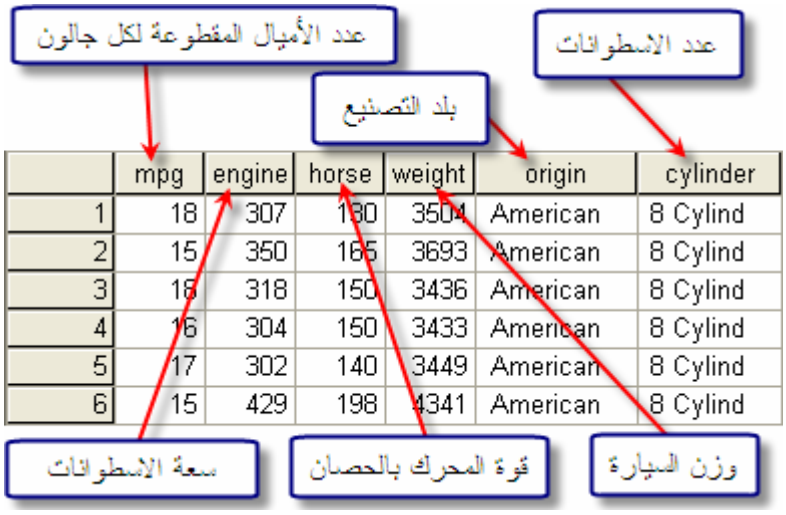
وصف المتغيرات الاسمية والكمية
Data Description

مقدمة

وصف المتغيرات الاسمية والكمية هي عملية تلخيص المتغيرات بحيث يتم الحصول على قيم عددية تعبر عن أو تصف الحالات في البيانات. ويمكن أن تكون هذه الملخصات تعداداً للحالات حسب متغير تصنيفي أو حسب عدة متغيرات تصنيفي، أو قد تكون هذه الملخصات مقياس لدالة إحصائية أو رياضية معينة مثل الوسط الحسابي أو الوسيط أو الانحراف المعياري. وسيتم في هذا الفصل التعرف على عدة طرائق لوصف المتغيرات الاسمية والكمية باستخدام SPSS، وكذلك التوسع في العرض البياني لتمثيل الملخصات بيانياً.

البيانات المستخدمة

سيتم استخدام بيانات Cars.sav لهذا الغرض، وهو أحد ملفات البيانات التي تأتي جاهزة عند تحميل البرنامج، ويحتوي هذا الملف على متغيرات تصف مجموعة من السيارات حسب المتغيرات التالية.



	mpg	engine	horse	weight	origin	cylinder
1	18	307	130	3504	American	8 Cylind
2	15	350	165	3693	American	8 Cylind
3	18	318	150	3436	American	8 Cylind
4	16	304	150	3433	American	8 Cylind
5	17	302	140	3449	American	8 Cylind
6	15	429	198	4341	American	8 Cylind

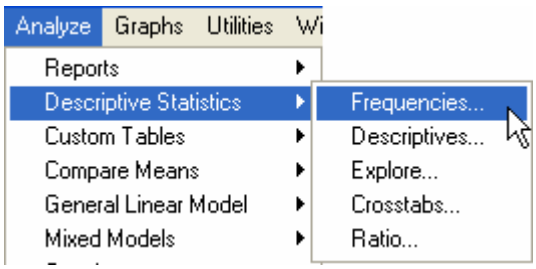
ويوجد في هذا الملف متغيرين تصنيفيين وهما Origin و Cylinder وأربع متغيرات كمية وهي mpg، engine، horse و weight.

البيانات الاسمية Qualitative Variables

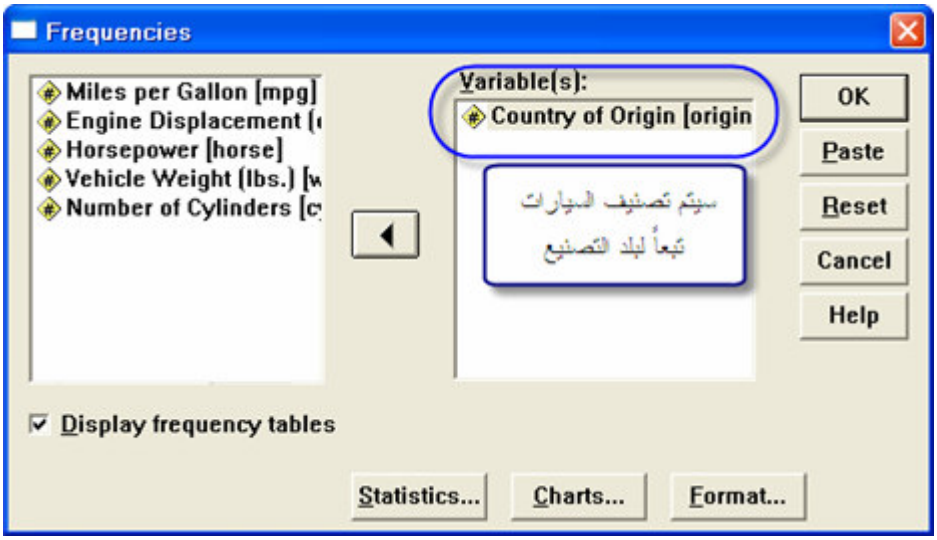
البيانات الاسمية هي جميع المتغيرات التصنيفية والتي تمثل كل قيمة من قيمها مستوى من مستويات التصنيف المختلفة، إلا أنه يجب التنبيه إلى أن قيم التصنيف ليس لها معنى بحد ذاتها وإنما هي قيم عددية تمكن الباحث من فصل الحالات تبعاً لمستويات التصنيف. فعلى سبيل المثال فإن لجنس الطالب مستويين وهما (ذكور، إناث)، وبذلك فإنه يمكن تصنيف الطلاب إلى مستويين بحيث تكون قيمة المتغير 1 عندما يكون جنس الطالب (ذكر) وتكون قيمة التغير 2 عندما يكون جنس الطالب أنثى. وبذلك فإن الأعداد 1 و 2 تمثل المستويات المختلفة للمتغير التصنيفي الجنس وليس لها مدلول كقيمة عددية. وتشمل عملية وصف المتغيرات الاسمية الحصول على تكرار لأعداد الحالات حسب التصنيف أو الحصول على جداول اقتران بحيث يتم تصنيف الحالات تبعاً لمتغيرين تصنيفيين أو أكثر.

وفيما يلي، سنقوم بتلخيص هذه البيانات وذلك عن طريق الإجابة على العديد من الاستفسارات عن السيارات.

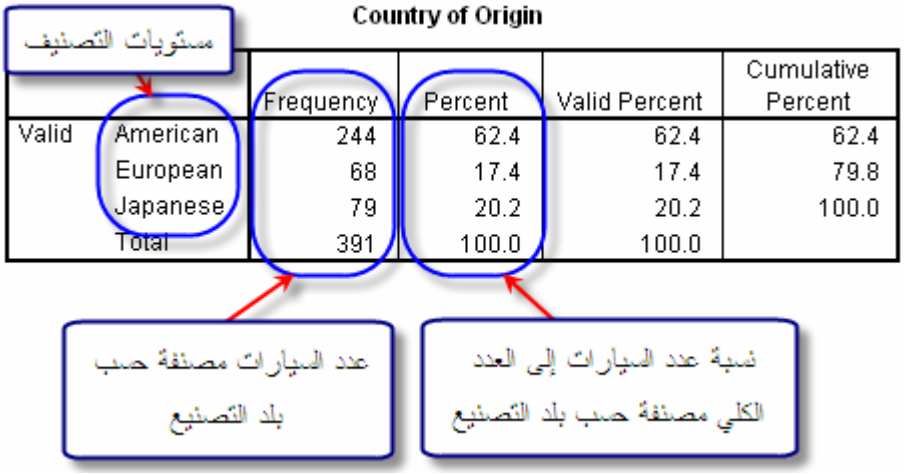
الاستفسار الأول: معرفة عدد السيارات مصنفة حسب بلد التصنيع، ويتم ذلك باختيار الأمر



ليظهر مربع الحوار التالي.

