الارتباط والانحدار الخطي البسيط

مقـدمة

في السابق تم عرض بعض المقاييس الوصفية، مثل مقاييس النزعة المركزية، والتشتت، ومقاييس الالتواء والتفرطح، وغيرها من المقاييس الأخرى والتي يمكن من خلالها وصف شكل توزيع البيانات التي تم جمعها عن متغير واحد، وننتقل من التعامل مع متغير واحد إلى التعامل مع متغيرين أو أكثر لدراسة وتحليل العلاقة بينها، ويتم ذلك باستخدام بعض طرق التحليل الإحصائي مثل تحليل الارتباط، والانحدار الخطي البسيط، فإذا كان اهتمام الباحث هو دراسة العلاقة بين متغيرين استخدم لذلك أسلوب تحليل الارتباط، وإذا كان اهتمامه بدراسة أثر أحد المتغيرين على الآخر استخدم لذلك أسلوب تحليل الانحدار الخطي البسيط ، ومن الأمثلة على ذلك:

1- الإنفاق، والدخل العائلي.

2- سعر السلعة، والكمية المطلوبة منها.

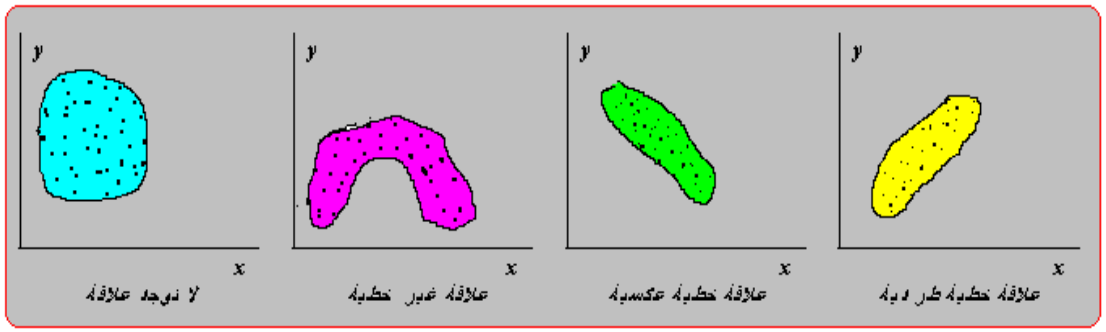
3- عدد مرات ممارسة نوع معين من الرياضة البدنية، ومستوى الكلسترول في الدم.

4- وزن الجسم، وضغط الدم.

والأمثلة على ذلك في المجال التطبيقي كثيرة، فإذا كان لدينا المتغيرين $(x\ ,\ y)$ ، وتم جمع بيانات عن أزواج قيم هذين المتغيرين، وتم تمثيلها بيانيا فيما يسمى بشكل الانتشار، فإن العلاقة بينها تأخذ أشكالا مختلفة على النحو التالي :

شكل (1)

شكل الانتشار لبيان نوع العلاقة بين **x , y**



الارتباط الخطى البسيط Simple Correlation

يهتم هذا الفصل بعرض أسلوب تحليل الارتباط الخطي البسيط، أي في حالة افتراض أن العلاقة بين المتغيرين تأخذ الشكل الخطي، وسوف يجرى حسابه في حالة البيانات الكمية **Quantitative**، وكذلك البيانات الوصفية الرتبية **Ordinal Qualitative**.

الغرض من تحليل الارتباط الخطى البسيط

الغرض من تحليل الارتباط الخطي البسيط هو تحديد نوع وقوة العلاقة بين متغيرين، ويرمز له في حالة المجتمع بالرمز ρ (رو)، وفي حالة العينة بالرمز r، وحيث أننا في كثير من النواحي التطبيقية نتعامل مع بيانات عينة مسحوبة من المجتمع، سوف نهتم بحساب معامل الارتباط في العينة r كتقدير لمعامل الارتباط في المجتمع، ومن التحديد السابق للغرض من معامل الارتباط، نجد أنه يركز على نقطتين هما:

- نوع العلاقة:. وتأخذ ثلاث أنواع حسب إشارة معامل الارتباط كما يلي:

1- إذا كانت إشارة معامل الارتباط سالبة ($r < 0$) توجد علاقة عكسية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه انخفاض في المتغير الثاني، والعكس صحيح.

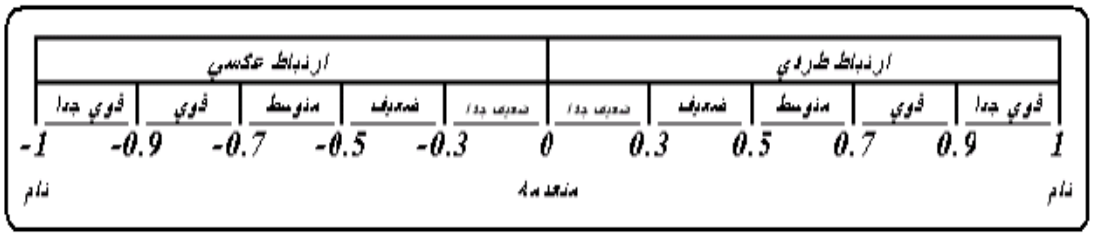
2- إذا كانت إشارة معامل الارتباط موجبة ($r > 0$) توجد علاقة طردية بين المتغيرين، بمعنى أن زيادة أحد المتغيرين يصاحبه زيادة في المتغير الثاني، والعكس صحيح .

