الفصل الرابع

العرض والتحليل الإحصائي لمتغير أحادي

View and Statistical Analysis of Univariate Variable

(4-1) مقدمة :

يبـدأ التحليل الإحصائـي بتحليل السمات الإحصائية لكل متغير على حده الذي نطلق عليه التحليل الأحادي Univariate Analysis، ومن خلال التحليل الإحصائي يمكننا الحصول على المعلومات الآتية:

- معرفة كيف تتوزع قيم المفردات- توزيعًا طبيعيًا، أسيًا، ذا الحدين...،
- حسـاب مقاييس النزعة المركزية، وهي مجموعة من المقاييس تصف المتغير مثل الوسط، الوسيط، المنوال...،
- معرفـة تشتت المفردات عن القيم المركزية المحسوبـة في الخطوة السابقة، وتقاس عادة باستخدام التباين، الانحراف المعياري، المدى ....،
- توضيح هل توجد قيم شاذة في المفردات من خلال مقياس معامل الاختلاف.
- إجراء اختبارات الفروض على البيانات.

في هذا الفصل نقوم بتوضيح الطرق المختلفة لعرض وتحليل متغير أحادي، ففي المبحث الأول نقوم بتوضيـح كيف يمكن عرض البيانات في أشكال بيانية أو مخططات رسومية للمتغير. ونستكمل فيما تبقى من الفصل كيفية إيجاد المقاييس الإحصائية المناسبة للمتغير.

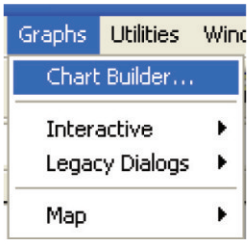
(4-2) عرض البيانات باستخدام

☜ الرسومات البيانية البسيطة:

الرسومـات البيانية نستطيـع أن نستخدمها للحصول على معلومات محددة عن المتغير مثل الوسط، الوسيط، التوزيع المتراكم ولعمل رسم بياني نتبع الخطوات الآتية:

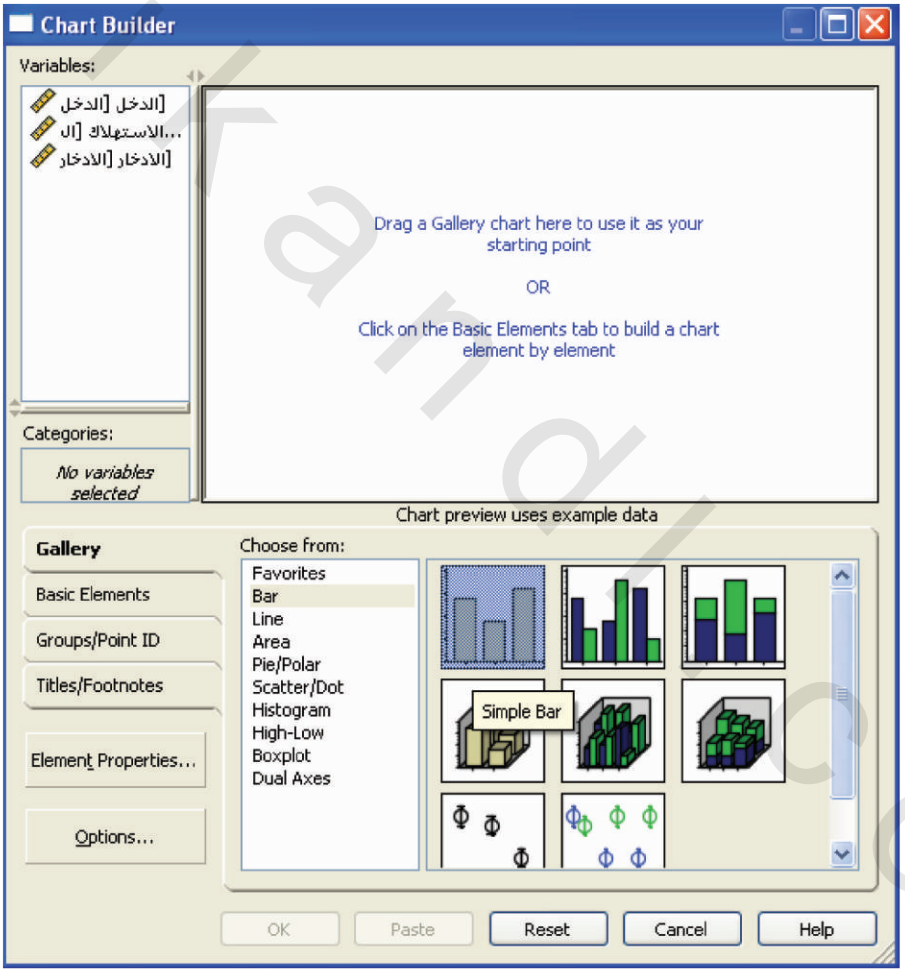
- نختار Graphs ثم نقوم باختيار Chart Builder... من الشكل الآتي:

obeikandl.com



شكل (4-1)

⊙ فيظهـر الشـكل (4-2) فنختار منه Bar من عمود Choose from: ثم نقـوم باختيار الشكل الأول وهو Simple Bar وذلك بالنقر عليه نقراً مزدوجًا:



شكل (4-2)

