الفصل الثاني
التحليل الاستكشافي للبيانات

(2,1) مقدمة في مقاييس التحليل الاستكشافي للبيانات

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، ومن حيث التعرف على مدى تجانس القيم التي يأخذها المتغير، وأيضا ما إذا كان هناك قيم شاذة أم لا. والاعتماد على العرض البياني وحده لا يكفى، ولذا يتناول هذا البند عرض بعض المقاييس الإحصائية التي يمكن من خلالها التعرف على خصائص الظاهرة محل البحث، وكذلك إمكانية مقارنة ظاهرتين أو أكثر، ومن أهم هذه المقاييس، مقاييس النزعة المركزية والتشتت.

(2,2) مقاييس النزعة المركزية Central Tendency

تسمى مقاييس النزعة المركزية بمقاييس الموضع أو المتوسطات، وهى القيم التي تتركز القيم حولها، ومن هذه المقاييس، الوسط الحسابي، والوسيط،والمنوال، والوسط الهندسي، وفيما يلي عرض لأهم هذه المقاييس.

المتوسط الحسابي Arithmetic Mean

وهو من أهم مقاييس النزعة المركزية، وأكثرها استخداما في النواحي التطبيقية، ويعرف الوسط الحسابي بشكل عام على أنه مجموع القيم مقسوما على عددها. فإذا كان لدينا n من القيم، ويرمز لها بالرموز :

$y_1, y_2, \ldots, y_n$.

فإن المتوسط الحسابي لهذه القيم ، ونرمز له بالرمز $\bar{y}$ يحسب بالمعادلة التالية :

$$\bar{y} = \frac{y_1 + y_2 + \ldots + y_n}{n} = \frac{\sum_{i=1}^{n} y_i}{n}$$

ويدل الرمز Σ **Summation** على المجموع

الوسيط Median

هو أحد مقاييس النزعة المركزية، والذي يأخذ في الاعتبار رتب القيم، ويعرف الوسيط بأنه القيمة التي يقل عنها نصف عدد القيم $(n/2)$ ، ويزيد عنها النصف الآخر $(n/2)$ ، أي أن 50 % من القيم أقل منه، 50 % من القيم أعلى منه. وفيما يلي خطوات حساب الوسيط.

1. ترتب القيم تصاعديا.
2. تحديد رتبة الوسيط

أ. إذا كان عدد القيم n عدد فردي فتكون رتبة الوسيط = $\left(\frac{n+1}{2}\right)$

ويكون الوسيط هو:

الوسيط = القيمة رقم $\left(\frac{n+1}{2}\right)$

ب. إذا كان عدد القيم n عدد زوجي، فإن الوسيط يقع بين القيمة رقم $(n/2)$، والقيمة رقم $((n/2)+1)$،

ومن ثم يكون الوسيط متوسط هاتين القيمتين، أي:

الوسيط = { القيمة رقم $\left(\frac{n}{2}\right)$ + القيمة رقم $\left(\frac{n}{2}+1\right)$ } مقسومة على 2

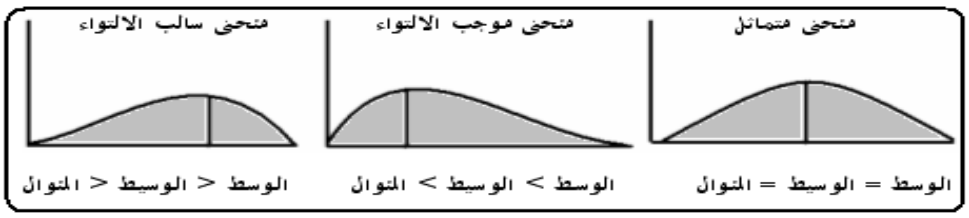
- **المنوال: Mode**

يعرف المنوال بأنه القيمة الأكثر شيوعا أو تكرارا بين القيم، أي القيمة التي تتكرر أكثر من غيرها، ويكثر استخدامه في حالة البيانات الوصفية، لمعرفة النمط (المستوى) الشائع

المنوال = القيمة (المستوى) الأكثر تكرارا

- **استخدام مقاييس النزعة المركزية في تحديد شكل توزيع البيانات**

يمكن استخدام الوسط الحسابي والوسيط والمنوال في وصف المنحنى التكراري، والذي يعبر عن شكل توزيع البيانات، كما يلي:



شكل (2,1) العلاقة بين الوسط الحسابي والوسيط والمنوال

- من شكل (2,1) يكون المنحنى متماثل إذا كان : الوسط = الوسيط = المنوال .
- يكون المنحنى موجب الالتواء (ملتوي جهة اليمين) إذا كان: الوسط > الوسيط > المنوال

- يكون المنحنى سالب الالتواء (ملتوي جهة اليسار) إذا كان: الوسط < الوسيط < المنوال

- **المتوسط الهندسي The Geometric Mean**

هو أحد المتوسطات أو المعدّلات التي تقيس النزعة المركزية أو القيمة النموذجية لمجموعة من البيانات، ويشبه المتوسط الهندسي نظيره، المتوسط الحسابي، إلا أنّه بدلاً من أن يتم جمع البيانات والقسمة على عدد ها، يتم حساب الجذر الـ n لحاصل ضرب هذه البيانات، حيث n هو عدد البيانات.

على سبيل المثال، فإنّ المتوسط الهندسي للعددين 8 ,2 ما هو إلاّ الجذر التربيعي لحاصل ضربهما (16)، أي 4.

وبشكل عام إذا كان لدينا n من القيم، ويرمز لها بالرمز : $y_1, y_2, \ldots, y_n$.
فإن المتوسط الهندسي لهذه القيم ، ونرمز له بالرمز G يحسب بالمعادلة التالية :

$$G.M. = \sqrt[n]{y_1 . y_2 \ldots y_n}$$

(2,3) مقاييس التشتت Dispersion Measurements

الاستناد إلى مقاييس النزعة المركزية وحدها عند وصف البيانات لايكفي لأنها لاتعطي صورة واضحة عن شكل توزيع هذه البيانات، فقد يكون الوسط الحسابي لمجموعتين متساوي، بينما يوجد اختلاف كبير بين المجموعتين من حيث مدى تقارب وتباعد البيانات من بعضها البعض،لذا تستخدم مقاييس التشتت في التعرف على مدى تجانس القيم التي أمكن جمعها حول ظاهرة معينة، ويعتمد عليها في مقارنة بيانات ظاهرتين لهما نفس القياس خاصة إذا كانت متوسطي الظاهرتين متساوي.

ومن أهم مقاييس التشتت التي تصف مدى تجانس البيانات المدى Rang، ومتوسط الانحرافات Mean of Deviations، والتباين Variance والانحراف المعياري Standard Deviation، وفيما يلي عرض بعض هذه المقاييس:

- **المدى Range**

هو أبسط مقاييس التشتت، ويحسب بتطبيق المعادلة التالية.

المدى = أكبر قيمة – أدنى قيمة

- **متوسط الانحرافات المطلقة Mean of Deviations**

وهو عبارة عن متوسط الانحرافات المطلقة للقيم عن المتوسط الحسابي، ومن ثم يحسب بتطبيق المعادلة:

$$MD = \Sigma \left| y - \bar{y} \right| / n$$

- **التباين Variance**

من أهم مقاييس التشتت استخداما، وهو عبارة عن متوسط مربعات انحرافات القيم عن وسطها الحسابي، ويحسب بتطبيق المعادلة:

$$S^2 = \Sigma (y - \bar{y})^2 / (n-1)$$

- **الانحراف المعياري Standard Deviation**

من أكثر مقاييس التشتت استخداما،لاعتماده في العديد من العمليات الاحصائية المتعلقة باجراء المقارنات واختبار الفرضيات. ويقيس الانحراف المعياري درجة الاختلاف بين القيم ووسطها الحسابي، وهو عبارة عن الجذر التربيعي الموجب للتباين، وهو أوقع استخداما من التباين، ويحسب بالمعادلة:

$$S = \sqrt{S^2} = \sqrt{\frac{\Sigma (y - \bar{y})^2}{n-1}}$$

(2,4) مقاييس الانتشار Spreading measurements

تعتمد مقاييس التشتت السابقة جميعها على الوحدات المستخدمة في القياس، وبالتالي لا يمكن استخدمها في المقارنة بين مجموعتين أو أكثر مقاسة بوحدات قياس مختلفة، مثل مقارنة الأطوال مع الأوزان لمجموعة من الطلبة، لذلك وجدت مقاييس أخرى لا تعتمد على الوحدات المستخدمة في القياس حيث تقيس الاختلاف النسبي دون وحدة تمييز، من أهم هذه المقاييس معامل الاختلاف النسبي، والدرجة المعيارية، وفيما يلي عرض لهذين المقياسين.

