

Advance Biostatistics Course

SPSS Practical (3)

Data Transformation

تحويل
البيانات

Hosham Mohamed Osman Karar

Health Education and Health Promotion Department College of Public

Health & Health Informatics

Lecture (4) content

- Introduction

- Recoding تسجيل

- Computing New Variable حساب متغير جديد

- Select Cases (If, Sample & Range) اختيار حالات

- Sort Cases تصنيف الحالات

- Split File فصل الملفات

Introduction

يعتمد SPSS أدوات للتدبير على المتغيرات الموجودة

In addition to statistical analysis, SPSS also provides facilities for modifying existing variables, creating new variables, rearranging cases and selection of subsets within the data set. ^{مجموعة جزئية} It is that these command are mastered before undertaking statistical analysis. The most important for these commands are listed below

التدبيرات تشمل

- Recoding Variable
- Computing New Variable
- Select Cases (If, Sample & Range)
- Sort Cases
- Split file

① تغيير (تدبير) المتغير الموجود

② إنشاء متغير جديد

③ اعاده توزيع الحالات

④ اختيار مجموعة جزئية

Recoding Variable

Record :- نتخدم من اجل تحويل قيمه الى قيمه اخرى

Sometimes a variable needs to be changed before it can be used by SPSS.

Recode is used to change the values of one variable into other values.

To recode or combine categories of a variable: *خوارزمتغير Value في متغير*

1. Display the Data Editor window (i.e., execute the following commands while in the Data Editor window displaying the data file you want to use to recode variables). *ط Data Editor* *١ تكون تاتيه الهدف هي نافذة* *وخلنا تطبيق الخطوات بعد الجدول الذي يزيد تعديله*
2. Choose Transform on the menu bar *ترابط العوائق*
3. Choose Recode *تسجيل على نفس المتغير*
4. Choose Into Same Variable... or Into Different Variable... *صانه خيارين ر تسجيل متغير جديد*
5. Select a variable to recode from the variable list on the left and then click on the arrow located in the middle of the window. This defines the input variable. *اختيار المتغير الذي يزيد ندفله من قائمه المتغيرات التي تظهر يار* *تم ضغط على الاسم لينقل الى النافذة الوسطى* *مدخل*
6. If recoding into a different variable, enter the new variable name in the box under Name:, then choose Change. This defines the output variable. *اذا افتدنا التسهيل في متغير جديد نغير نافذه ادخل الاسم الجديد*

Recoding Variable (Cont.)

اختر القيم القديمه و الجديده

7. Choose Old and New Values... ① خيار اما قيمه او صدى عند القيم
8. Choose Value or Range under Old Value and enter old value(s).
9. Choose New Value and enter new value, then choose **Add**.
10. Repeat the process until all old values have been redefined.
11. Choose **Continue** ② خيار القيم الجديده و نزلها تم اضافه
12. Choose **OK** ③ نقره التالى لجميع القيم

بعد انشاء المتغير الجديده لحفظ المتغير عند فلال save

After creating a new variable(s), you will probably want to save the new variable(s) by re-saving your data using the Save command under File box on the menu bar (See Saving Data as an SPSS Save File).

Wintergreen Study - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

Compute...
Random Number Seed...
Count...
Recode ▶
Categorize Variables...
Rank Cases...
Automatic Recode...
Create Time Series...
Replace Missing Values...

Into Same Variables...
Into Different Variables...

1	2
1	1
2	1
2	1
n	1

Recode into Different Variables

Input Variable -> Output Variable:

Output Variable
Name:
Label:

If...
Old and New Values...

OK Paste Reset Cancel Help

Recode into Different Variables

Numeric Variable -> Output Variable:

aa -> ?
*المتغيرات
التي تغيرت*

Output Variable
Name:
Label:

If...
Old and New Values...