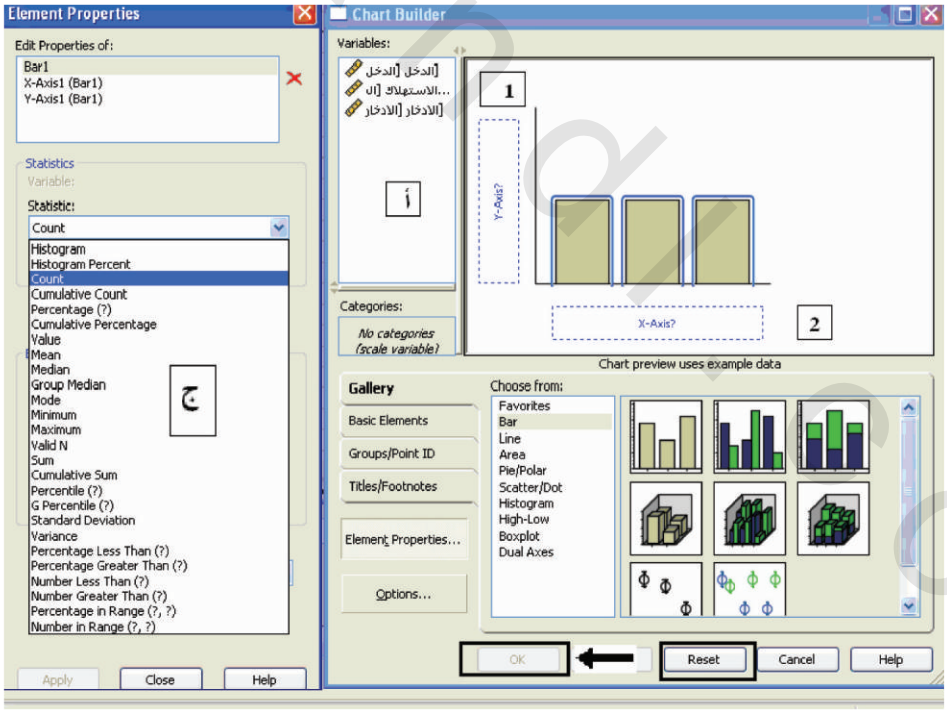
⊙ بعـد النقر على الشكل نقـراً مزدوجًا يظهر الشـكل (4-3) الآتي الذي يتكون مـن جزأين بالشكل الآتي:

أ - الجزء الأيمن هو شكل الرسم البياني والمتغيرات المتاحة للرسم.

ب - الجزء الأيسر وهو شكل خصائص المحاور المتاحة.

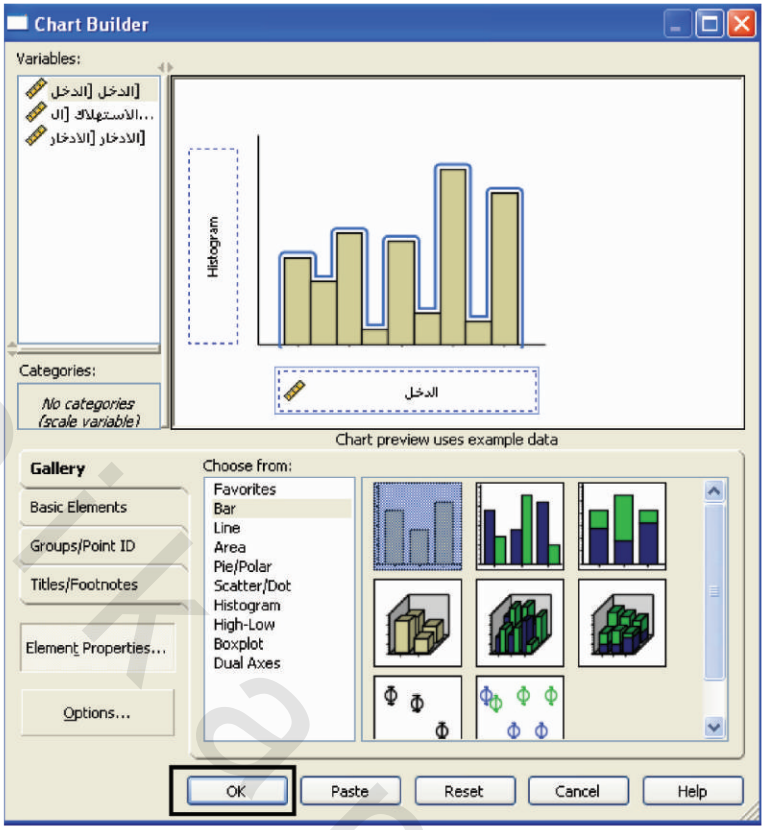
وهنا يوجد داخل شاشة الرسم ما يأتي:

- "أ" وهو عمود المتغيرات، وهنا نفترض أننا نريد رسم المتغير الدخل عن بيانات المثال (3-1).
- "1" Y-Axis? وهو محور العمليات التي نوضح عليها الشكل، وهنا نفترض أننا نريد أن نوضح الشكل التكراري للمتغير Histogram فإننا ننتقل إلى الجزء الثاني من الشكل (3-4).
- "2" وهو المحور X في الشكل البياني المراد رسمه، ونفترض أننا نريد رسم المتغير الدخل على المحورX فإننا ننقر عليه ونسحبه إلى محور X والمحدد في الشكل برقم 2.
- "ج" وهو عمود يحمل اسم Statistics، وهو عمود العمليات التي نريد إظهارها على المحور Y، ونستطيع الاختيار المتعدد لها ولنفرض أننا نريد إظهار تكرار قيم متغير الدخل فإننا نختار Histogram من (ج).