3- إذا كان معامل الارتباط قيمته صفرا ($r = 0$) دل ذلك على انعدام العلاقة بين المتغيرين.

- قوة العلاقة:. ويمكن الحكم على قوة العلاقة من حيث درجة قربها أو بعدها عن (± 1)، حيث أن قيمة معامل الارتباط تقع في المدى($-1 < r < 1$)، وقد صنف بعض الإحصائيين درجات لقوة العلاقة يمكن تمثيلها على الشكل التالي:

شكل (2)

درجات معامل الارتباط



معامل الارتباط الخطى البسيط " لبيرسون" **Pearson**

في حالة جمع بيانات عن متغيرين كميين $(x \ , \ y)$، يمكن قياس الارتباط بينهما، باستخدام طريقة "بيرسون" **Pearson**، ومن الأمثلة على ذلك: قياس العلاقة بين الوزن والطول، والعلاقة بين الإنتاج والتكلفة، والعلاقة بين الإنفاق الاستهلاكي والدخل، والعلاقة بن الدرجة التي حصل عليها الطالب وعدد ساعات الاستذكار، وهكذا الأمثلة على ذلك كثيرة.

ولحساب معامل الارتباط في العينة ، تستخدم صيغة " بيرسون" التالية :

$$r = \frac{S_{xy}}{S_x S_y} = \frac{\dfrac{\sum (x - \bar{x})(y - \bar{y})}{(n-1)}}{\sqrt{\dfrac{\sum (x - \bar{x})^2}{(n-1)}}\sqrt{\dfrac{\sum (y - \bar{y})^2}{(n-1)}}}$$

حيث أن :

$S_{xy} = \sum (x - \bar{x})(y - \bar{y}) \big/ (n-1)$: هو التغاير بين $(x \ , \ y)$،

$S_x = \sqrt{\sum (x - \bar{x})^2 / (n-1)}$: هو الانحراف المعياري لقيم (x)،

$S_y = \sqrt{\sum (y - \bar{y})^2 / (n-1)}$: هو الانحراف المعياري لقيم (y).

ويمكن اختصار الصيغة السابقة على النحو التالي:

$$r = \frac{S_{xy}}{S_x S_y} = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

تطبيقات على الارتباط باستخدام برنامج SPSS

تطبيق:1 (محلول)

فيما يلي متوسطات أسعار كيلو لحوم الأغنام المحلية والمستوردة بالريال خلال الفترة 2007م – 2016م،

السنوات	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
أسعار محلية	26.2	25.5	26.2	27.1	26.4	26.1	25.7	26.2	31.6	32.1
أسعار مستورد	17.2	16.9	17.4	18.1	18.5	20.1	20.5	20.6	24.1	24.6

والمطلوب

1- أدخل البيانات باستخدام برنامج **SPSS**.

2- ارسم نقط الانتشار وحدد الاتجاه العام للعلاقة بين أسعار اللحوم المحلية، وأسعار اللحوم المستوردة.

3- استخدم البرنامج في الحصول على قيمة معامل الارتباط الخطي لبيرسون، وما هو مدلوله؟

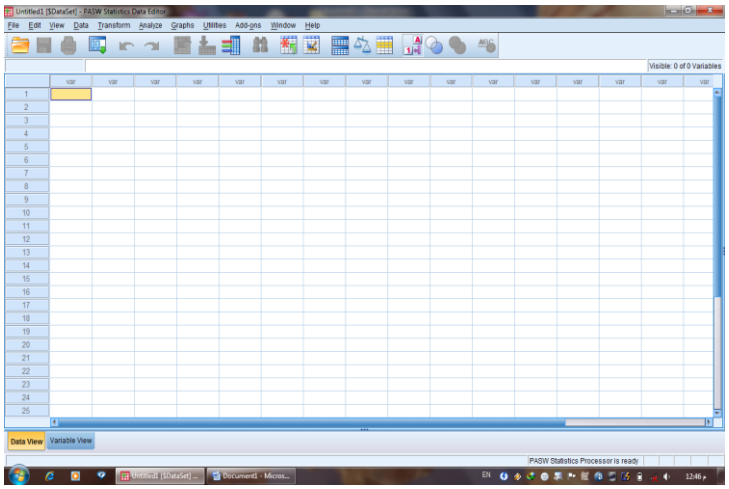
الحل

1- إدخال البيانات

قبل الشروع في حل التطبيق لابد أن تعرف المتغيرات وعددها فنلاحظ أنه يوجد متغيران وهما المتغير الأول الأسعار المحلية وسوف ندخلها باسم **Local_p** ، والمتغير الثاني وهو الأسعار المستوردة وسوف ندخلها باسم **Import_p**