- **معامل الاختلاف النسبي** (CV) **Coefficient of Variation**

وهو عبارة عن الانحراف المعياري كنسبة مئوية من المتوسط الحسابي، وكلما كان هذا المعامل صغيرا كلما دل ذلك على انتشار البيانات في مدى ضيق ويستدل منه على أن البيانات أكثر تجانسا، ويحسب هذا المعامل بتطبيق المعادلة التالية.

$$CV = \frac{S}{\bar{y}} \times 100$$

- **الدرجة المعيارية** (Z) **Standardized Value**

ويقيس مقدار انخفاض أو زيادة القيم عن وسطها الحسابي بعدد من وحدات انحراف معياري، ويطبق باستخدام التحويلة التالية.

$$Z = \frac{Value \quad x - \bar{y}}{S}$$

(2,5) حساب مقاييس الاحصاء الوصفي باستخدام برنامج SPSS

يمكن استخدم برنامج **SPSS** للحصول على مقاييس الإحصاء الوصفي للبيانات باستخدام أمر **Statistics Descriptive** من قائمة[Analyze] ولتوضيح ذلك نعرض التطبيق التالي:

تطبيق (2,1)

البيانات التالية تمثل أوزان المواليد بالكيلوجرام لعينة عشوائية من الأمهات حجمها 12 .

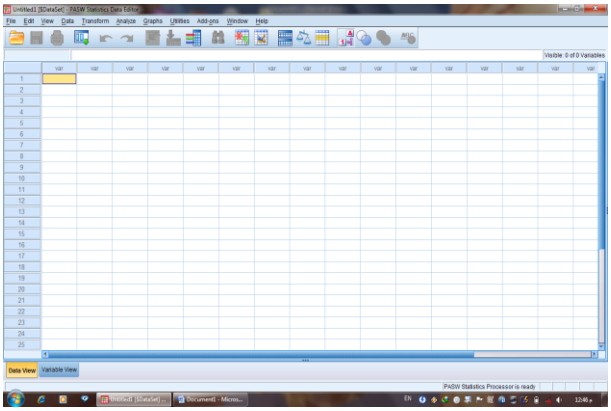
المولود	1	2	3	4	5	6	7	8	9	10	11	12
الوزن	2.2	2.9	3.3	3.1	2.1	3.2	3.5	2	3.7	3.2	1.7	2.7

والمطلوب:

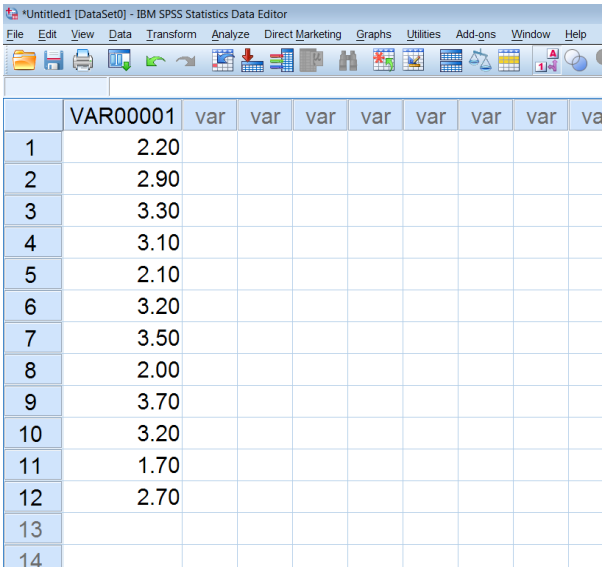
استخدم برنامج **SPSS** للحصول على مقاييس الإحصاء الوصفي لبيانات وزن الطفل عند الميلاد على أن يكون اسم ملف البيانات **Bweight** واسم المتغير الذي يعبر عن الوزن **BW**.

حل التطبيق

- لحساب مقاييس الاحصاء الوصفي في تطبيق (2,1) باستخدام برنامج SPSS 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:

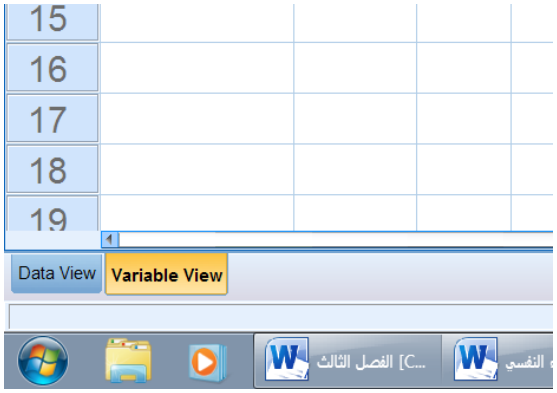


- إدخال البيانات المستهدفة (أوزان المواليد بالكيلوجرام) في أول أعمدة الملف الفارغ من جهة اليسار وذلك بالوقوف بالسهم على أول خانة من خانات العمود الخاص بالدرجات لبدء عمليات الإدخال.

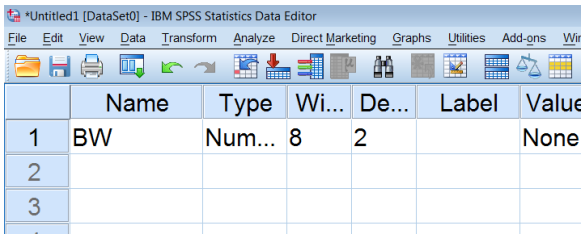


	VAR00001	var	var	var	var	var	var	var	va
1	2.20								
2	2.90								
3	3.30								
4	3.10								
5	2.10								
6	3.20								
7	3.50								
8	2.00								
9	3.70								
10	3.20								
11	1.70								
12	2.70								
13									
14									

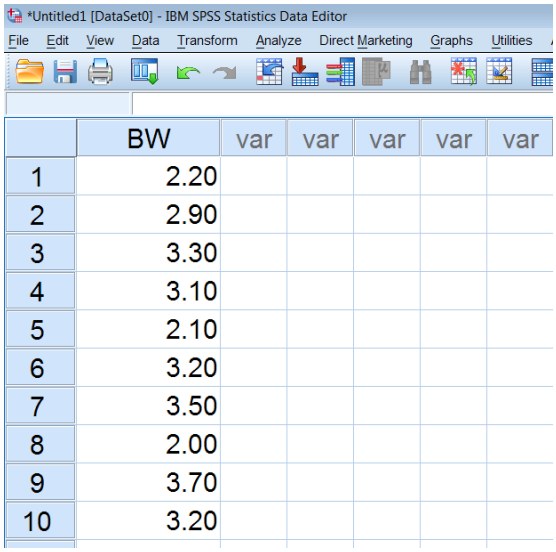
- تسمية المتغير الذي تم إدخاله وذلك بالضغط على **[Variable View]** (رؤية المتغير)، بشريط التعليمات السفلي.



- يظهر لنا جدول يحتوي على كل الخصائص الممكنة للمتغير الذي تم إدخاله.
- يتم إدخال الاسم (BW) للمتغير في الخانة الأولى تحت[Name] .

