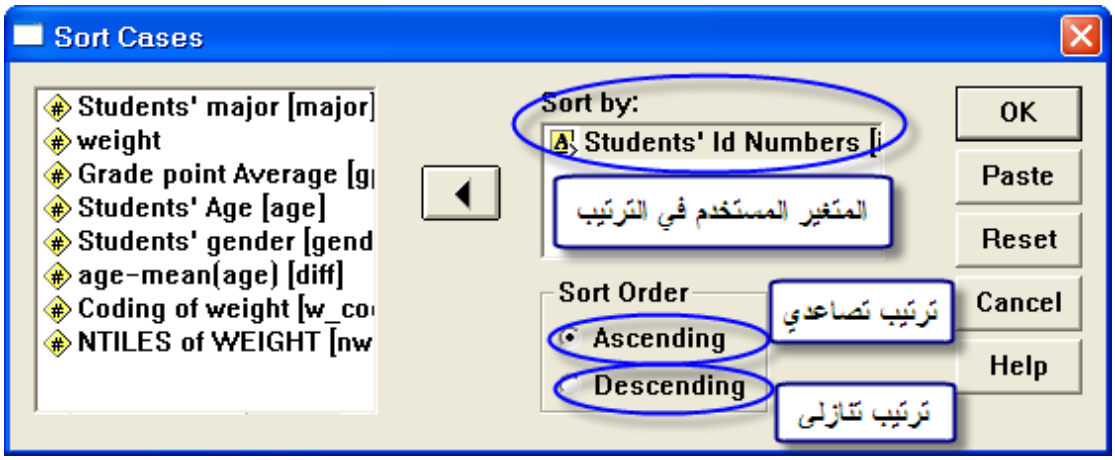
ترتيب واختيار الحالات

تشمل عمليات التعامل مع البيانات على العديد من الأوامر والتي لا يمكن التحدث عن جميعها بالتفصيل، ولكن يوجد بعض الأوامر التي يمكن الاستفادة منها في تكوين ملفات بيانات بالشكل التي تناسب التحليل.