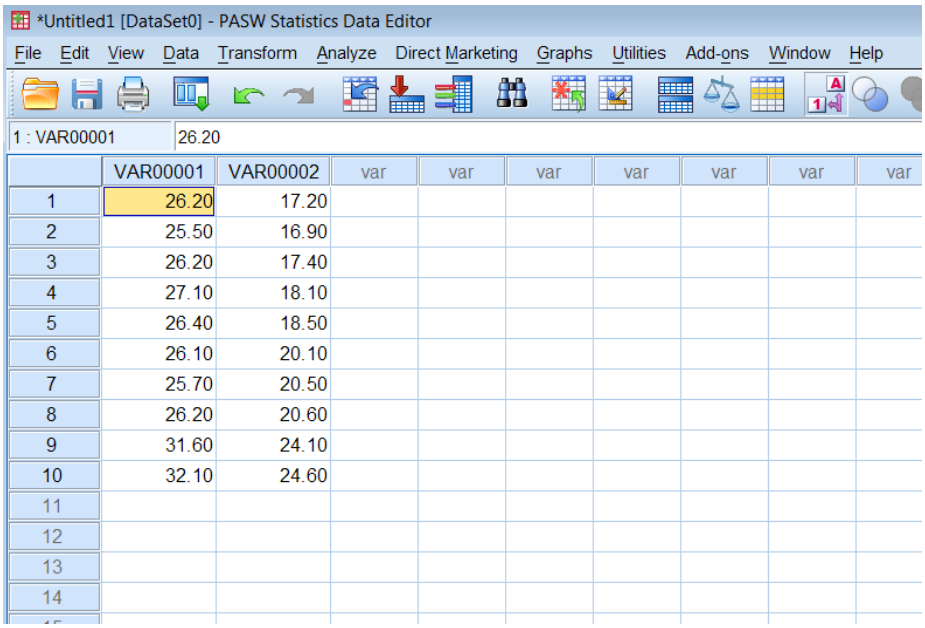
ثانياً: للحصول على شكل الانتشار، وقيمة معامل الارتباط الخطي لبيرسون، يتم ادخال البيانات على البرنامج من خلال إتباع التالي:

- ندخل البيانات في صفحة **Data view** وذلك عن طريق النقر مرتين **Double-click** بزر الماوس على أيقونة أو رمز البرنامج في حال وجوده على سطح المكتب، أو من خلال سطح المكتب نضغط بالفأرة على كلمة **Start** تظهر قائمة نختار منها **Programs** ومنها نختار برنامج **SPSS**.فتظهر صفحة **Data view** كالتالي:



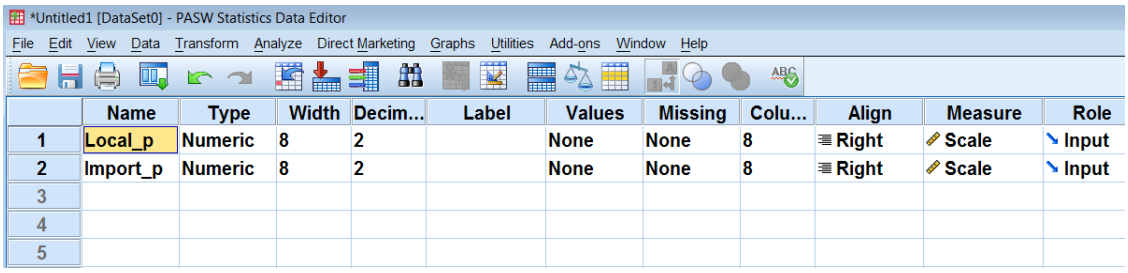
- إدخال البيانات المستهدفة الأسعار المحلية في ثاني أعمدة الملف الفارغ من جهة اليسار وذلك بالوقوف بالسهم على أول خانة من خانات العمود الخاص بالأسعار المحلية لبدء عمليات الإدخال.

- إدخال المتغير الثاني الأسعار المستوردة في ثالث أعمدة الملف الفارغ من جهة اليسار وذلك بالوقوف بالسهم على أول خانة من خانات العمود الخاص بالأسعار المستوردة لبدء عمليات الإدخال.



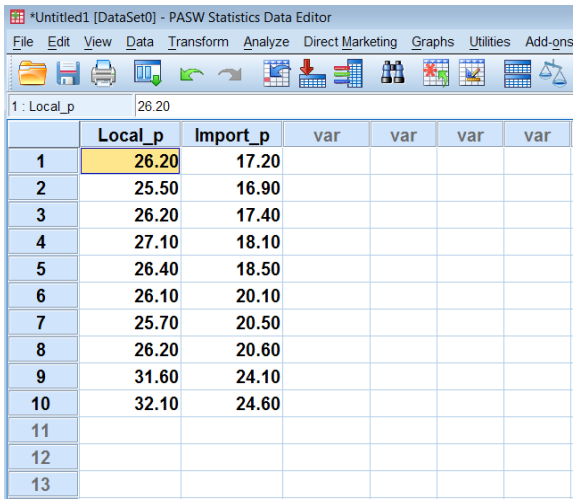
	VAR00001	VAR00002	var	var	var	var	var	var	var
1	26.20	17.20							
2	25.50	16.90							
3	26.20	17.40							
4	27.10	18.10							
5	26.40	18.50							
6	26.10	20.10							
7	25.70	20.50							
8	26.20	20.60							
9	31.60	24.10							
10	32.10	24.60							
11									
12									
13									
14									

- تسمية المتغيرات التي تم إدخالها وذلك بالضغط على [**Variable View**] (رؤية المتغير)، بشريط التعليمات السفلي، يظهر لنا جدول يحتوي على كل الخصائص الممكنة للمتغيرات التي تم إدخالها.
- يتم إدخال الاسم (**Local_p**) للمتغير الأول في الخانة الأولى تحت[Name] والاسم **Import_p** للمتغير الثاني.



	Name	Type	Width	Decim...	Label	Values	Missing	Colu...	Align	Measure	Role
1	Local_p	Numeric	8	2		None	None	8	Right	Scale	Input
2	Import_p	Numeric	8	2		None	None	8	Right	Scale	Input
3											
4											
5											

نضغط أسفل الصفحة على **Data View** فتظهر البيانات كالتالي:.