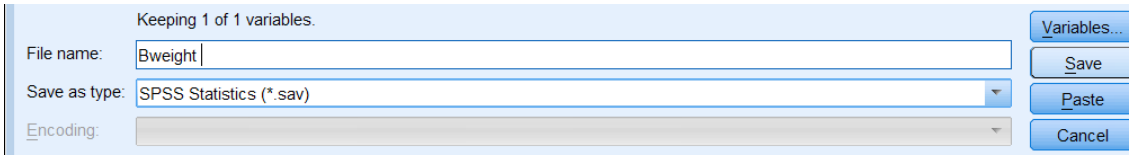


- ثم الضغط على [Data View] للعودة الى صفحة البيانات.



- وبالتالي تم تهيئة البيانات تمهيداً لعرض النتائج.

- كما يفضل حفظ الملف: وذلك بالضغط على علامة حفظ
- ثم نكتب اسم الملف **Bweight** ونضغط **Save**

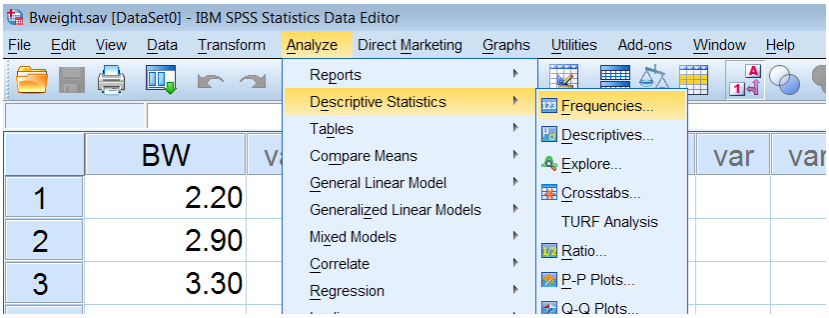


- وبذلك يكون تم حفظ الملف باسم Bweight

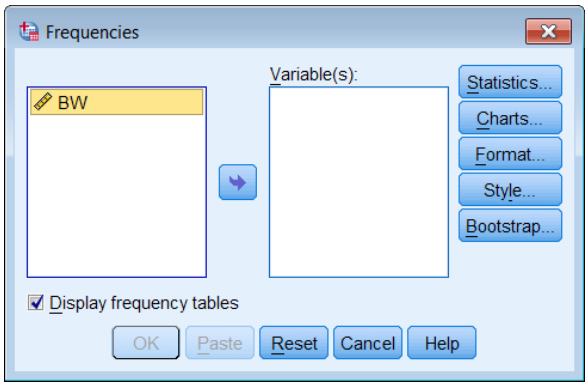
– ايجاد المقاييس الاحصائية المطلوبة

- اختيار الأمر [Analyze] من شريط القوائم.

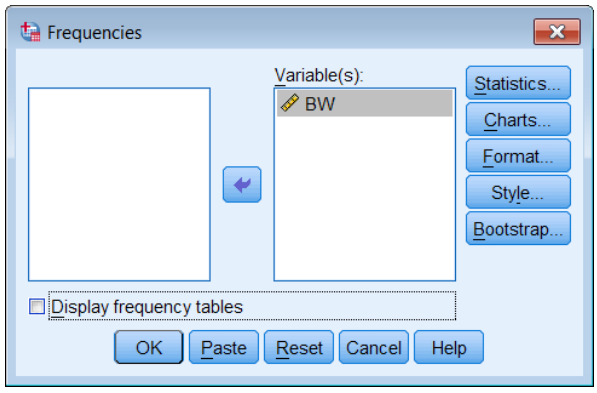
- يظهر لنا قائمة تحتوي على كل الأساليب الإحصائية ، نختار من بينها [Descriptive Statistics] ثم اختيار الأمر [Frequencies].



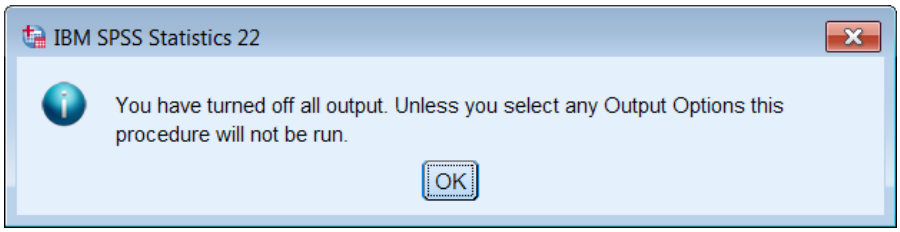
فيظهر المربع التالي:



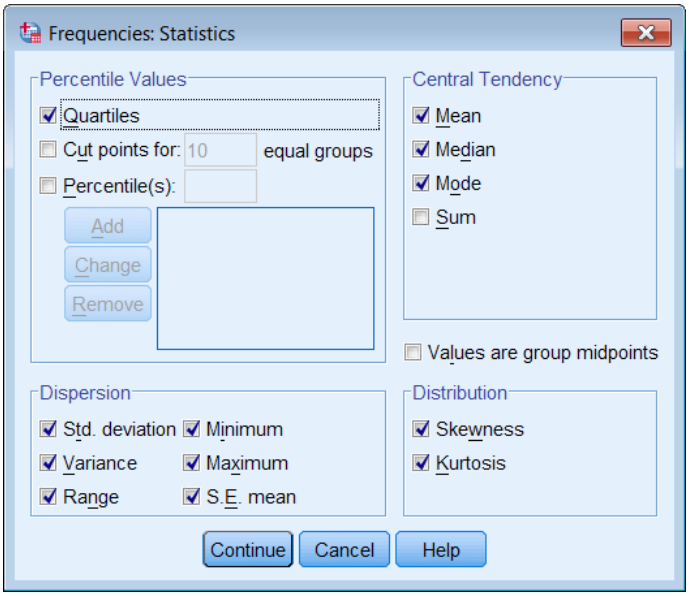
- التظليل على المتغير المستهدف تحليله وهو [BW] ، ثم نقل هذا المتغير إلى المربع الأيمن الخاص بقائمة المتغيرات [Variables].



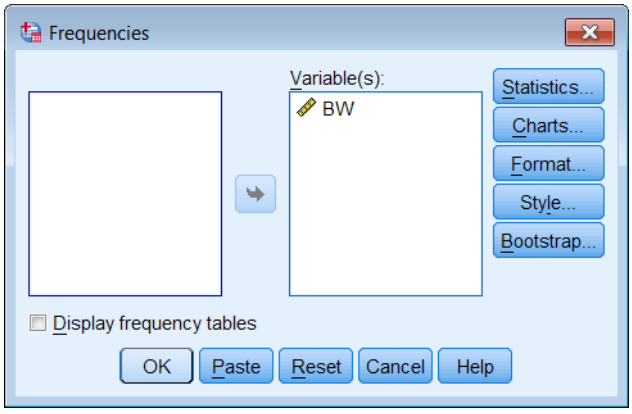
- إزالة التنشيط الخاص بعرض الجداول التكرارية [Display Frequency Tables] فتظهر رسالة خاصة بالموافقة بعدم عرض الجداول التكرارية يجب الموافقة عليها بالضغط OK



- الضغط على الأمر [Statistics]، لاختيار المقاييس الإحصائية المطلوبة.
- يظهر لنا مربع حواري يتضمن جميع الإحصاءات الوصفية، نختار من بينها المطلوب



- الضغط على كلمة [Continue]فيظهر المربع التالي



- الضغط على كلمة [OK]فيظهر المقاييس الاحصائية المطلوبة:

Statistics		
BW		
N	Valid	12

	Missing	0
Mean		2.8000
Std. Error of Mean		.18829
Median		3.0000
Mode		3.20
Std. Deviation		.65227
Variance		.425
Skewness		-.413
Std. Error of Skewness		.637
Kurtosis		-1.194
Std. Error of Kurtosis		1.232
Range		2.00
Minimum		1.70
Maximum		3.70
Percentiles	25	2.1250
	50	3.0000
	75	3.2750

تظهر النتائج أن وزن المولود يتراوح في العينة بين حد أدنى 1.7 كجم، وحد أعلي 3.7 كجم، كما بلغ متوسط وزن الطفل عند الميلاد 2.8 كجم بانحراف معياري 0.65 كجم، ومن ثم معامل اختلاف نسبي قيمته 23.3% ، ويدل على أن وجود تشتت في بيانات الوزن ولكن بدرجة صغيرة.