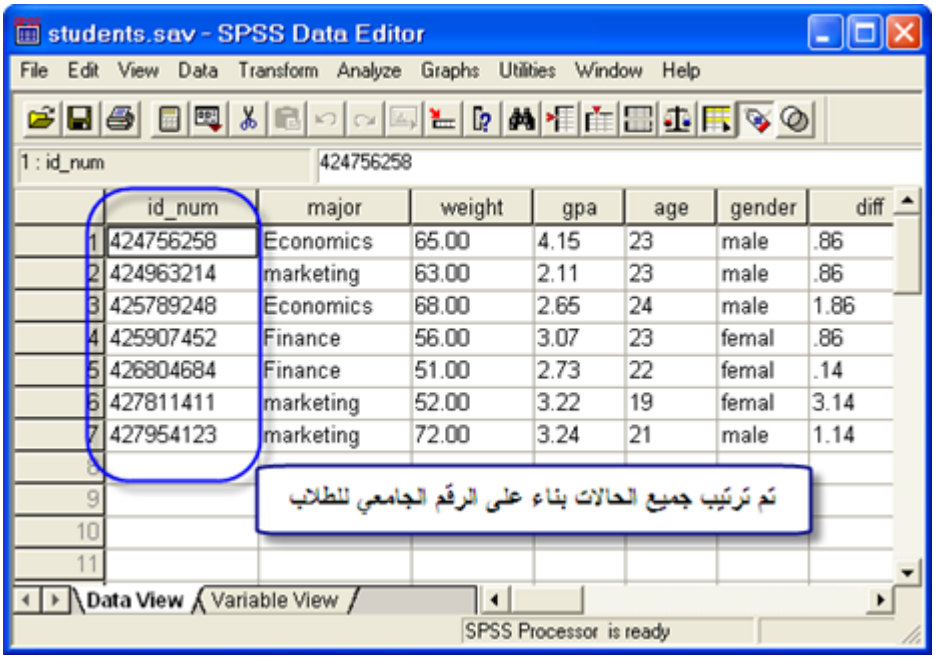
ترتيب الحالات Sorting Cases

تكمن أهمية ترتيب الحالات بأنها وسيلة لعرض البيانات وإعادة عرضها بطريقة تلائم التحليل، ويمكن بواسطة الأمر Sort Cases ترتيب الحالات بناء على قيم متغير أو عدة متغيرات، فعلى سبيل المثال، يمكن ترتيب الطلاب باستخدام متغير الوزن "weight"، وقد يكون الترتيب تصاعدياً بحيث تظهر الحالات الأقل وزناُ في البداية والحالات الأكثر وزناُ في النهاية، أو ترتيب الحالات تنازلياً بحيث تظهر الحالات الأكثر وزناً في البداية والأقل وزناُ في النهاية. كذلك فإنه يمكن ترتيب البيانات بناء على أكثر من متغير، أي أن الباحث يستخدم أكثر من معيار لترتيب الحالات.

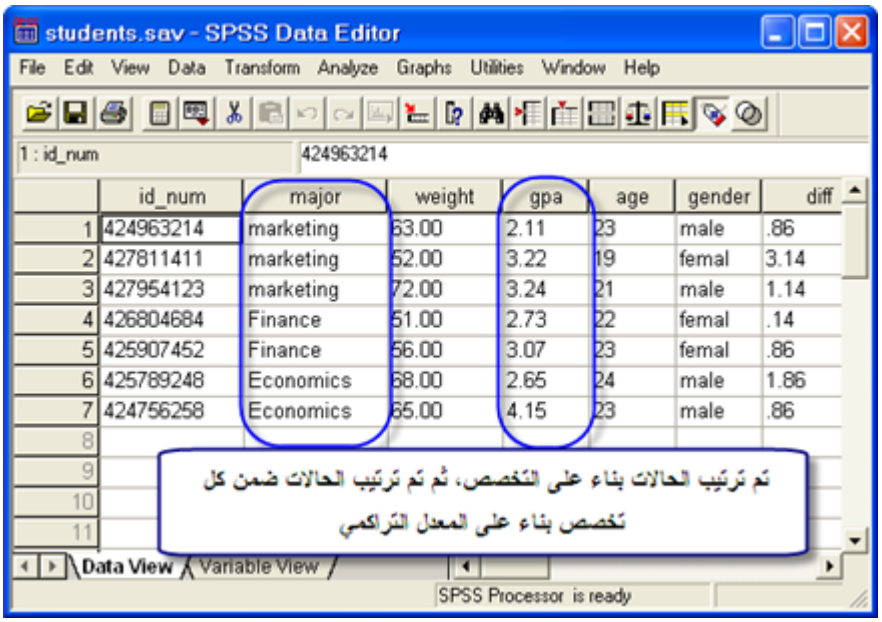
لنفرض أن لدينا الرغبة في ترتيب الحالات الموجودة في ملف الطلاب وذلك تبعاً للرقم الجامعي للطلاب، نستخدم الأمرِ Sort Cases من قائمة البيانات، وعندها يظهر مربع الحوار التالي.



ومن قائمة المتغيرات نختار المتغير "id_num" ثم نختار طريقة الترتيب وذلك إما تصاعدية أو تنازلية، وبعد النقر على OK، تظهر لنا البيانات مرتبة بحب الرقم الجامعي للطالب.



لنفرض أن لدينا الرغبة في ترتيب الحالات بناء التخصص ثم المعدل التراكمي، نتبع نفس الخطوات السابقة إلا أننا نضيف "major" ثم "gpa" إلى قائمة المتغيرات المستخدمة في الترتيب، وبذلك تظهر لنا البيانات على النحو التالي.