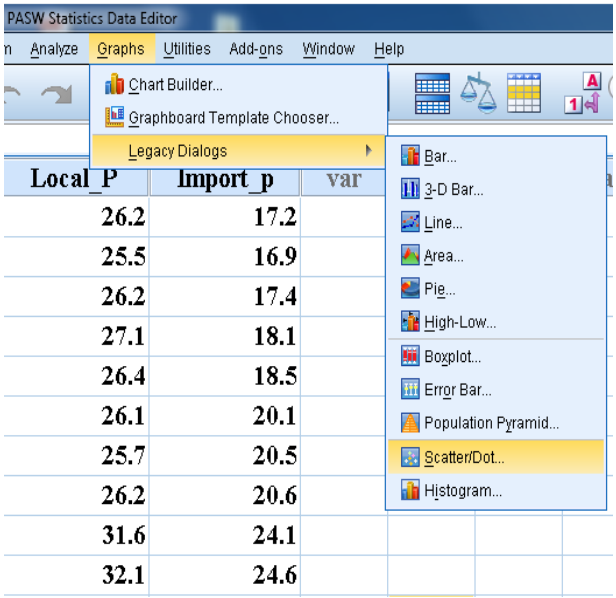


	Local_p	Import_p	var	var	var	var
1	26.20	17.20				
2	25.50	16.90				
3	26.20	17.40				
4	27.10	18.10				
5	26.40	18.50				
6	26.10	20.10				
7	25.70	20.50				
8	26.20	20.60				
9	31.60	24.10				
10	32.10	24.60				
11						
12						
13						

1- رسم نقط الانتشار.

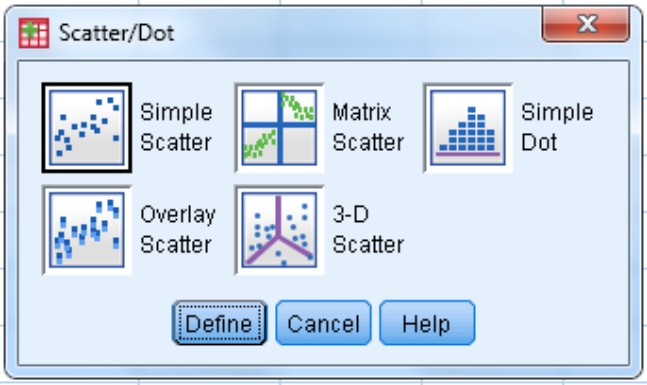
- فتح قائمة الأشكال البيانية Graphs من شريط القوائم، ثم اختيار النافذة Legacy Dialogs ، ثم اختيار اسم الرسم Scatter/Dot...

كما هو موضح بالشكل التالي.



Local_P	Import_p	var
26.2	17.2	
25.5	16.9	
26.2	17.4	
27.1	18.1	
26.4	18.5	
26.1	20.1	
25.7	20.5	
26.2	20.6	
31.6	24.1	
32.1	24.6	

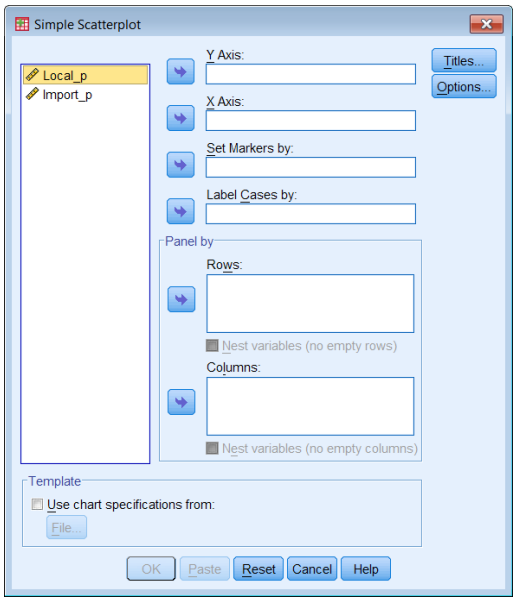
وعند النقر على Scatter/Dot... تظهر لنا النافذة التالية.

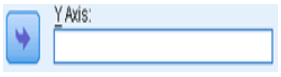


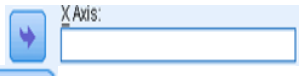
- يتم تنشيط نافذة الانتشار البسيط

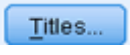
ثم النقر على نافذة

يظهر لنا الشكل التالي.