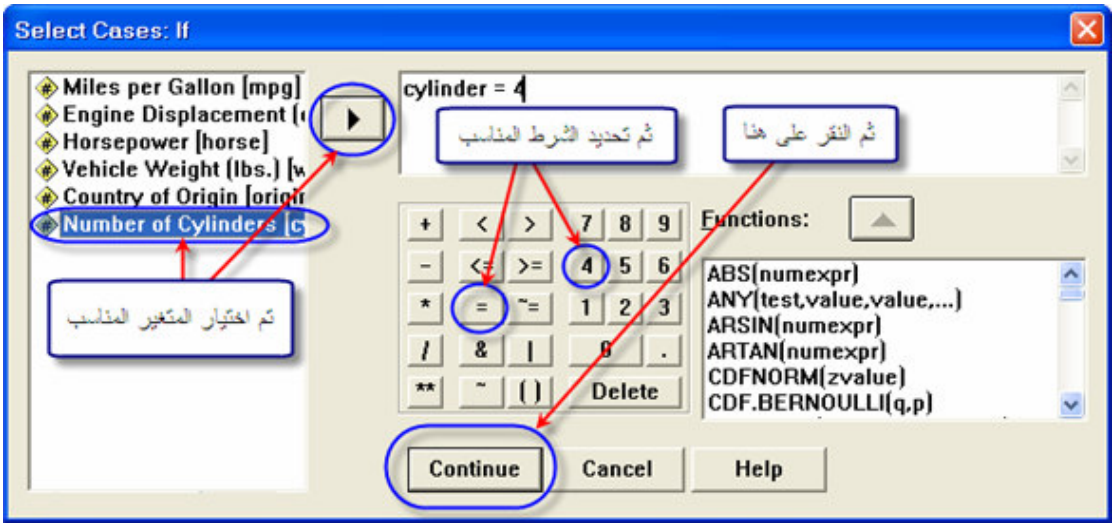


وبتحديد المتغير المناسب ثم النقر على OK، تظهر النتائج في شاشة عارض النتائج.



ويوضح الجدول أعداد السيارات ونسبة عدد السيارات إلى العدد الكلي وذلك حسب مستويات المتغير التصنيفي Origin والذي يحتوي على ثلاث مستويات. ويوفر الأمر Frequencies إمكانية تمثيل أعداد السيارات أو نسب السيارات بيانياً وذلك بالنقر على Charts ثم اختيار نوع الرسم المناسب.

الاستفسار الثاني: معرفة عدد السيارات ذات الأربع اسطوانات مصنفة حسب بلد التصنيع، ويمكن عمل ذلك بإتباع نفس الخطوات السابقة وذلك باستخدام الأمر Frequencies. ولكن قبل تنفيذ الأمر، يجب اختيار مجموعة جزئية من البيانات بحيث تكون جميع السيارات المختارة هي سيارات ذات الأربع اسطوانات. ويتم ذلك باستخدام الأمر Select Cases من قائمة Data ثم وضع الشرط المناسب وهو اختيار السيارات ذوات الأربع اسطوانات.



وبذلك تكون البيانات جاهزة للإجابة على الاستفسار الثاني حيث تم استبعاد جميع السيارات التي تكون عدد اسطواناتها تختلف عن الأربع اسطوانات، وباستخدام الأمر Frequencies يتم تصنيف السيارات حسب بلد التصنيع.

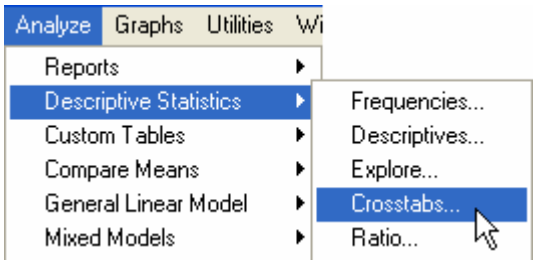
Country of Origin

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	American	69	34.7	34.7	34.7
	European	61	30.7	30.7	65.3
	Japanese	69	34.7	34.7	100.0
	Total	199	100.0	100.0	

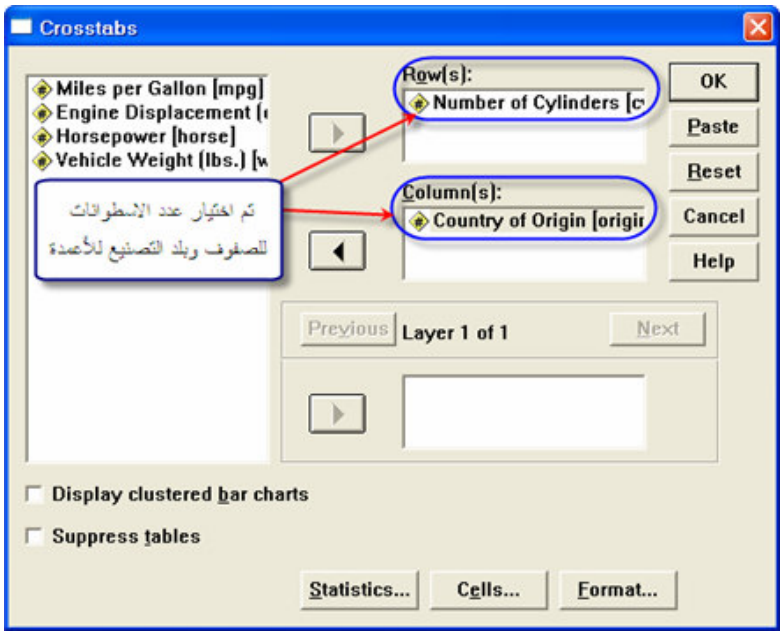
ويوضح الجدول السابق أن العدد الكلي للسيارات ذوات الأربع اسطوانات هو 199، كذلك يبين الجدول توزيع هذه السيارات حسب بلد التصنيع.

ويجب التنبيه هنا إلى أنه يجب إعادة تضمين جميع البيانات وذلك باختيار Select Cases من قائمة Data ثم النقر على Reset وذلك لإلغاء عملية اختيار مجموعة جزئية من البيانات.

الاستفسار الثالث هو معرفة أعداد السيارات مصنفة حسب بلد التصنيع وعدد الاسطوانات، أي أنه سيتم تكوين جدول اقتران بحيث يتم تصنيف السيارات تبعاً لمتغيرين تصنيفيين هما Origin و Cylinder. وباستخدام الأمر



وبذلك يظهر مربع الحوار التالي.