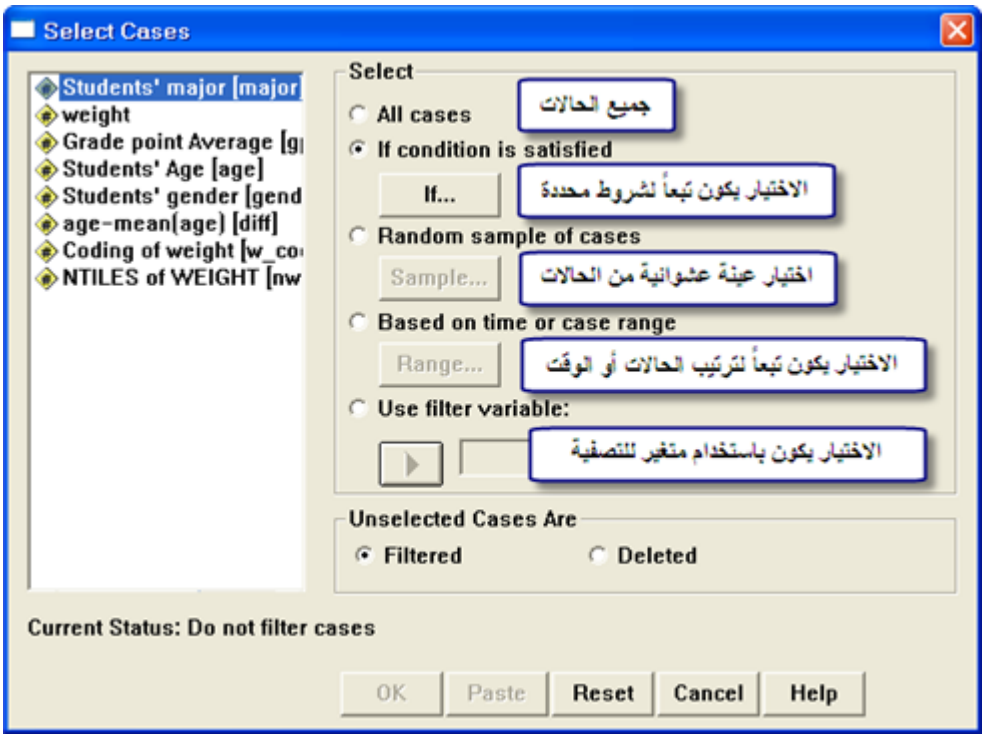


ويلاحظ أن تخصص "marketing" كان الأول في الترتيب وذلك بسبب أن قيمة المتغير "major" لهذا التخصص هو 1 والقيمة 2 لتخصص "finance" والقيمة 3 لتخصص "economics".

اختيار الحالات Selecting Cases

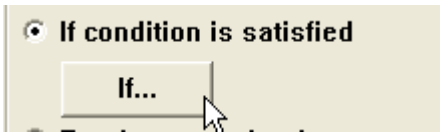
يستخدم الأمر Select Cases لاختيار مجموعة جزئية من الحالات وذلك ليتم استخدام الحالات المختارة فقط في التحليل. لنفرض أن لدينا الرغبة في حساب قيمة متوسط عمر الطلاب في تخصصات التسويق والمالية، نتبع الخطوات التالية:

بالنقر على الأمر Select Cases من قائمة البيانات يظهر لنا مربع الحوار التالي.

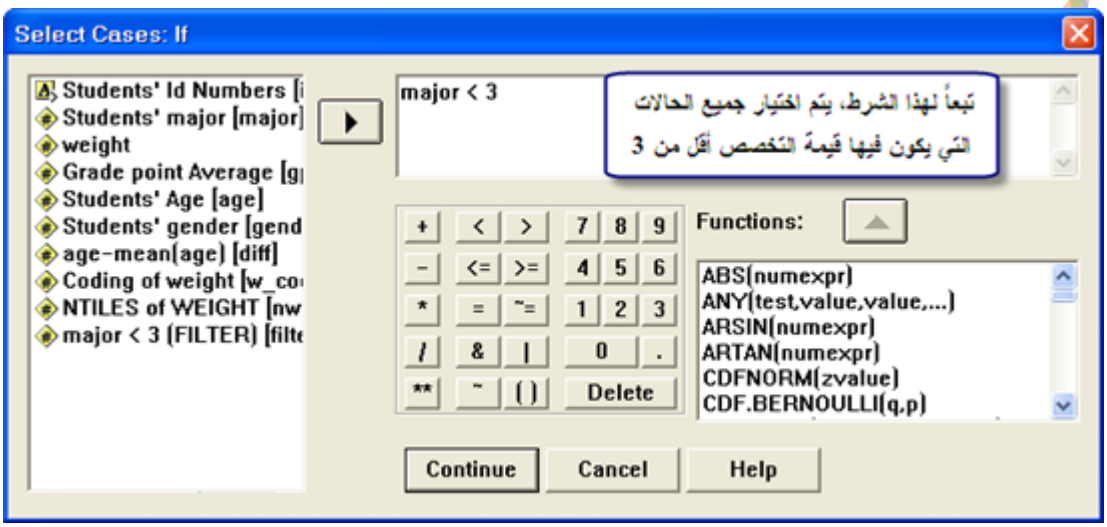


ويتوفر في مربع الحوار السابق عدة طرائق لاختيار الحالات حيث يمكن أن يكون الاختيار تبعاً لشروط محددة أو اختيار عينة عشوائية أو يكون الاختيار تبعاً لترتيب الحالات أو يكون الاختيار باستخدام متغير لتصفية البيانات.