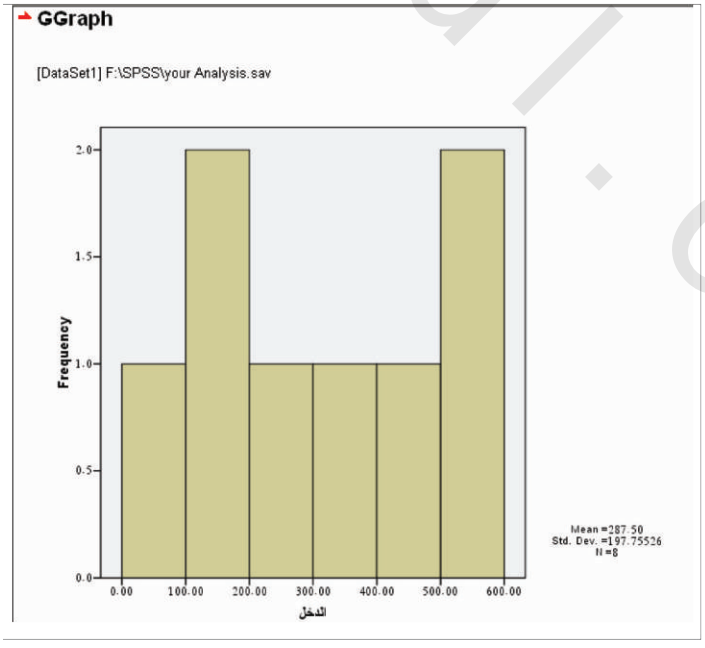
شكل (3-4)

- بعد الانتهاء من العمليات السابقة يصبح الشكل (3-4) السابق الصورة الآتية:



شكل (4-4)

ومن ثم النقر على OK



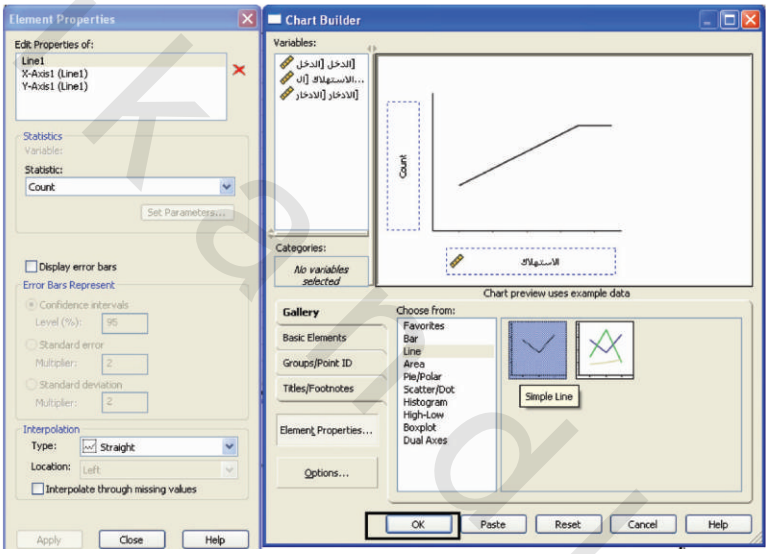
شكل (4-5)

obeikandl.com

⊙ ونلاحظ أن الشكل السابق يوضح الرسم البياني لمتغير الدخل وبجوار الرسم بعض المقاييس الإحصائية عن المتغير مثل الوسط الحسابي Mean يساوي 287.4، الانحراف المعياري يساوي 197.74426 وعدد البيانات يساوي 8.

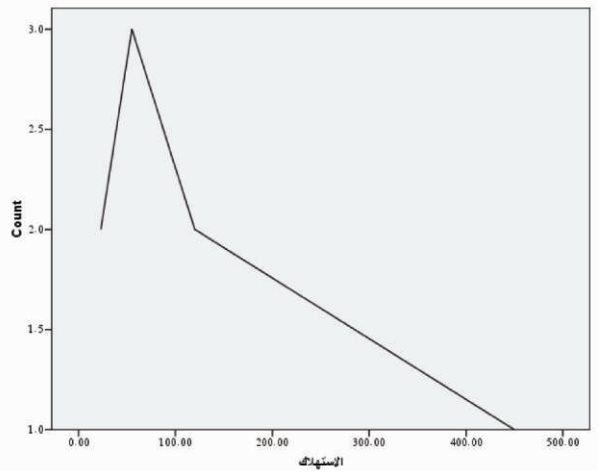
Line Graphs

إذا أردنا عمل شكل بياني من نوع line لمتغير الاستهلاك فإننا نختار Graphs ثم نقوم باختيار Chart Builder... فيظهر الشكل (4-2) فنختار منه Line من عمود Choose from: ثم نقوم باختيار الشكل الأول وهو Simple Line، وذلك بالنقر عليه نقراً مزدوجًا، ومن ثم اختيار المتغير الاستهلاك من عمود المتغيرات بطريقة اختيار الدخل السابق نفسها، وذلك بسحبها إلى عمود X فيظهر الشكل (4-6) الآتي:



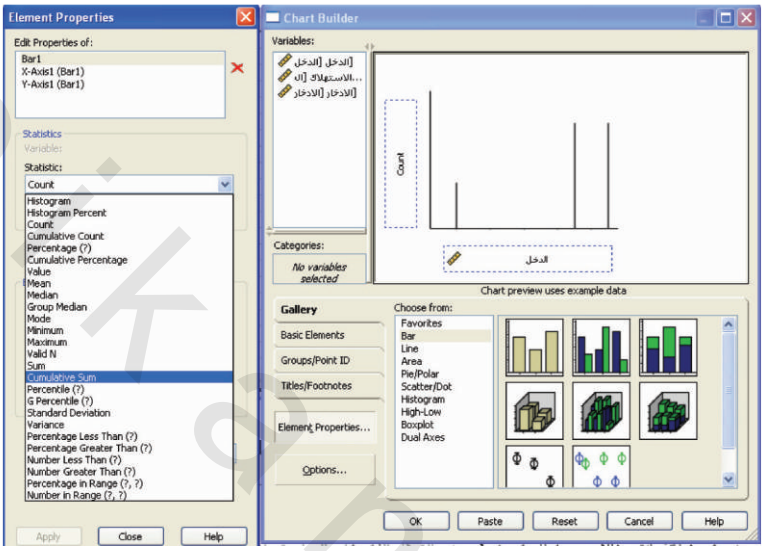
شكل (4-6)

⊙ وبعد النقر على ok لتأكيد الاختيار فإن الشكل (4-7) يظهر في نافذة output.