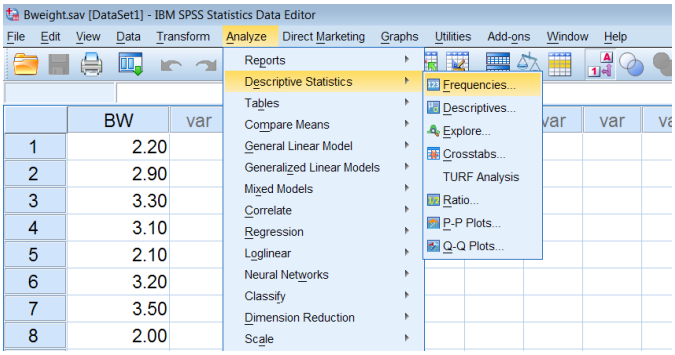
بما أن (المتوسط < الوسيط < المنوال) إذا توزيع بيانات وزن المولود سالب الالتواء (ملتوي جهة اليسار) ويؤيد ذلك أيضا اشارة معامل الالتواء (413.-). وحيث أن قيمة معامل التفرطح (1.19-) أقل من 3 إذا منحنى توزيع بيانات وزن المولود مفرطح.

تطبيق (2,2)

من بيانات التطبيق السابق أوجد الإعشاريات

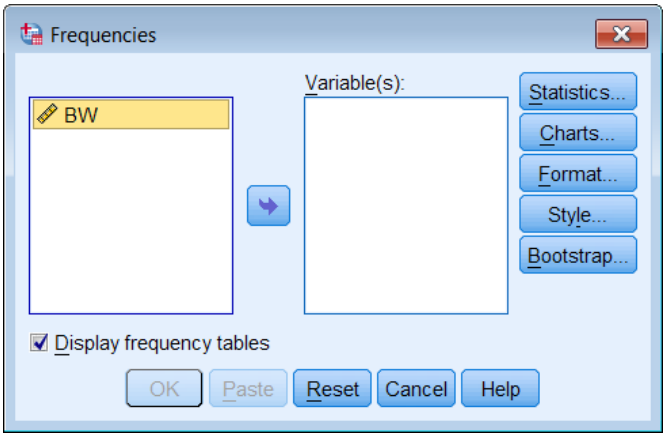
الإعشاريات: هي تسعة نقاط تقسم توزيع الدرجات إلى عشرة أجزاء متساوية ولايجاد باستخدام برنامج SPSS نتبع التالي:

1. نختار الأمر. [Analyze]
2. نختار الإحصاءات الوصفية [Descriptive Statistics] ومنه نختار [Frequencies] ثم الضغط عليه لمرة واحدة.

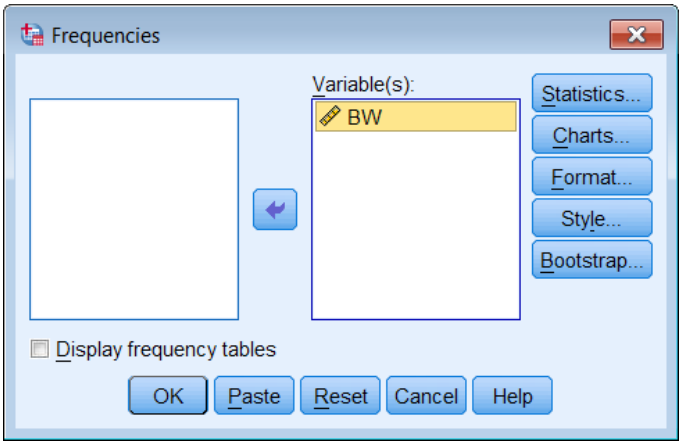


3. يظهر لنا مربع حواري يحتوي في يساره على قائمة بأسماء المتغيرات المدرجة بملف البيانات المفتوح نختار منها متغير التحصيل. [BW]

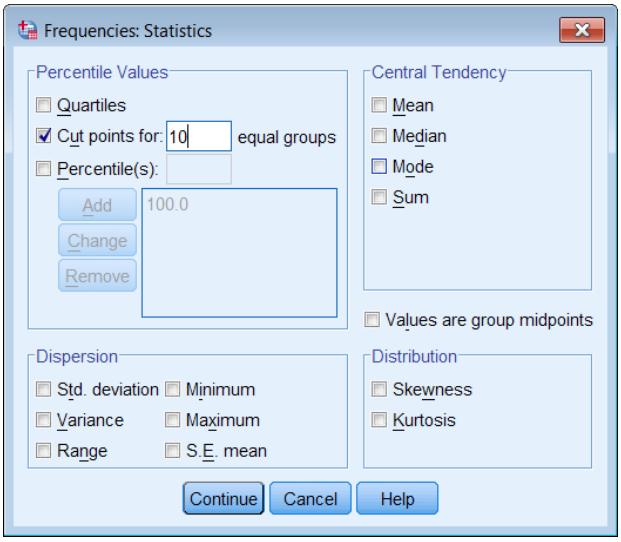
4. الضغط على السهم الأوسط لنقل المتغير المستهدف إلى مربع [Variables]



5. إلغاء خيار عرض الجداول التكرارية [Display Frequency Tables] ، فتبدو لنا رسالة بالموافقة على الإلغاء ثم نضغط على موافق لتختفي هذه الرسالة.



6. الضغط على خيار الإحصاءات [Statistics] على يمين المربع الحواري ، فيبدو لنا مربع حواري آخر يحتوي في أعلى يساره على قائمة بالقيم المئوية [Percentile Valuse] نختار منها Cut Points For ونكتب رقم (10) في المربع الصغير .



7. الضغط على استمر [Continue] .

8. الضغط على [Ok] ، للحصول على النتائج كما يلي:

Statistics		
BW		
N	Valid	12
	Missing	0
Percentiles	10	1.7900
	20	2.0600
	30	2.1900
	40	2.7400
	50	3.0000
	60	3.1800
	70	3.2100
	80	3.3800
	90	3.6400

تطبيق (2,3)

البيانات التالية تمثل درجات الطلاب في مقرر الإحصاء

63	59	66	71	71	74	52	64	70	60
62	57	56	60	61	54	62	63	66	63
63	63	72	65	64	62	75	69	78	62
73	57	68	70	75	65	62	68	63	69

والمطلوب استخدام برنامج **SPSS** في الحصول على النتائج التالية:

1- مقاييس الإحصاء الوصفي.

2- الجدول التكراري

3- المدرج والمنحنى التكراري.

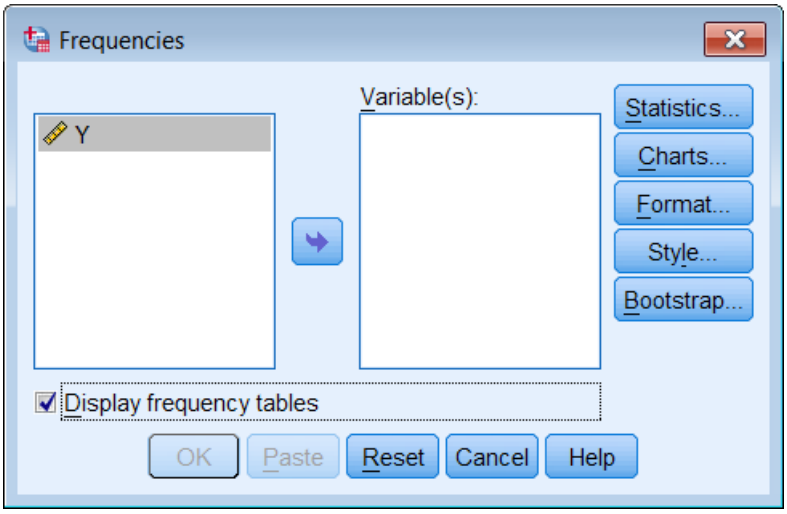
4- المدرج والمنحنى التكراري مستخدما مراكز فئات 50، 55، 60، 65، 70، 75، 80 .