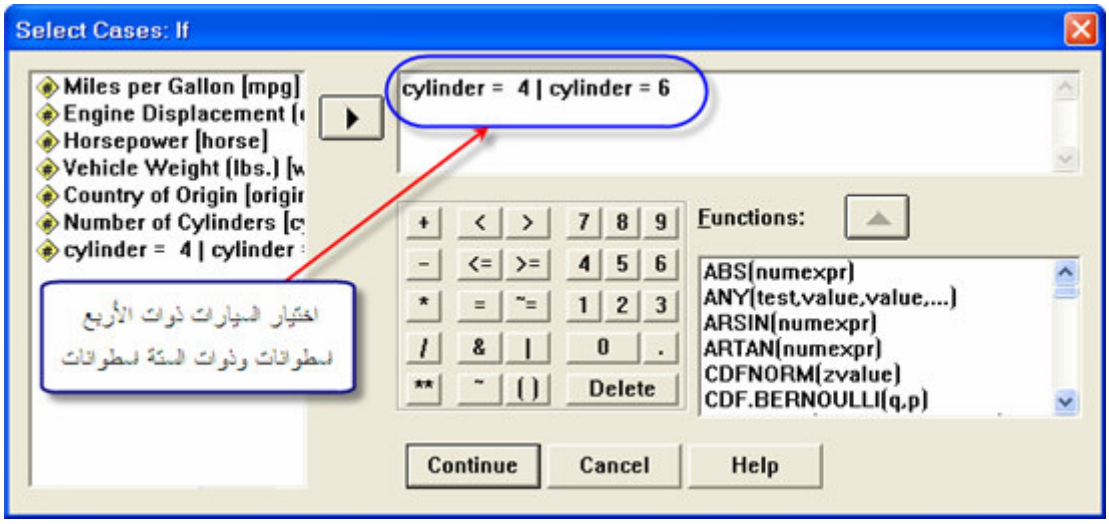


وبالنقر على OK يظهر لنا جدول الاقتران الذي يوضح عدد السيارات بعد تصنيفها.

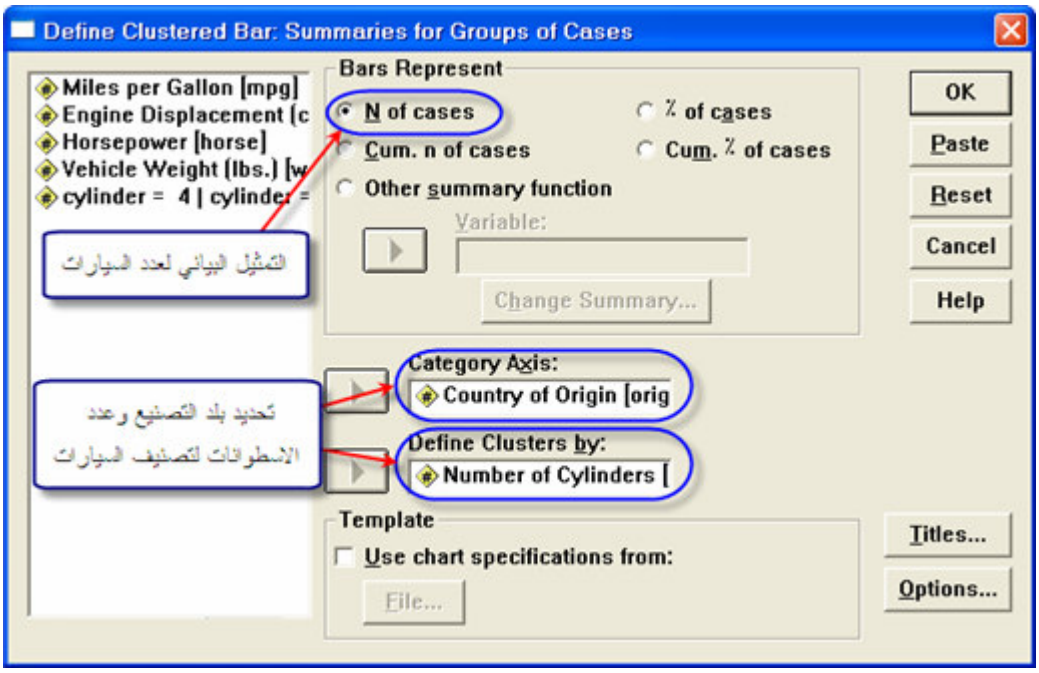
Count

		Country of Origin			
		American	European	Japanese	Total
Number of Cylinders	3 Cylinders			4	4
	4 Cylinders	69	61	69	199
	5 Cylinders		3		3
	6 Cylinders	73	4	6	83
	8 Cylinders	102			102
Total		244	68	79	391

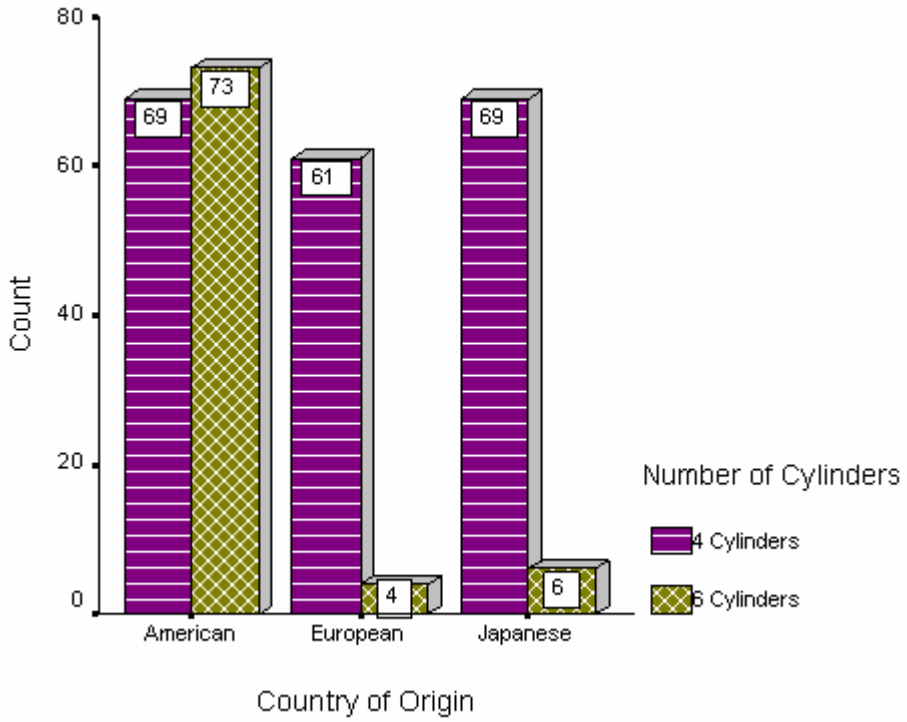
ويلاحظ أن جميع السيارات ذوات الثمانية اسطوانات هي سيارات أمريكية وأن جميع السيارات ذوات الخمس اسطوانات هي سيارات أوروبية وأن جميع السيارات ذوات الثلاث اسطوانات هي سيارات يابانية، كذلك يمكن الإجابة على الاستفسار الثاني وذلك بالنظر إلى السطر الثاني والذي يصنف السيارات ذوات الأربع اسطوانات حسب بلد التصنيع. ويمكن تمثيل أعداد السيارات ذوات الأربع والستة اسطوانات بيانيا وذلك باختيار مجموعة جزئية من البيانات باستخدام الأمر Select Cases.



وباختيار الأمر Bar من قائمة Graphs ثم اختيار Clustered و Summaries for groups of cases ثم النقر على Define، يظهر مربع الحوار التالي.



وبالنقر على OK يظهر الرسم البياني التالي على شاشة عارض النتائج.