OK Paste Reset Cancel Help

Recode into Different Variables

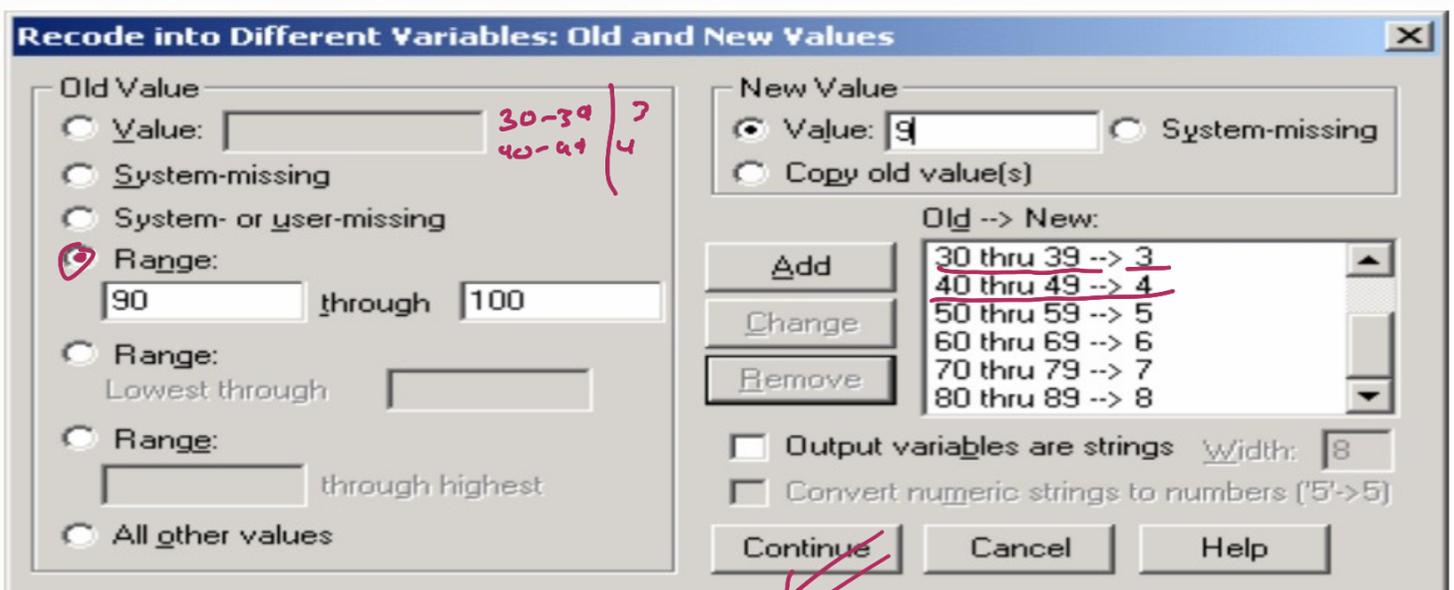
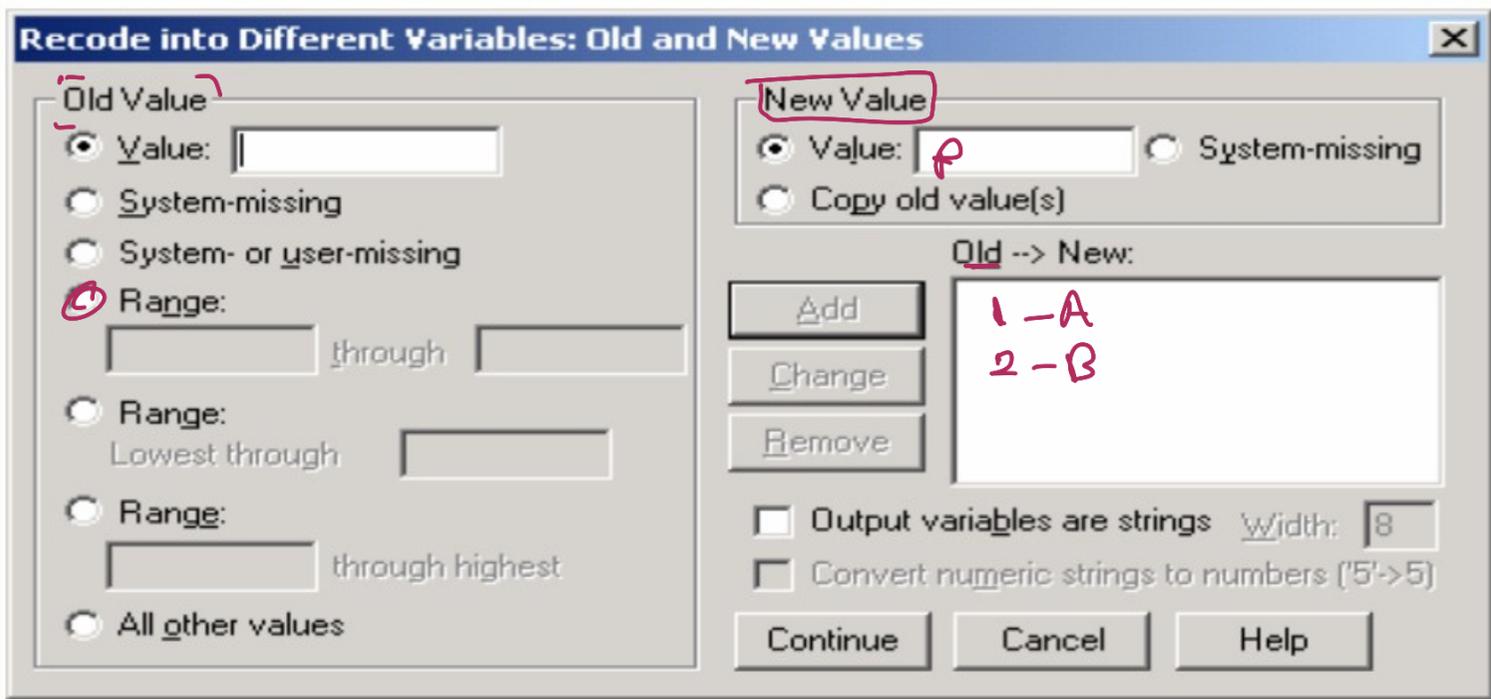
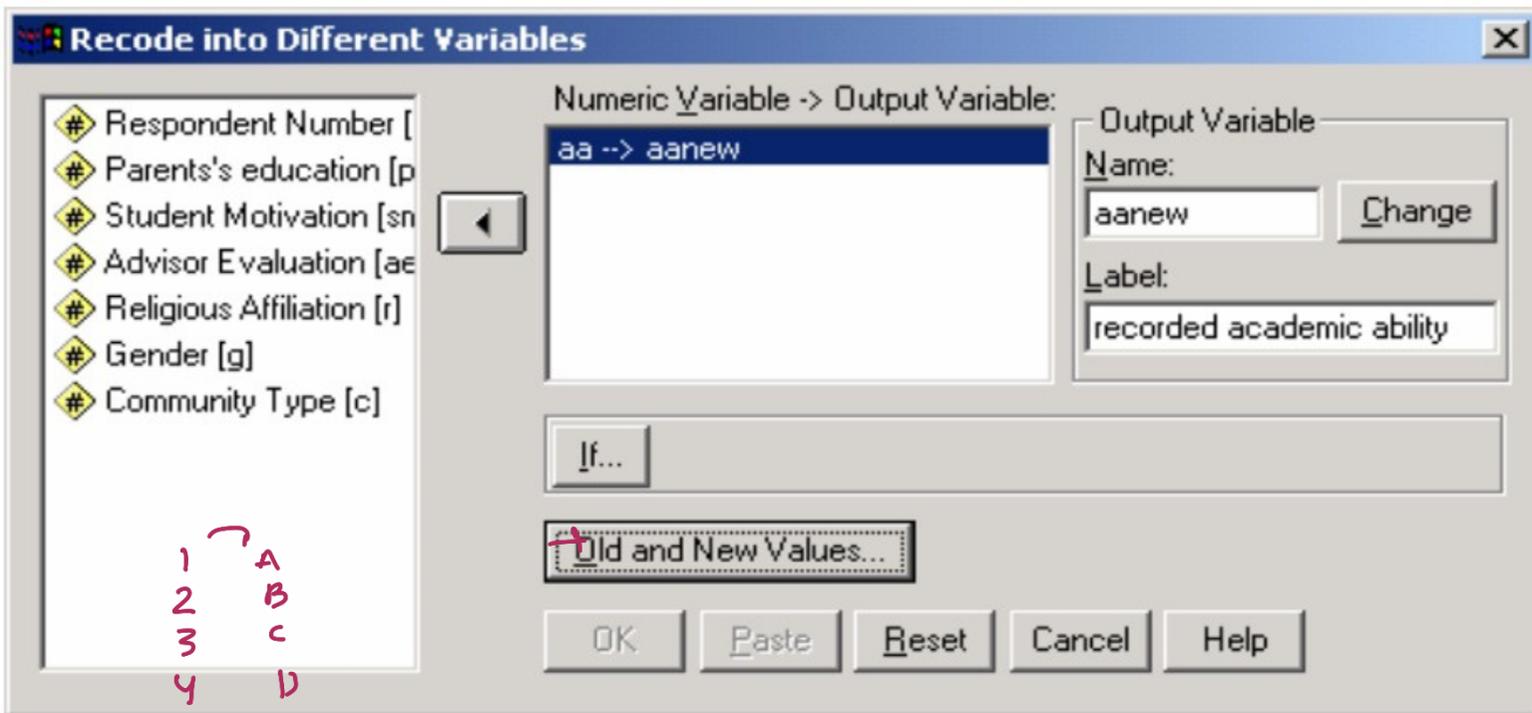
Numeric Variable -> Output Variable:

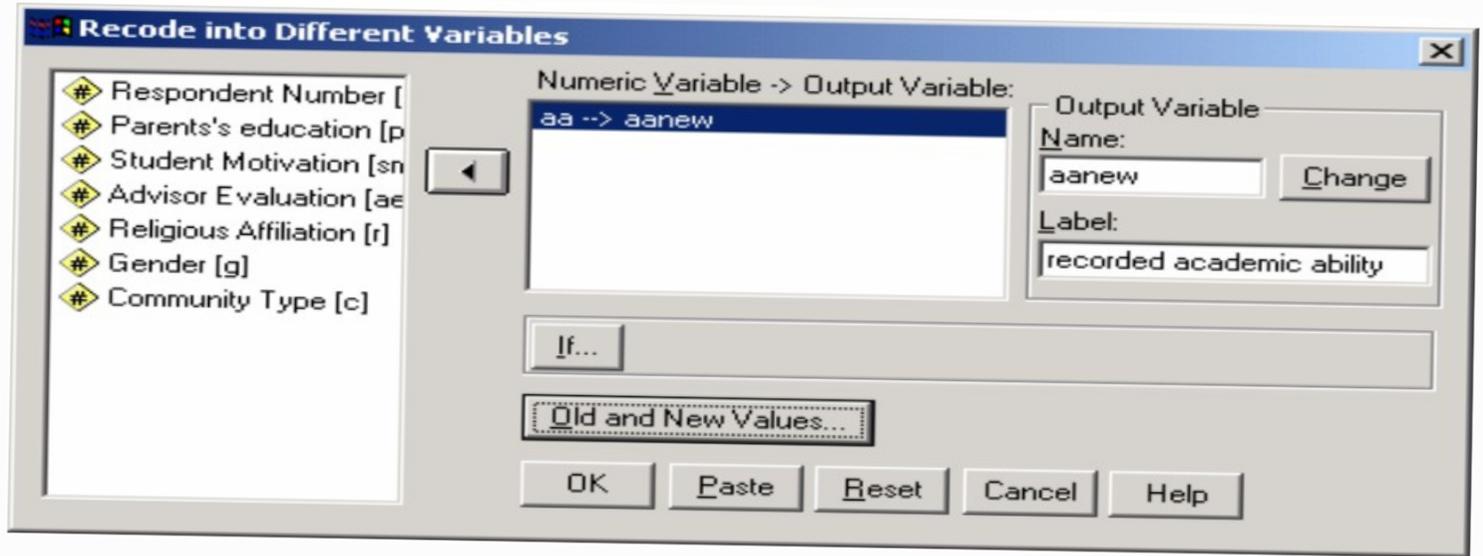
aa -> ?