على أن يكون اسم ملف البيانات Scores، واسم المتغير Y

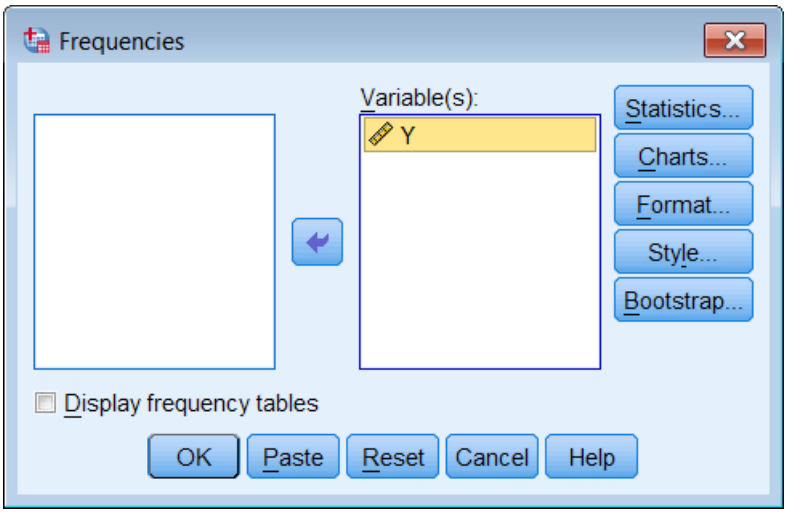
حل التطبيق:

للحصول على مقاييس الإحصاء الوصفي بعد ادخال البيانات وحفظها كما سبق نتبع التالي:

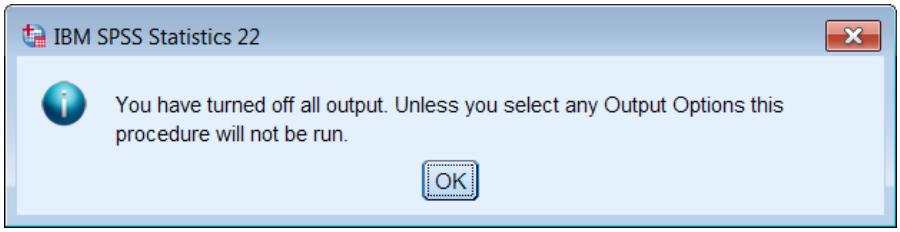
- اختيار الأمر **[Analyze]** من شريط القوائم.
- يظهر لنا قائمة تحتوي على كل الأساليب الإحصائية ، نختار من بينها **[Descriptive Statistics]** ثم اختيار الأمر **[Frequencies]** فيظهر المربع التالي:



- التظليل على المتغير المستهدف تحليله وهو [Y] ، ثم ننقل هذا المتغير إلى المربع الأيمن الخاص بقائمة المتغيرات [Variables].

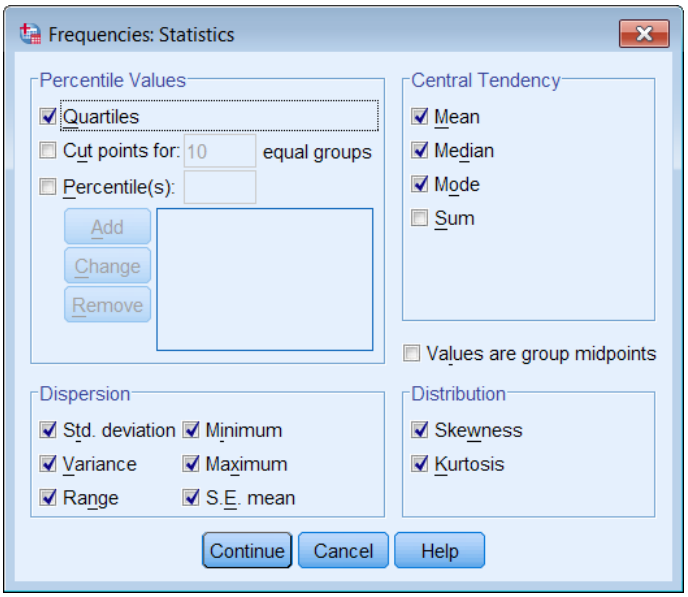


- إزالة التنشيط الخاص بعرض الجداول التكرارية [Display Frequency Tables] فتظهر رسالة خاصة بالموافقة بعدم عرض الجداول التكرارية يجب الموافقة عليها بالضغط OK



- الضغط على الأمر [Statistics]، لاختيار المقاييس الإحصائية المطلوبة.

يظهر لنا مربع حواري يتضمن جميع الإحصاءات الوصفية، نختار من بينها المطلوب



- الضغط على كلمة [OK]فيظهر المقاييس الاحصائية المطلوبة:

Statistics

Y		
N	Valid	40
	Missing	0
Mean		64.93
Std. Error of Mean		.950
Median		63.50
Mode		63
Std. Deviation		6.010
Variance		36.122
Skewness		.120
Std. Error of Skewness		.374
Kurtosis		-.320
Std. Error of Kurtosis		.733
Range		26
Minimum		52
Maximum		78
Percentiles	25	62.00
	50	63.50
	75	69.75

تظهر النتائج أن درجة الطالب في الاختبار تتراوح بين حد أدنى 52 درجة، وحد أعلي 78 درجة، كما بلغ متوسط درجة الطالب 64.9 درجة بانحراف معياري 6.01 درجة، ومن ثم معامل اختلاف نسبي قيمته 9.26% ، ويدل على أن وجود تشتت في درجات الطلاب ولكن بدرجة صغيرة.

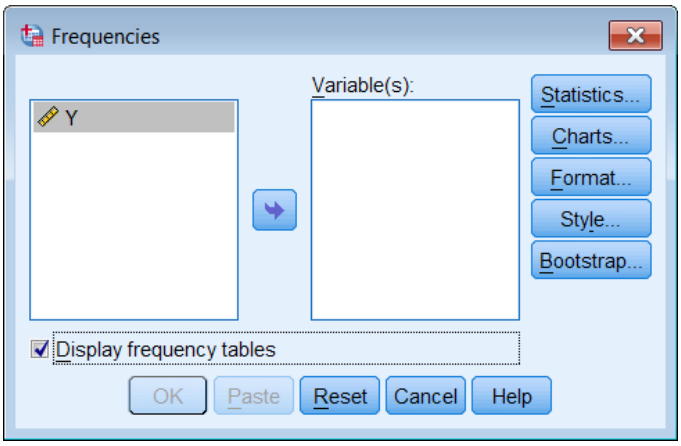
بما أن (المتوسط > الوسيط > المنوال) إذا توزيع درجات الطلاب موجب الالتواء (ملتوي جهة اليمين بدرجة صغيرة) ويؤيد ذلك أيضا قيمة واشارة معامل الالتواء (0.12) وكذلك قرب الوسيط من الربيع الأدنى، وشكل المدرج التكراري . وحيث أن قيمة معامل التفرطح (0.32-) أقل من 3 إذا منحنى توزيع درجات الطلاب مفرطح.

- لايجاد المدرج والمنحنى التكراري مستخدما الفئات (50 – 55، 55- 60 ، 60 – 65 ، 65 – 70 ، 70- 75 ، 75 – 80) نتبع التالي

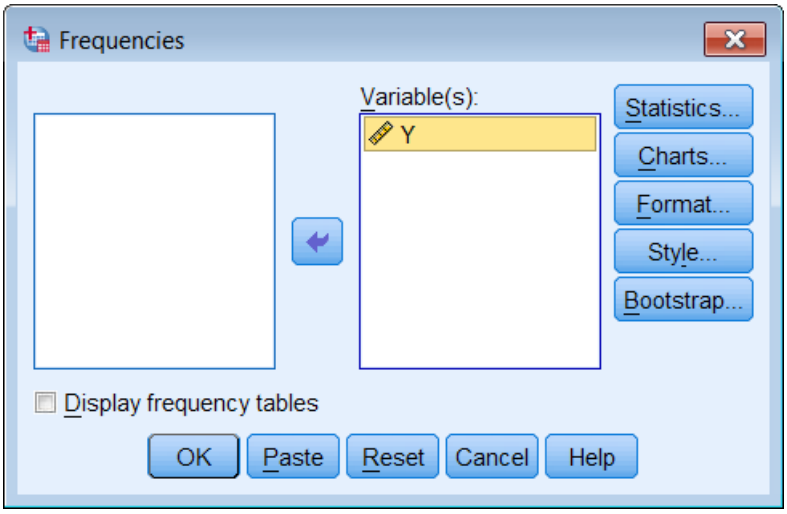
اختيار الأمر **[Analyze]** من شريط التعليمات.