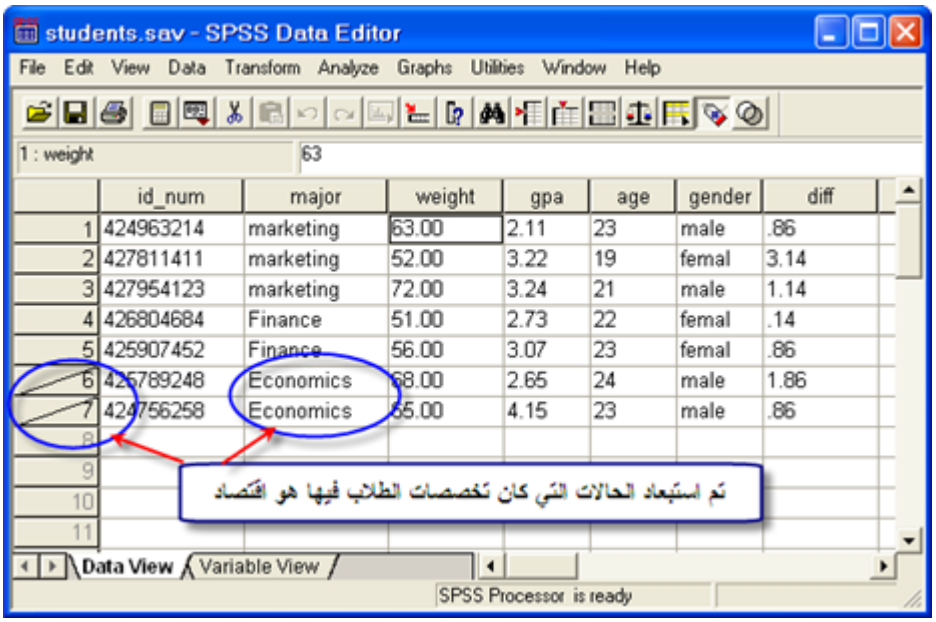
وباختيار If condition is satisfied والنقر على:



يظهر لنا مربع الحوار التالي.



ويمكن في مربع الحوار السابق تحديد الشرط المناسب باستخدام أحد المتغيرات الموجودة في قائمة المتغيرات، وقد كان الشرط المناسب لاختيار الطلاب المتخصصين تسويق ومالية هو أن تكون قيمة المتغير "major" اقل من 3، حيث أن قيم المتغير "major" لتخصصات التسويق والمالية هي 1 و 2 على التوالي. وعند النقر على Continue ثم Ok نشاهد البيانات وقد تم استبعاد الطلاب المتخصصين اقتصاد.



وبذلك فإن التحليل التالي سيشمل فقط طلاب التسويق والمالية. وعند حساب متوسط عمر طلاب التسويق والمالية باستخدام الأمر Descriptive Statistics من قائمة التحليل، نتج أن متوسط عمر الطلاب هو 21.6 سنة.

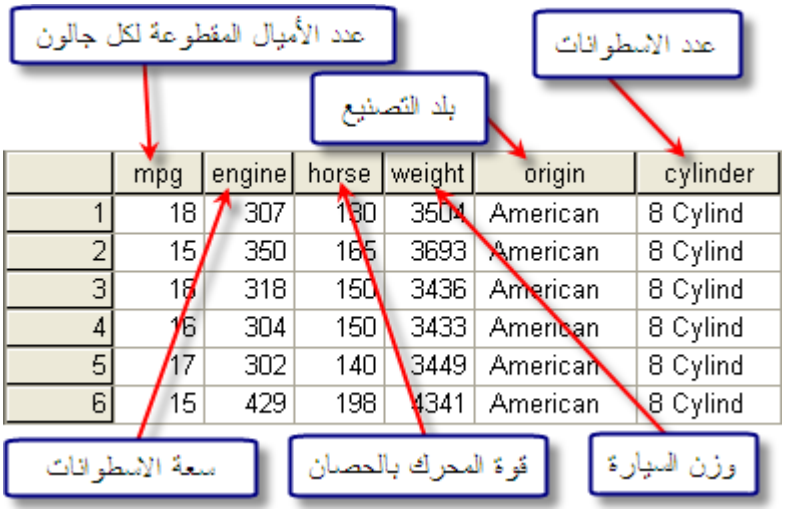
وصف المتغيرات الاسمية والكمية
Data Description

