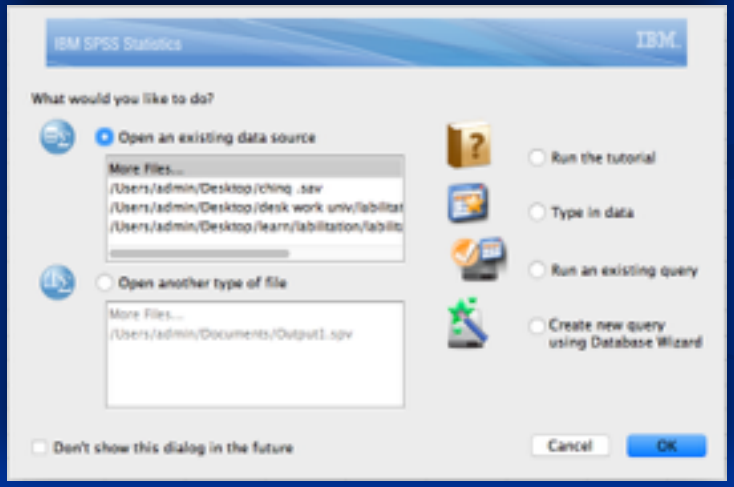
المركز الجامعي لميلة

دروس مكثفة في تحليل قواعد البيانات



الأستاذ: غيشي عبد العالي

السنة الدراسية 2019 - 2020

الدرس الثاني

المحتوى:

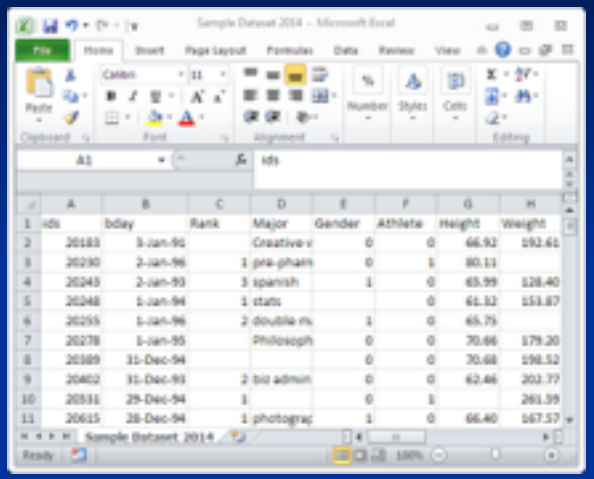
- إدخال البيانات من برامج أخرى
- استحداث/ حساب متغير جديد
- إعادة ترميز المتغيرات
- أوامر وصف البيانات عبر برنامج (SPSS)
- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي

- Inputting data from other application programs
- إدخال البيانات من برامج أخرى

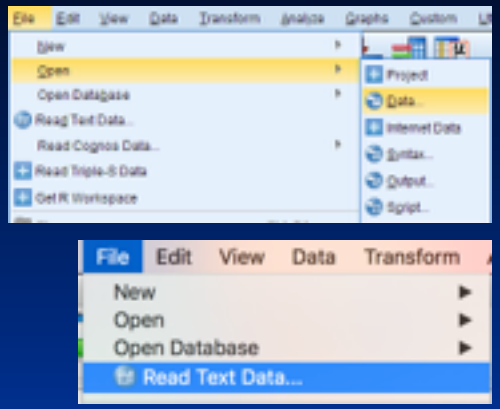
Import data from an Excel file

Here is an example of what properly formatted data looks like in Excel 2010:

0

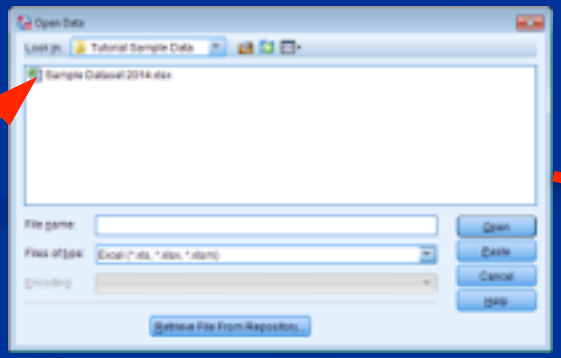


1

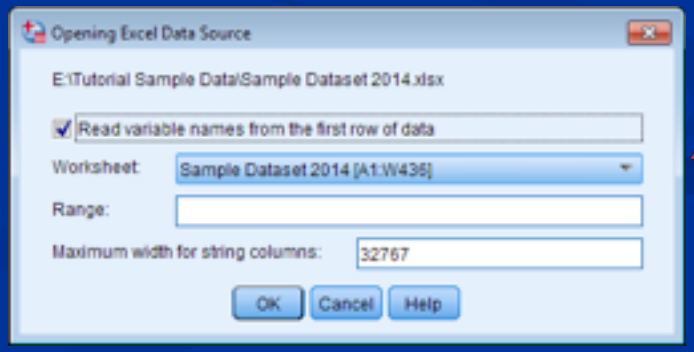


2

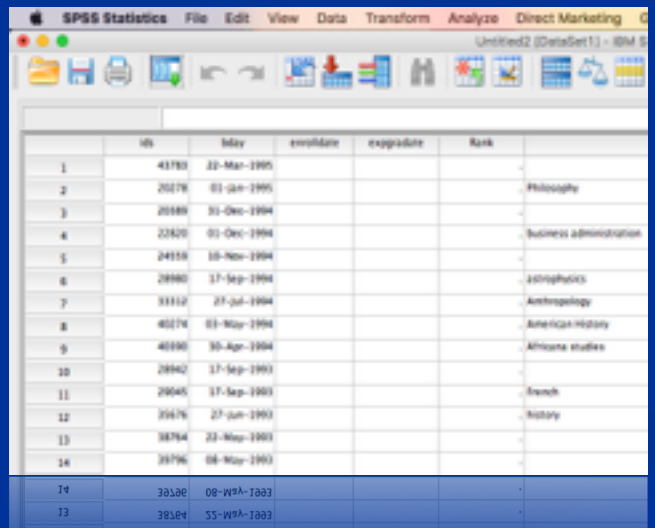
تعيين مكان الملف



3



4



- Creating/calculating a new variable
- استحداث/ حساب متغير جديد

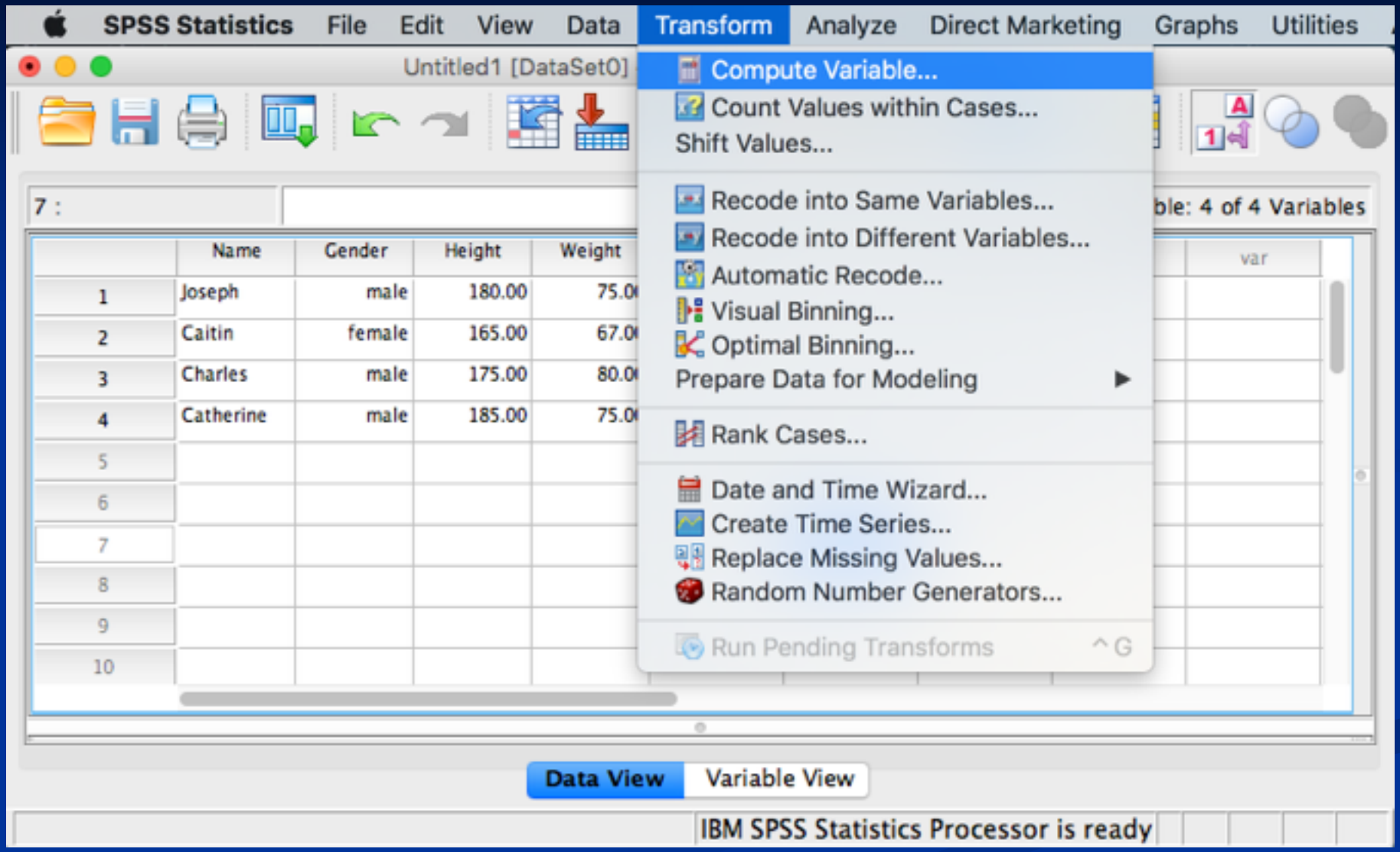
Suppose that the researcher would like to include an extra column in the example which indicates the 'body-mass index (BMI)'. The BMI is defined as the body weight in kilograms divided by the square of the height in metres. let use the table below:

Table 1.1

Name	*Gender*	*Height (cm)*	*Weight (kg)*
Joseph	1	180	75
Caitlin	0	165	67
Charles	1	175	80
Catherine	0	170	70
Peter	1	185	75

The path to be followed to calculate an additional variable is Transform/Compute Variable