- يظهر لنا قائمة تحتوي على كل الأساليب الإحصائية ، نختار من بينها

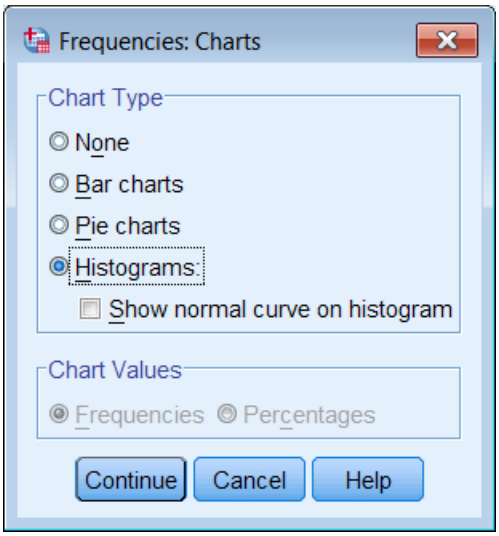
[Descriptive Statistics]ثم اختيار الأمر **[Frequencies]** فيظهر المربع التالي:



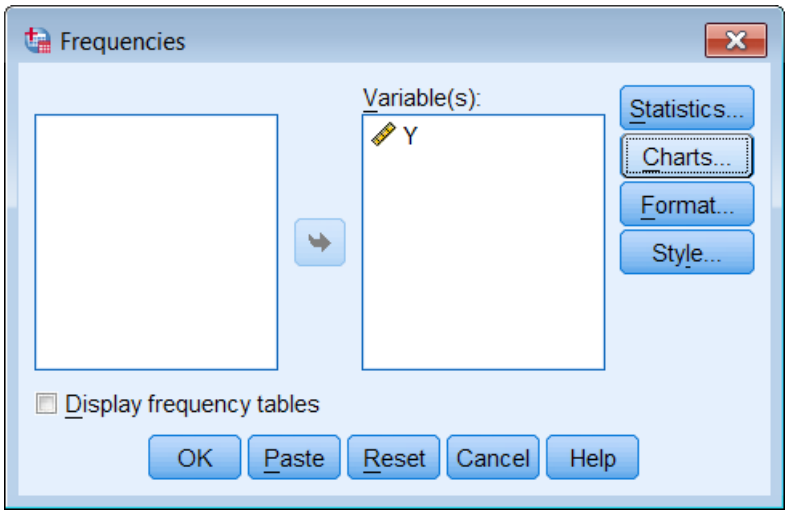
- التظليل على المتغير المستهدف تحليله وهو [Y] ، ثم ننقل هذا المتغير إلى المربع الأيمن الخاص بقائمة المتغيرات [Variables].



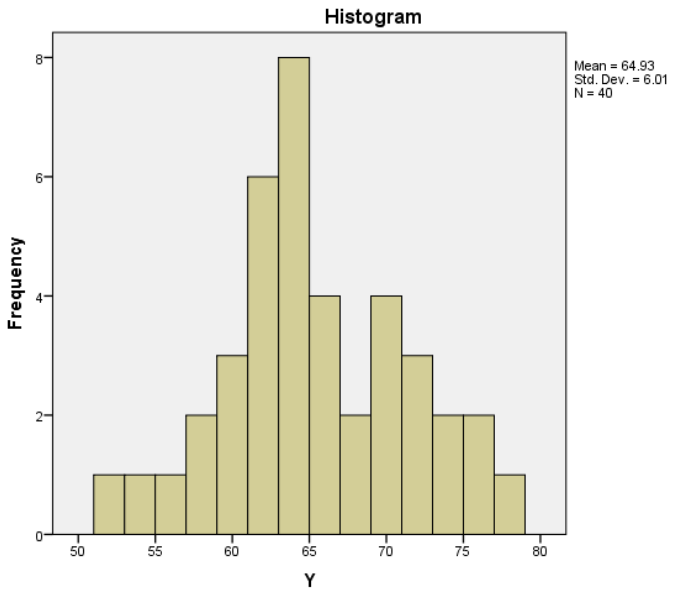
- إزالة التنشيط الخاص بعرض الجداول التكرارية [Display Frequency Tables] فتظهر رسالة خاصة بالموافقة بعدم عرض الجداول التكرارية يجب الموافقة عليها بالضغط OK
- الضغط على الأمر [Charts]، لاختيار الرسم المطلوبة.
- يظهر لنا مربع حواري يتضمن الرسوم، نختار من بينها المطلوب (Histograms)



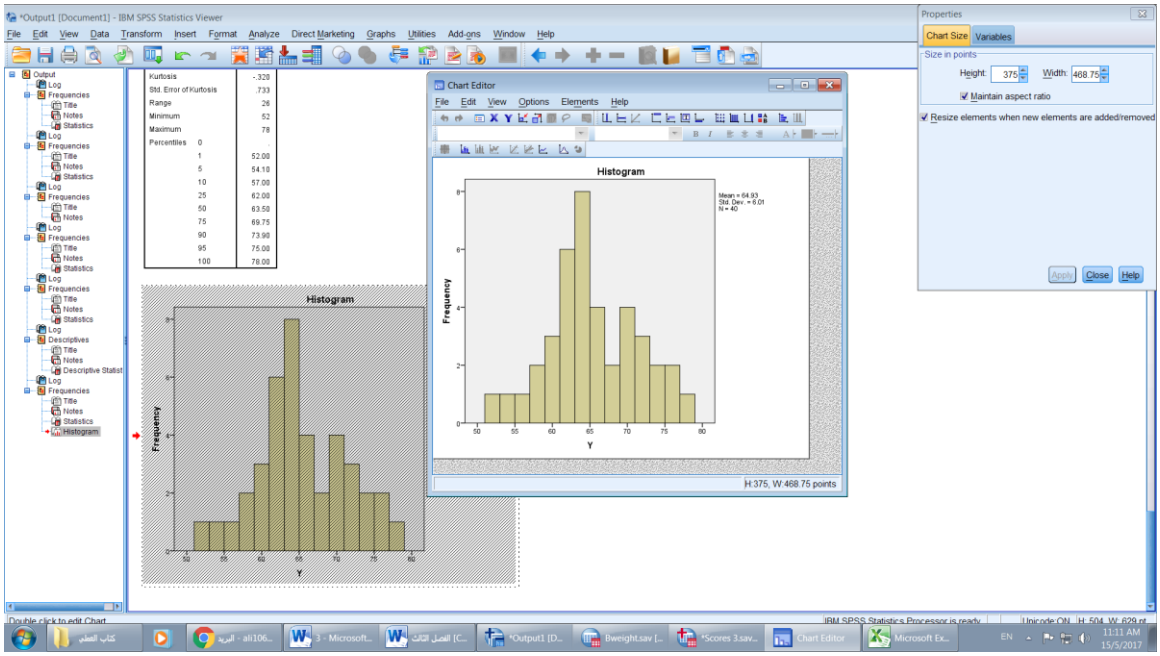
- الضغط على كلمة [Continue]فيظهر المربع التالي