Output Variable
Name:
Label:

If...
Old and New Values... ✓

OK Paste Reset Cancel Help





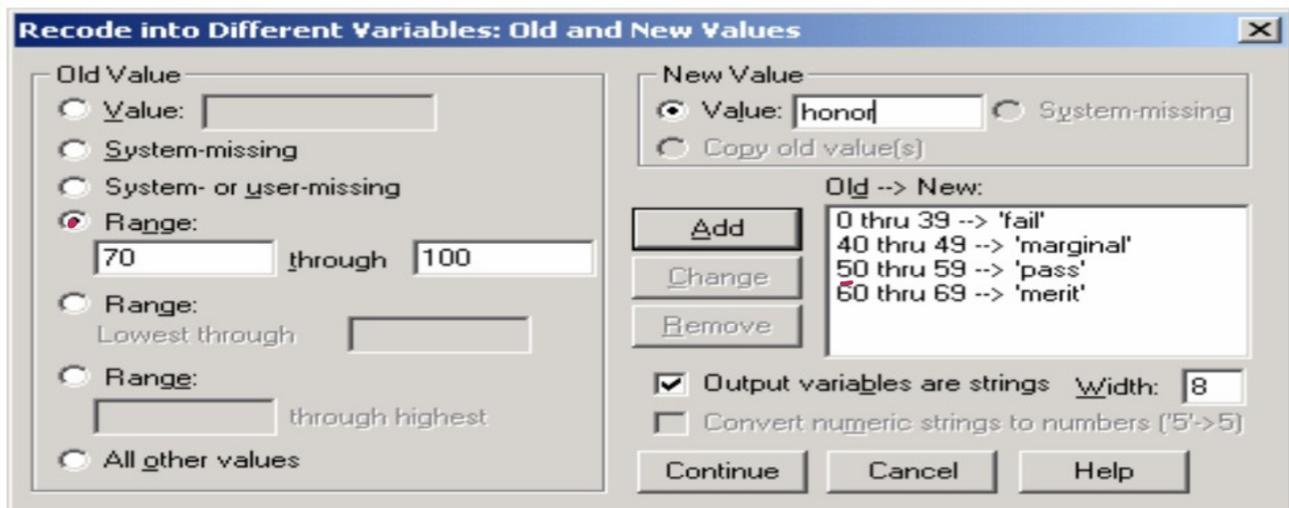
Wintergreen Study_1 SPSS Data Editor

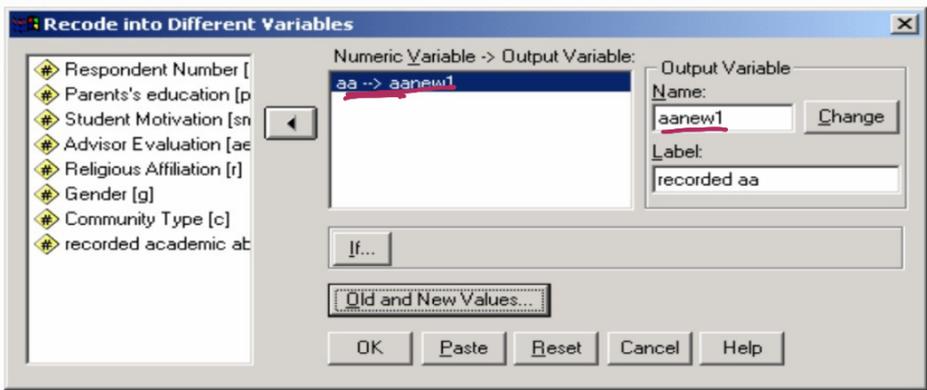
File Edit View Data Transform Analyze Graphs Utilities Window Help

SI : aa

	resp_num	aa	pe	sm	ae	r	g	c	aanew	Var
1	1	43	19	1	2	11	11	1	9.111	
2	2	46	12	0	0	0	0	0	4.000	
3	3	67	16	1	1	0	0	0	6.000	
4	4	94	10	2	2	1	1	1	9.000	
5	5	02	13	2	1	1	1	1	0.000	
6	6	59	12	0	0	2	0	0	5.000	
7	7	61	12	1	2	0	0	0	6.000	
8	8	29	9	0	0	1	1	0	2.000	
9	9	36	13	1	1	0	0	0	3.000	
10	10	91	16	2	2	1	1	0	9.000	
11	11	66	111	11	11	1	11	11	6.111	
12	12	68	11	0	1	0	0	0	6.000	
13	13	67	14	1	1	0	1	1	6.000	
14	14	77	14	1	2	2	1	0	7.000	
15	15	71	12	0	0	2	1	0	7.000	
16	16	03	16	2	2	1	0	1	0.000	
17	17	96	15	2	2	2	0	1	9.000	
18	18	87	12	1	1	0	0	1	8.000	
19	19	62	11	0	0	0	0	0	6.000	
20	20	62	9	0	1	2	1	0	6.000	
21	21	46	10	1	0	0	1	0	4.000	
22	22	91	11	2	2	1	11	11	9.111	

Data View Variable View SPSS Processor is ready





Wintergreen Study_1 - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

51 : ae

	r	g	c	aanew	aanew1
1	0	0	1	9.00	honor
2	0	0	0	4.00	marginal
3	0	0	0	5.00	pass
4	1	1	1	9.00	honor
5	1	1	1	8.00	honor
6	2	0	0	5.00	pass
7	0	0	0	6.00	merit
8	1	1	0	2.00	fail
9	0	0	0	3.00	fail

Recoding Variable (Cont.)