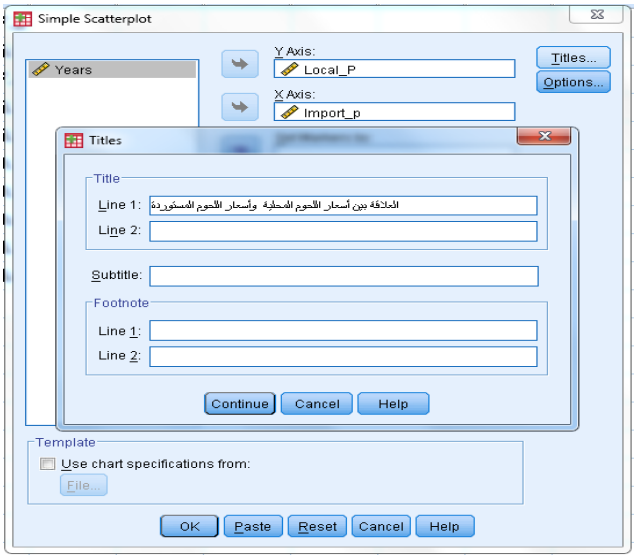


- يتم إدخال بيانات **Local_p** لتمثل على المحور الرأسي 

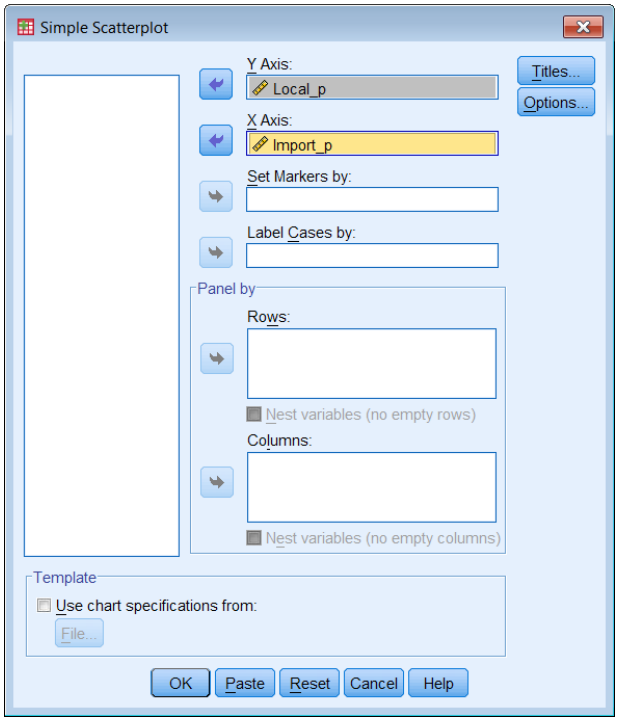
- يتم إدخال بيانات **import_p** لتمثل على المحور الأفقي 

- يتم النقر على نافذة 

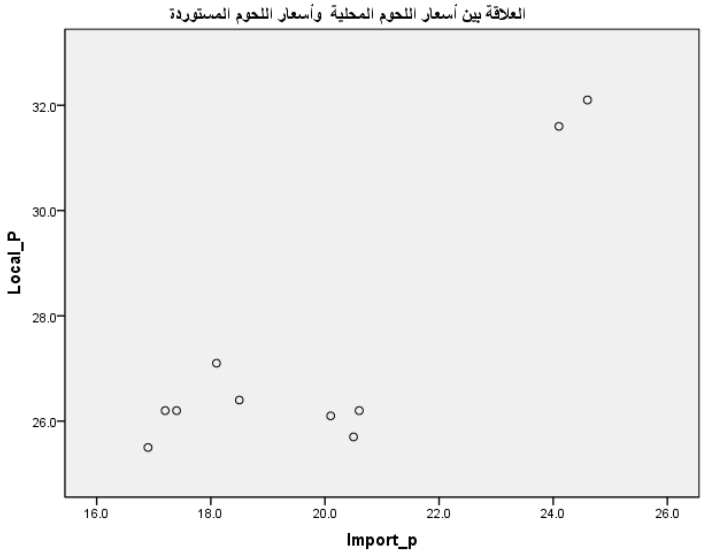
لكتابة عنوان للشكل. تظهر لنا النافذة التالية:



- يتم الانتقال للشاشة السابقة بالنقر على النقر على استمرار Continue



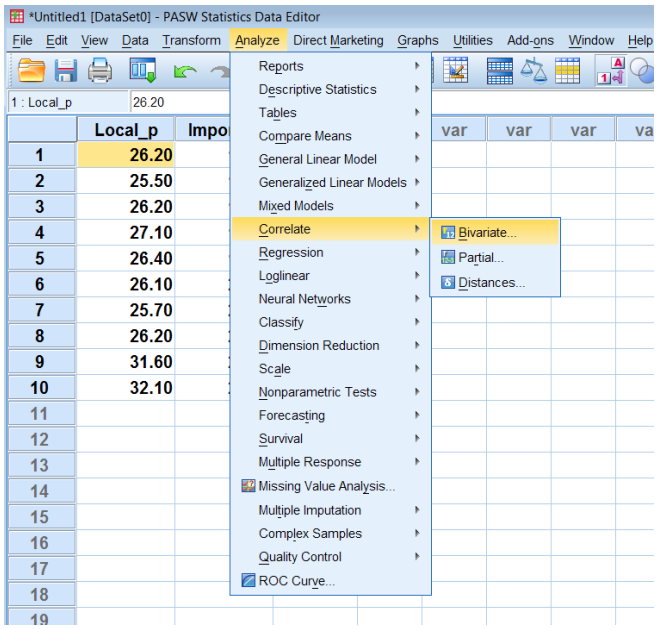
- وأخيرا يتم النقر على OK يظهر شكل الانتشار التالي.

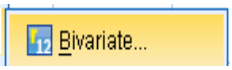


- يدل الشكل على وجود اتجاه عام طردي بين أسعار اللحوم المحلية وأسعار اللحوم المستوردة.

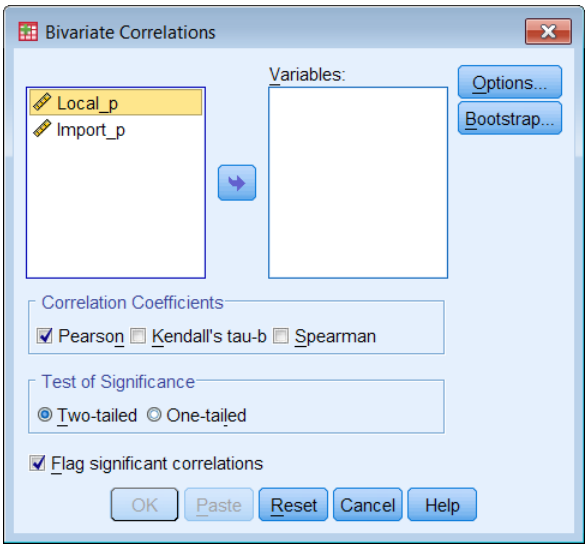
2- استخدام البرنامج في الحصول على قيمة معامل الارتباط الخطي لبيرسون.