- Creating/calculating a new variable
- استحداث/ حساب متغير جديد

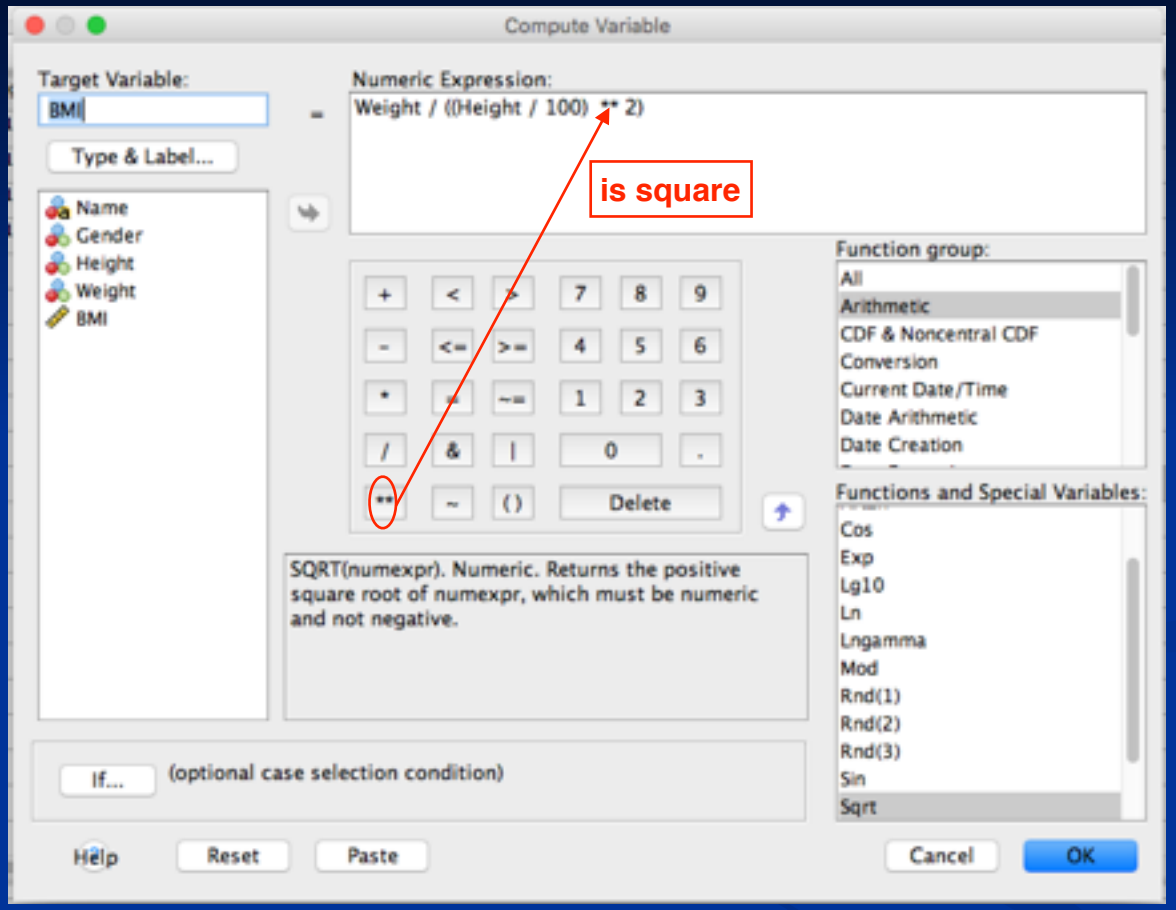


A dialogue window will be displayed

• Creating/calculating a new variable • استحداث/ حساب متغير جديد

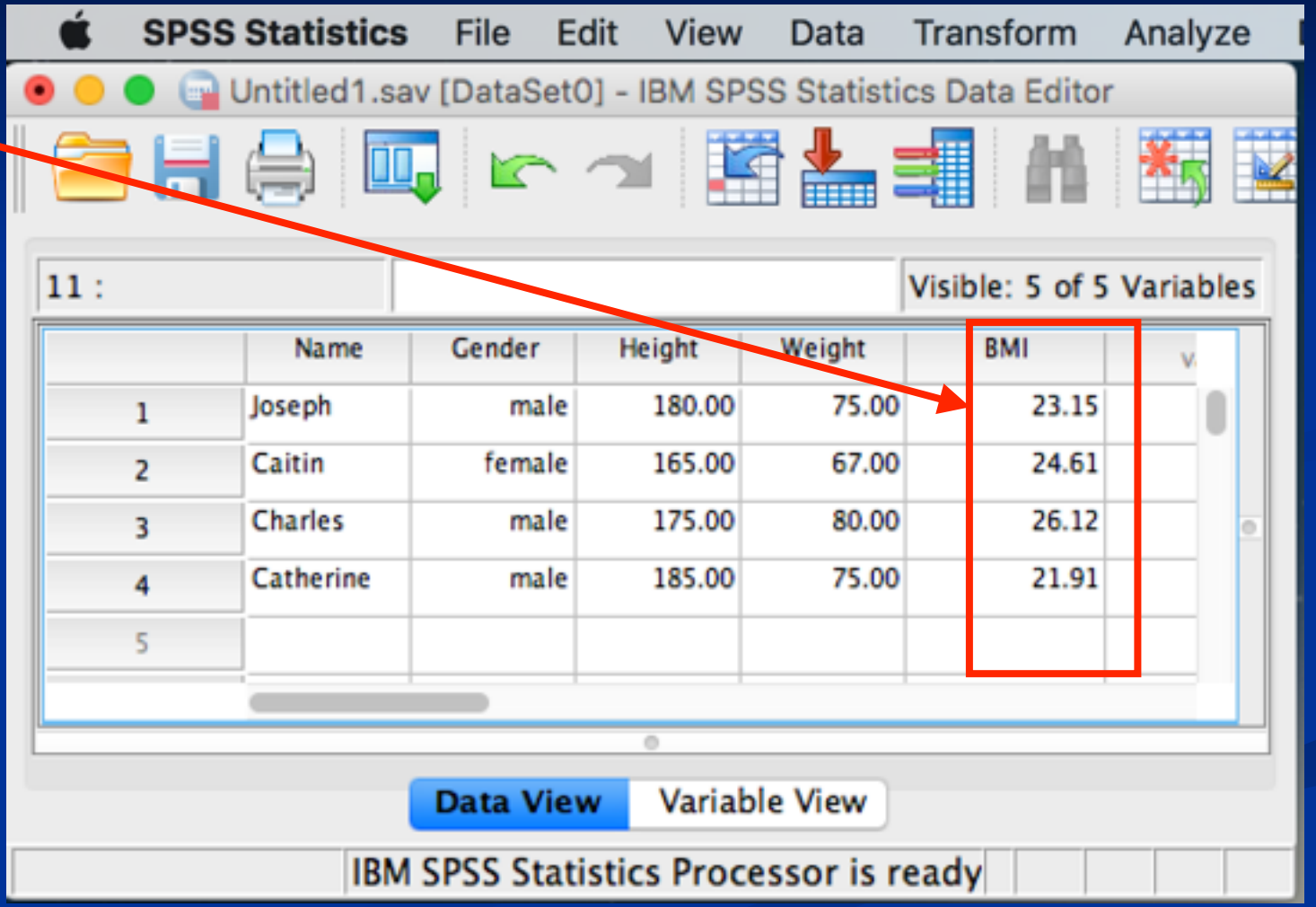
1. type the name of the new variable (BMI in this case)
2. In the 'Numeric Expression' field, type the formula .
3. you may also select the variable names from the left box and click the button).
4. click OK

5. The new variable will now be shown in the 'Data View' screen



- Creating/calculating a new variable
- استحداث/ حساب متغير جديد

The new variable

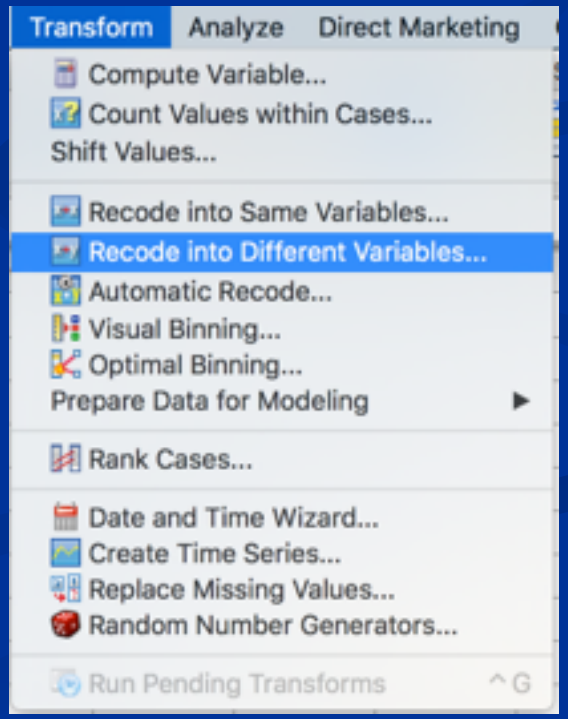


Recoding variables:

Sometimes you will want to transform a variable by grouping its categories or values together. For example, you may want to change a continuous variable into a categorical variable, or you may want to merge the categories of a nominal variable. In SPSS, this type of transform is called recoding.

In SPSS, there are three basic options for recoding variables:

1. Recode into Different Variables
2. Recode into Same Variables
3. DO IF syntax



Recode into Different Variables: Let's say that these five people must complete a questionnaire. we assume that this questionnaire consists of three questions (statements) in which their preferences regarding candy are being studied. The three statements must be evaluated on a 7-point scale, ranging from 'totally disagree (1)' to 'totally agree (7)'.

Question 1: When I watch television in the evening, I eat candy on a regular basis.
Question 2: If I'm hungry between meals, I will eat fruit more often than candy.
Question 3: I always like to add extra sugar to my dessert.

Their answers are shown in this table:

Name	*Question 1*	*Question 2*	*Question 3*
Joseph	7	3	6
Caitlin	2	5	3
Charles	5	1	5
Catherine	3	6	2
Peter	6	2	6

- Recoding variables
- إعادة ترميز المتغيرات

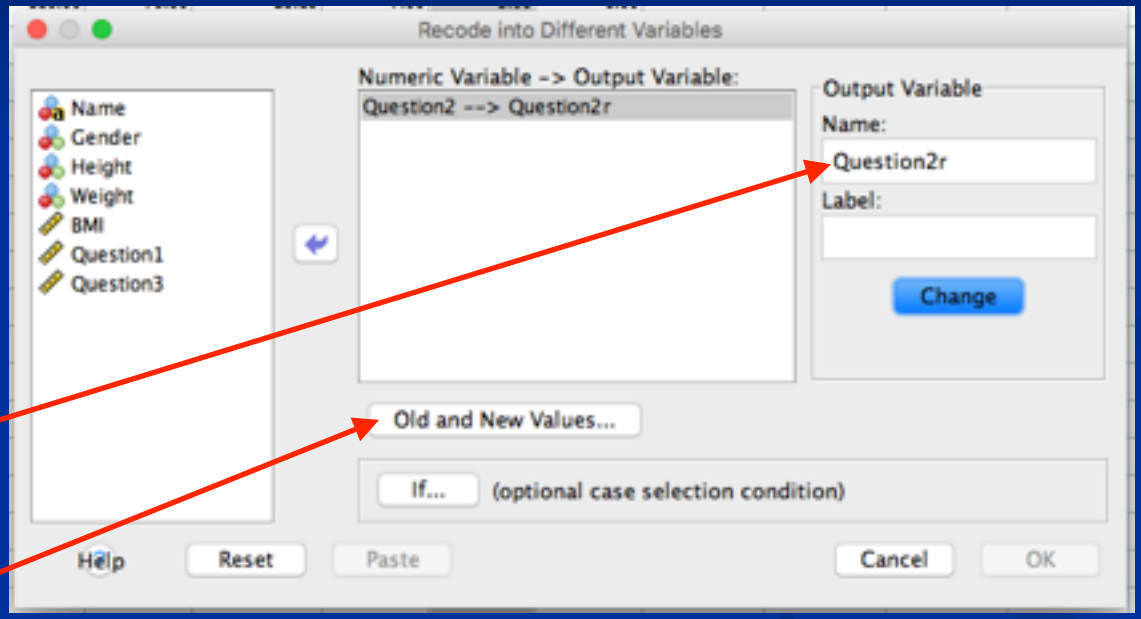
If the researcher wishes to perform an analysis of this data (e.g. calculate an 'average for candy preference'), he must first determine whether the questions were all scaled 'in the same direction'.

	Name	Gender	Height	Weight	BMI	Question1	Question2	Question3
1	Joseph	1.00	180.00	75.00	23.15	7.00	3.00	6.00
2	Caitin	.00	165.00	67.00	24.61	2.00	5.00	3.00
3	Charles	1.00	175.00	80.00	26.12	5.00	1.00	5.00
4	Catherine	.00	185.00	70.00	24.22	3.00	6.00	2.00
5	peter	1.00	185.00	75.00	21.91	6.00	2.00	6.00

- Take question 2 for example.
- Question 2: If I'm hungry between meals, I will eat fruit more often than candy.
- question 2 is not scaled in the same direction as questions 1 and 3 and for this reason needs to be recoded.

Go to Transform/Recode into Different Variables which will bring up this subscreen.

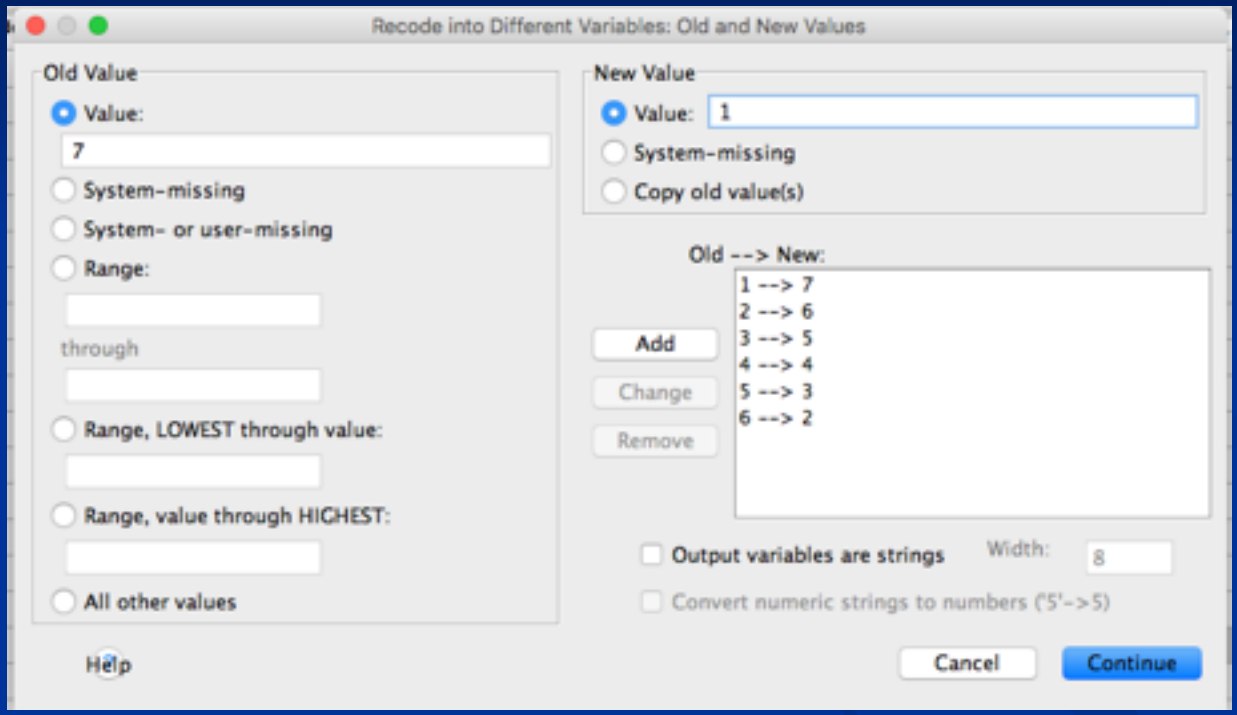
1. Click on 'question2' in the list of variables on the left and click
2. Under 'Output Variable' enter the name of the recoded variable (question2r) and click 'Change'
3. Click on the 'Old and New Values' button
4. you see a dialogue window
5. see in the next slide



• Recoding variables • إعادة ترميز المتغيرات

For each value to be recoded, the researcher must input the old and the new value.

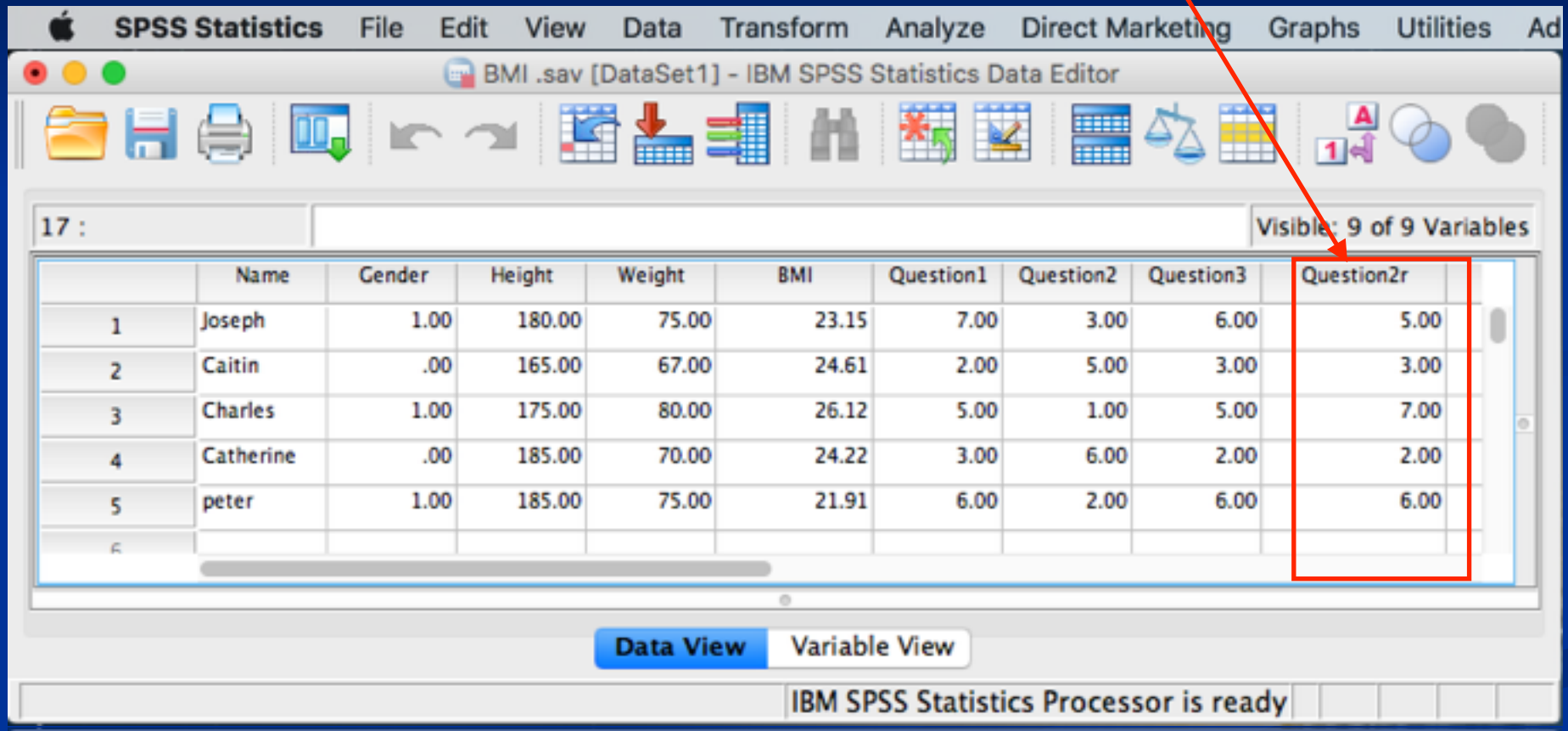
1. For ‘Old Value’, fill in the value to be changed and under ‘New Value’, type the new value (1). Next, click on the ‘Add’ button
2. Click Continue
3. then Click Ok