امثلة تطبيق

Using the Epidemiology Books SPSS file to find out the following:

القدر، لا كاربس
aa

قم بالاضوات

1. Modify data values of academic ability into 10 categories that ranges from 1-10. [For instance, student who scored in the ranges '0-9' into 1, '10-19' into 2, and so on. The last one will assign the values '90-100' into 10.]

0-9 → 1
10-19 → 2
20-29 → 3

30-39 → 4

2. Modify the academic ability of students who scored in the ranges "0-39" into fail, '40-49' into "marginal fail", '50-59' into pass, '60-69' into merit" and '70-100' into honor categories.

0-39 - Fail
40-49 - marginal fail
50-59 - it is pass

3. If it is a matter of "pass" or "fail", it can also be recoded into fail for range '0-49' and pass for range '50-100'.

0-49 → fail
50-100 → pass

لبيع SPSS لحساب متغير جديد صنفًا وضعه نصيًا عليه الرافيه او
 منطقة لكتابة الحرف
 حساب متغير جديد

100
 150
 500

Computing New Variable

Computing is similar to **recoding** except that computing **allows** you to **specify a mathematical or logical expression** that is used to recode the **data**. This feature is often used to transform the data in order to meet an assumption of a **statistical test**. It can also be used to **collapse** several variables that measure the same concept into a **single variable**, thus increasing the reliability of the measure.

هذه الخاصية تسمح بتجهيز البيانات من أجل التحليل الإحصائية
 تسمح أيضا باختصار المتغيرات العديدة من متغير واحد ليبدو المفهوم

Exercise: Using the Epidemiology Books SPSS file to find out the following:
 مما يزيد من موضوع القياس

مجموعًا حساب جميع الحالات

- Compute the total score for all cases
- Compute the average for all cases **حساب متوسط الحالات**
- Recode the average to A, B, C, D & F **تحويل متوسط الحالات من الأعداد**

Computing New Variable (Cont.)

① Transform \rightarrow Compute

حساب المتوسط

Compute Variable

Target Variable: =

Numeric Expression:

Functions:

- ABS(numexpr)
- ANY(test,value,value,...)
- ARSIN(numexpr)
- ARTAN(numexpr)
- CDFNORM(zvalue)
- CDF.BERNOULLI(q,p)

Handwritten notes: *Sum(- و ->)* and *sum average*

Compute Variable

Target Variable: =

Numeric Expression:

Functions:

- LG10(numexpr)**
- LN(numexpr)
- LNGAMMA(numexpr)
- LOWER(strexp)
- LPAD(strexp,length)
- LPAD(strexp,length,char)

IF شرط

Select Cases (If, Sample & Range)

If Condition: کدیر شرط

① First, from the Data pull-down menu choose Select Case. Then select If condition is satisfied and click the If button. Click on the variable from the list on the left, then type a condition. Then click the Continue button, and finally click the OK button to select the cases.

Sample: اختیار عینہ کوانہ

① First, from the Data pull-down menu choose Select Case. Then select random sample of cases and click the sample button. Write the percentage you want to be selected in the sample. Then click the Continue button, and finally click the OK button to select the cases.

مردی

Range:

① First, from the Data pull-down menu choose Select Case. Then select based on time or case range and click the range button. Write the range you want to be selected. Then click the Continue button, and finally click the OK button to select the cases.

افتریا معتمدی کا زمن او مردی

کتاب مردی

Select Cases (If, Sample & Range) (Cont.)

بعد ظهور جدول Data يظهر علامه / كذا البيان - التي تم اختيارها
When you select cases, you will notice that in the Data editor window a slash mark / appears through the record number of those cases that were not selected. Here you will also notice that a new column labeled filter_\$, and containing 1 and 0 has appeared. These two values represent the selected and unselected cases, respectively.