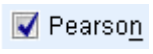
- إدخال البيانات ثم فتح قائمة Analyze من شريط القوائم، ثم اختيار قائمة الارتباط Correlate ومنها اختار Bivariate... للحصول على الارتباط بين قيم متزاوجة لمتغيرين، ويظهر ذلك في الشكل التالي.



- وبفتح نافذة 

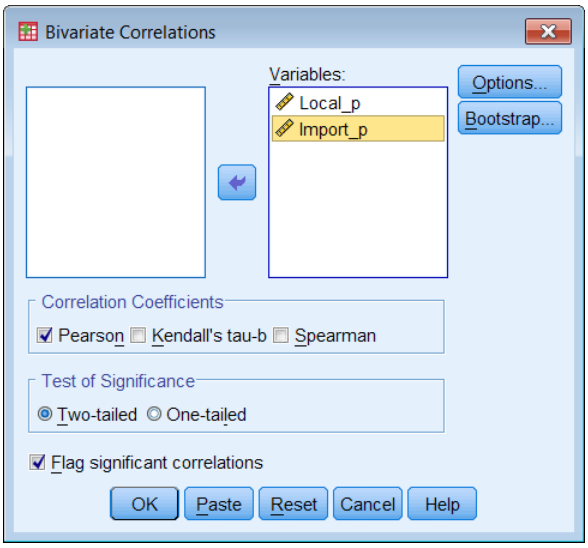
يظهر لنا الشكل التالي.



- يتم إدخال المتغيرين **Local_p** ، و **import_p** من القائمة اليسرى إلى قائمة المتغيرات اليمنى، ثم اختيار نوع الارتباط 

من الثلاثة أنواع للإرتباط

كما هو موضح بالشكل التالي:



- يتم النقر بعد ذلك على

، يظهر لنا النتائج التالية.

Correlations			
		Local_P	Import_p
Local_P	Pearson Correlation Sig. (2-tailed) N	1 10	.853** .002 10
Import_p	Pearson Correlation Sig. (2-tailed) N	.853** .002 10	1 10
**. Correlation is significant at the 0.01 level (2-tailed).			

- يتكون الجدول من جدولين جزئيين، كل جدول خاص بمتغير، ويوضح فيه معامل الإرتباط بين هذا المتغير وجميع المتغيرات، كما يوضح فيه معنوية كل معامل ارتباط (**2–tailed**) .**Sig** ، وعدد المشاهدات.
- ففي الجدول الخاص بأسعار لحوم الأغنام المحلية **Local_p**، نجد أن:

1– معامل الإرتباط بين أسعار لحوم الأغنام المحلية، ونفس الأأسعار هو $r_{Local_p,Local_p} = 1$ ، ويدل على وجود ارتباط تام.

2– أما معامل الارتباط بين أسعار لحوم الأغنام المحلية **Local_P** وأسعار اللحوم المستوردة **Import_p** قيمته هي: $r_{Local_p,Import_p} = 0.853$ ويدل ذلك على وجود علاقة طردية قوية بين النوعين من الأسعار.

3– قيمة المعامل **0.853** مصاحبة بنجمتين (**0.853****) ، يدل ذلك على أن معامل الإرتباط بين النوعين من الأسعار معنوي وذو دلالة عند مستوى معنوية 1%، ويلاحظ ذلك أيضا من المعنوية المحسوبة $Sig.(2-tailed) = 0.002$ ، والتي تقل قيمتها عن **0.01** .

4– قيمة **N=10** وتدل على أن عدد أزواج القيم يساوي **10**، أو أن ع\حجم العينة يساوي **10**.

تطبيق:2 (غير محلول)

البيانات التالية تم الحصول عليها من عينة عشوائية قوامها 25 موظف يعملون في شركة ما.

رقم المشاهدة	الراتب السنوي (ألف دولار)	سنوات الخدمة
1	35	15

17	27	2
25	45	3
12	22	4
2	25	5
10	30	6
17	37	7
17	25	8
1	17	9
4	28	10
25	43	11
15	25	12
1	22	13
6	28	14
20	29	15
3	19	16
21	29	17
19	38	18
5	19	19
1	22	20
20	39	21
22	40	22
10	21	23
7	28	24
8	30	25

المطلوب

1- أدخل البيانات باستخدام برنامج **SPSS**.

2- ارسم نقط الانتشار وحدد الاتجاه العام للعلاقة بين الراتب السنوي، سنوات الخدمة.

3- استخدم البرنامج في الحصول على قيمة معامل الارتباط الخطي لبيرسون، وما هو مدلوله؟

الانحدار الخطى 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:

1- دراسة أثر كمية السماد على إنتاجية الدونم من محصول زراعي معين .

2- دراسة أثر كمية البروتين التي يتناولها الطلاب على الزيادة في الوزن.

3- أثر الدخل على الإنفاق الاستهلاكي.

وهكذا الأمثلة على ذلك في النواحي التطبيقية، مثل العلوم الاقتصادية، والزراعية، والتجارية، والعلوم السلوكية، وغيرها من المجالات الأخرى كثيرة.