الجداول المحورية Pivot Tables

الجداول المحورية هي جداول تفاعلية تتكون من أعمدة (Columns) وصفوف (Rows) وطبقات (Layers)، وبحيث تحتوي الخلايا المتكونة من تقاطعات الأعمدة والصفوف والطبقات على نتائج من أوامر برنامج SPSS. ومن الخصائص المهمة للجداول المحورية أنه يمكن ترتيب الصفوف والأعمدة والطبقات بالطريقة التي يراها الباحث مناسبة لعرض النتائج. ولتعديل طريقة عرض أي جدول في شاشة عارض النتائج، يتم النقر على الجدول مرتين لتفعيل محرر الجداول المحورية Pivot Tables Editor أو بالنقر على الجدول بواسطة زر الفارة الأيمن واختيار SPSS Pivot Table Object ثم اختيار Edit.

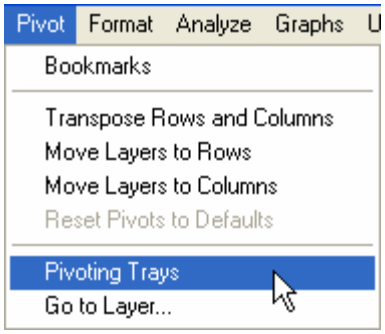
لنفرض أنه تم تصنيف السيارات حسب بلد التصنيع وعدد الاسطوانات بواسطة الأمر Crosstabs، وفي ما يلي نتائج الأمر

Country of Origin * Number of Cylinders Crosstabulation

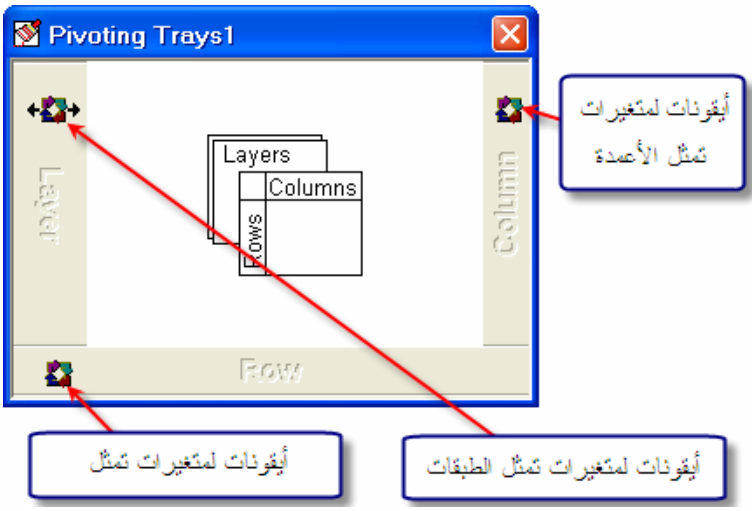
Count

		Number of Cylinders					Total
		3 Cylinders	4 Cylinders	5 Cylinders	6 Cylinders	8 Cylinders	
Country of Origin	American		69		73	102	244
	European		61	3	4		68
	Japanese	4	69		6		79
Total		4	199	3	83	102	391

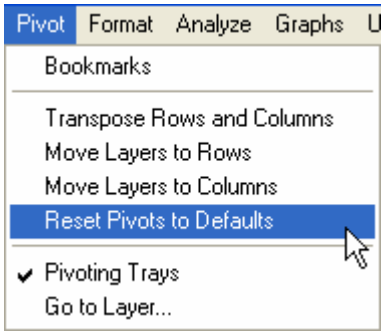
ولإعادة طريقة عرض النتائج، يتم النقر مرتين على الجدول ثم اختيار الأمر



وبذلك يظهر محرر الجداول المحورية.



ويمكن الآن سحب أيقونات المتغيرات في الصفوف أو الأعمدة أو الطبقات وتبـــديل مواقعهـــا، كذلك يمكن وضع جميع المتغيرات ضمن الطبقات أو ضمن الأعمدة أو ضمن الصـــفوف، إلا أنه يجب ملاحظة ترتيب أيقونة المتغيرات ضمن الصفوف أو الأعمدة أو الطبقـــات ومعرفـــة كيف يتم إعادة ترتيب الجدول. إن الخيارات الممكنة لترتيب المتغيرات في الجداول المحورية تعتبر كبيرة، ويمكن للقارئ تجربة طرائق العرض بإبدال مواقع المتغيـــرات، ويمكـــن دائمـــا الرجوع إلى طريقة العرض الافتراضية باستخدام الأمر



ويمكن إخفاء احد الصفوف أو الأعمدة بواسطة Ctrl+Alt والنقر علـــى الصـــف أو العمـــود وذلك لتحديد الصف أو العمود، ثم اختيار الأمر

ويتم إعادة إظهار الصف أو العمود المخفي باختيار Show All من قائمة View.

وكمثال على تعديل الجداول المحورية فقد تم إبدال الصفوف مع الأعمدة في الجـــدول الســـابق لتظهر النتائج على النحو التالي:

Country of Origin * Number of Cylinders Crosstabulation

Statistics: Count

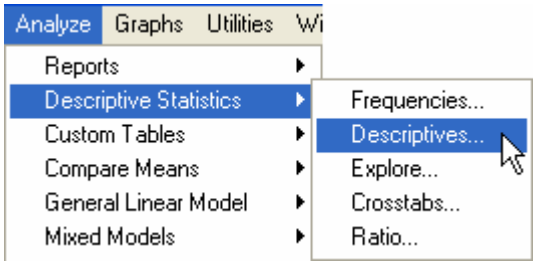
		Country of Origin			
		Country of Origin			
Number of Cylinders		American	European	Japanese	Total
Number of Cylinders	3 Cylinders			4	4
	4 Cylinders	69	61	69	199
	5 Cylinders		3		3
	6 Cylinders	73	4	6	83
	8 Cylinders	102			102
Total		244	68	79	391

البيانات الكمية Quantitative Variables

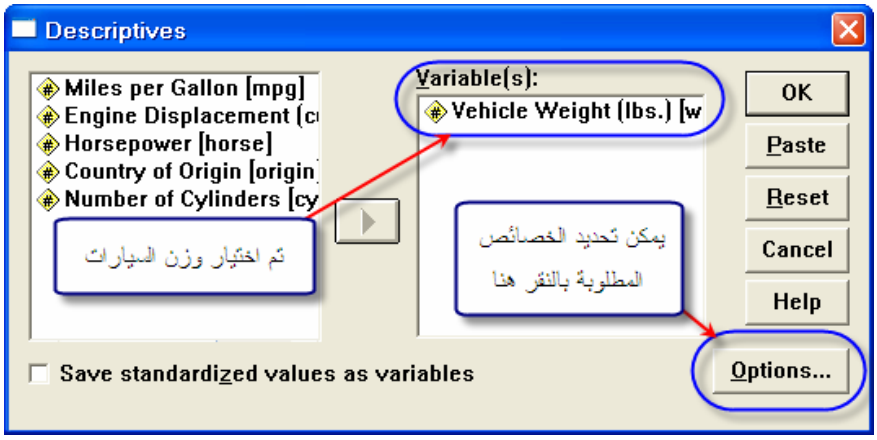
المتغيرات الكمية هي جميع المتغيرات التي يمكن قياسها مثل عمر الموظف أو وزن السيارة أو المعدل التراكمي لطالب. ويوفر برنامج SPSS العديد من الأوامر التي تستخدم لوصف واستكشاف المتغيرات الكمية نذكر منها مقاييس النزعة المركزية مثل الوسط الحسابي أو مقاييس التشتت مثل الانحراف المعياري. وسيتم التطرق إلى العديد من الأوامر التي تخدم هذا الغرض بالإضافة إلى عرض كيفية تكوين مخططات بيانية لوصف المتغيرات الكمية مثل Histograms و Box plot. وسيتم استخدام بيانات Cars.sav للإجابة على بعض الاستفسارات وذلك بهدف عرض ملخصات عن بعض المتغيرات.