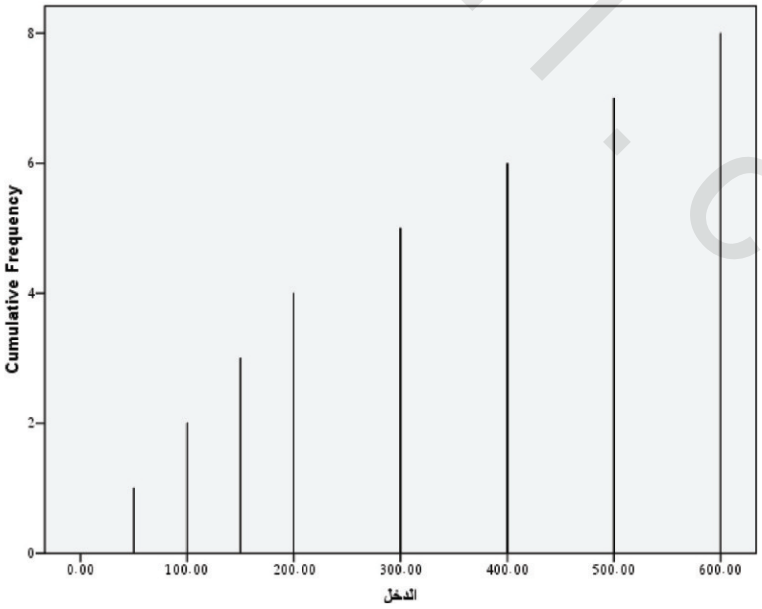
شكل (4-7)

إذا أردنـا عمـل شكل بياني من نـوع Graphs for cumulative frequency للمتغـير الدخل فإننا نختار Graphs ثـم نقـوم باختيـار Chart Builder... فيظهـر الشكل (4-2) فنختار منـه Bar من عمود Choose from: ثـم نقـوم باختيـار الشكل الأول وهـو Simple Bar وذلك بالنقر عليه نقراً مزدوجًـا، ومن ثم اختيار المتغـير الدخـل من عمود المتغيرات، وذلـك بسحبه إلى عمود X ومن الجزء الذي علـى اليسار يتم اختيار cumulative sum من عمود Statistic فيظهر الشكل (4-8) الآتي:



شكل (4-8)

وبعد ذلك النقر على Ok فيظهر الشكل (4-9) الآتي:



شكل (4-9)

الحصول على مقياس إحصائي:

في كثير من النواحي التطبيقية يكون الباحث في حاجة إلى حساب بعض المؤشرات التي يمكن الاعتماد عليها في وصف الظاهرة من حيث القيمة التي تتوسط القيم أو تنزع إليها القيم، وتسمى مقاييس النزعة المركزية، ومن حيث التعرف على مدى تجانس القيم التي يأخذها المتغير وتسمى مقاييس التشتت، وأيضًا ما إذا كانت هناك قيم شاذة أم لا من خلال مقياس يسمى معامل الاختلاف. والاعتماد على العرض البياني وحده لا يكفي، ولذا يتناول هذا الفصل أيضاً كيف يمكن وصف المتغير باستخدام المقاييس الإحصائية المناسبة، حيث توجد مجموعة رئيسة من المقاييس تعبر عن البيان الإحصائي منها:

(3-4) مقاييس النزعة المركزية:

⊙ الوسط الحسابي:

يعد الوسط الحسابي من أهم المقاييس في التحليل الإحصائي بشكل عام، إلا أنه يناسب الظواهر الكمية فقط، ونستطيع تعريفه على أنه القيمة الناتجة من خارج قسمة قيمة مجموع المشاهدات على عددها، وتحسب بشكل رياضي من العلاقة الآتية:

$$\overline{X} = \sum_{i=1}^{n} x_i / n, \; i = 1, 2, \ldots n$$

حيث:

$\overline{X}$ رمز للمتوسط وتقرأ "إكس بار"،

$\sum_{i=1}^{n}$ رمز للمجموع وتقرأ "سيجما من i=1 إلى n"،

n عدد المفردات محل الدراسة.

⊙ الوسيط:

الوسيط هو أحد أهم مقاييس النزعة المركزية، ويمكن أن تكون الظاهرة أو المتغير كميًا أو ترتيبيًا، ويعرف بأنه القيمة التي تقع في منتصف البيانات بعد الترتيب (تصاعديًا أو تنازليًا) في حالة البيانات المفردة.

⊙ المنوال:

وهو القيمة الأكثر تكرارًا في البيانات، وقد يكون للبيانات منوال وحيد أو أكثر من منوال، وأحياناً ينعدم المنوال لبعض البيانات.

(4-4) مقاييس التشتت المطلق والنسبي:

⊙ المدى:

ويعرف بأنه الفرق بين أكبر رقم وأصغر رقم للبيانات وهو كالآتي:

المدىR = أكبر قراءة – أصغر قراءة.

⊙ الانحراف المعياري:

يقيس تشتت البيانات حول المتوسط، وهو الجذر التربيعي الموجب للتباين σ^2 ويحسب كالآتي:

$$\sigma = \sigma^2 = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

حيث:

σ رمز للانحراف المعياري وتقرأ "سيجما".