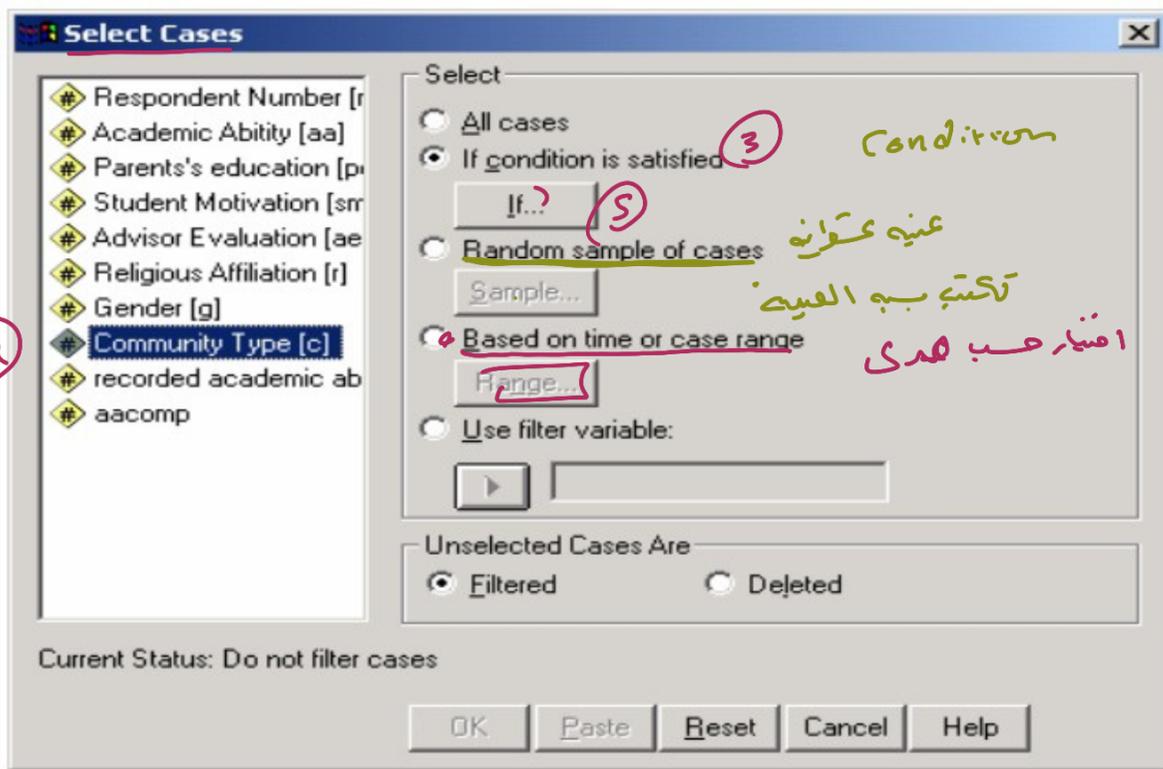
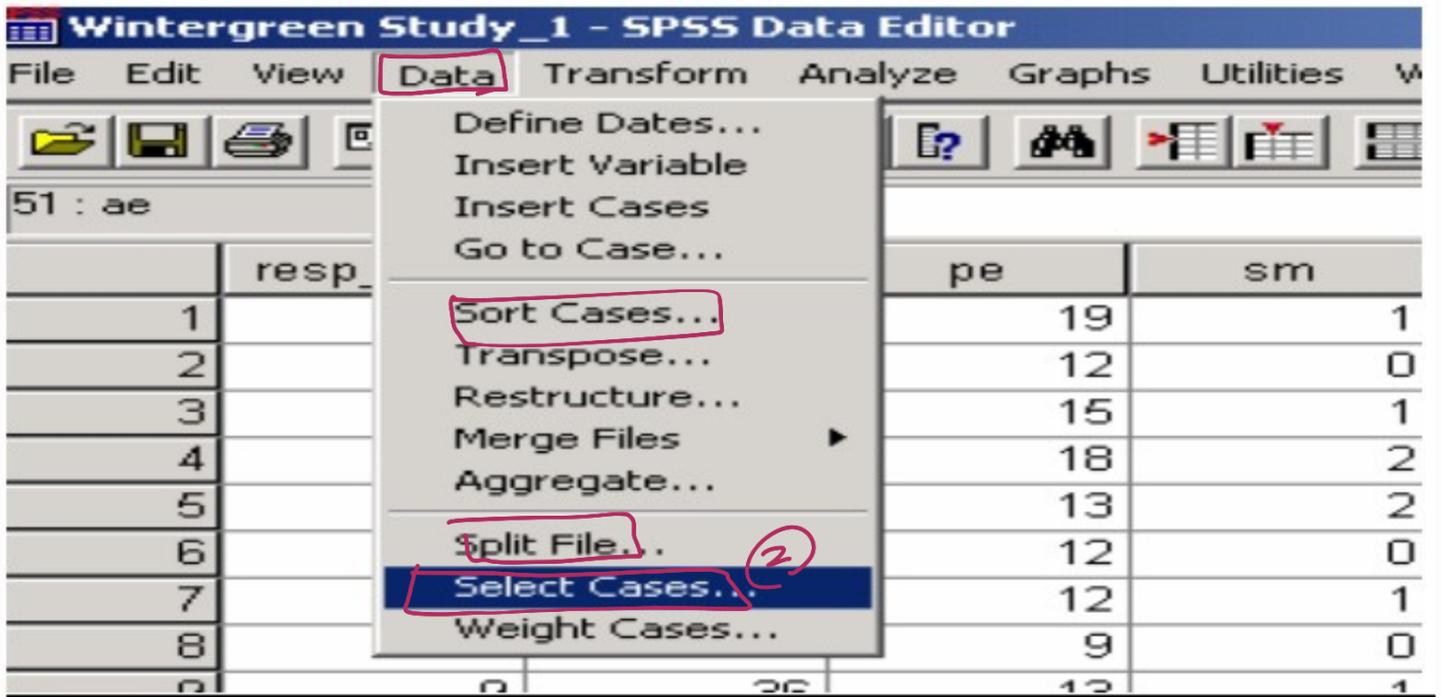
ويعبر عمود جديد Filter مكتوب بالقيم 1 و 0
1 مينا 0 غير مينا