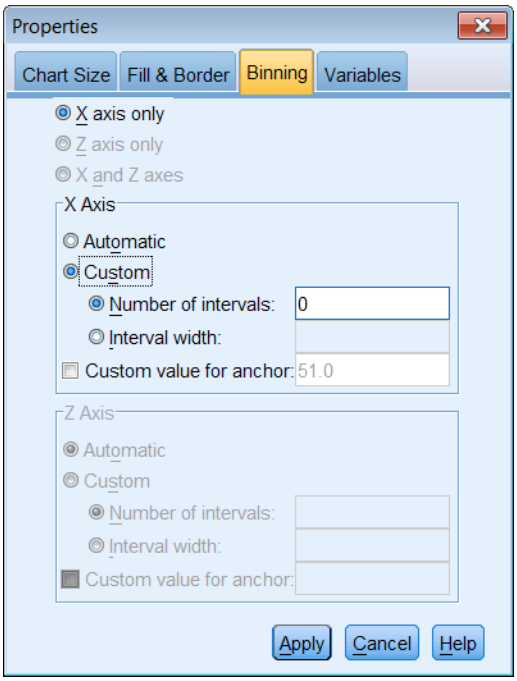
- الضغط على كلمة [OK]فيظهر الرسم التالي:



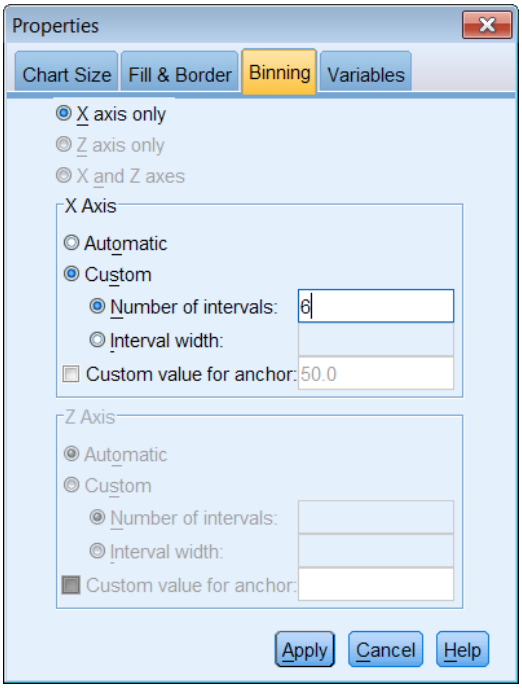
- لتحسين الرسم اضغط بالفارة مرتين على الرسم فيظهر لك نافذة أخرى باسم Chart Editor



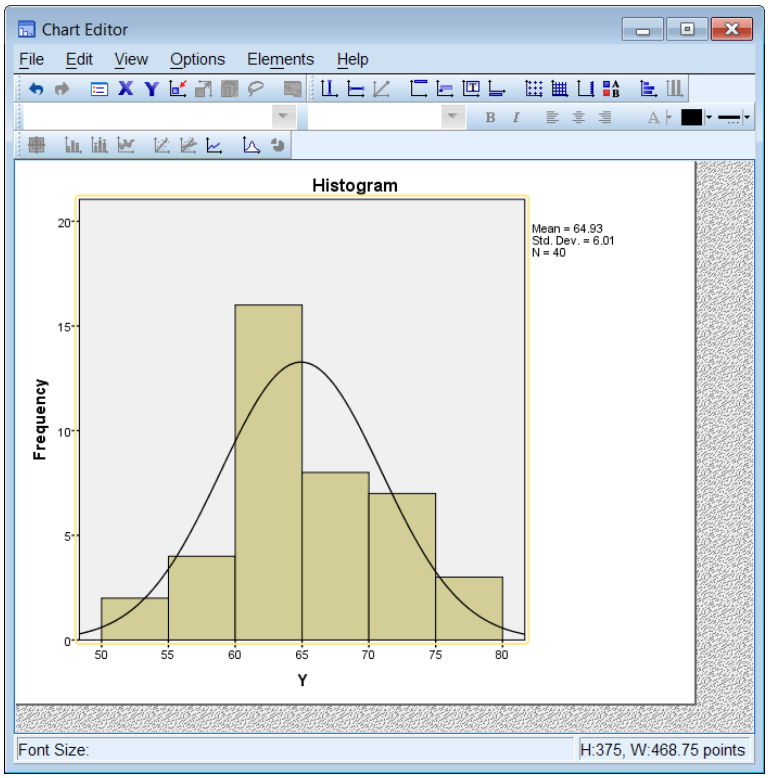
- نضغط على الأعمدة فيظهر نافذة أخرى باسم properties نضغط على Binning لتحديد عدد الفئات



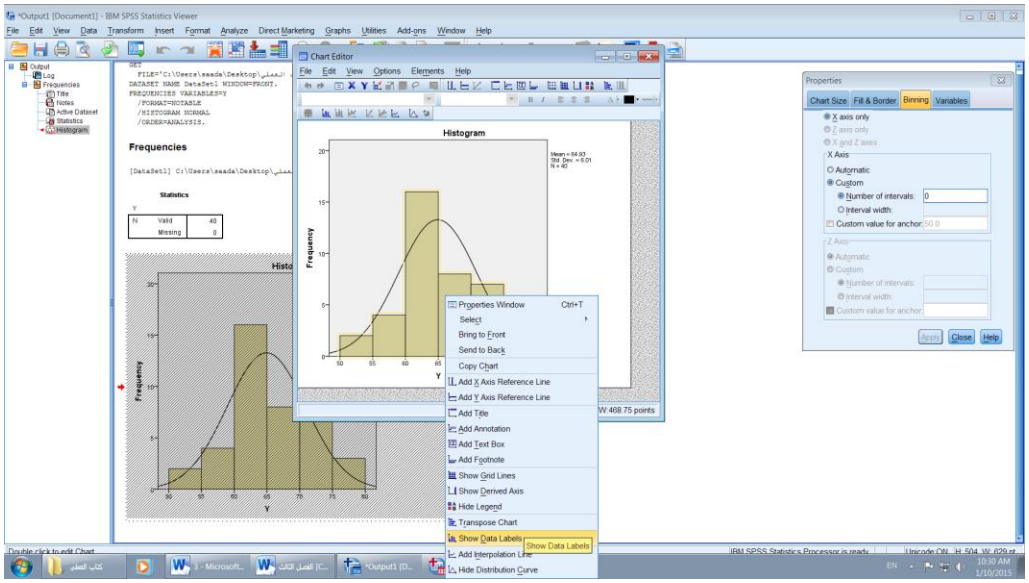
- نختار Custom ثم نختار Number of intervals ونضع عدد الفئات وهو رقم 6 كما بالمربع التالي:



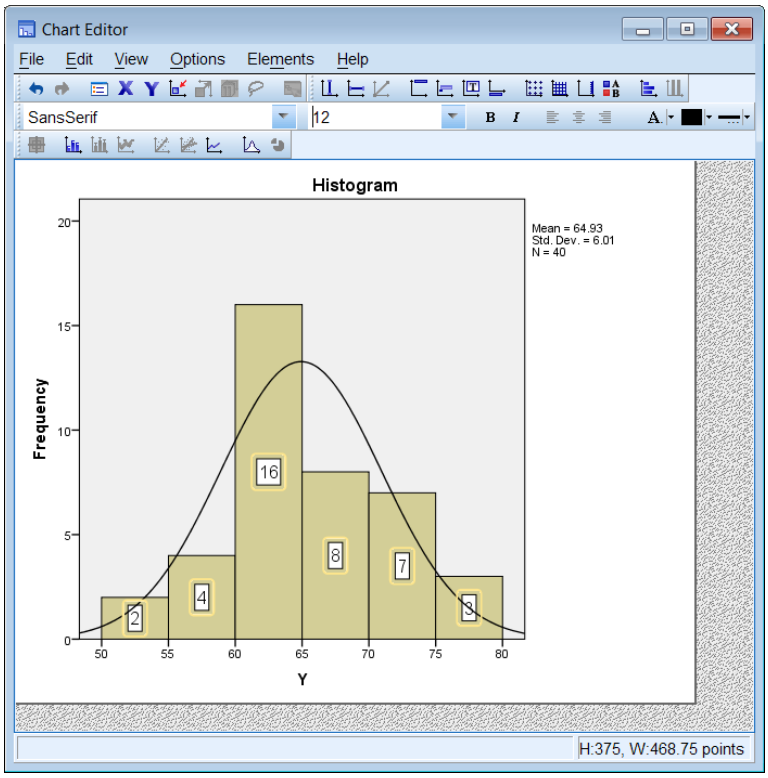
- نضغط Apply فيظهر الشكل التالي:



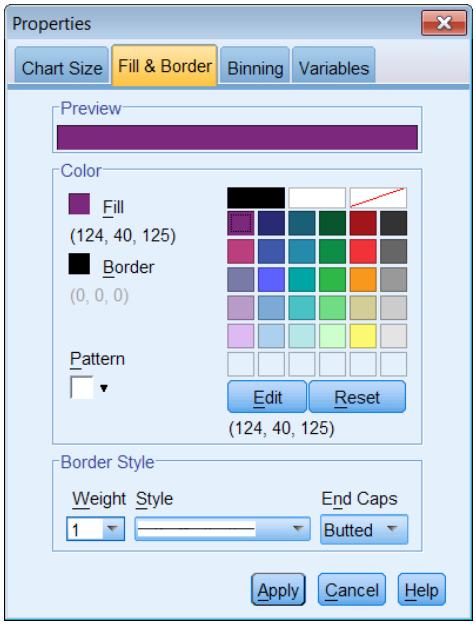
- لاظهار تكرار الفئات على الرسم نضغط على الأعمدة، ثم نضغط بيمين الفارة فيظهر قائمة نختار منها Show Data Lapels كما بالشكل التالي:



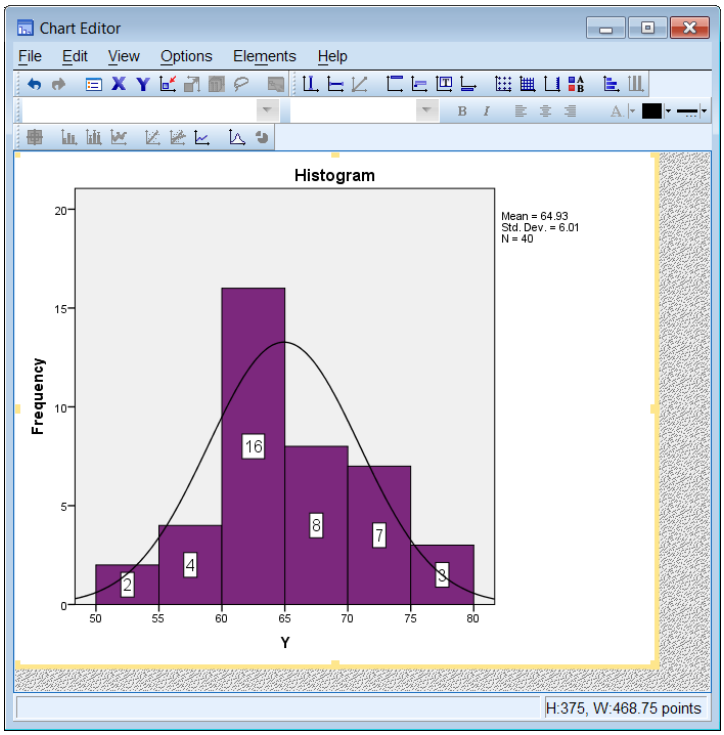
- فيظهر الشكل التالي:



- ولتغيير لون الأعمدة نختار Fill & Border

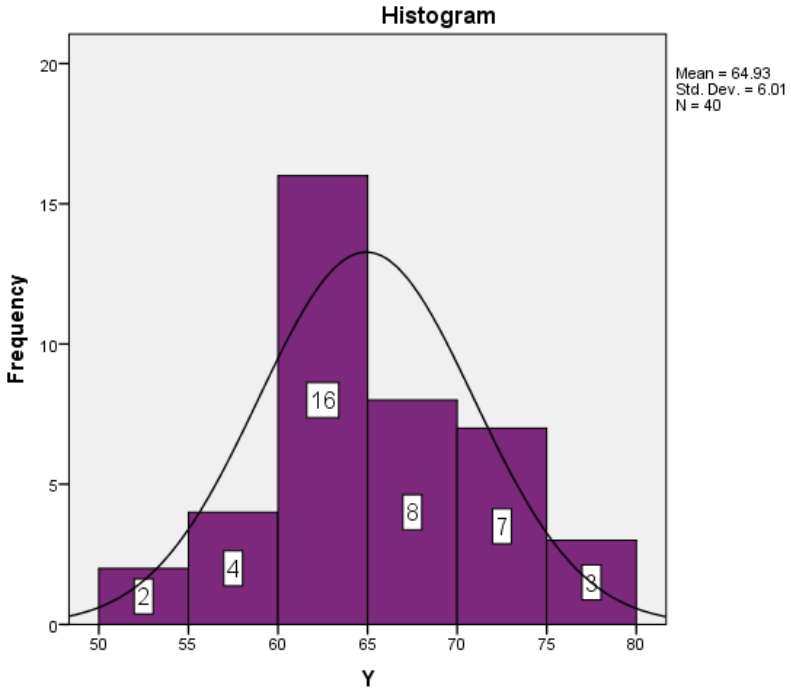


- نختار أي لون ونضغط Apply فيظهر الشكل التالي:



- نغلق نافذة Chart Editor فيظهر الرسم على نافذة Output كالتالي:

وبالتالي يكون حصلنا على المدرج والمنحنى التكراري والتوزيع التكراري.



تطبيقات

1- إذا كانت لديك البيانات التالية المتعلقة بأوزان مجموعة من الاطفال والبيانات هي :

12 ,10 ,13 ,14 , 11,16 , 15, 9,12 ,11, 13 ,15,16,12, 12

المطلوب: باستخدام برنامج SPSS أوجد:

1. إيجاد مقايس النزعة المركزية ومقاييس التشتت.
2. إيجاد القيمة التي تتوسط القيم والقيمة الاكثر تكراراً .
3. إيجاد القيم المعيارية.
4. إيجادالانحراف الربيعي وتمثيل البيانات عبر صندوق Box plots .

2- في دراسة قام بها أحد الباحثين لمعرفة أكثر أنواع الألبان ومنتجاتها مبيعات في أسواق مدينة الرياض تم اختيار 50 مركز تجاري وتم الحصول على النتائج الآتية:

زبادي	لبن	قشطه	زبادي	قشطه	لبنة	زبادي	حليب
لبن	زبادي	قشطه	قشطه	زبادي	قشطه	حليب	زبادي
حليب	قشطه	لبن	لبنة	قشطه	حليب	لبن	قشطه
زبادي	حليب	زبادي	قشطه	حليب	زبادي	لبنة	زبادي
قشطه	زبادي	حليب	قشطه	لبن	حليب	قشطه	حليب
لبنة	قشطه	لبنة	حليب	لبنة	لبنة	قشطه	قشطه
زبادي	لبن						

والمطلوب:

ما هو نوع المتغير؟، وما هو المعيار المستخدم في قياس البيانات؟.

اعرض البيانات في شكل جدول تكراري باستخدام برنامج SPSS.

كون التوزيع التكراري النسبي، ثم علق على النتائج.

3- فيما يلي درجات 40 طالب :

18	72	64	82	71	66	60	40	97	44
19	79	68	65	60	88	78	77	69	55
62	55	70	83	75	61	66	71	64	62
50	60	66	67	62	74	62	61	72	57

باستخدام برنامج SPSS المطلوب:

أ-كون جدول توزيع تكراري مستخدماً حدود الفئات العليا التالية :

42 ، 47 ، .....، 97 .

ب- ارسم المدرج التكراري.

4- البيانات التالية تمثل أعمار مجموعة من المتدربين في أحد البرامج التدريبية باحدى المؤسسات التعليمية :

30	40	25	27	43	37	38	39	33	33
41	32	36	28	27	23	30	31	36	22
32	36	37	34	35	29	32	36	34	44
34	26	27	33	34	32	28	24	35	34
39	30	37	38	34	34	39	28	27	31

المطلوب:

أ- حساب مقاييس النزعة المركزية باستخدام برنامج SPSS

ب- حساب مقاييس التشتت باستخدام برنامج SPSS