نموذج الانحدار الخطي

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:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

حيث أن:

y : هو المتغير التابع (الذي يتأثر)

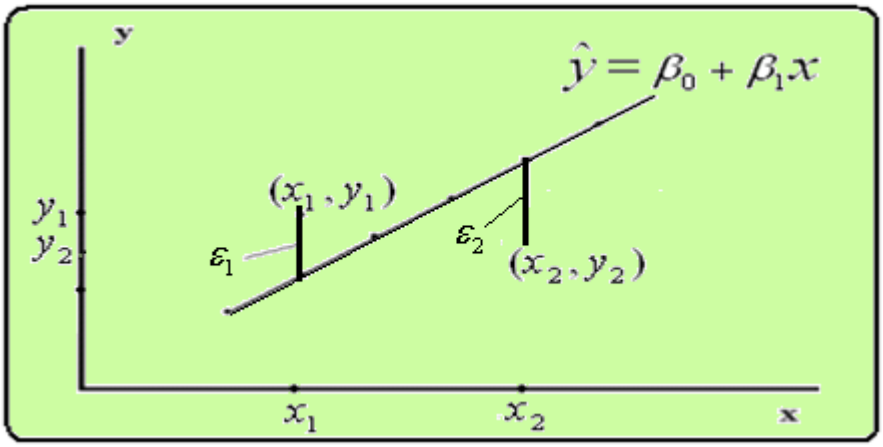
x : هو المتغير المستقل (الذي يؤثر)

β_0 : هو الجزء المقطوع من المحور الرأسي y ، وهو يعكس قيمة المتغير التابع في حالة انعدام قيمة المتغير المستقل x ، أي في حالة $x = 0$

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

ε : هو الخطأ العشوائي، والذي يعبر عن الفرق بين القيمة الفعلية y ، والقيمة المقدرة $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ ، أي أن : $\varepsilon = y - (\hat{\beta}_0 + \hat{\beta}_1 x)$ ، ويمكن توضيح هذا الخطأ على الشكل التالي لنقط الانتشار.

توضيح كيفية تحديد الخطأ العشوائي



تقدير نموذج الانحدار الخطي البسيط

يمكن تقدير معاملات الانحدار $(\beta_1 \ , \ \beta_0)$ في النموذج (5) باستخدام طريقة المربعات الصغرى، وهذا التقدير هو الذي يجعل مجموع مربعات الأخطاء العشوائية $\sum \varepsilon^2 = \sum \left(y - (\beta_0 + \beta_1 x)\right)^2$ أقل ما يمكن، ويحسب هذا التقدير بالمعادلة التالية:

$$\hat{\beta}_1 = \frac{\sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}}{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}$$

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

حيث أن $\bar{x}$ هو الوسط الحسابي لقيم x، $\bar{y}$ هو الوسط الحسابي لقيم y، ويكون تقدير قيمة المتغير التابع y عند قيمة محددة لـ x هي: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$، ويطلق عليها تقدير معادلة انحدار y على x، وهي معادلة خطية من الدرجة الأولى.

تطبيق: 3 (محلول)

لدراسة العلاقة بين الدخل والاستهلاك بالريال في مدينة الرياض، أخذت عينة مكونة من عشرة أسر فأعطت النتائج التالية:

الدخل	3000	3500	5000	6000	9000	10000	9000	12000	10500	2500
الاستهلاك	2800	3400	5000	5500	8000	7500	8500	10500	10000	2500

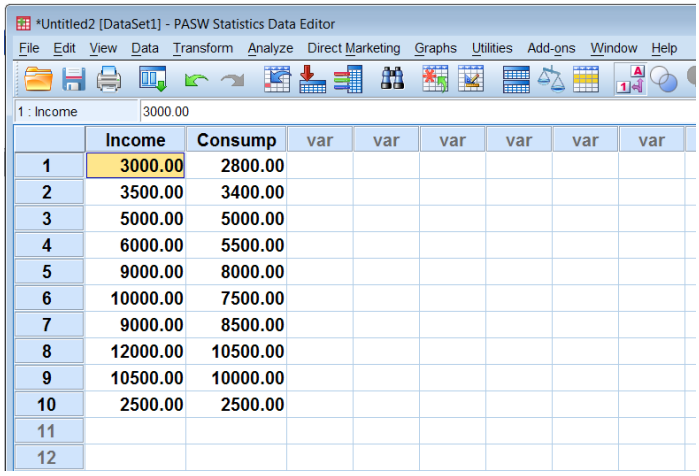
المطلوب: إيجاد نموذج انحدار الاستهلاك على الدخل.

الحل

1- إدخال البيانات

قبل الشروع في حل التطبيق لابد أن تعرف المتغيرات وعددها فنلاحظ أنه يوجد متغيران وهما المتغير الأول الدخل وسوف ندخلها باسم Income ، والمتغير الثاني وهو الاستهلاك وسوف ندخلها باسم Consump

ندخل البيانات كما سبق ذكره فتكون كالتالي:



	Income	Consump	var	var	var	var	var	var
1	3000.00	2800.00						
2	3500.00	3400.00						
3	5000.00	5000.00						
4	6000.00	5500.00						
5	9000.00	8000.00						
6	10000.00	7500.00						
7	9000.00	8500.00						
8	12000.00	10500.00						
9	10500.00	10000.00						
10	2500.00	2500.00						
11								
12								