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

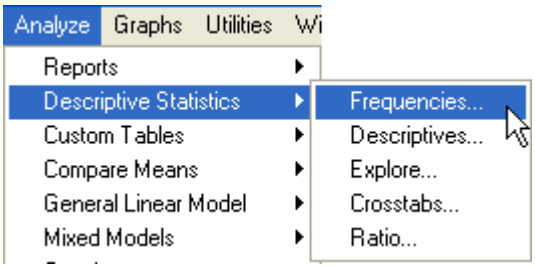
ويوجد في هذا الملف متغيرين تصنيفيين وهما Origin و Cylinder وأربع متغيرات كمية وهي mpg، engine، horse و weight.

البيانات الاسمية Qualitative Variables

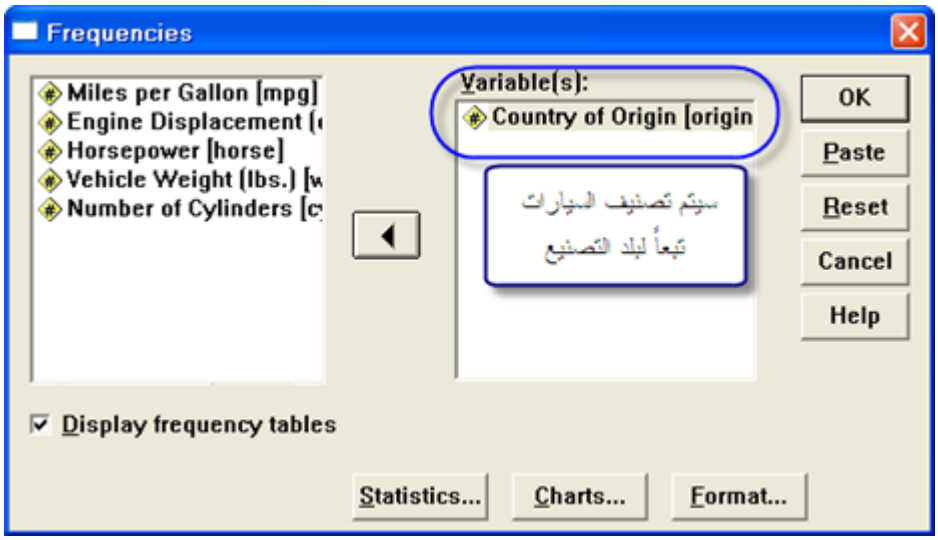
البيانات الاسمية هي جميع المتغيرات التصنيفية والتي تمثل كل قيمة من قيمها مستوى من مستويات التصنيف المختلفة، إلا أنه يجب التنبيه إلى أن قيم التصنيف ليس لها معنى بحد ذاتها وإنما هي قيم عددية تمكن الباحث من فصل الحالات تبعاً لمستويات التصنيف. فعلى سبيل المثال فإن لجنس الطالب مستويين وهما (ذكور، إناث)، وبذلك 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.