σ^2 رمز التباين وتقرأ "سيجما تربيع".

⊙ معامل الاختلاف:

وهو من المقاييس النسبية للتشتت، ويعرف بأنه خارج قسمة الانحراف المعياري على الوسط أي

$$\gamma = \frac{\sigma}{\mu}$$

وهو من أفضل مقاييس التشتت النسبية، ويستخدم للتعرف على وجود قيم شاذة للبيانات أم لا.

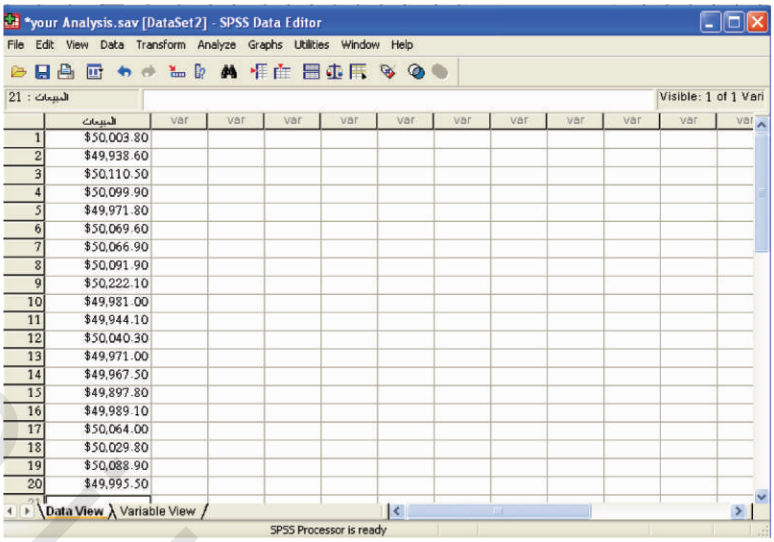
الآن نستطيع إيجاد المقاييس الإحصائية السابقة وغيرها من خلال برنامجنا SPSS وما يحويه من أوامر تمكننا من تنفيذ وتحديد كل المقاييس السابقة، فضلاً عن إيجاد توزيع تكراري للبيانات لمتغير واحد أو أكثر.

☜ الأمر Frequencies:

يستعمل هذا الأمر لعرض تكرار كل قيمة لمتغير ما وحساب النزعة المركزية ومقاييس التشتت والربيعيات والمئينات مع عرض بعض المخططات البيانية. يمكن الحصول على هذا الأمر من القائمة Analyze ثم الأمر Frequencies.

إذا افترضنا أننا لدينا بيانات عن قيمة مبيعات منتج السيارات لشركة ما وتم إدخالها في شاشة data view وكانت كالآتي:

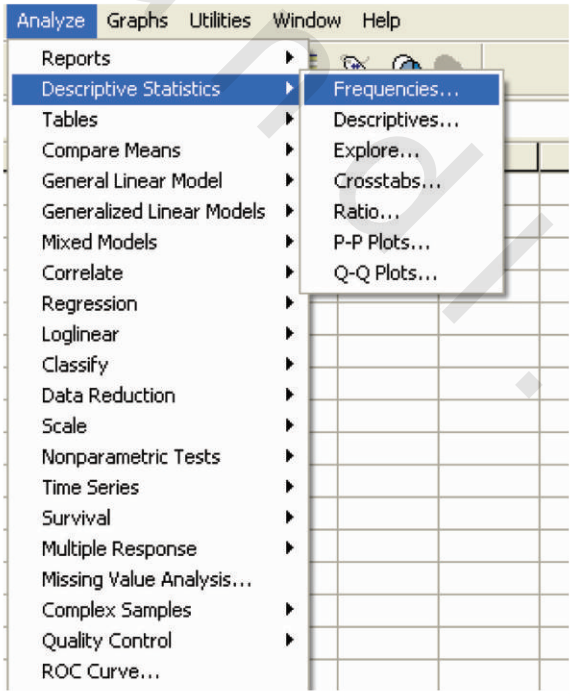
obeikandl.com



شكل (4-10)

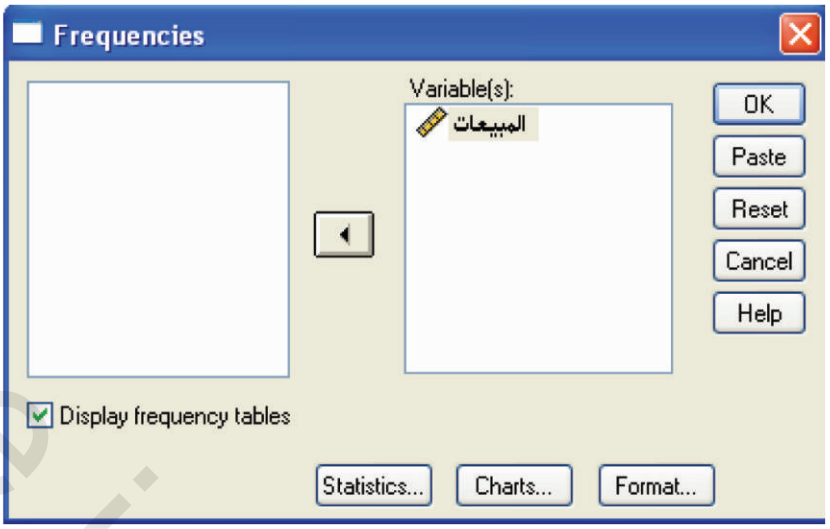
ونريد إيجاد توزيع تكراري للبيانات، وإيجاد مقاييس النزعة المركزية ومقاييس التشتت كالآتي:

من القائمة Analyze الظاهر في الشكل (4-11) الآتي:



شكل (4-11)

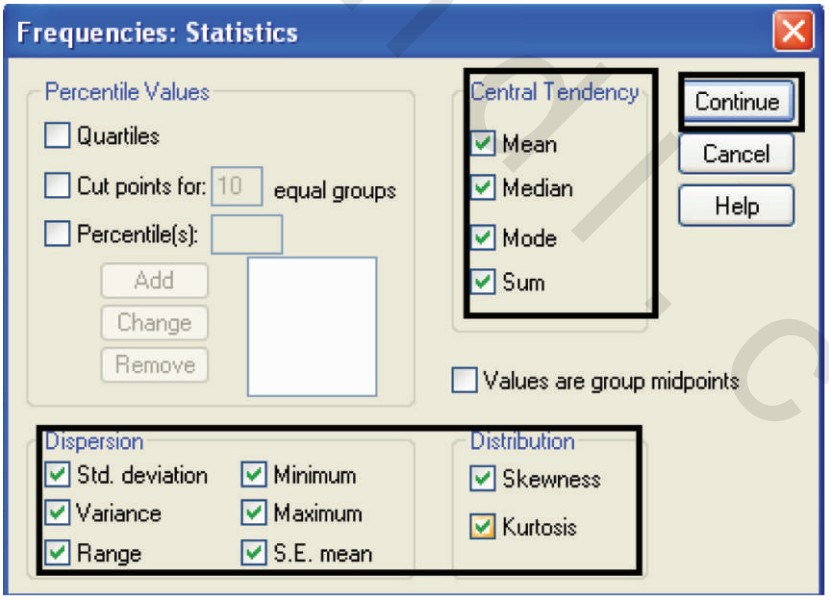
- انقـر على Descriptive statistics ثـم انقر على الأمر الفرعـي Frequencies فيظهر صندوق حوار Frequencies الذي يظهر الشكل الآتي بعد إدخال المبيعات عن طريق تحديدها ثم النقر على علامة الانتقال ▶ الذي يرمز للقيم في قائمة Variables انظر الشكل (4-12):



شكل (4-12)

حيث إن:

- Display Frequency Tables :- يعرض الجدول التكراري عند تأشيرة المربع المجاور له.
- Statistics :- يعرض بعض المؤشرات الإحصائية حسب حاجة المستفيد، عند نقره يظهر صندوق حواري Statistics وقد اخترنا المؤشرات الآتية للعرض كما يلي انظر الشكل (4-13):

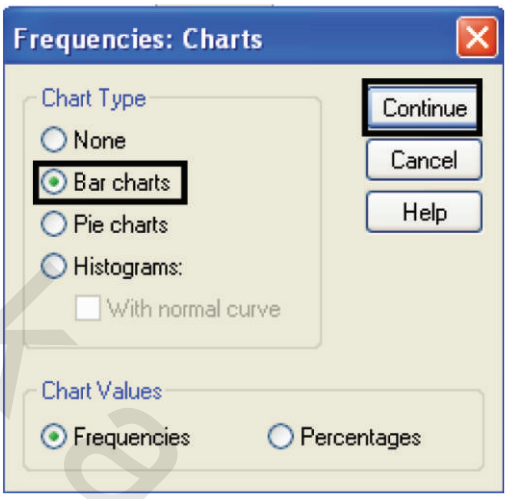


شكل (4-13)

Percentile Values :- تشمل الربيعات Quartile والمئينات Percentiles، ولتحديد المئين نكتبه في المربع المجاور لـ Percentiles ثم ننقر زر Add فيتم إضافة المئين الذي نريد إلى المستطيل أسفل Percentiles، ويستعمل الزر Change لتغيير قيمة المئين والزر Remove لإزالته.

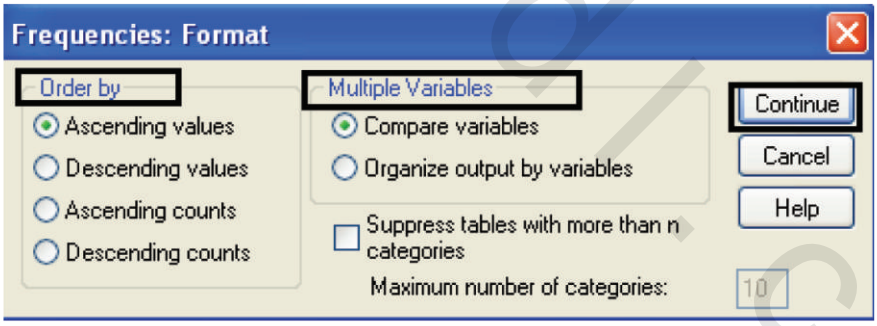
- ❑ Dispersion: لعرض مقاييس التشتت.
- ❑ Central Tendency: لعرض مقاييس النزعة المركزية.
- ⊙ وبعد تحديد ما نريد من مقاييس ننقر على continue.

❑ Chart: لعرض المخططات Bar، Pie وغيرها من المخططات ونفرض أننا نريد Bar فإننا نحددها ثم ننقر على continue كما في الشكل (4-14) الآتي:



شكل (4-14)

❑ Format: عند النقرعليه يظهر صندوق الحوار انظر الشكل (4-15):



شكل (4-15)

حيث إن:-

❑ Ordered By: لترتيب المشاهدات في الجدول التكرار تصاعديًا أو تنازليًا حسب القيم Values أو التكرارات Counts، وفي هذا المثال اخترنا الترتيب حسب القيم.