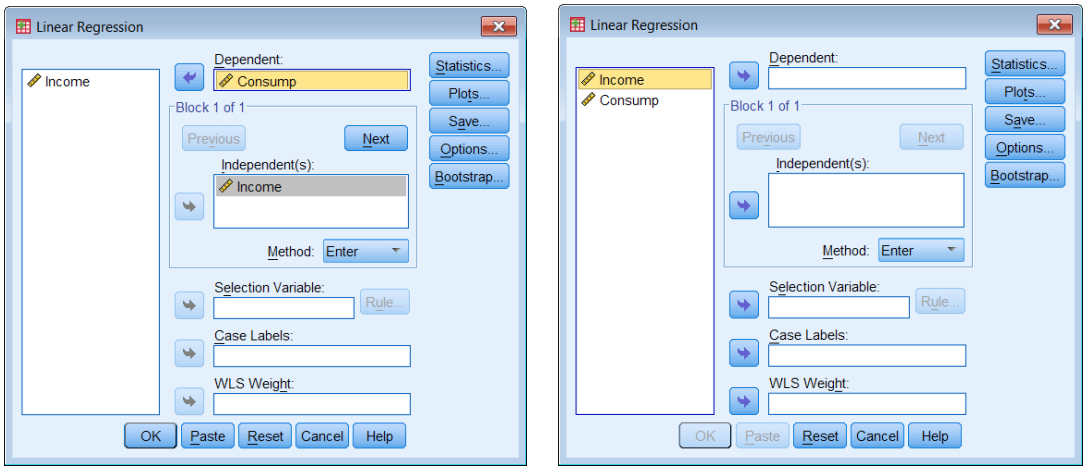
لإيجاد نموذج انحدار الاستهلاك على الدخل باستخدام SPSS نتبع التالي:

من قائمة Analyze نختار Regression ومنها نختارLinear

Analyze ⇒ Regression ⇒ Linear

فيظهر المربع الحواري التالي:

- ثم التظليل على المتغير المستهدف وهو **Consump** ثم ننقل هذا المتغير إلى المربع الأيمن الخاص بقائمة **Dependent** والتظليل على المتغير **Income** وننقله إلى المربع الأيمن الخاص بقائمة **Independent**



ثم الضغط على OK فيظهر الجداول التالية:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.982[a]	.965	.960	58.60899

a. Predictors: (Constant), income

ANOVA [b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	751329.9	1	751329.889	218.727	.000[a]
	Residual	27480.111	8	3435.014		
	Total	778810.0	9			

a. Predictors: (Constant), income

b. Dependent Variable: consump

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	48.229	43.913		1.098	.304
	income	.835	.056	.982	14.789	.000

a. Dependent Variable: consump

من النتائج السابقة يمكن استنتاج ما يلي:

1- نموذج انحدار الاستهلاك على الدخل هو:

$$Consump. = 48.229 + 0.835 * Income$$

2- معامل الارتباط بين الدخل والاستهلاك = 0.982 وهو يدل على وجود ارتباط طردي قوي بينهما،($Sig. = 0.000 < \alpha = 0.05$)

3- معامل التحديد $R^2 = 0.965$ ، ومعامل التحديد المعدَّل= 0.96، ويدل على أن 96.5% من التغير في قيمة الاستهلاك (المتغير التابع) يرجع الى التغير في الدخل، والنسبة المتبقية 3.5% ترجع إلى عوامل أخرى تؤثر على قيمة الاستهلاك.

4- الخطأ المعياري للتقدير = 58.6190.

5- F= 218.727، والقيمة الاحتمالية تساوي ($Sig. = 0.000$) وهذا يدل على وجود علاقة معنوية بين الدخل والاستهلاك وأن نموذج الانحدار السابق يصلح للتنبؤ.

6- القيمة الاحتمالية لثابت الانحدار تساوي Sig.=0.304 وتعني أن ثابت الانحدار لايختلف عن الصفر أي أن نموذج انحدار الاستهلاك على الدخل يمر بنقطة الأصل.

7- القيمة الاحتمالية لمعامل الانحدار تساوي Sig.=0.000 وهذا يدل على أن الدخل متغير مؤثر معنويا في تقدير قيمة الاستهلاك ويجب أن يكون ضمن نموذج خط الانحدار.

تطبيق: 4(غير محلول)

فيما يلي بيانات عن كمية البروتين اليومي بالجرام التي يحتاجها العجل الرضيع أقل من شهرين، ومقدار وزن العجل بعد مرور شهر بالكجم، وذلك لعينة من العجول الرضيعة حجمها 10.

كمية البروتين	114	113	103	102	101	99	98	96	91	90
الوزن بعد شهر	168	158	157	156	151	151	147	146	142	138

والمطلوب :

1- ارسم شكل الانتشار، وما هو توقعاتك لشكل العلاقة ؟

2- قدر معادلة انحدار الوزن على كمية البروتين.

3- فسر معادلة الانحدار.

4- ما هو مقدار الوزن عند إعطاء عجل رضيع أقل من شهرين **101** جرام من البروتين ؟ وما هو مقدار الخطأ العشوائي؟

ملاحظة:

التنبؤ :بقيمة ومتوسط المتغير التابع عند قيمة محددة للمتغير المستقل

	التنبؤ بقيمة $Y \mid x_o$	التنبؤ بقيمة متوسط $\mu_{y \mid x_o}$
القيمة المتنبأ بها	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_0$	$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_0$
فترة تنبؤ $(1-\alpha)\%$	$\hat{Y} \pm S.E_{\hat{y}}.t_{(\alpha/2,\, df_{error})}$	$\hat{Y} \pm S.E_{\mu_{\hat{y}}}.t_{(\alpha/2,\, df_{error})}$