الاستفسار الرابع: معرفة متوسط وزن السيارات مع عرض بعض خصائص المتغير.

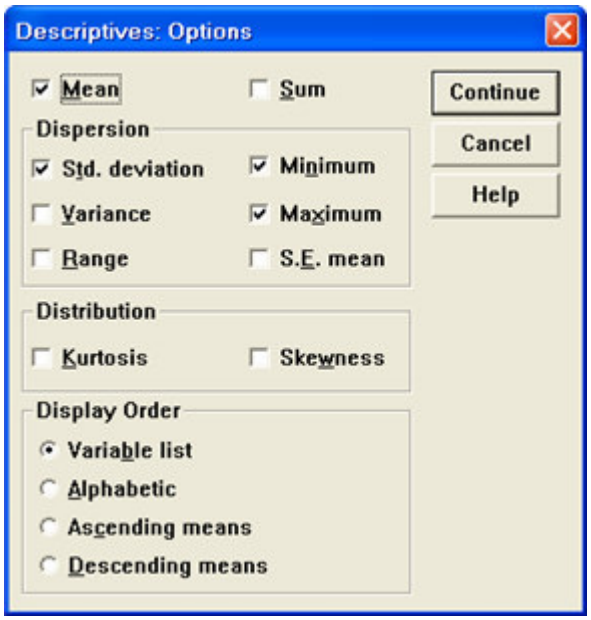
ويتم ذلك باختيار الأمر



ليظهر مربع الحوار التالي



وباختيار المتغير weight ثم النقر على زر Options، يظهر مربع الحوار التالي.



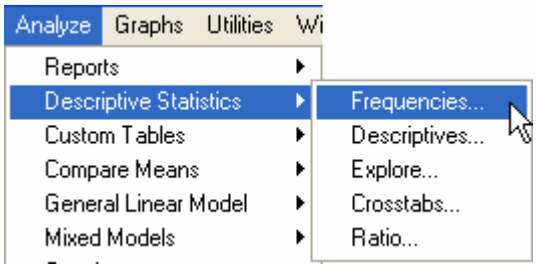
ويوفر هذا المربع إمكانية تحديد خصائص المتغير والتي يرغب الباحث في عرضها مثل الوسط الحسابي والانحراف المعياري. وعند النقر على Continue ثم OK، تظهر النتائج على شاشة عارض النتائج على النحو التالي.

Descriptive Statistics

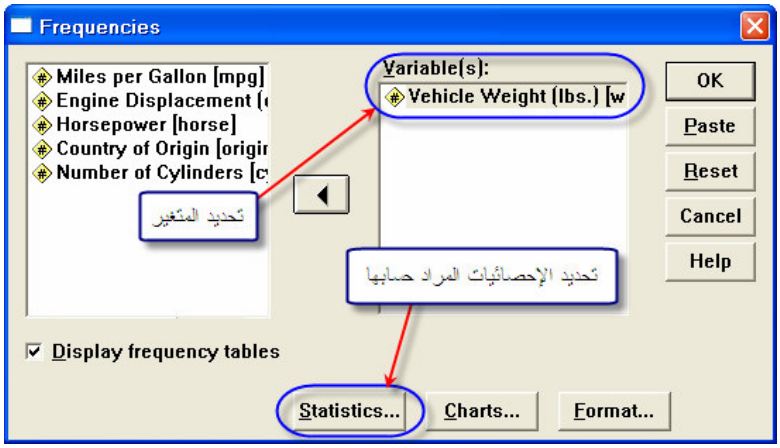
	N	Minimum	Maximum	Mean	Std. Deviation
Vehicle Weight (lbs.)	391	1613	5140	2973.10	845.826
Valid N (listwise)	391				

وبهذا فإن متوسط وزن السيارات هو 2973.10 وبانحراف معياري 845.83 وأكثر السيارات وزناً تزن 5140 وأقلها وزناً تزن 1613.

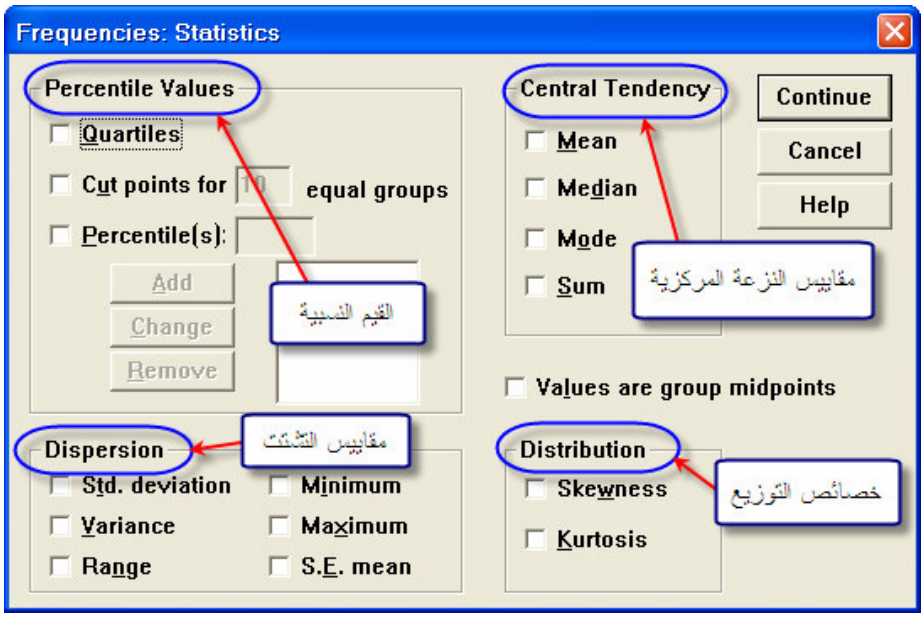
ويمكن التوسع في عرض خصائص المتغير weight وذلك باستخدام الأمر



ليظهر مربع الحوار التالي.



وبالنقر على الزر Statistics يظهر مربع الحوار التالي.



وبذلك فإن مربع الحوار السابق يوفر إمكانية حساب خصائص أكثر للمتغيرات محل الدراسة، حيث يمكن حساب التالي:

مقاييس النزعة المركزية Central Tendency

1. الوسط الحسابي Mean، ويمثل المتوسط الحسابي لقيم المتغير.
2. الوسيط Median، ويمثل القيمة التي تقع في منتصف البيانات، بحث يكون 50% من القيم أقل من قيمة الوسيط و 50% من القيم أعلى من قيمة الوسيط.
3. المنوال Mode، وهي القيمة الأكثر تكراراً من بين جميع القيم.

مقاييس التشتت Dispersion

1. التباين Variance، ويمثل مقدار التشتت في القيم عن الوسط الحسابي.

2. المدى Range، ويمثل الفرق بين أكبر قيمة وأقل قيمة.
3. الخطأ المعياري S.E. mean، ويمثل الانحراف المعياري للوسط الحسابي.