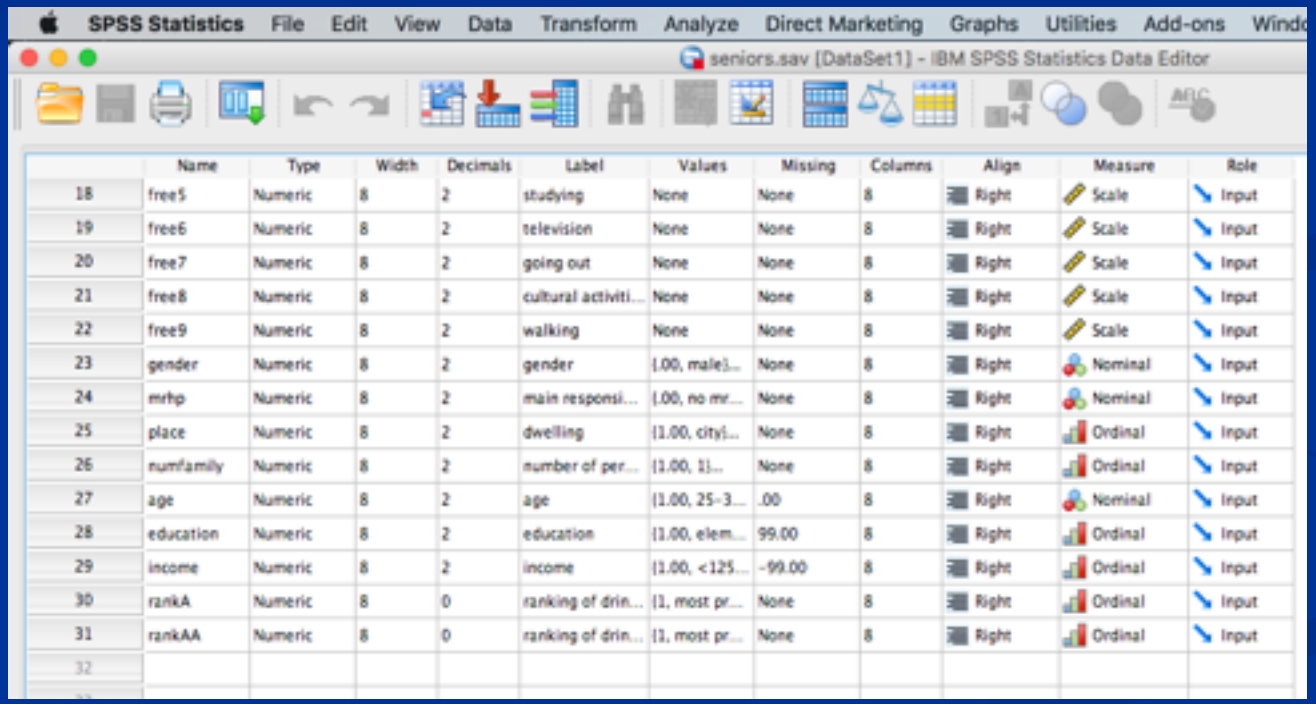
4. An extra variable with the recoded values has been created in the ‘Data View’ tab. See in the next slide

- Recoding variables
- إعادة ترميز المتغيرات

Now, a useful average of the variables 'question1', 'question2r' and 'question3' may be calculated if desired.



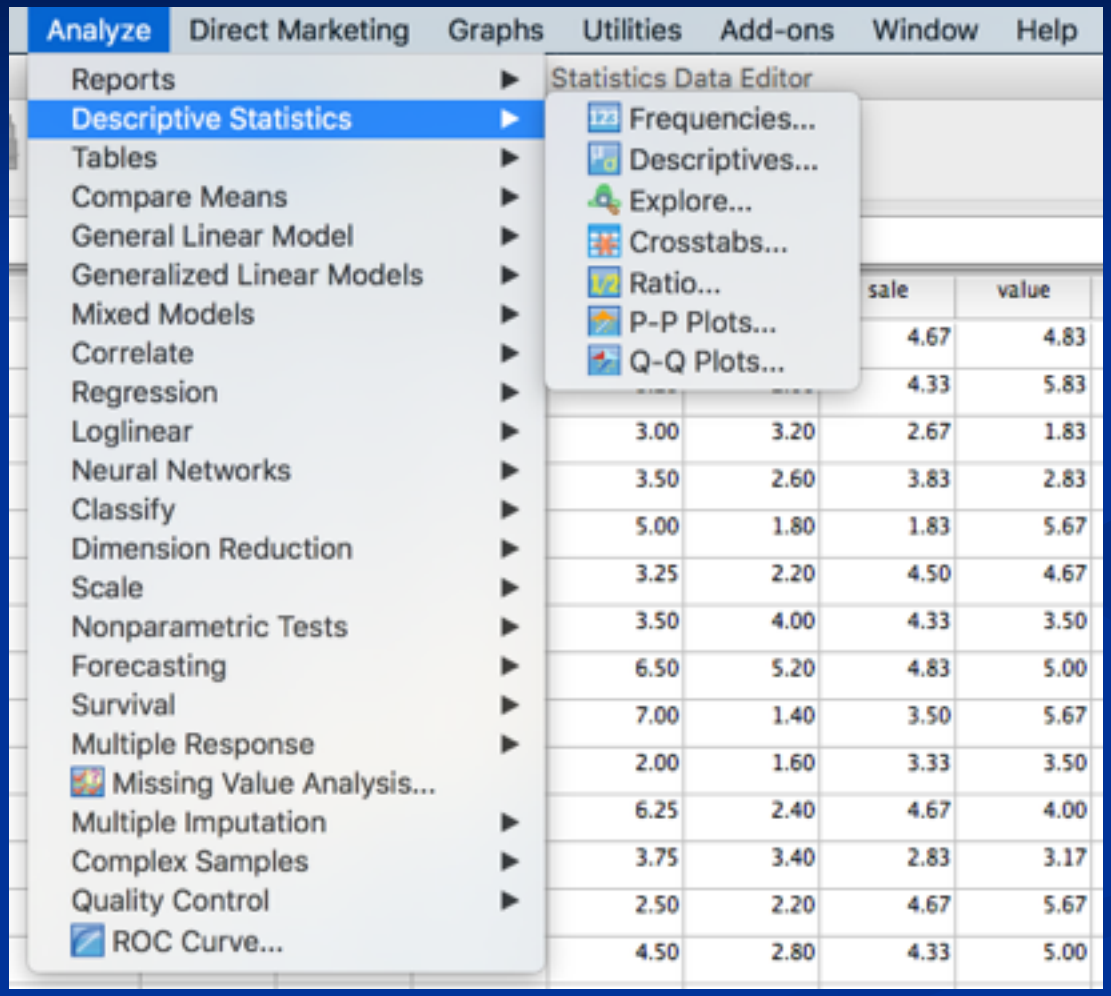
سنعتمد في تقديم هذه الأوامر الثلاثة (Frequency, Descriptive, Explore) على قاعدة البيانات في الملف المسمى (Seniors.sav)، ويشتمل هذا الملف على العديد من المتغيرات كما يبين الشكل في الأسفل:



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
18	free5	Numeric	8	2	studying	None	None	8	Right	Scale	Input
19	free6	Numeric	8	2	television	None	None	8	Right	Scale	Input
20	free7	Numeric	8	2	going out	None	None	8	Right	Scale	Input
21	free8	Numeric	8	2	cultural activiti...	None	None	8	Right	Scale	Input
22	free9	Numeric	8	2	walking	None	None	8	Right	Scale	Input
23	gender	Numeric	8	2	gender	{.00, male}...	None	8	Right	Nominal	Input
24	mrhp	Numeric	8	2	main responsi...	{.00, no mr...	None	8	Right	Nominal	Input
25	place	Numeric	8	2	dwelling	{1.00, city}...	None	8	Right	Ordinal	Input
26	numfamily	Numeric	8	2	number of per...	{1.00, 1}...	None	8	Right	Ordinal	Input
27	age	Numeric	8	2	age	{1.00, 25-3...	.00	8	Right	Nominal	Input
28	education	Numeric	8	2	education	{1.00, elem...	99.00	8	Right	Ordinal	Input
29	income	Numeric	8	2	income	{1.00, <125...	-99.00	8	Right	Ordinal	Input
30	rankA	Numeric	8	0	ranking of drin...	{1, most pr...	None	8	Right	Ordinal	Input
31	rankAA	Numeric	8	0	ranking of drin...	{1, most pr...	None	8	Right	Ordinal	Input
32											

تمثل هذه القاعدة من البيانات قياسات للعديد من السلوكيات الشرائية لعينة مكونة من 310 مستهلك. كما تقدم أيضا خياراتهم في أوقات الراحة والفراغ، أين تم استخدام سلم ليكرت السباعي في تقييم مواقفهم.

تشتمل القائمة (Descriptive Statistics) على قائمة الأوامر التالية:



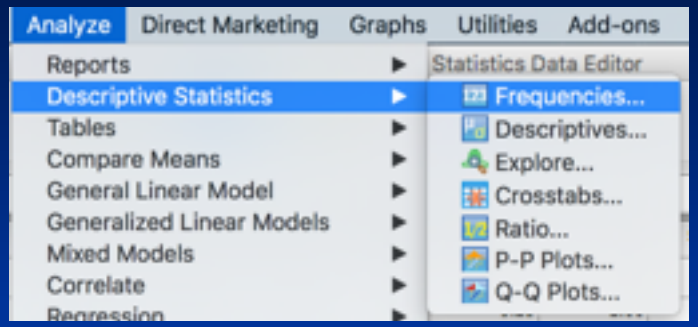
يتم الوصول إلى القائمة من خلال المسار التالي:

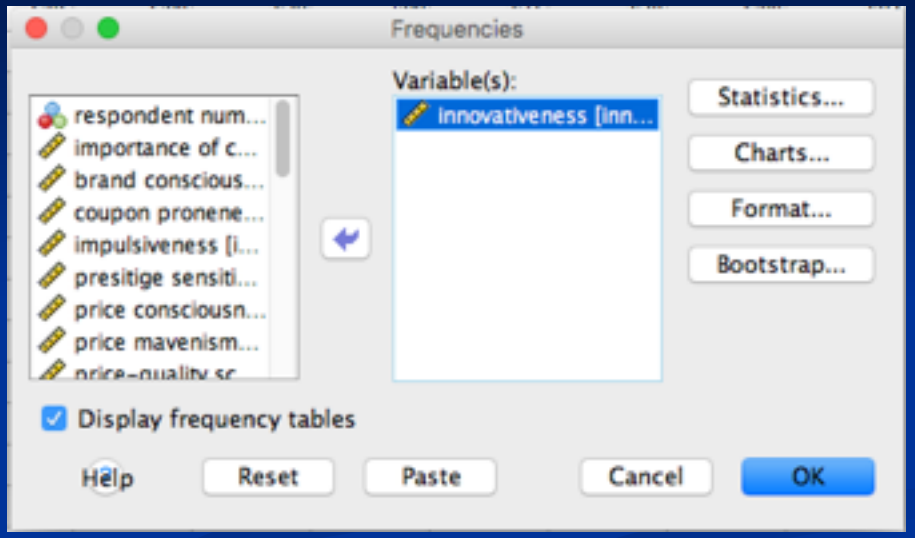
Analyse/Descriptive Statistics/

ثم ننتقي الأمر المناسب

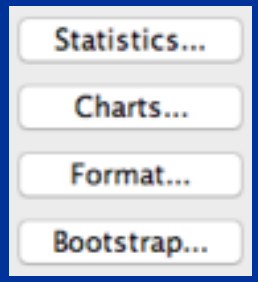
1. الأمر: (Frequencies)
2. الأمر: (Descriptive)
3. الأمر: (Explore)
4. الأمر: (Crosstabls)
5. الأمر: (Ratio)
6. الأمر: (P-P Plots)
7. الأمر: (Q-Q plots)

1. الأمر: (Frequencies) عند اختيار هذا الأمر يظهر مربع الحوار التالي:





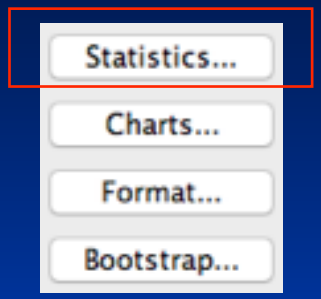
بعد إدخال المتغير يتم انتقاء المقاييس الملائمة من خلال هذه القائمة بالضغط على كل على حدى وملء ما يلائم الباحث



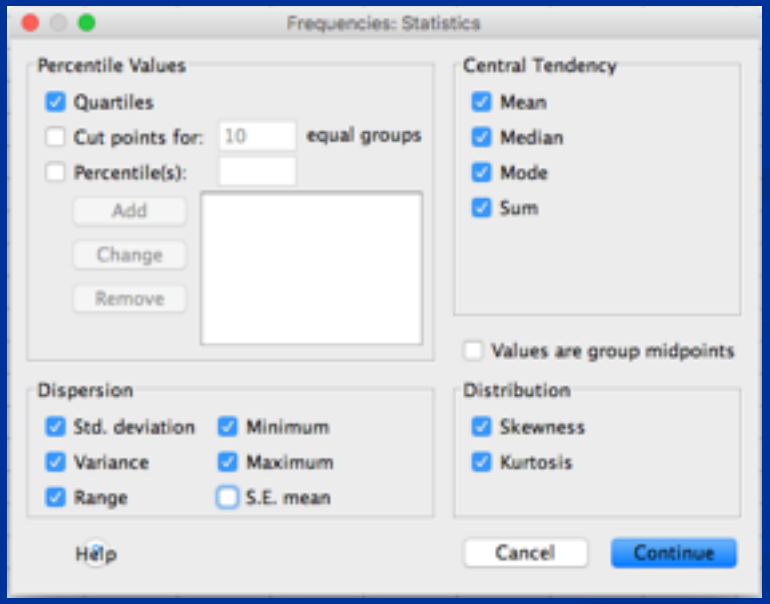
ثم ندخل المتغير المراد تحليله إلى مربع التحليل المسمى (Variables) من قائمة المتغيرات عبر السهم التالي:

مثلا في هذه الحالة تم إدخال المتغير (Innovativeness)

1. الأمر: (Frequencies):



من خلال الضغط على إيقونة (Statistics) يمكنك الحصول على كل من مقاييس النزعة المركزية (1) ومقاييس التشتت (2) و الشكل (3)، ويتم ذلك بتأشير الباحث في الخانات الملائمة كما يظهر في مربع الحوار أسفله:



نضغط على (Continue) ثم على (OK)

أنظر المخرجات في الصفحة الموالية

1. الأمر: (Frequencies):

بالتطبيق على المثال السابق للملف (Seniors.sav) والمتغير (Innovativeness) نتحصل على الجدولين التاليين:

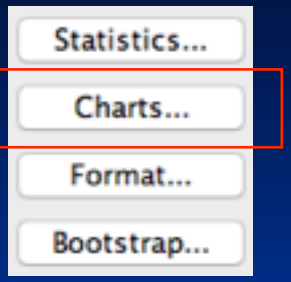
innovativeness

N	Valid	310
	Missing	0
Mean		3.6826
Median		3.6000
Mode		4.00
Std. Deviation		.90315
Variance		.816
Skewness		.068
Std. Error of Skewness		.138
Kurtosis		.266
Std. Error of Kurtosis		.276
Range		5.60
Minimum		1.00
Maximum		6.60
Sum		1141.60
Percentiles	25	3.0000
	50	3.6000
	75	4.2000

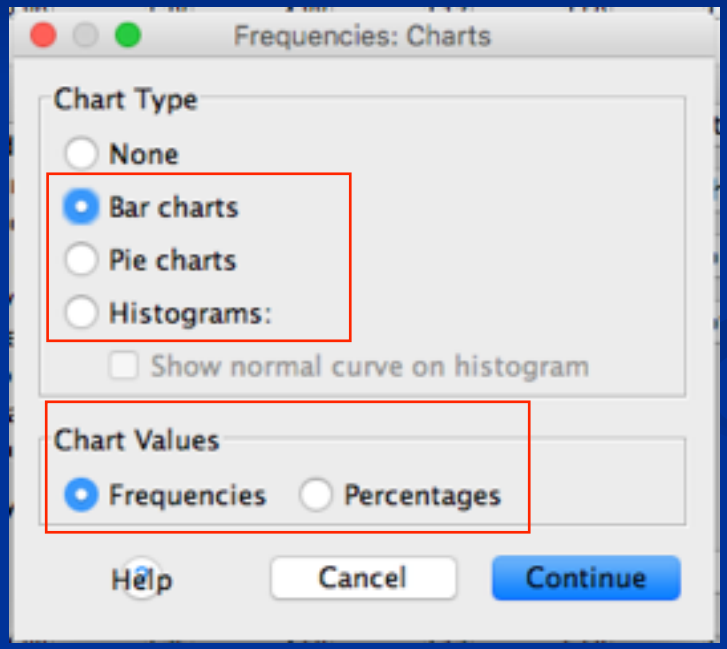
innovativeness

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	1	.3	.3	.3
	1.20	1	.3	.3	.6
	1.40	1	.3	.3	1.0
	1.60	3	1.0	1.0	1.9
	1.80	1	.3	.3	2.3
	2.20	12	3.9	3.9	6.1
	2.40	12	3.9	3.9	10.0
	2.60	9	2.9	2.9	12.9
	2.80	22	7.1	7.1	20.0
	3.00	16	5.2	5.2	25.2
	3.20	24	7.7	7.7	32.9
	3.40	30	9.7	9.7	42.6
	3.60	25	8.1	8.1	50.6
	3.80	22	7.1	7.1	57.7
	4.00	31	10.0	10.0	67.7
	4.20	30	9.7	9.7	77.4
	4.40	21	6.8	6.8	84.2
	4.60	13	4.2	4.2	88.4
	4.80	11	3.5	3.5	91.9
	5.00	6	1.9	1.9	93.9
	5.20	5	1.6	1.6	95.5
	5.40	6	1.9	1.9	97.4
	5.60	2	.6	.6	98.1
	5.80	2	.6	.6	98.7
	6.00	3	1.0	1.0	99.7
	6.60	1	.3	.3	100.0
	Total	310	100.0	100.0	

1. الأمر: (Frequencies):



من خلال الضغط على إيقونة (Chart) يمكنك الحصول على كل من التمثيل البياني في شكل أعمدة أو دوائر نسبية أو مدرج تكراري (1)، وكما يمكن إختيار أساس عرض القيم في الأشكال إما عبر التكرارات أو النسب المئوية (2)، ويتم ذلك بتأشير الباحث في الخانات الملائمة كما يظهر في مربع الحوار أسفله:



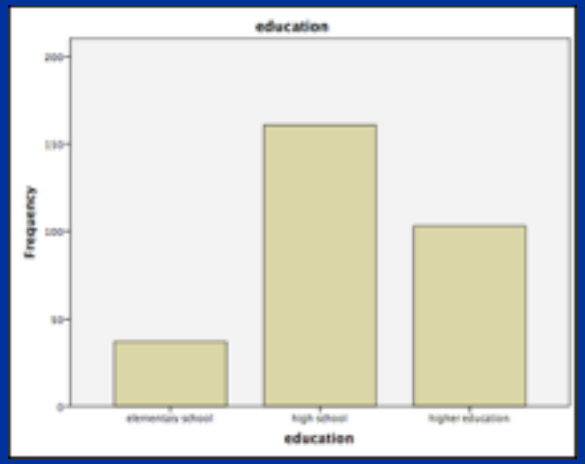
نضغط على (Continue) ثم على (OK)

أنظر المخرجات في الصفحة الموالية

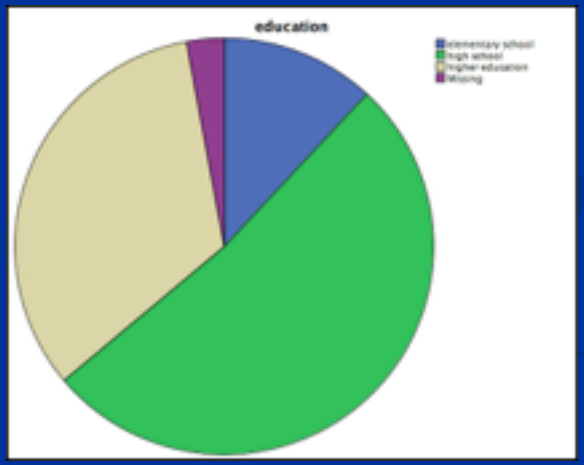
1. الأمر: (Frequencies):

بالتطبيق على المثال السابق للملف (Seniors.sav) والمتغيرين (Education, value)
نتحصل على الرسومات البيانية التالية:

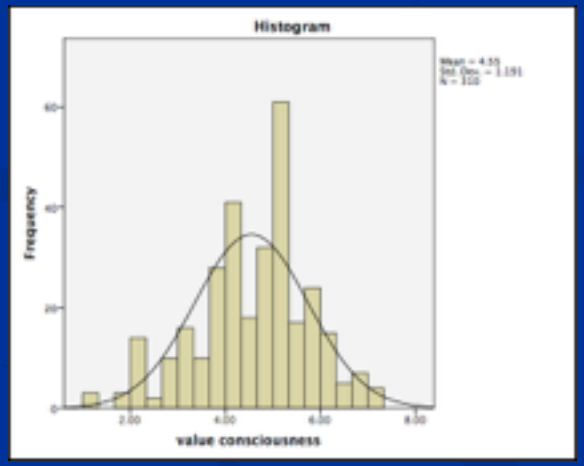
Bar charts



Pie charts



Histograms:
Show normal curve on histogram

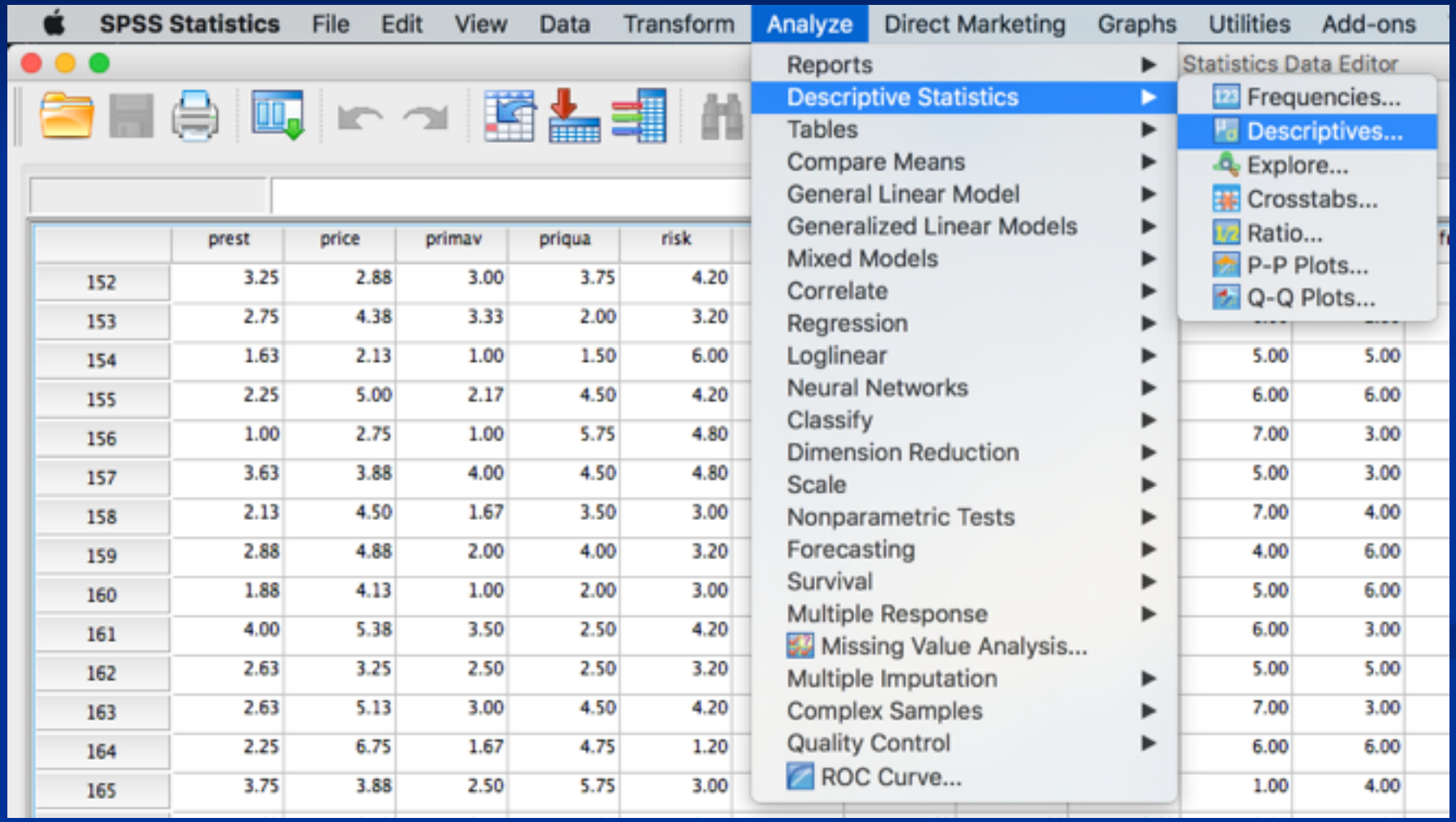


لإدخال المزيد من التحسينات على الشكل نضغط مرتين على الشكل ونتعامل مع مربع الحوار لمزيد من العرض في جوانب المعلومات والشكل على حد سواء.

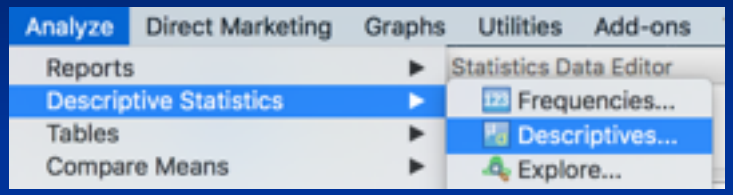
- الأمر الوصف (Descriptive)

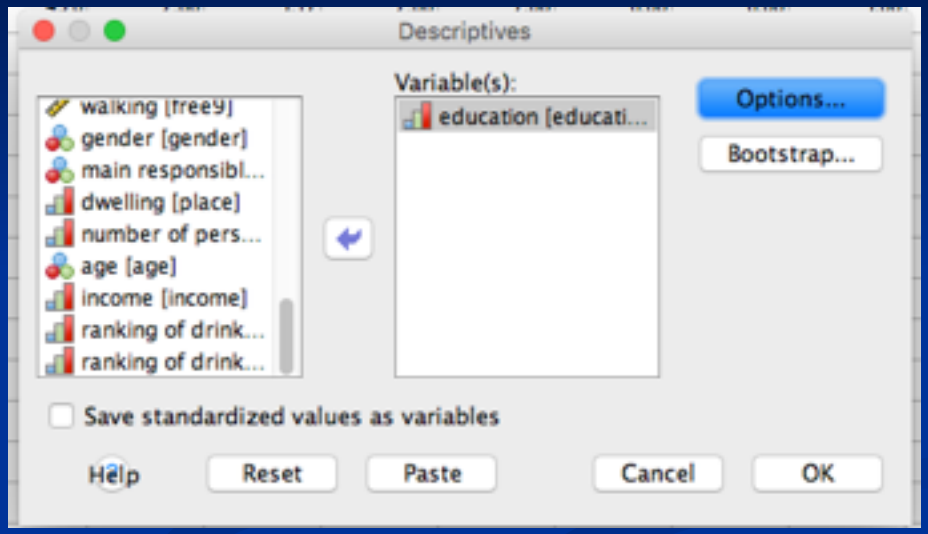
يتم الوصول إلى الأمر (Descriptive) من خلال المسار التالي:

Analyse/Descriptive Statistics/Descriptive



2. الأمر: (Descriptive) عند اختيار هذا الأمر يظهر مربع الحوار التالي:

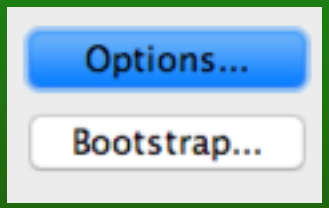




ثم ندخل المتغير المراد تحليله إلى مربع التحليل المسمى (Variables) من قائمة المتغيرات عبر السهم التالي:

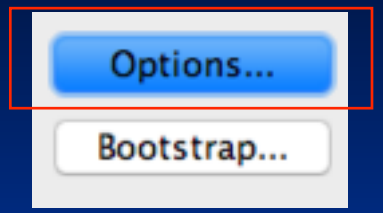
مثلا في هذه الحالة تم إدخال المتغير (Education)

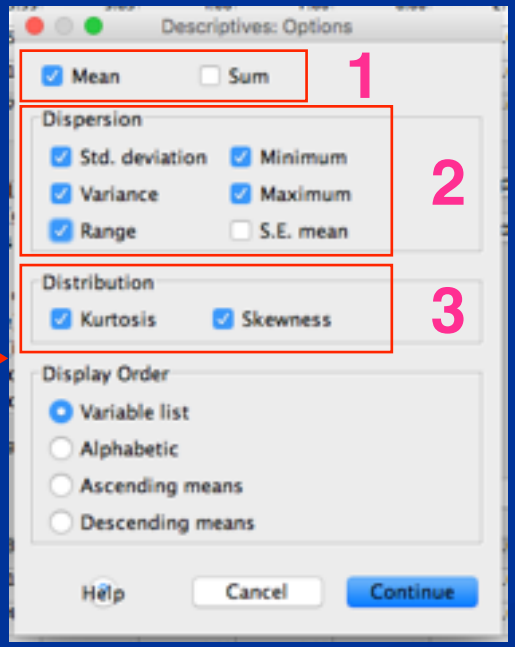
بعد إدخال المتغير يتم انتقاء المقاييس الملائمة من خلال هذه القائمة بالضغط على كل على حدى وملء ما يلائم الباحث



2. الأمر: (Descriptive):

من خلال الضغط على إيقونة (Options) يمكنك الحصول على كل من مقاييس النزعة المركزية (1) ومقاييس التشتت (2) و الشكل (3)، ويتم ذلك بتأشير الباحث في الخانات الملائمة كما يظهر في مربع الحوار أسفله:





نضغط على (Continue) ثم على (OK)

أنظر المخرجات في الصفحة الموالية

2. الأمر: (Descriptive):

بالتطبيق على المثال السابق للملف (Seniors.sav) والمتغير (Education) نتحصل على الجدول التالي:

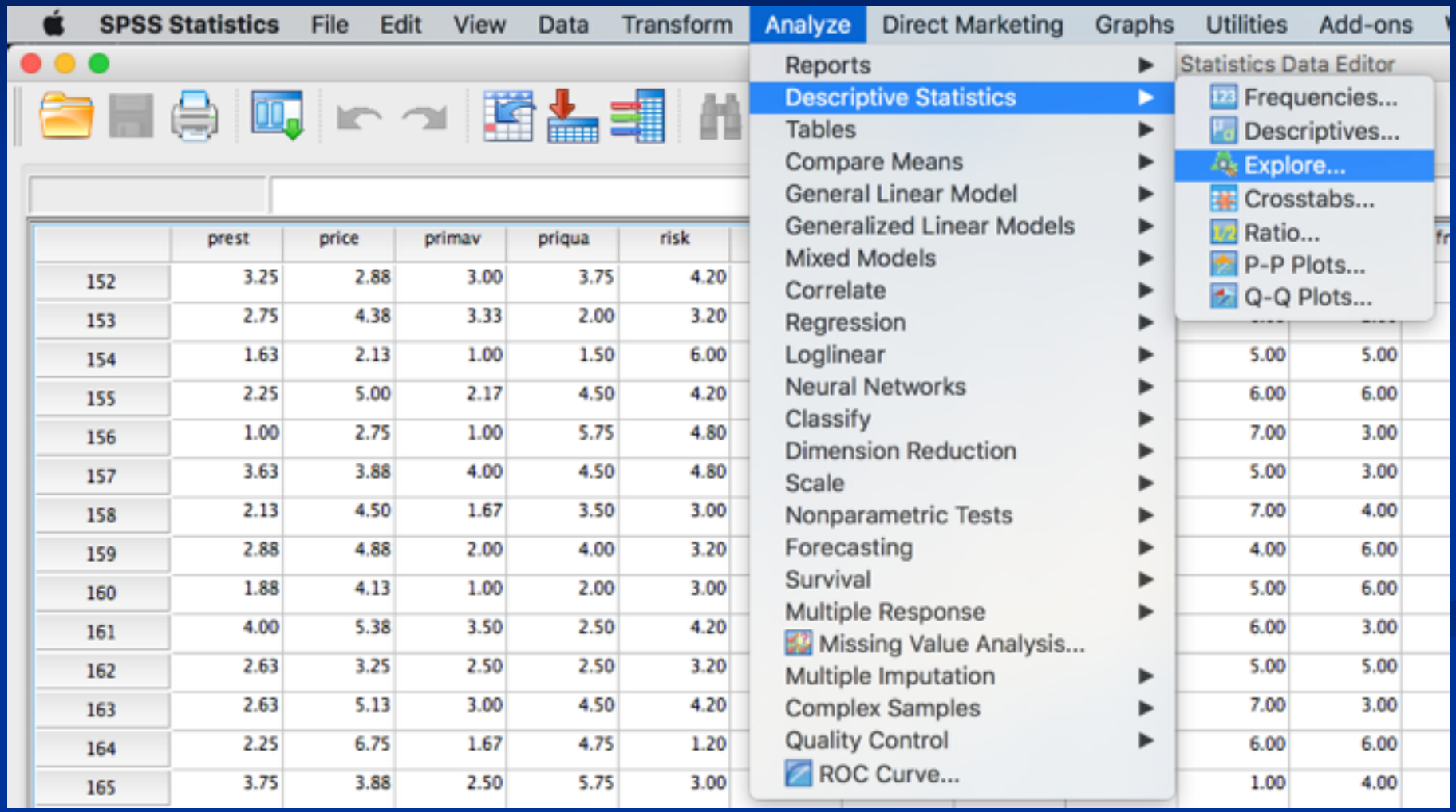
Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
education	301	2.00	1.00	3.00	2.2193	.64686	.418	-.245	.140	-.691	.280
Valid N (listwise)	301										

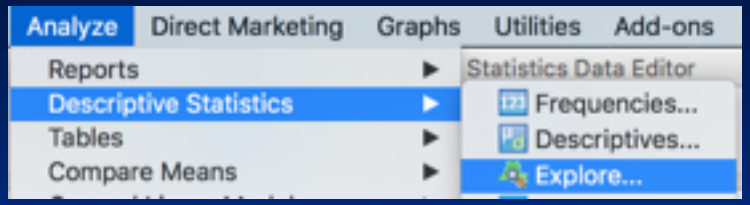
- الأمر إستكشاف (Explore)

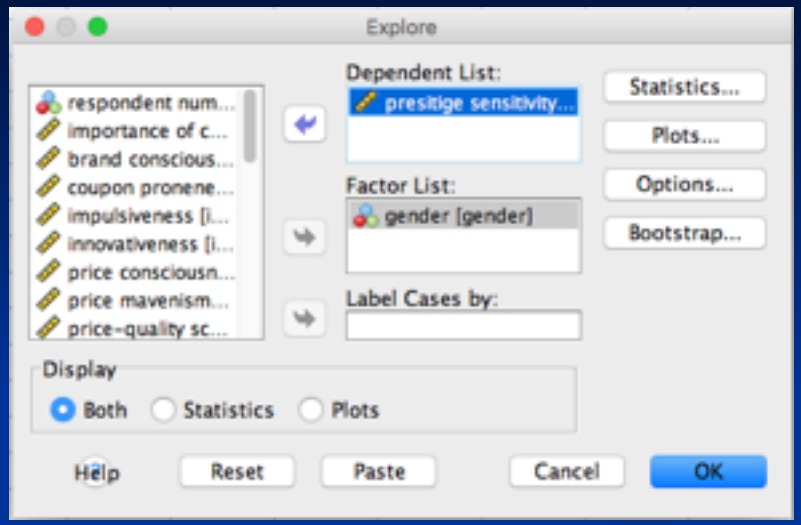
يتم الوصول إلى الأمر (Explore) من خلال المسار التالي:

Analyse/Descriptive Statistics/Explore



3. الأمر: (Explore) عند اختيار هذا الأمر يظهر مربع الحوار التالي:





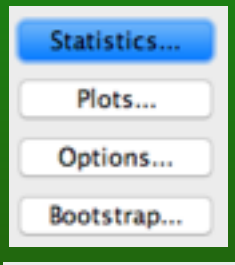
ثم ندخل المتغير المراد تحليله إلى(Dependent List)

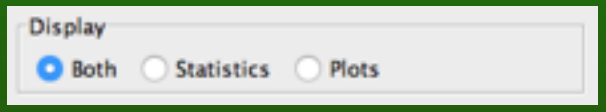
فمتغير تقسيم عرض المعلومات في (Factor List)

عبر الضغط دوما على السهم التالي:

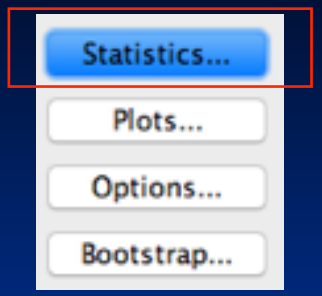
مثلا في هذه الحالة تم إدخال المتغير (prestige) والمتغير (gender)

بعد إدخال المتغير يتم انتقاء المقاييس الملائمة من خلال هذه القائمة ، وطريقة العرض الملائمة، ويم ذلك بملء المطلوب حسب كل مربع حوار لكل من الإيقونات التالي:

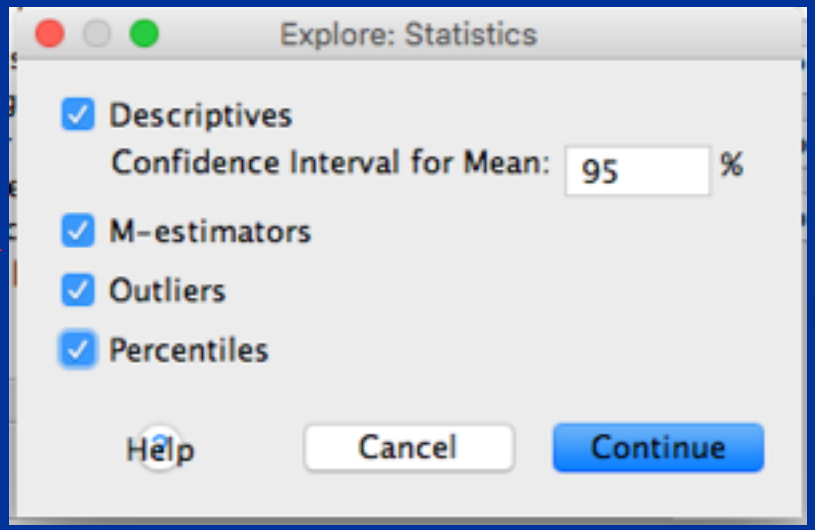




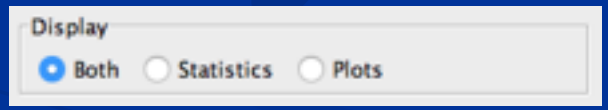
3. الأمر: (Explore):



من خلال الضغط على إيقونة (Statistics) يمكنك تحديد مجال الثقة المناسب للدراسة (1)، كما يمكنك تحديد متوسط العينة التقديري، القيم الشادة والمئينات.. (2)، ويتم ذلك بتأشير الباحث في الخانات الملائمة كما يظهر في مربع الحوار أسفله:



نترك التعيين في مكانه لعرض النتائج والرسوم البيانية معا



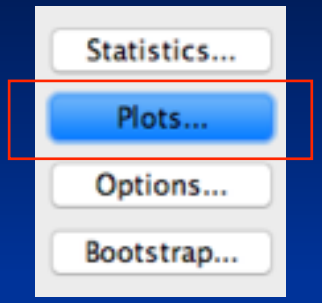
ثم نضغط على

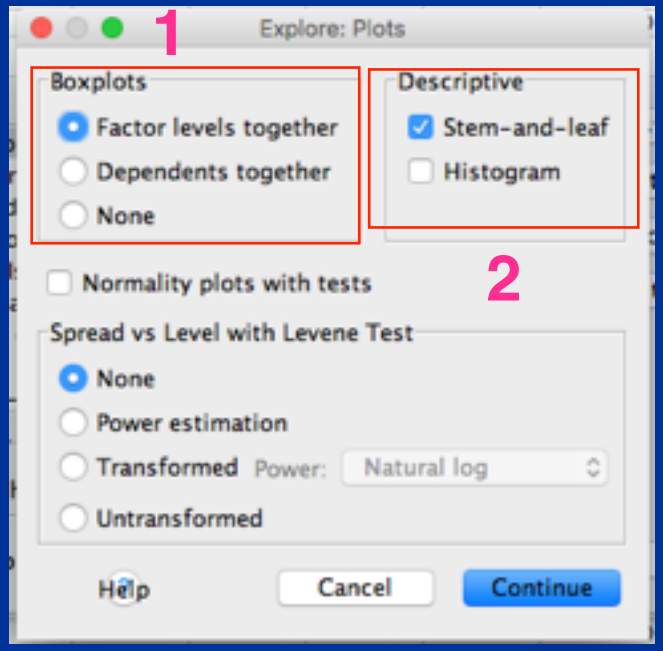
(Continue)

يتبع في الصفحة الموالية

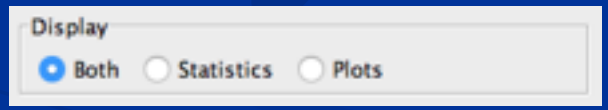
3. الأمر: (Explore):

من خلال الضغط على إيقونة (Plots) يمكنك تحديد عرض النتائج الاحصائية للمتغيرات منفصلة أو للتراكيب مجتمعة (1)، كما يمكنك عرض البيانات من خلال تكرارات الريشة أو الوريقات(2)، ويتم ذلك بتأشير الباحث في الخانات الملائمة كما يظهر في مربع الحوار أسفله:





نترك التعيين في مكانه لعرض النتائج والرسوم البيانية معا



ثم نضغط على (Continue)

يتبع

3. الأمر: (Explore):

بالتطبيق على المثال السابق للملف (Seniors.sav) والمتغيرين ((prestige) والمتغير (gender)) نتحصل على النتائج التالي:

Explore

[DataSet1] /Users/arbah/Desktop/desk work univ/data base analysis/seniors.sav

gender

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	gender	N	Percent	N	Percent	N	Percent
presitige sensitivity	male	109	100.0%	0	0.0%	109	100.0%
	female	201	100.0%	0	0.0%	201	100.0%

يتبع

3. الأمر: (Explore):

بالتطبيق على المثال السابق للملف (Seniors.sav) والمتغيرين ((prestige) والمتغير (gender)) نتحصل على النتائج التالي:

المتوسط الحسابي(mean)، وموقع المتوسط الحسابي بين الحدين الأدنى والأعلى حسب مجال الثقة المقدم، الوسيط(median)، التباين(variance)، الإنحراف المعياري(syd deviation)، القيمة الدنيا والأعلى، المدى(range)، الربيع(interquartile range)، الإلتواء (skewness)، والتحدب(kurtosis).