Country of Origin

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	American	244	62.4	62.4	62.4
	European	68	17.4	17.4	79.8
	Japanese	79	20.2	20.2	100.0
	Total	391	100.0	100.0	

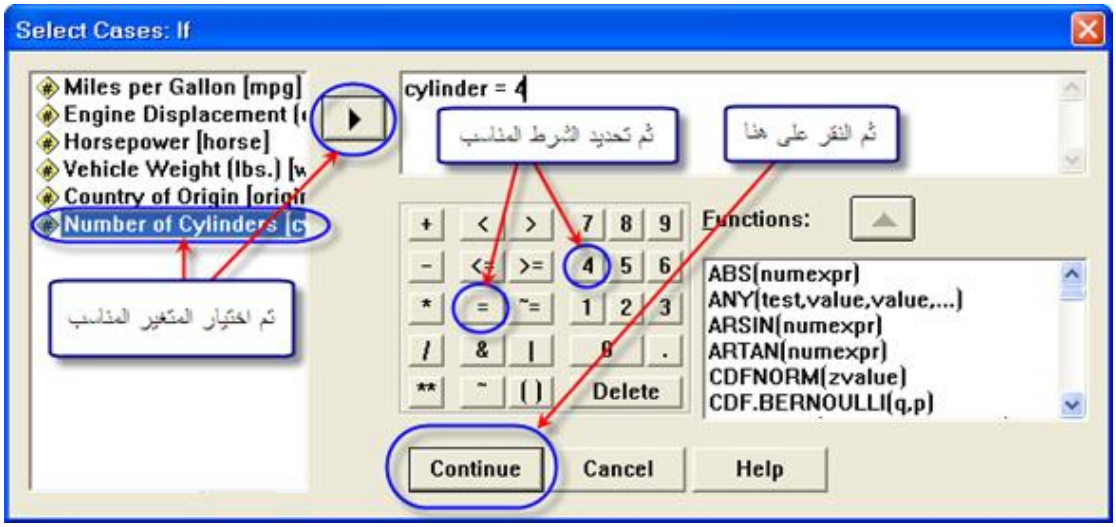
مستويات التصنيف

عدد السيارات مصنفة حسب بلد التصنيع

نسبة عدد السيارات إلى العدد الكلي مصنفة حسب بلد التصنيع

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.

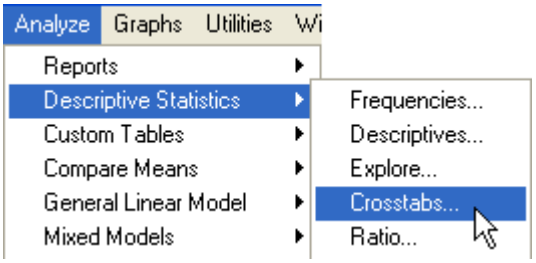
Country of Origin

		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	

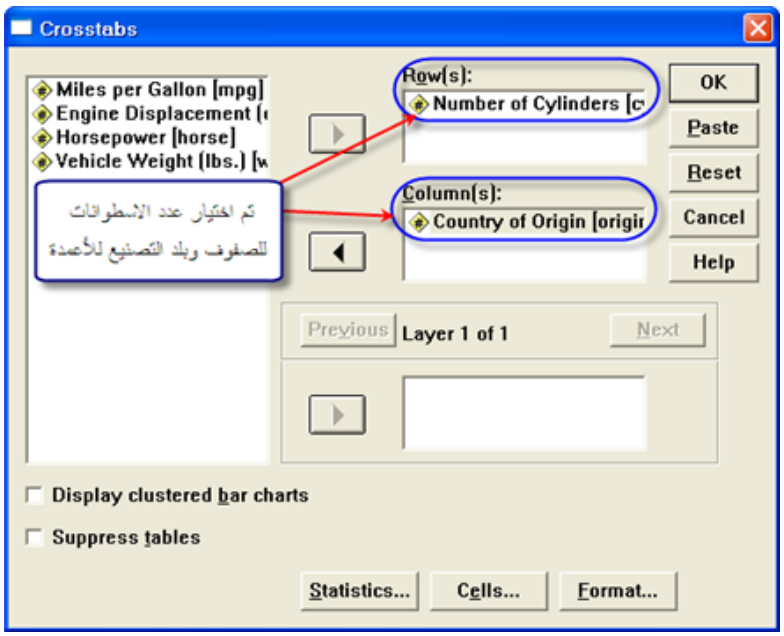
ويوضح الجدول السابق أن العدد الكلي للسيارات ذوات الأربع اسطوانات هو 199، كذلك يبين الجدول توزيع هذه السيارات حسب بلد التصنيع.

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.