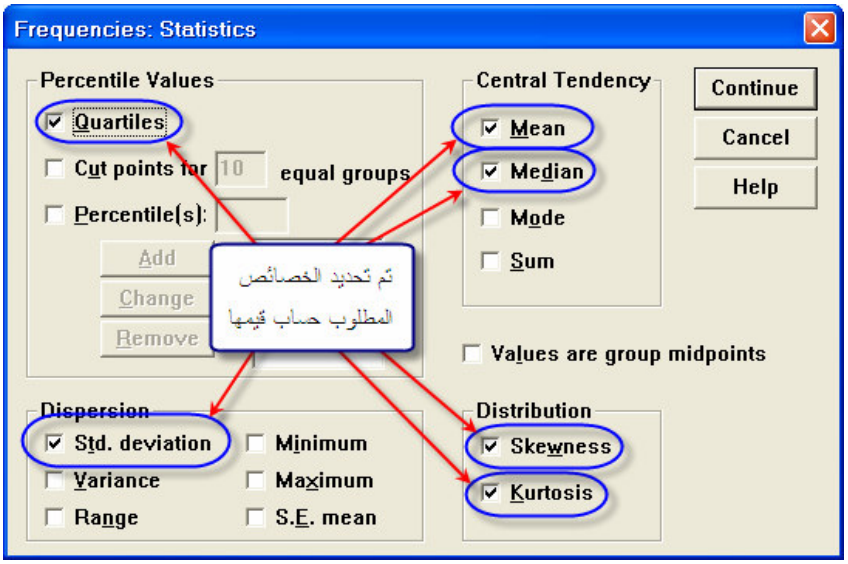
خصائص التوزيع Distribution

1. الالتواء Skewness، وهو مقياس لتمركز البيانات حيث يكون التواء المتغير موجب إذا كانت البيانات تتمركز حول القيم الصغير للمتغير وبذلك تكون القيم المتطرفة إلى اليمين، ويكون الالتواء سالب إذا كانت البيانات تتمركز حول القيم الكبيرة للمتغير وبذلك تكون القيم المتطرفة إلى اليسار. ويكون الالتواء على العلاقة بين الوسط الحسابي والوسيط، فإذا كان الالتواء موجب كان الوسط الحسابي أكبر من الوسيط، وإذا كان الالتواء سالب كان الوسط الحسابي أصغر من الوسيط.
2. التفلطح Kurtosis، وهو مقياس لتكرارات القيم على طرفي توزيع المتغير، فإذا كانت قيمة التفلطح كبيرة كانت تكرارات القيم اكبر على طرفي التوزيع أما إذا كانت تكرارات القيم أقل على طرفي التوزيع كان التفلطح أقل.

القيم النسبية Percentile Values

1. الربيعيات Quartiles، وهي:
   A. الربيع الأول : وهو القيمة التي تكون أكبر من 25% من القيم.
   B. الربيع الثاني : وهو الوسيط أو القيمة التي يقل عنها 50% من القيم.
   C. الربيع الثالث : وهي القيمة التي تكون أكبر من 75% من القيم.
2. نقاط الفصل للقيم Cut points، وهي النقاط التي تقسم قيم المتغير إلى مجموعات بحيث تحتوي كل مجموعة على نفس العدد من القيم.
3. القيم النسبية المحددة Percentiles، وهي قيم مشابهة للربيعيات يحددها الباحث بحيث تكون النسبة المحدد مساوية لنسبة القيم التي تقل عن الناتج.

الاستفسار الخامس: معرفة خصائص المتغير horse للسيارات ذوات الأربع اسطوانات. ويتم ذلك باختيار السيارات ذوات الأربع اسطوانات باستخدام الأمر Select Cases من قائمة Data، ثم استخدام الأمر Frequencies وتحديد الخصائص المطلوب حسابها كما يلي.



وبالنقر على Continue ثم OK، تظهر النتائج التالية على شاشة عارض النتائج.

Statistics

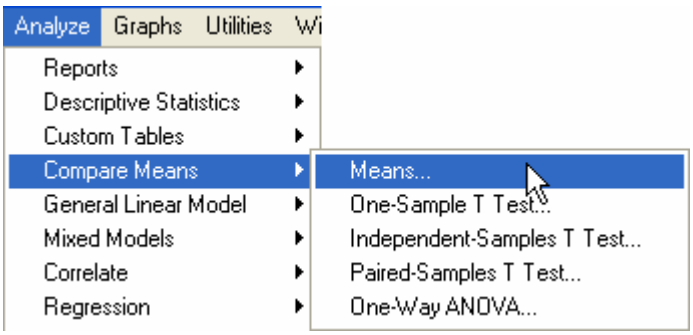
Horsepower

N	Valid	391
	Missing	0
Mean		104.24
Median		93.00
Std. Deviation		38.278
Skewness		1.094
Std. Error of Skewness		.123
Kurtosis		.742
Std. Error of Kurtosis		.246
Percentiles	25	75.00
	50	93.00
	75	125.00

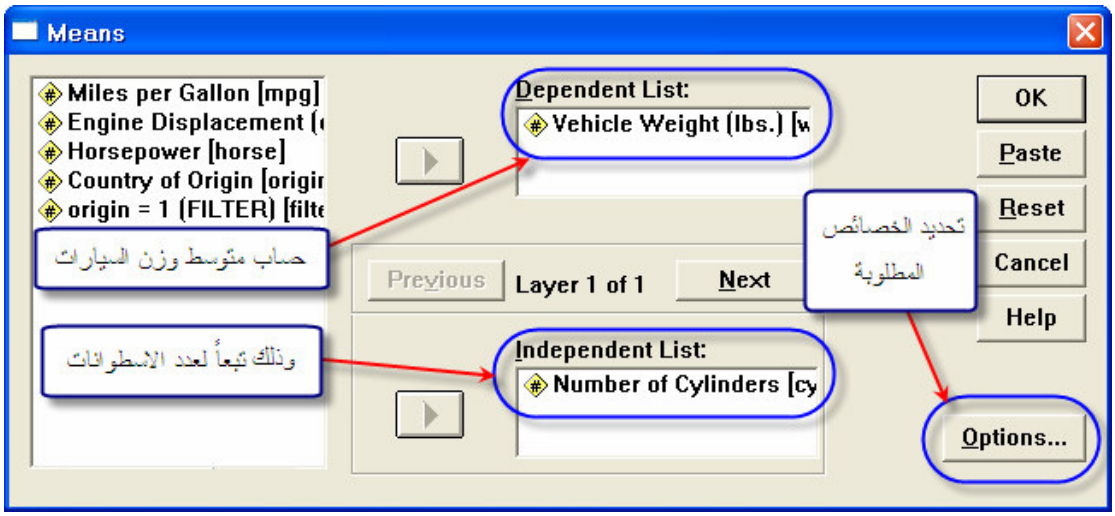
وبهذا يتم الحصول على الإحصائيات المطلوبة.

الاستفسار السادس : معرفة متوسط وزن السيارات الأمريكية وذلك حسب عدد الاسطوانات مع التمثيل البياني للبيانات الخام لمعرفة خصائص المتغير.

ويتم ذلك باختيار مجموعة السيارات الأمريكية أولاً وذلك باستخدام الأمر Select Cases ثم استخدام الأمر



ويوفر الأمر Means إمكانية حساب متوسطات لمتغير تبعاً لمتغير تصنيفي أو عدة متغيرات تصنيفية. وعند النقر على الأمر Means يظهر مربع الحوار التالي.



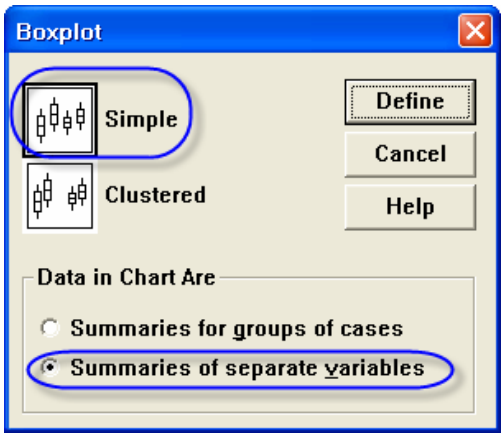
وعند النقر على OK تظهر النتائج التالية

Vehicle Weight (lbs.)