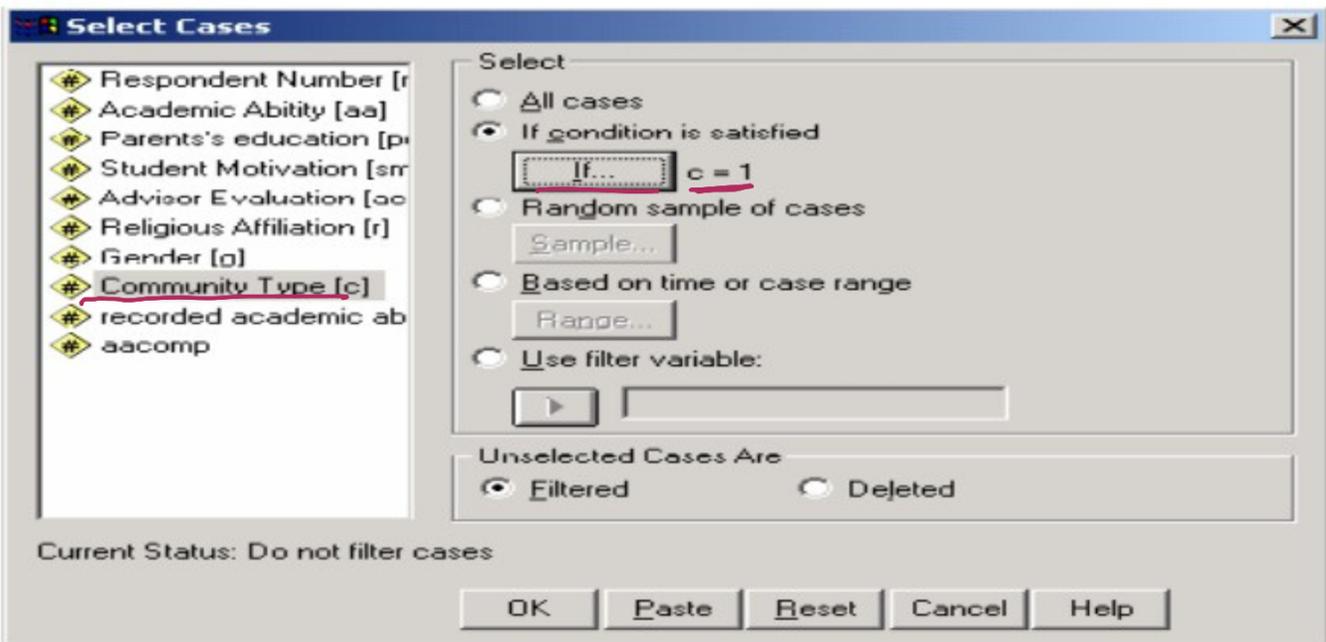
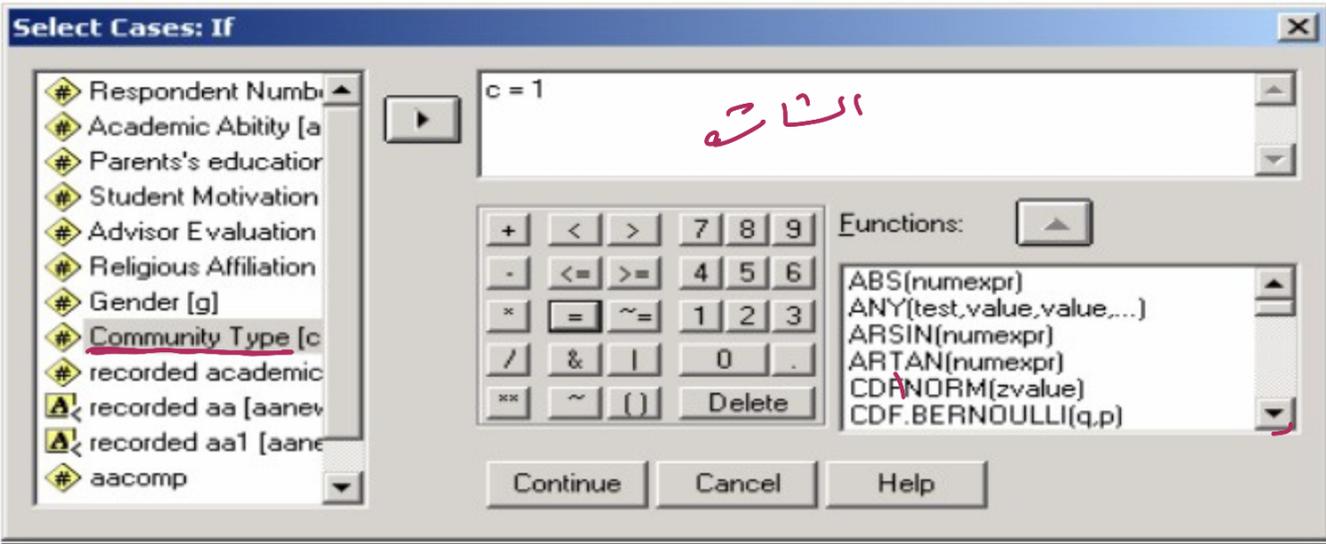
If you wish to include all the cases in later analysis, simply return to the Data pull-down menu and choose Select Cases dialog box and select All cases.

لدرجتي مره اخرى لجميع البيانات -
Data (1)
Select cases (2)
All cases (3)

Exercise: Using the Epidemiology Books SPSS file to find out the following

- Select students if the average is greater than 80 variable > 80
- Select students if the gender is female femar
- Select student ranged from 5 to 25
- Select random sample of cases (20%)