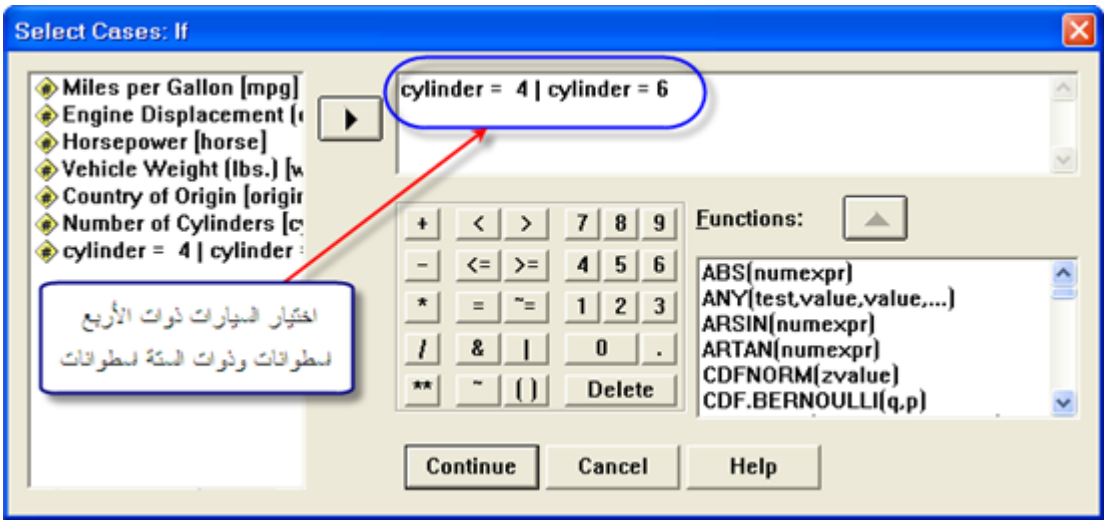


وبالنقر على OK يظهر لنا جدول الاقتران الذي يوضح عدد السيارات بعد تصنيفها.

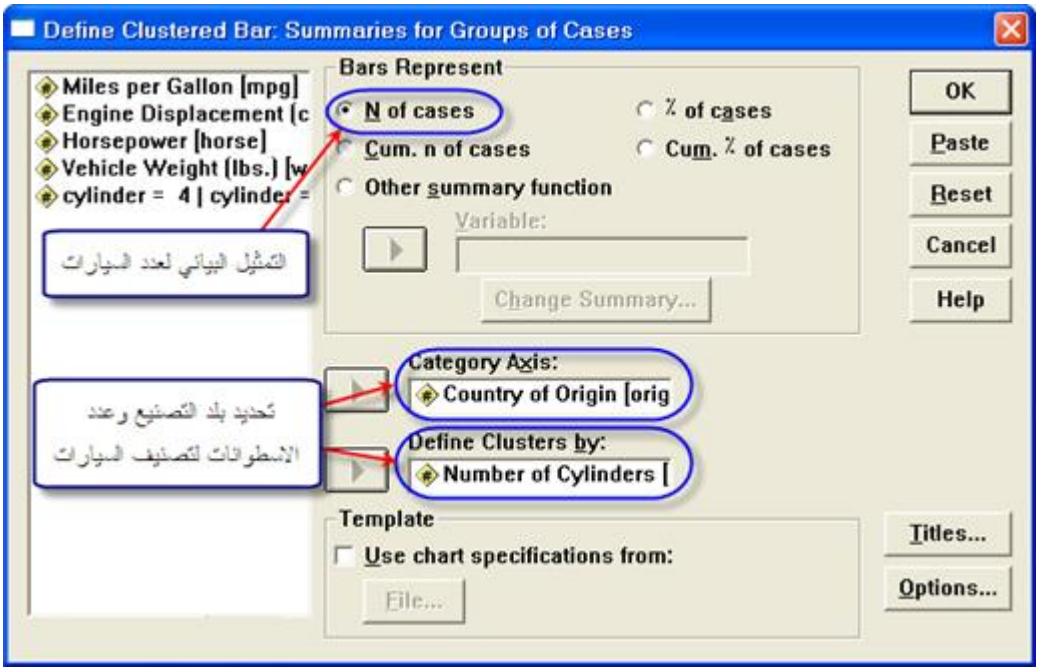
Count

		Country of Origin			Total
		American	European	Japanese	
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.