Descriptives

	gender			Statistic	Std. Error
presitige sensitivity	male	Mean		2.8267	.09898
		95% Confidence Interval for Mean	Lower Bound	2.6305	
			Upper Bound	3.0229	
		5% Trimmed Mean		2.8048	
		Median		2.7500	
		Variance		1.068	
		Std. Deviation		1.03340	
		Minimum		1.00	
		Maximum		5.63	
		Range		4.63	
		Interquartile Range		1.57	
		Skewness		.244	.231
		Kurtosis		-.478	.459
	female	Mean		2.6889	.07993
		95% Confidence Interval for Mean	Lower Bound	2.5312	
			Upper Bound	2.8465	
		5% Trimmed Mean		2.6322	
		Median		2.6300	
		Variance		1.284	
		Std. Deviation		1.13325	
		Minimum		1.00	
		Maximum		6.88	
		Range		5.88	
		Interquartile Range		1.57	
		Skewness		.609	.172
		Kurtosis		.263	.341

يتبع

3. الأمر: (Explore): بالتطبيق على المثال السابق للملف (Seniors.sav) والمتغيرين (prestige) والمتغير (gender)) نتحصل على النتائج التالي:

M-Estimators

	gender	Huber's M-Estimator[a]	Tukey's Biweight[b]	Hampel's M-Estimator[c]	Andrews' Wave[d]
presitige sensitivity	male	2.7955	2.7765	2.7905	2.7764
	female	2.6076	2.5593	2.6079	2.5582

a. The weighting constant is 1.339.

b. The weighting constant is 4.685.

c. The weighting constants are 1.700, 3.400, and 8.500

d. The weighting constant is 1.340*pi.

M-estimators. Robust alternatives to the sample mean and median for estimating the location. The estimators calculated differ in the weights they apply to cases. Huber's M-estimator, Andrews' wave estimator, Hampel's redescending M-estimator, and Tukey's biweight estimator are displayed.

يتبع

3. الأمر: (Explore):

بالتطبيق على المثال السابق للملف (Seniors.sav) والمتغيرين ((prestige) والمتغير (gender)) نتحصل على النتائج التالي:

يقدم لنا الجدول القيم الشادة الخمسة الأعلى والأدنى، مع مواقعها الترتيبية والقيم التي تحملها.

Extreme Values

	gender			Case Number	Value
presitige sensitivity	male	Highest	1	21	5.63
			2	209	5.13
			3	43	4.75
			4	191	4.75
			5	196	4.75[a]
		Lowest	1	252	1.00
			2	250	1.00
			3	185	1.00
			4	149	1.00
			5	235	1.25[b]
	female	Highest	1	298	6.88
			2	48	5.75
			3	258	5.38
			4	293	5.38
			5	7	5.25[c]
		Lowest	1	307	1.00
			2	281	1.00
			3	253	1.00
			4	232	1.00
			5	205	1.00[d]

a. Only a partial list of cases with the value 4.75 are shown in the table of upper extremes.

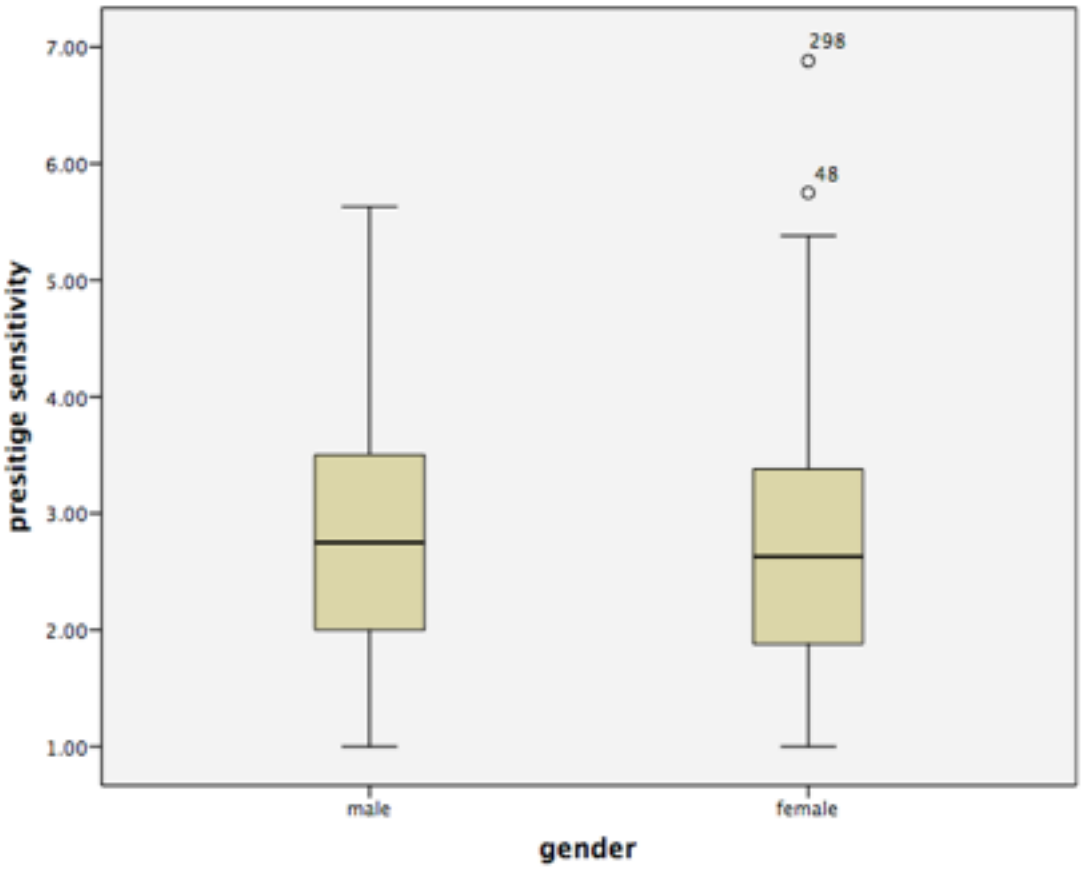
b. Only a partial list of cases with the value 1.25 are shown in the table of lower extremes.

c. Only a partial list of cases with the value 5.25 are shown in the table of upper extremes.

d. Only a partial list of cases with the value 1.00 are shown in the table of lower extremes.

يتبع

3. الأمر: (Explore):



بالتطبيق على المثال السابق للملف (Seniors.sav) والمتغيرين ((prestige) والمتغير (gender))

نتحصل على النتائج التالي:

يعطينا رسم الصندوق معلومات حول المدى، موقع الوسيط في الصندوق والنقاط الشادة ومواقعها على طرفي ديل الصندوق، كما يعطينا مواقعها الترتيبية في قاعدة البيانات، بالإضافة للذلك يحدد موقع الخط على الصندوق إتجاه الإلتواء في توزيع البيانات، كما سيشرح لاحقا.

يتبع

3. الأمر: (Explore): بالتطبيق على المثال السابق للملف (Seniors.sav) والمتغيرين ((prestige) والمتغير (gender)) نتحصل على النتائج التالي:

Percentiles

			Percentiles						
		gender	5	10	25	50	75	90	95
Weighted Average (Definition 1)	presitige sensitivity	male	1.2500	1.3800	2.0000	2.7500	3.5650	4.2500	4.7500
		female	1.0000	1.2500	1.8150	2.6300	3.3800	4.1040	4.9630
Tukey's Hinges	presitige sensitivity	male			2.0000	2.7500	3.5000		
		female			1.8800	2.6300	3.3800		

يعطينا الجدول التالي معلومات حول تقسيم سلسلة البيانات المجموعة إلى جزئياتها المختلفة (المئينات) من خلال تقديم قيم المواقع التالية في السلسلة: (5 بالمئة الأولى)، فـ (10 بالمئة)، فالربيع (25 بالمئة)، فالوسيط (50 بالمئة)، (75 بالمئة)، (90 بالمئة)، و(95 بالمئة).

يتبع

3. الأمر: (Explore):

لقد قمنا بعرض ساق وأوراق الذكور فقط لأجل الشرح، أولا هذه طريقة تعطينا مؤشر أولى حول شكل التوزيع، ومدى مطابقة شكله بالتوزيع الطبيعي الجرسي، وتتألف من الساق الذي يمثل قيم المقياس المستخدم، والأوراق التي تمثل الحالات، ويمثل ما هو أكبر من أو يساوي (6.9=<) القيم الشادة في البيانات، هذا النوع من العرض يصلح في العينات الصغيرة.

بالتطبيق على المثال السابق للملف (Seniors.sav) والمتغير ((prestige)) لنتحصل على النتائج التالي:

```
presitige sensitivity Stem-and-Leaf Plot

 Frequency     Stem &  Leaf

    43.00        1 .  00000000000000000111111222222233333333333333
    39.00        1 .  55555556666666777777777777777888888888888
    50.00        2 .  00000000001111111111111222222222222223333333333
3
    48.00        2 .  55555555555666666666666666667777777777778888888
    50.00        3 .  00000000000001111111111112222222222233333333333
3
    38.00        3 .  55555555555666666677777777777778888888
    19.00        4 .  0000000000012222333
    11.00        4 .  55556667777
     9.00        5 .  000112233
     2.00        5 .  67
     1.00 Extremes    (>=6.9)

 Stem width:      1.00
 Each leaf:       1 case(s)
```

• الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

لأجل إختبار التوزيع الطبيعي للبيانات، لدينا العديد من الوسائل يمكن أن نلخصها في الشكل البياني التالي:

إختبار التوزيع الطبيعي للبيانات
ينظر في الجانبين معا ولا تغني إحداهما على الأخرى على العموم

بالنظر لشكل التوزيع

- histograms
- box plots
- normal Q-Q plot
- detrended normal Q-Q plot

بالنظر للإختبار الإحصائي

- Kolmogorov-Smirnov (K-S) and Shapiro-Wilk

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

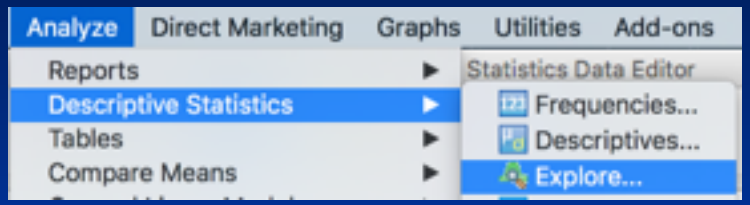
بعد تعرفنا على الإتجاهات الواجب سلوكها في تقييم مدى إتباع البيانات للتوزيع الطبيعي، نقدم المثال التالي لأجل تطبيق ما سبق تعريفه نظريا:

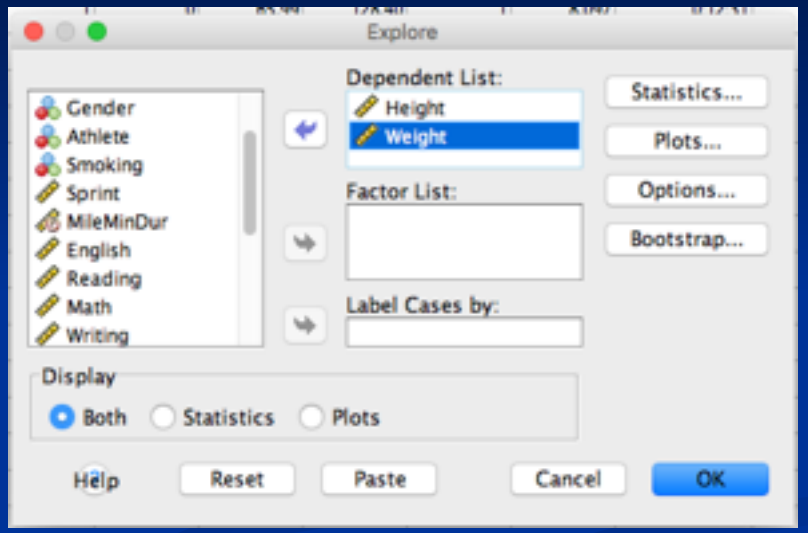
مثال: في العموم تخضع أطوال البشر إلى التوزيع الطبيعي، بينما تميل أوزانهم إلى الإلتواء، سنحاول من خلال بيانات الملف (Sample_Dataset_2014.sav) التأكد من اتباع هذه العينة لمنحنى التوزيع الطبيعي، وذلك من خلال القيام بالاختبار الإحصائي والرسومات البيانية لمتغيري الطول (heights) والوزن (weights) كما سبق وقدمنا في الأعلى.

سنستخدم الأمر (Explore) للقيام بهذا الإختبار الإحصائي.