❑ Multiple Variables: يستخدم هذا الأمر في حالة وجود أكثر من متغير في القائمة Variables في صندوق حوار Frequencies، وتشمل ما يلي:

✓ Compare Variables: لعرض المؤشرات الإحصائية للمتغيرات كافة في جدول واحد.

✓ Organize Out By Variables: لعرض مؤشرات كل متغير في جدول مستقل، وفي هذا المثال لا يهم اختيار أي من هذين المتغيرين لوجود متغير واحد فقط.

✓ Suppress Tables with More Than n Categories: لإخفاء الجدول التكراري للمتغيرات التي يزيد عدد فئاتها على العدد المحدد من قبل المستفيد، ونلاحظ عند تنشيطها يتم تنشيط خانة بعنوان Maximum numbers of categories عندها نستطيع ما هو العدد المحدد الذي نريد إظهاره فقط.

⊙ عند النقر على زر Ok في صندوق حوار Frequencies تظهر النتائج الآتية:

جدول (4-1)

Statistics

المبيعات

N	Valid	20
	Missing	0
Mean		########
Std. Error of Mean		$16.97829
Median		########
Mode		########[a]
Std. Deviation		$75.92924
Variance		5765.250
Skewness		.650
Std. Error of Skewness		.512
Kurtosis		.782
Std. Error of Kurtosis		.992
Range		$324.30
Minimum		########
Maximum		########
Sum		$1000544

a. Multiple modes exist. The smallest value is shown

حيث إن:

❑ Valid: تمثل القيم الصحيحة (غير المفقودة).

❑ Missing: تمثل القيم المفقودة.

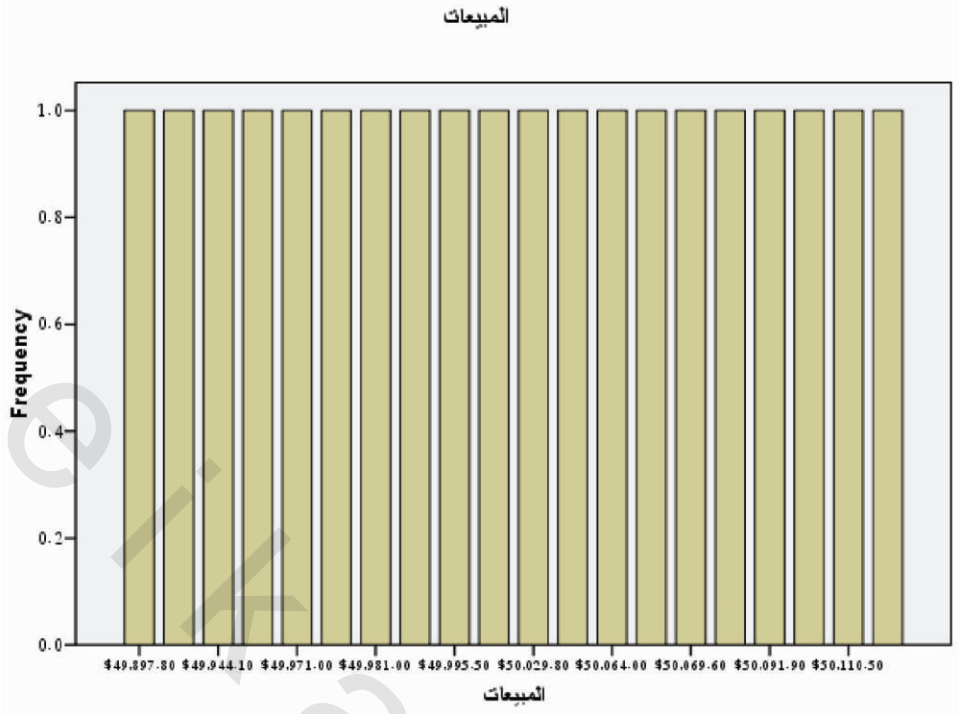
جدول (4-2)

المبيعات

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	$49,897.80	1	5.0	5.0	5.0
	$49,938.60	1	5.0	5.0	10.0
	$49,944.10	1	5.0	5.0	15.0
	$49,967.50	1	5.0	5.0	20.0
	$49,971.00	1	5.0	5.0	25.0
	$49,971.80	1	5.0	5.0	30.0
	$49,981.00	1	5.0	5.0	35.0
	$49,989.10	1	5.0	5.0	40.0
	$49,995.50	1	5.0	5.0	45.0
	$50,003.80	1	5.0	5.0	50.0
	$50,029.80	1	5.0	5.0	55.0
	$50,040.30	1	5.0	5.0	60.0
	$50,064.00	1	5.0	5.0	65.0
	$50,066.90	1	5.0	5.0	70.0
	$50,069.60	1	5.0	5.0	75.0
	$50,088.90	1	5.0	5.0	80.0
	$50,091.90	1	5.0	5.0	85.0
	$50,099.90	1	5.0	5.0	90.0
	$50,110.50	1	5.0	5.0	95.0
	$50,222.10	1	5.0	5.0	100.0
	Total	20	100.0	100.0	

نلاحظ أن المشاهدات في الجدول قد رتبت تصاعديًا حسب قيم المتغير X.

obeikandl.com



ونلاحظ أن كل البيانات لها تكرار واحد، ولا توجد قيمة من البيانات تم تكرارها مرتين أو أكثر.