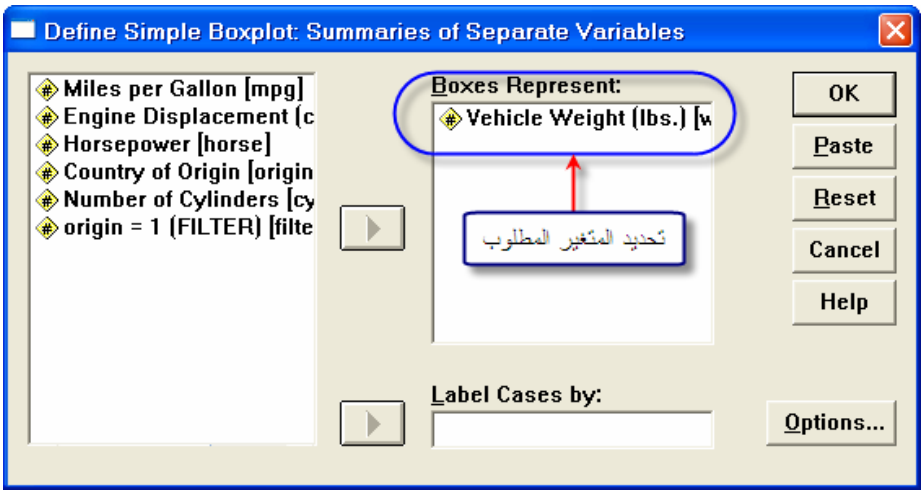
Number of Cylinders	Mean	N	Std. Deviation
4 Cylinders	2427.39	69	289.974
6 Cylinders	3218.55	73	332.880
8 Cylinders	4108.67	102	446.807
Total	3366.92	244	792.176

وتمثل هذه النتائج متوسطات أوزان السيارات الأمريكية بعد تصنيفها تبعاً لعدد الاسطوانات، فعلى سبيل المثال فإن متوسط وزن السيارات الأمريكية ذوات الأربع اسطوانات هو 2427.39 رطل.

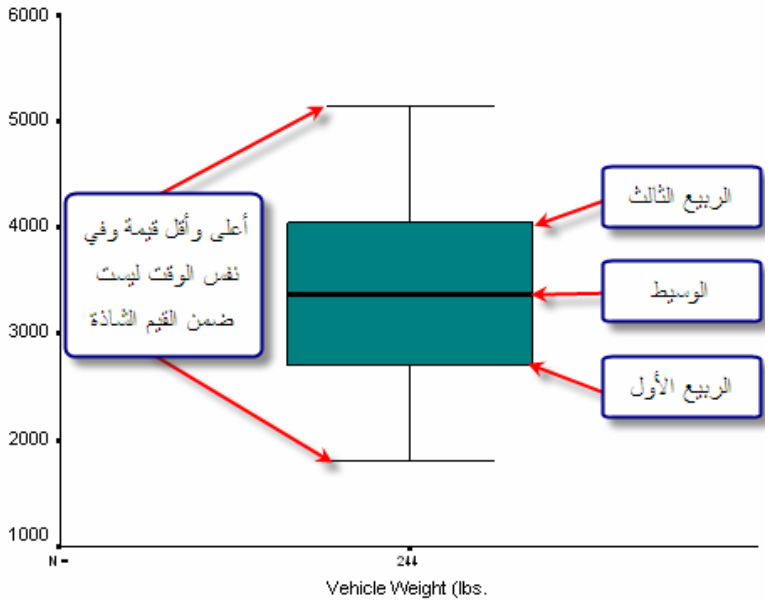
ويمكن دراسة البيانات الخام بيانياً لمعرفة خصائص التوزيع للمتغير، وتستخدم لهذا الغرض الأوامر Boxplot، الأمر Histograms. وسيتم عرض التمثيل البياني لوزن السيارات الأمريكية باستخدام الأوامر السابقة مع التعليق عليها. ويتم الحصول على التمثيل البياني باستخدام Boxplot باختيار الأمر Boxplot من قائمة Graphs ليظهر مربع الحوار التالي.



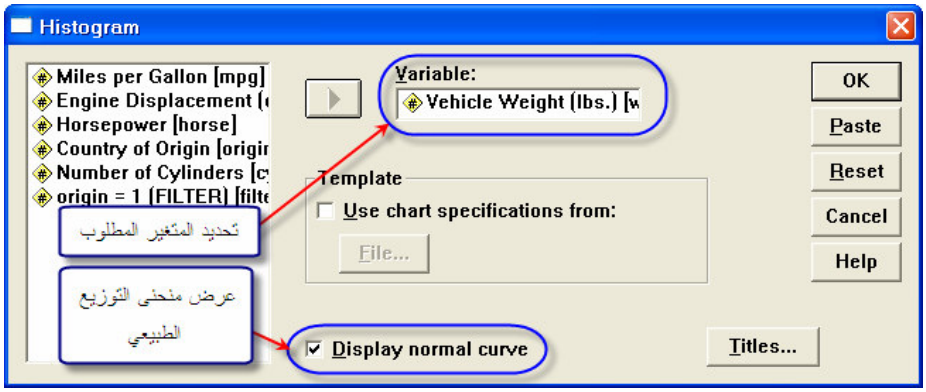
وبالنقر على Define يظهر مربع الحوار التالي.



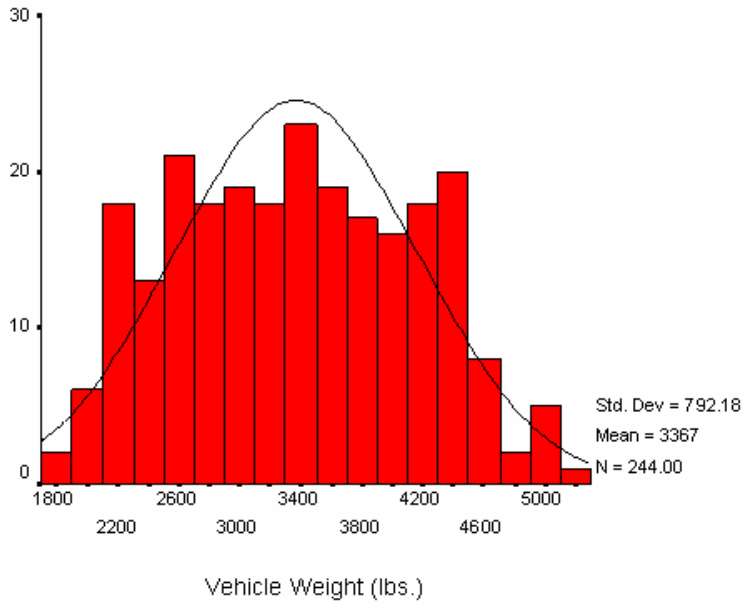
وبعد النقر على OK، تظهر الرسم على شاشة عارض النتائج.



ويوضح الرسم أنه لا توجد قيم شاذة في البيانات وأن الوسيط يقع في منتصف البيانات وذلك لا يوجد التواء في توزيع المتغير. ولعرض البيانات باستخدام الأمر Histogram، يتم اختيار الأمر Histograms من قائمة Graphs ليظهر مربع الحوار التالي.