رابطه
ببین
مکنه

Wintergreen Study_1 - SPSS Data Editor

	aanow	aanow1	aanow2	aacomp	filter_\$	var	var	var	var	var
1	9.00	honor	more	9.40	1					
2	4.00	marginal	less	4.70	0					
3	5.00	pass	more	5.80	0					
4	9.00	honor	more	9.40	1					
5	8.00	honor	more	8.30	1					
6	5.00	pass	more	5.00	0					
7	6.00	merit	more	6.20	0					
8	2.00	fail	less	3.00	0					
9	3.00	fail	less	3.70	0					
10	9.00	honor	more	9.20	0					
11	5.00	pass	more	5.00	0					
12	5.00	pass	more	5.90	0					
13	6.00	merit	more	6.00	1					
14	7.00	honor	more	7.80	0					
15	7.00	honor	more	7.20	0					
16	8.00	honor	more	8.40	1					
17	9.00	honor	more	9.70	1					
18	8.00	honor	more	8.80	1					
19	6.00	merit	more	6.30	0					
20	5.00	pass	more	5.00	0					
21	4.00	marginal	less	4.70	0					
22	9.00	honor	more	9.20	0					

Microsoft Office

Microsoft

Start | week2 Microsoft Word | session6na Microsoft W... | Wintergreen Study_1 ... | 16:27

تصنيف

Sort Cases

احاده ترتيب المتغير والبيانات المرتبطة به ترتيب تصاعدي او تنازلي حسب المتغير محددًا

The Sort Function Sorting involves rearranging subjects and their associated data in order of increasing or decreasing values for a chosen variable. One may sort data for a variety of reasons including the need to easily find particular data points.

الترتيب سهل الوصول الى البيانات =

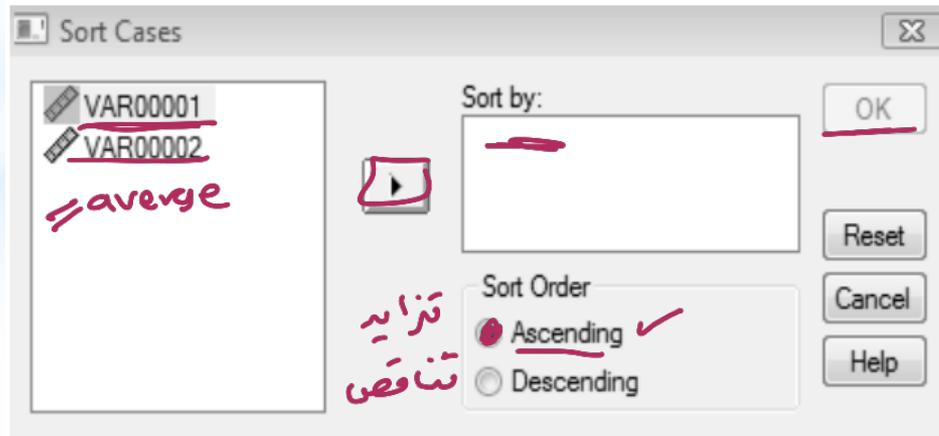
الترتيب

4

The process of sorting data in SPSS requires four steps:

1. Begin by selecting Data a pull-down menu should appear.
2. From the pull-down menu, select Sort Cases. A new window entitled Sort Cases should appear

ترتيب



Sort Cases (Cont.)

3. The user sorts' data by selecting the appropriate variable from those listed in the box above. The designation in the "Sort Order" portion of the window instructs SPSS whether to arrange the values in increasing or decreasing order.

4. Click "OK."

✓ بعد اتمام العملية تظهر نسخة من البيانات وبنوعه حسب المتغير
Upon completing this process, the Data View screen should reappear with data sorted according to the values of the variable specified.

Exercise: Using the Epidemiology Books SPSS file to find out the following:

- Sort the cases by the average from the lowest to the highest value
- Sort the cases by the total from the highest to the lowest value
- Sort the cases by serial *الرقم التسلسلي*

فصل الملفات

Split file

يوزر دسps تنفيذ الادامر
بتلك فنفسه متغير محدد

With SPSS's split file function enabled, SPSS executes all commands separately for each category of a particular variable. Researchers should use the following procedure to split a file.

1. Begin by selecting "Data" from the options at the top of the Data View or Variable View screen. A pull-down menu should appear.
2. From the pull-down menu, select "Split File" A new window entitled Split File should appear.



Split file (Cont.)

3. Select “Organize output by groups” from the options in the center of the Split Files window. The originally faded box marked “Groups Based on,” should become clearly visible.

4. An untitled box in the Split Files window contains the names of all variables for which data exists in the file. Indicate the variable that should serve as the basis for splitting the file by clicking on its name and clicking on the arrow to the right of the box. The name of the variable should move from its original position to the box marked “Groups Based on.”

5. Click OK. *بعد اختيار المتغيرات التي نريد فصلها نتم كـ OK وعندما نعود
الى الكمبيوتر الاصل لا نرى اي اثر لعرض البيانات ونظر فصل البيان
عند اصدار المخرجات تظهر البيانات ليتم تحليل منفصل من*

Upon returning to the Data View page, the user sees no immediate evidence of splitting a file, as he or she sees after sorting or filtering data. The results of having split a file do not become apparent until the user creates SPSS output, when he or she can see separate statistics or graphics for each identified category of subjects.

Exercise: Using the Epidemiology Books SPSS file to find out the following:

- Split the file by the gender.

Further study – Internet sites

http://www.gla.ac.uk/media/media_212500_en.pdf

<https://developer.ibm.com/predictiveanalytics/wp-content/uploads/sites/48/2015/04/Programming-and-Data-Management-for-IBM-SPSS-Statistics-23.pdf>

<http://www.spsstools.net/en/resources/spss-programming-book/>

http://www.uni-muenster.de/imperia/md/content/ziv/service/software/spss/handbuecher/englisch/spss_programming_and_data_management_4th_edition.pdf

