أنه تم اختيار المتغير X3 فقط في النموذج، أي إن دخل الأسرة هو المتغير الوحيد المؤثر في المصروفات المعيشية من بين المغيرات التي قمنا باختيارها.

بعض مقاييس النموذج:

جدول (9-20)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.570(a)	.325	.301	6.11488

a، 3x Predictors: (Constant)

b Dependent Variable: y

نلاحظ من قيمة R Square المعدلة أن المتغير X3 يفسر تقريباً 30% من Y.

جدول تحليل التباين:

ANOVA (b)

جدول (9-21)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	505.129	1	505.129	13.509	.001(a)
	Residual	1046.970	28	37.392		
	Total	1552.099	29			

a، 3x Predictors: (Constant)

b Dependent Variable: y

obeikandl.com

من هذا الجدول نستطيع اختبار معنوية الانحدار، وذلك باستخدام الاختبار الآتي:

الفروض الإحصائية:

H0: الانحدار غير معنوي (غير دال).

H1: الانحدار معنوي (دال).

جدول (9-22)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.553	2.801		.197	.845	-5.184	6.290
	x3	.845	.230	.570	3.675	.001	.374	1.317

خطوات الاختبار:

نقارن P-value = .001 بمستوى المعنوية 0.05، نلاحظ أن القيمة P-value أقل من مستوى المعنوية.

القرار الإحصائي:

نقوم برفض الفرض العدمي، وهذا معناه أن الانحدار معنوي.

جدول المعاملات:

Coefficients **(a)**

a Dependent Variable: y

من جدول المعاملات نجد أن معادلة الانحدار المقدرة هي:

$$Y=0.553 + 0.845X_3$$

حيث تكون قيمة $\hat{\beta}_0 = 0.553$, $\hat{\beta}_3 = 0.845$ فلنختبر معنوية β_3 عند مستوى معنوية 0.05:

الفروض الإحصائية:

فرض العدم هو: $H_0 : \beta_3 = 0$

ضد الفرض البديل: $H_A : \beta_3 \neq 0$

خطوات الاختبار:

نقارن *P-value =.001* بمستوى المعنوية *0.05* نلاحظ أن القيمة *P-value* أقل من مستوى المعنوية.

obeikandl.com

القرار الإحصائي:

نقوم برفض الفرض العدمي وهذا معناه أن β_3 معنوي.

جدول المتغيرات المستبعدة من النموذج:

Excluded Variables **(b)**

جدول (9-23)

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	x1	.221(a)	1.120	.273	.211	.614
	x2	.312(a)	1.859	.074	.337	.786
	x4	.218(a)	1.211	.237	.227	.729

a Predictors in the Model: (Constant)، 3x

b Dependent Variable: y

جدول البواقي:

Residuals Statistics **(a)**

جدول (9-24)

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	4.4418	21.6882	9.9933	4.17352	30
Std. Predicted Value	-1.330	2.802	.000	1.000	30
Standard Error of Predicted Value	1.116	3.372	1.514	.457	30
Adjusted Predicted Value	4.4878	25.7033	10.0856	4.57232	30
Residual	-9.18822	29.72967	00000.	6.00853	30
Std. Residual	-1.503	4.862	000.	.983	30
Stud. Residual	-1.801	5.123	007.-	1.048	30
Deleted Residual	-13.20330	33.01107	09227.-	6.84940	30
Stud. Deleted Residual	-1.881	20.104	.491	3.727	30
Mahal. Distance	.000	7.852	.967	1.485	30
Cook's Distance	.000	1.448	.076	.290	30
Centered Leverage Value	.000	.271	.033	.051	30

a Dependent Variable: y

obeikandl.com