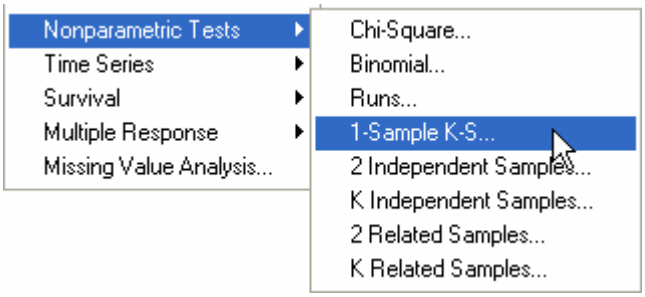
وبالنقر على زر OK يظهر الرسم على شاشة عارض النتائج.



ويوضح الرسم أن توزيع البيانات قريب من التوزيع الطبيعي.

ويمكن كذلك معرفة مدى التجانس بين توزيع البيانات والتوزيع الطبيعي باستخدام اختبار Kolmogorov-Smirnov وهو أحد الاختبارات اللامعلمية بالإضافة إلى نوعين من المخططات البيانية وهما Normal P-P Plot و Detrended Normal P-P Plot.

ولإجراء اختبار Kolmogorov-Smirnov، يتم اختيار الأمر



من قائمة Analysis، ليظهر مربع الحوار التالي:

وبالنقر على زر OK، تظهر النتائج التالية:

One-Sample Kolmogorov-Smirnov Test

		Vehicle Weight (lbs.)
N		244
Normal Parameters[a,b]	Mean	3366.92
	Std. Deviation	792.176
Most Extreme Differences	Absolute	.061
	Positive	.059
	Negative	-.061
Kolmogorov-Smirnov Z		.956
Asymp. Sig. (2-tailed)		.320

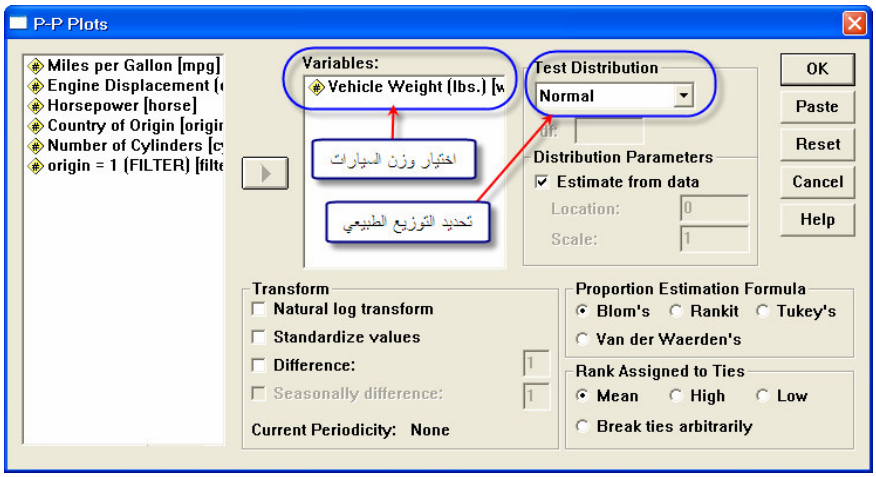
a. Test distribution is Normal.

b. Calculated from data.

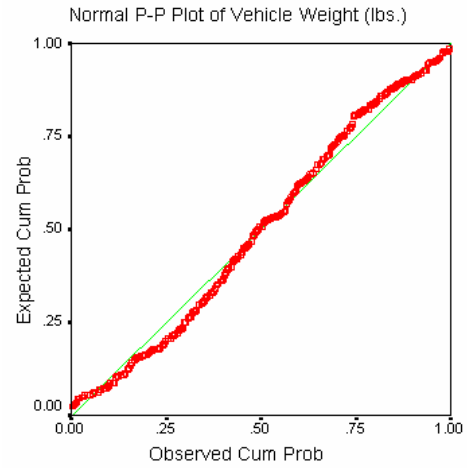
نتائج الاختبار

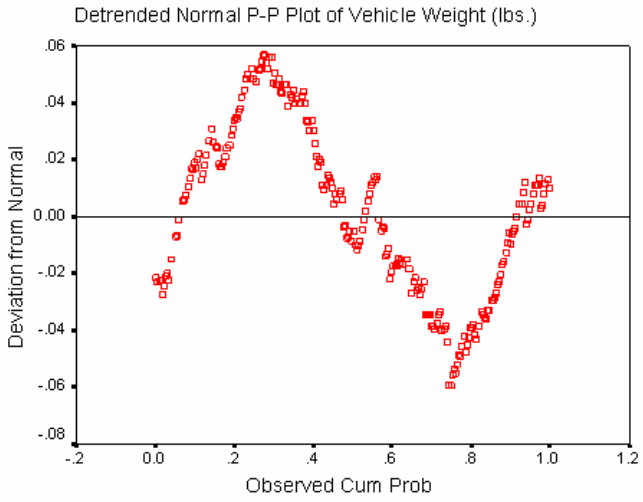
ويختبر الأمر 1-Sample K-S فرضية تجانس توزيع وزن السيارات الأمريكية مع التوزيع الطبيعي، حيث تكون فرضية العدم بأن المتغير يتبع التوزيع الطبيعي وتكون الفرضية البديلة أن المتغير لا يتبع التوزيع الطبيعي. ويقوم الاختبار بحساب قيمة Z المعيارية والتي تتبع التوزيع الطبيعي المعياري وقيمة P-Value والتي يتم مقارنتها بقيمة $\alpha/2$ والتي يحددها الباحث. وبناء على نتائج الاختبار فإنه لا يمكن رفض فرضية العدم وهذا يدل على أن توزيع وزن السيارات الأمريكية يتبع التوزيع الطبيعي.

كذلك يمكن التأكد بيانياً من أن المتغير السابق يتبع التوزيع الطبيعي باستخدام الأمر P-P من قائمة Graphs، وعند اختيار الأمر يظهر مربع الحوار التالي:



وعند النقر على OK تظهر النتائج التالية:





وينتج من الأمر الرسمين السابقين، حيث يمثل الرسم انتشار الاحتمال التراكمي للمتغير (المحور الأفقي) مع الاحتمال التراكمي المتوقع تبعاً للتوزيع الطبيعي المعياري (المحور الرأسي). وعندما يكون توزيع المتغير متوافق مع التوزيع الطبيعي فإن الانتشار يكون قريب جداُ من الخط القطري. أما إذا كان المتغير غير.متوافق مع التوزيع الطبيعي فإن الانتشار

يبتعد عن الخط القطري. ويعرض الرسم الثاني الفروقات بين الاحتمال التراكمي للمتغير مع الاحتمال التراكمي المتوقع تبعاً للتوزيع الطبيعي المعياري. ويكون المتغير متوافق مع التوزيع الطبيعي إذا كانت الفروقات قريبة من الصفر.

ومن الأوامر التي تفيد الباحث عند الرغبة تحليل بيانات كمية استخدام الأمر Explore وهو احد الأوامر من Descriptive Statistics من قائمة Analysis والذي يوفر نتائج إحصائية شاملة للتعرف على خصائص المتغير قبل إجراء التحليلات الإحصائية على المتغير محل الدراسة. كذلك يوفر الأمر Explore بعض الرسومات البيانية والتي تستخدم لغرض التعرف على البيانات الخام ومحاولة تصحيح الأخطاء ومعالجة القيم الشاذة وعلى توزيع المتغير مثل الأمر Boxplot والأمر Histogram.