☜ الأمر Descriptive:

من القائمة Analyze يمكن الحصول على الأمر Descriptive، حيث يمكن إجراء التحليل الإحصائي للبيانات عن طريقه أيضًا، وذلك بالحصول على المقاييس الإحصائية المختلفة وعلى بعض الأشكال البيانية، هذا الأمر يسمح أيضًا بالحصول على ما يسمى القيم المعيارية، ويتم حساب القيم المعيارية وفق العلاقة:

$$Z = \frac{x_i - \overline{X}}{S}$$

حيث x_i هي القيمة المدخلة، $\overline{X}$ هي المتوسط الحسابي، S هو الانحراف المعياري، حيث تظهر القيم المعيارية عادة في نهاية ملف البيانات.

مثال (4-1):

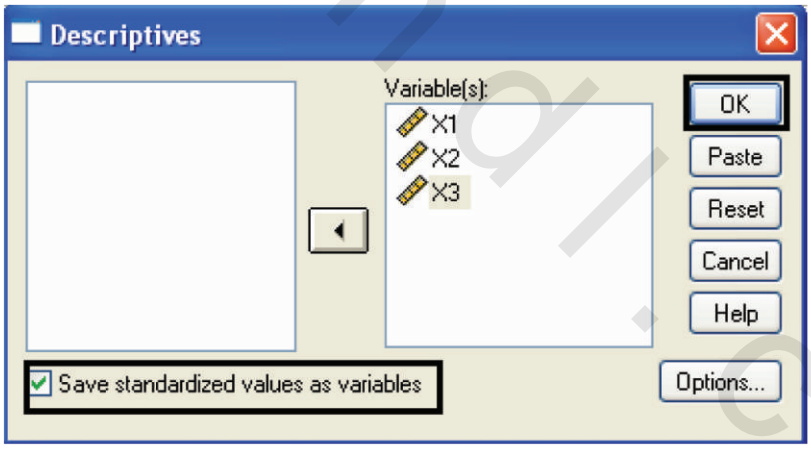
الجدول الآتي يتضمن المتغيرات X_1, X_2, X_3 التي تم إدخالها في Data editor للبرنامج:

جدول (4-3)

X_1	X_2	X_3
90	50	12
70	52	14
56	55	19
65	60	22
85	65	20
60		
69		
57		
50		
75		
62		
51		
85		

لغرض استخراج المقاييس الوصفية بالأمر Descriptive نتبع الخطوات الآتية:

- نختار من القائمة Analyze الخيار Descriptive Statistics، ومن القائمة الفرعية الخيار Descriptive يظهر مربع الحوار الذي نقوم من خلاله بإدخال المتغيرات على الشكل انظر الشكل (4-16):

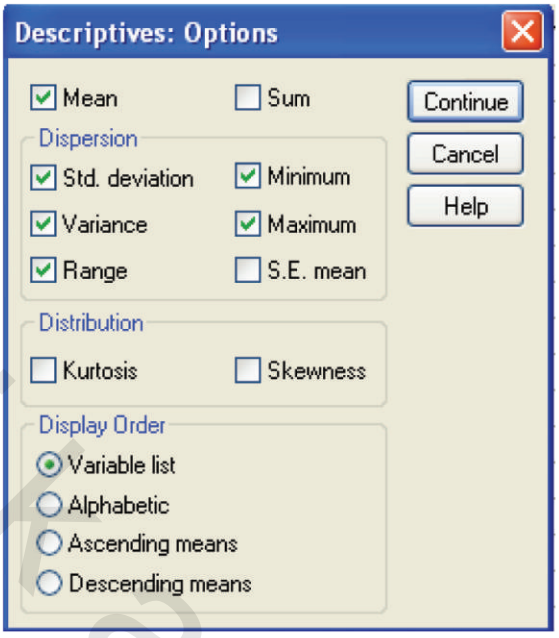


شكل (4-16)

حيث إن:

Save Standardized Values As Variables: لإضافة المتغيرات المعيارية z إلى Data editor عند التأشير عليها في مربع الحوار.

Options: عند نقره يظهر صندوق الحوار انظر الشكل (4-17):



شكل (4-17)

اختر المقاييس المطلوبة.

✓ Display Order: يمكننا من خلالها ترتيب عرض المقاييس الوصفية للمتغيرات، وقد اخترنا Variable List: الـذي يعرض المقاييس الوصفية حسب تسلسل المتغيرات الـوارد في خانة Variables في صندوق حوار Descriptive.

✓ ثم انقر على Continue لنعود لمربع الحوار Descriptive.

✓ ثم ننقر على زر Ok في صندوق حوار Descriptive يعرض المخرجات الآتية:

جدول (4-4)

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
X1	13	40.00	50.00	90.00	67.3077	13.21858	174.731
X2	5	15.00	50.00	65.00	56.4000	6.10737	37.300
X3	5	10.00	12.00	22.00	17.6000	4.03733	16.300
Valid N (listwise)	5						

obeikandl.com

حيـث تم عرض المقاييس الإحصائية المختارة حسب ترتيب المتغـيرات في خانة Variables في صندوق حوار Descriptive، كما تمت إضافة القيم المعيارية للمتغيرات إلى Data editor كما في الشكل الآتي:

X1	X2	X3	ZX1	ZX2	ZX3
90.00	50.00	12.00	1.71670	-1.04791	-1.38706
70.00	52.00	15.00	.20368	-.72044	-.64399
56.00	55.00	19.00	-.85544	-.22923	.34676
65.00	60.00	22.00	-.17458	.58945	1.08983
85.00	65.00	20.00	1.33844	1.40813	.59445
60.00	.	.	-.55284	.	.
69.00	.	.	.12802	.	.
57.00	.	.	-.77979	.	.
50.00	.	.	-1.30935	.	.
75.00	.	.	.58193	.	.
62.00	.	.	-.40153	.	.
51.00	.	.	-1.23370	.	.
85.00	.	.	1.33844	.	.

شكل (4-18)

obeikandl.com