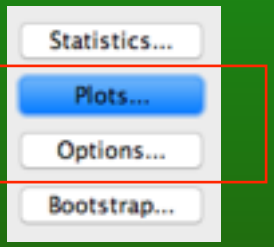
- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

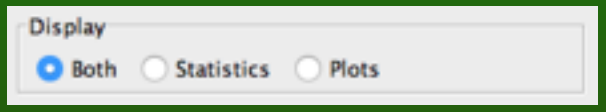
1. الأمر: (Explore) عند اختيار هذا الأمر يظهر مربع الحوار التالي:





في هذه الحالة نترك (statistics) كما هي ونتعامل فقط مع (Plots) و (Options) لما لهم من علاقة مع رسوم التخطيط الطبيعي





ثم ندخل المتغير المراد تحليله إلى(Dependent List)

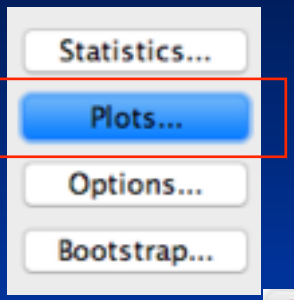
فمتغير تقسيم عرض المعلومات في (Factor List)

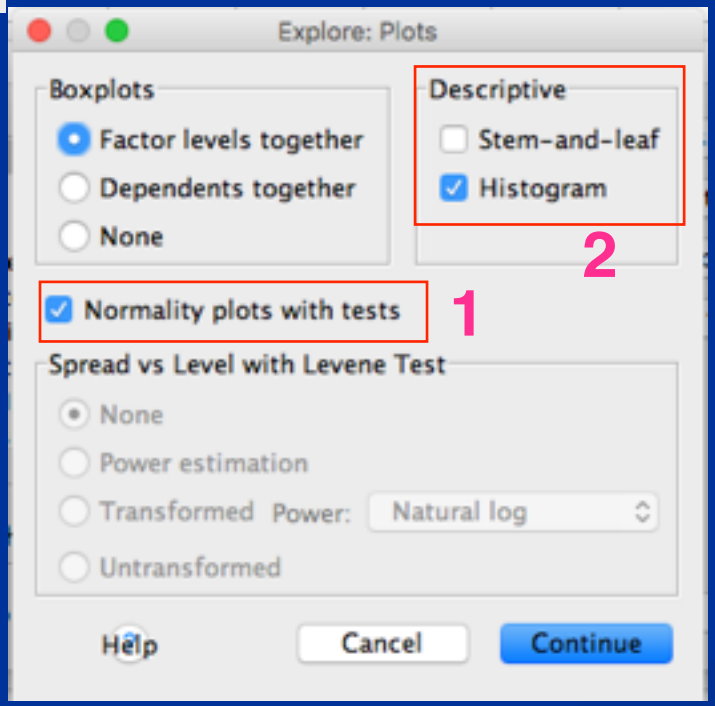
عبر الضغط دوما على السهم التالي:

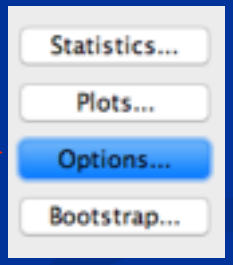
مثلا في هذه الحالة تم إدخال المتغير (heights) والمتغير (weights)

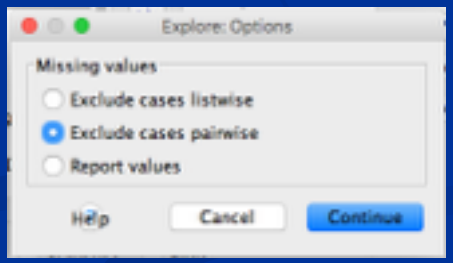
- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

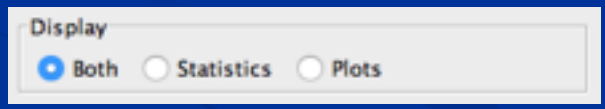
من خلال الضغط على إيقونة (Plots) نقوم بتحديد إختبار طبيعية البيانات من المربع (1)، كما نحدد المدرج التكراري في المربع (2)، ويتم ذلك بتأشير الباحث في الخانات الملائمة كما يظهر في مربع الحوار أسفله:





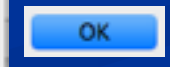






نترك التعيين في مكانه لعرض النتائج والرسوم البيانية معا

يتبع



- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

بالتطبيق نتحصل على النتائج التالي:

مختصر بيانات العينة (للمتغيرات المدروسة):

Explore

[DataSet1] /Users/arbah/Desktop/desk work univ/data base analysis/Sample_Dataset_2014.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Height	408	93.8%	27	6.2%	435	100.0%
Weight	376	86.4%	59	13.6%	435	100.0%

يقدم لنا الجدول معلومات حول عدد مفردات عينة الإختبار، وكذا البيانات الناقصة في كل من متغير (heights) ومتغير (weights) حيث تتمثل هذه المعلومات في النسب المئوية وعدد مفردات العينة.

يتبع

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

Descriptives

			Statistic	Std. Error
Height	Mean		68.0318	.26366
	95% Confidence Interval for Mean	Lower Bound	67.5135	
		Upper Bound	68.5501	
	5% Trimmed Mean		67.9687	
	Median		67.5700	
	Variance		28.363	
	Std. Deviation		5.32566	
	Minimum		55.00	
	Maximum		84.41	
	Range		29.41	
	Interquartile Range		6.78	
	Skewness		.230	.121
	Kurtosis		.113	.241
Weight	Mean		181.0316	2.20465
	95% Confidence Interval for Mean	Lower Bound	176.6966	
		Upper Bound	185.3666	
	5% Trimmed Mean		178.4763	
	Median		172.9600	
	Variance		1827.535	
	Std. Deviation		42.74968	
	Minimum		101.71	
	Maximum		350.07	
	Range		248.36	
	Interquartile Range		50.62	
	Skewness		1.005	.126
	Kurtosis		1.502	.251

النتائج الوصفية لتحليل البيانات:

في هذا الجدول نجد معلومات حول كل من المتوسط الحسابي، الحدود الدنيا والعليا لموقع المتوسط الحسابي حسب مجال الثقة المختار، في حالتنا هذه (95 بالمئة)، كما نجد أيضا معلومات حول التباين، الانحراف المعياري، المدى، الربيعيات وبيانات الإلتواء والتحدب

ملاحظة: القيم المعيارية لما يكون التوزيع طبيعي بالنسبة لقيم إحصائيات الإلتواء والتحدب هي (0). في حالتنا هذه قيمة الالتواء لمتغير الطول تساوي (0.23) وقيمة التحدب تساوي (0.113) وهي قيم تبتعد قليلا عن القيمة المعيارية، وعلى العكس من ذلك في متغير الوزن أين تبتعد كثيرا عن الصفر، هذه مؤشرات ابتدائية فقط .

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

نتائج إختبار (Kolmogorov-Smirnov):

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Height	.045	408	.049	.993	408	.070
Weight	.084	376	.000	.944	376	.000

a. Lilliefors Significance Correction

ملاحظة: من أجل إختبار مدى طبيعية توزيع بيانات المتغيرات المعروضة على اختبار (K-S)، لابد أولا من إختيار التوزيع الطبيعي لأن هذا الاختبار صالح للعديد من أنواع التوزيعات كما سنبين فيما بعد، كما يجب طرح فرضية التوزيع بالطريقة الملائمة، حيث دوما ننطلق من أن البيانات تتوزع طبيعيا أي أن فرضية العدم تقترح أن البيانات تتوزع طبيعيا، على عكس باقي أنواع الفرضيات المنفية الأخرى.

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

إختبار (Kolmogorov–Smirnov):

يعد أختبار (K–S) من الإختبارات اللامعلمية، تقنيا يمكنه اختبار مرجعية البيانات إلى العديد من التوزيعات المعروفة كتوزيع (Student, Pareto,..)، وليس فقط التوزيع الطبيعي الجرسي. إن الأمر (Explore) يتعامل مباشرة مع التوزيع الطبيعي، ولذا فمن غير المطلوب منك تحديد التوزيع المراد إختباره.

يدمج مع أختبار (K–S)، إختبار (Shapiro–Wilk) وهو إختبار معلمي، يعتمد نفس طريقة إختبار (K–S) من حيث طرح فرضية الاختبار على أن التوزيع هنا المراد إختباره يكون طبيعيا، وليكن في المعلوم أن إختبار (S–W) جد حساس لأبسط التغيرات بخصوص توزيع البيانات طبيعيا خصوصا في العينات ذات الأحجام الكبيرة. في كثير من الأحيان تتعارض نتائج الاختبارين وهما غير كافيتان للحكم على نوع التوزيع دون الإستعانة بالنظر إلى الأشكال البيانية المكملة، فعلى الباحث التعامل مع الاثنين بقدم المساوات.

طرح الفرضيات:

الفرضية الصفرية (H0): بيانات العينة مسحوبة من توزيع طبيعي

الفرضية البديلة (H1): بيانات العينة مسحوبة من توزيع غير طبيعي

يتبع

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

الحكم على الفرضية: إن الحكم على الفرضية يتم من خلال قيمة (P-Value) التي يصدرها برنامج (SPSS) في جدول مخرجات الأمر (Explore):

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Height	.045	408	.049	.993	408	.070
Weight	.084	376	.000	.944	376	.000

a. Lilliefors Significance Correction

- فإذا كانت قيمة (P-Value) أقل من مستوى المعنوية غالبا (0.05) عند مجال ثقة بـ (95 بالمئة)، حينها نرفض (H0)، ونقبل الفرضية البديلة (H1)، وبالتالي فإن التوزيع الذي تم سحب بيانات العينة منه غير طبيعي.
- أما إذا كانت قيمة (P-Value) أكبر من مستوى المعنوية غالبا (0.05) عند مجال ثقة بـ (95 بالمئة)، وبالتالي لا نرفض فرضية العدم (H0)، ونقول حينها أنه لا توجد دلائل بالشكل الكافي للتوصل إلى إستنتاج أن البيانات غير طبيعية.

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

التعليق على نتائج الإختبار :

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Height	.045	408	.049	.993	408	.070
Weight	.084	376	.000	.944	376	.000

a. Lilliefors Significance Correction

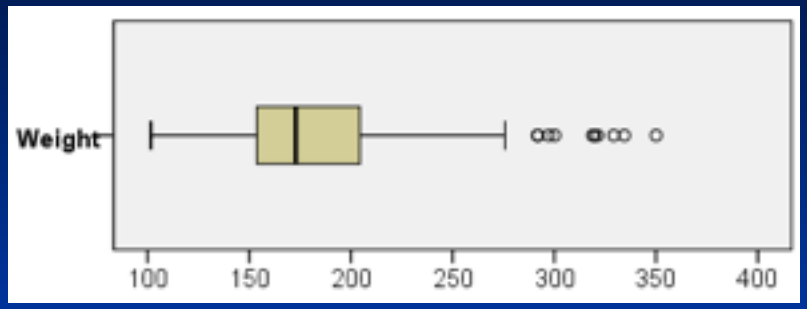
من خلال نتائج الجدول نلاحظ أن قيم (K-S and Shapiro-Wilk test p-values) بالنسبة للوزن هي أقل من (p < 0.001)، ومنه فرفض فرضية العدم أمر واضح، أما بالنسبة لمتغير الطول، فتبين قيمة (p-values) لاختبار (K-S) عن قيمة (0.049) وهي تكاد تقترب من مستوى المعنوية (0.05)، وعلى العكس من نتيجة اختبار (K-S)، يظهر إختبار (Shapiro-Wilk) عن قيمة (0.070) وهي قيمة أكبر من مستوى المعنوية (0.05)، يتوصل الأختباران إلى نتائج متضادة، أين يبين إختبار (K-S) أن البيانات غير طبيعية، بينما يرى إختبار (Shapiro-Wilk) أن البيانات طبيعية.

فكيف نتعامل مع هذا التناقض؟ هنا وجب النظر في مخرجات الرسوم البيانية الأخرى.

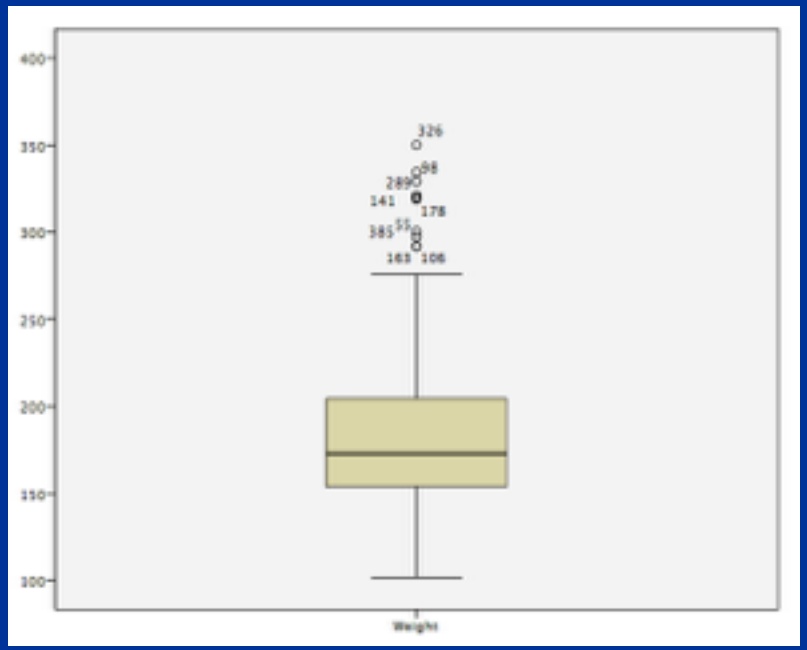
يتبع

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

أولا: النظر إلى رسمة الصندوق لمتغير الوزن



من رسمة الصندوق لمتغير الوزن في الشكل المقابل نفهم لما كانت نتائج اختبار طبيعة البيانات سلبية، فالملاحظ يرى أن توزيع البيانات هنا ملتوي نحو اليمين، ويظهر ذلك من خلال ما يلي:



طول جناح الصندوق الأيمن على الأيسر

1. الوسيط يميل إلى الجانب الأيسر منه إلى الأيمن
2. النقاط الشادة في الجهة العليا من نهاية التوزيع

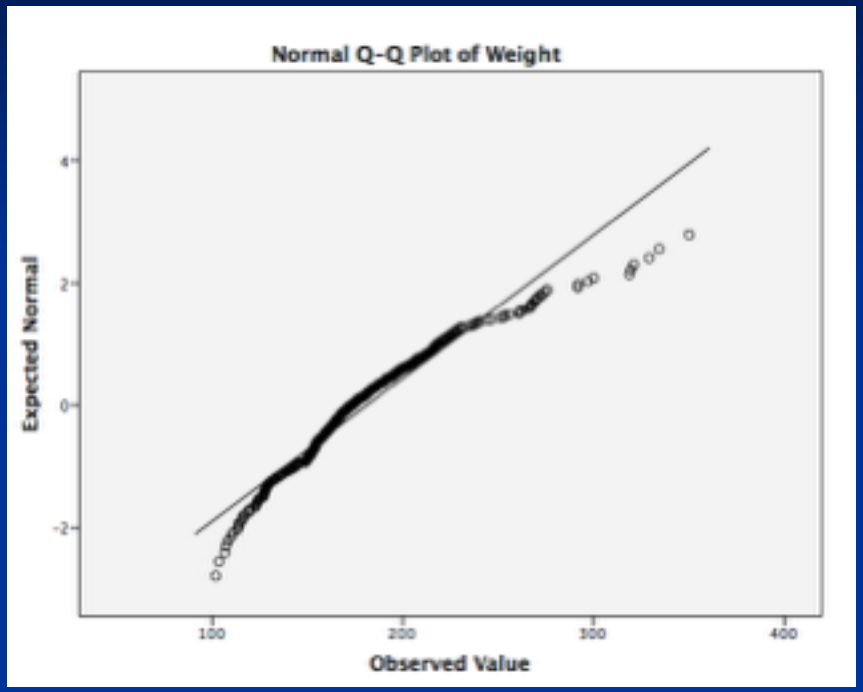
بعدها يمكننا النظر في كل من الرسم البياني لـ:

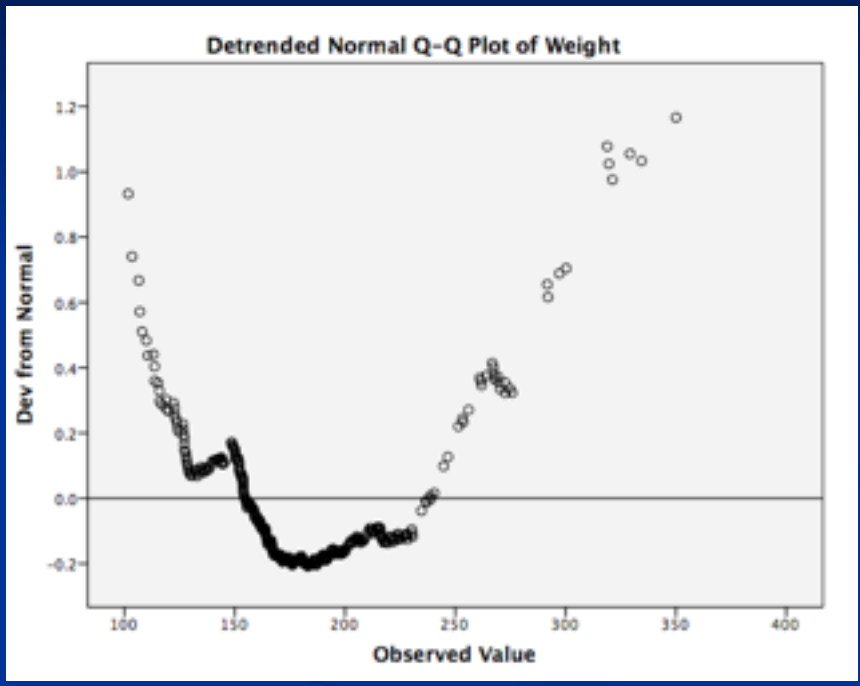
- رسم خط انتشار (Normal Q-Q plots)

ورسم خط انتشار (Detrended Q-Q plots)

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

- رسم خط انتشار (Q–Q and detrended Q–Q plots) لمتغير الوزن

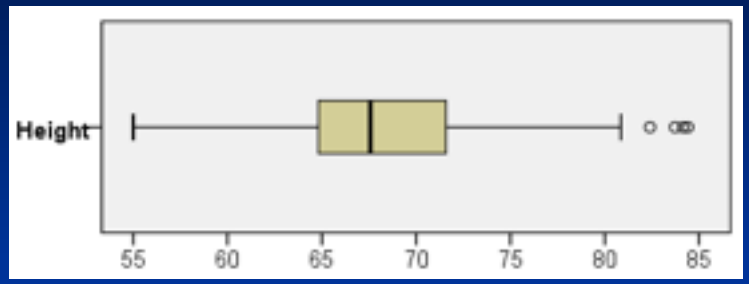


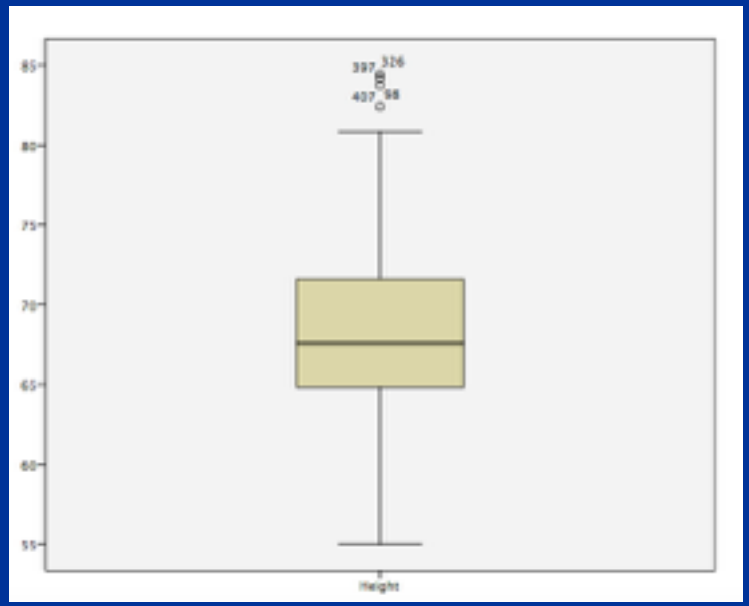


من الشكلين في الأعلى لكل من (Q–Q and detrended Q–Q plots) نلاحظ إنحرافات منتظمة عن خط التوزيع الطبيعي، ويمكن ملاحظة أن شكل توزيع بيانات (detrended Q–Q) مقعر (برابولي)، حيث تتغير إنحرافات التوزيع على محور العينات (y–axis) بمقدار يتراوح بين (0.2– حتى 1.2).

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

ثانيا: لننظر إلى رسمة الصندوق الطول



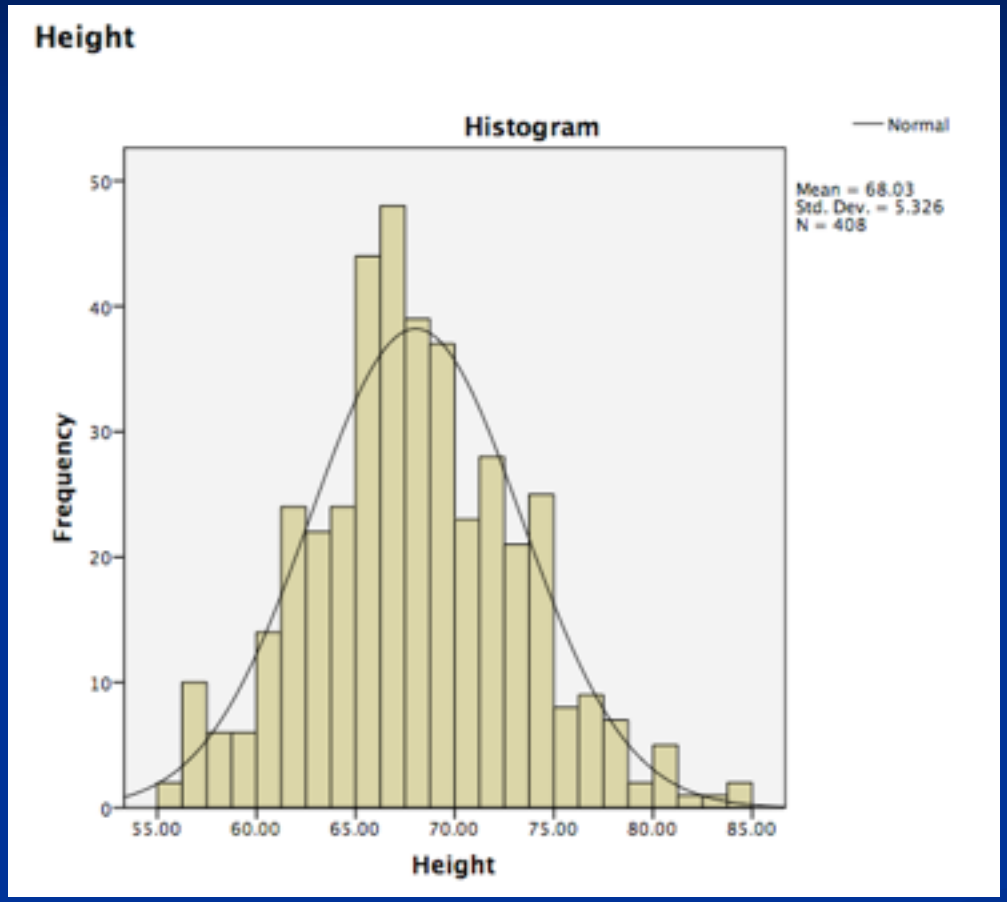


من رسمة الصندوق لمتغير الطول في الشكل المقابل نلاحظ أن توزيع البيانات هنا ملتوي نحو اليمين بشيء يسير، ويظهر ذلك من إقتراب الوسيط من متصف الصندوق، والنتيجه ليست مثلها في متغير الوزن، فالتوزيع في حالة الطول يظهر أقرب لأن يكون منتظم نحو المركز منه في حالة متغير الوزن.

ولأجل التاكد أكثر لابد من النظر في باقي الرسوم البيانية الأخرى ومنها (Histogram) وأيضا بالطبع مخرجات الأمر (Explore) المتمثلة في كل من (Normal Q-Q plots) و(detrended Q-Q plots).

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

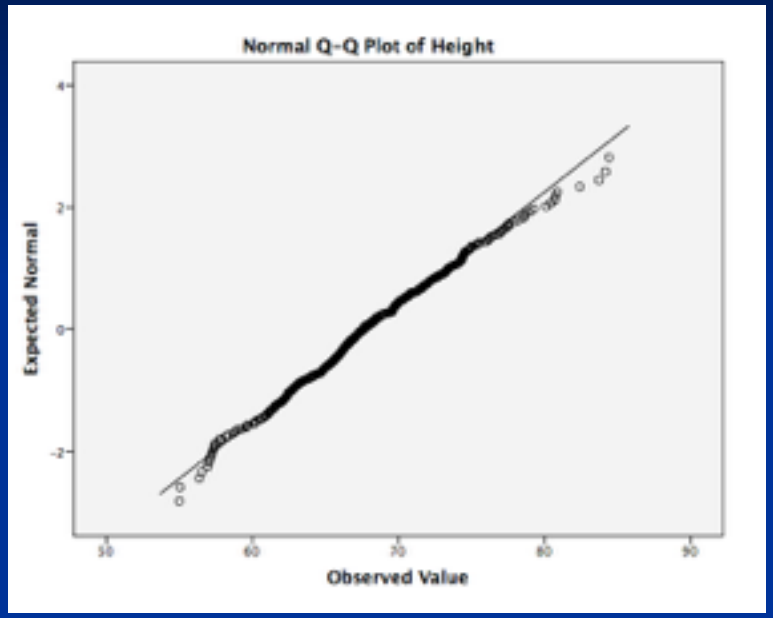
- المدرج التكراري (Histogram) لمتغير الطول

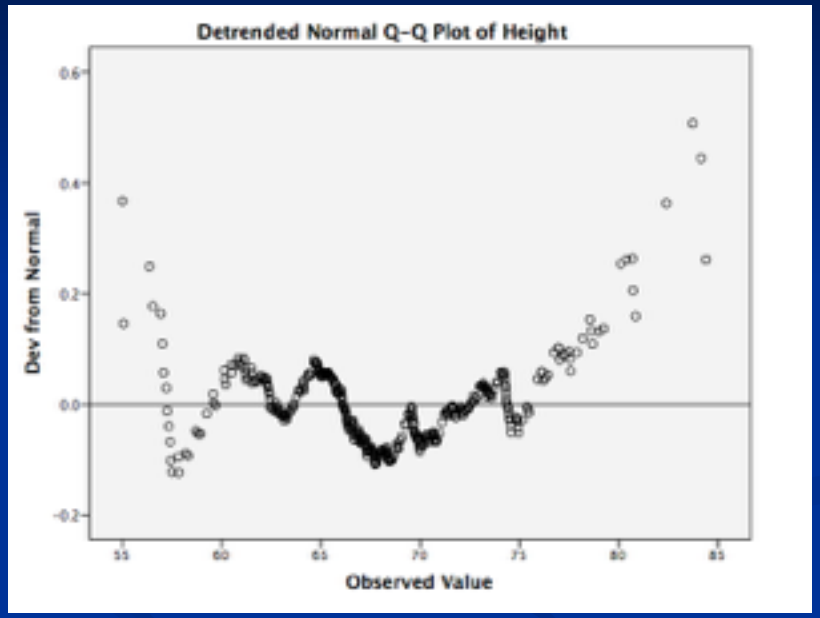


إن الهدف من النظر إلى المدرج التكراري هو معرفة مدى تماثل الشكل الجرسي، وليس بالضرورة التاكد من مصداقية التوزيع، لكنه إشارة فقط شكلية لمدى تطابق المنحنى البياني مع مخرجات البيانات والشكل الجرسي المعروف للتوزيع الطبيعي، كما يمكن معرفة مدى تمركز مؤشرات النزعة المركزية حول وسط المنحنى الجرسي والمدرج التكراري على حد السواء، فكلما كان الشكل معتدل ومتطابق على الجانبين وتوسطه احصائيات النزعة المركزية كلما كان ميل الباحث للحكم على أن البيانات تخضع للتوزيع الطبيعي.

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

- رسم خط انتشار (Q-Q and detrended Q-Q plots) لمتغير الطول





من الشكلين في الأعلى لكل من (normal Q-Q plot) نلاحظ أن توزيع النقاط ينتشر بالتطابق على خط التوزيع الطبيعي، والانحرافات طفيفة ولا تظهر إلا على الديلين فقط. أما في شكل توزيع بيانات (detrended Q-Q) فيمكن ملاحظة الإنحرافات عن الخط الطبيعي بوضوح إلا أنها ليست كبيرة مثل ما تم مشاهدته في حالة متغير الوزن، وقد كانت بين (-0.2 و0.6) وهي قيم معقولة خصوصا أنها قيم شادة.

- الأوامر المختلفة لاختبار مدى خضوع البيانات للتوزيع الطبيعي (Normality Test)

نتائج وتوصيات:

من خلال الإلتواء في صندوق العرض، والنحرافات المنتظمة في (Q-Q plots)، ونتائج الإختبارين أين تم تسجيل قيمة أقل بكثير عن مستوى المعنوية ($p < 0.001$)، تبين العديد من الدلائل على أن بيانات متغير الوزن لا تخضع للتوزيع الطبيعي.

بالنسبة لمتغير الطول، تظهر الدلائل ضعيفة، حيث أن أحد الإختبارات كان أقل (K-S test p = 0.049) بينما الآخر لا (Shapiro-Wilk p = 0.070). وبعد النظر في الرسوم البيانية، وجدنا أن هناك بعض الإنحرافات عن الخط الطبيعي، ولكن هذه الانحرافات لم تكن تظهر بالشكل الكبير، ولأجل العمل التطبيقي، سيكون من غير اللائق إعتبار توزيع بيانات متغير الطول غير طبيعية.

نـهايـة الدرس الثانـي

IBM
IBM® SPSS® Statistics
Version 20
Licensed Materials - Property of IBM Corp. © Copyright IBM Corporation and its licensors 1989, 2011. IBM, IBM logo, ibm.com, and SPSS are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at www.ibm.com/legal/copytrade.shtml. Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates. Other product and service names might be trademarks of IBM or other companies. This Program is licensed under the terms of the license agreement accompanying the Program. This license agreement may be either located in a Program directory folder or library identified as License or Non_IBM_License, if applicable, or provided as a printed license agreement. Please read the agreement carefully before using the Program. By using the Program you agree to these terms.
Java
Release 20.0.0