

**TEESSIDE UNIVERSITY
SCHOOL OF HEALTH & SOCIAL CARE**

SPSS Workbook 2 - Descriptive Statistics

Research, Audit and Data

RMH 2023-N

Module Leader: Sylvia Storey

Phone: 01642 0384969

s.storey@tees.ac.uk

SPSS– Workbook 2 – Descriptive Statistics

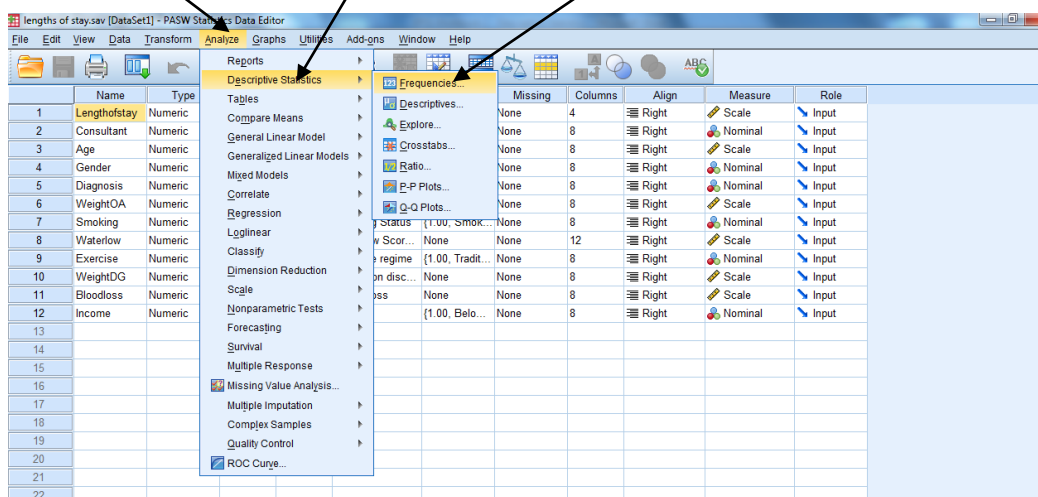
Accuracy of data input

Once you have entered your data into SPSS, you need to “clean-up” the data. This involves ensuring that data has been correctly entered. A good starting point for this is to run a frequency distribution for each of the variables eg:

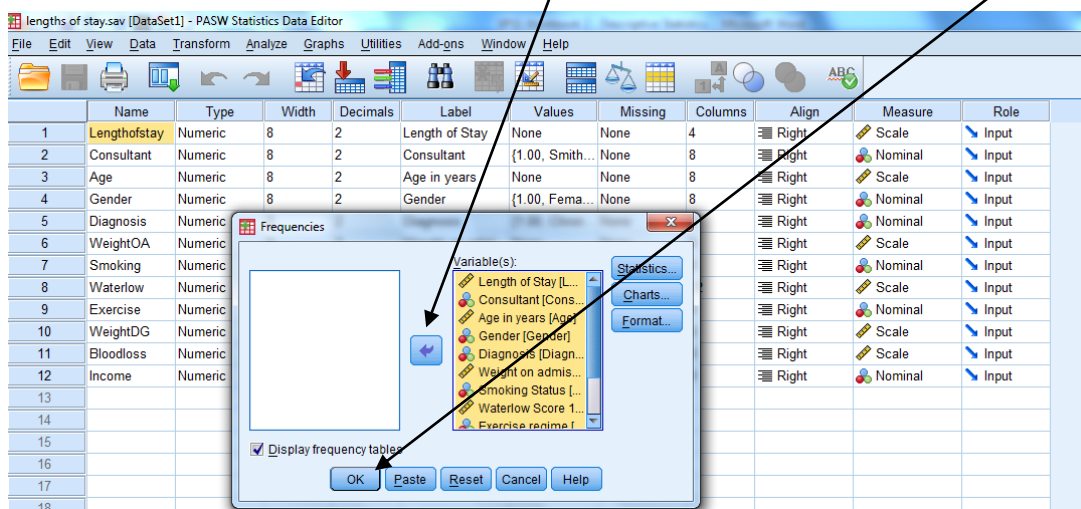
If you consider the variable “Gender” (1=female, 2=male) – then any other values entered would not fall into these categories eg 11, 12, 22. This is one of the most common mistakes when entering data into SPSS.

To check frequencies for all the variables :

Select **Analyse – Descriptive Statistics – Frequencies**



Move variables from the left hand column to the right hand column by highlighting the **variable name** and then clicking on the **arrow** to move the variable. Click on **OK** to finish.



When you have done this the frequencies will be displayed in a new output window. Look down the frequency tables and make a note of any variables that fall outside of the expected range:

Q1. Would this identify all mistakes? If not what other mistakes may be present?

Descriptive Statistics – Mean and Standard deviation

The following variables are all interval/ratio level data

Length of stay, Age, Weight OA, Blood Loss

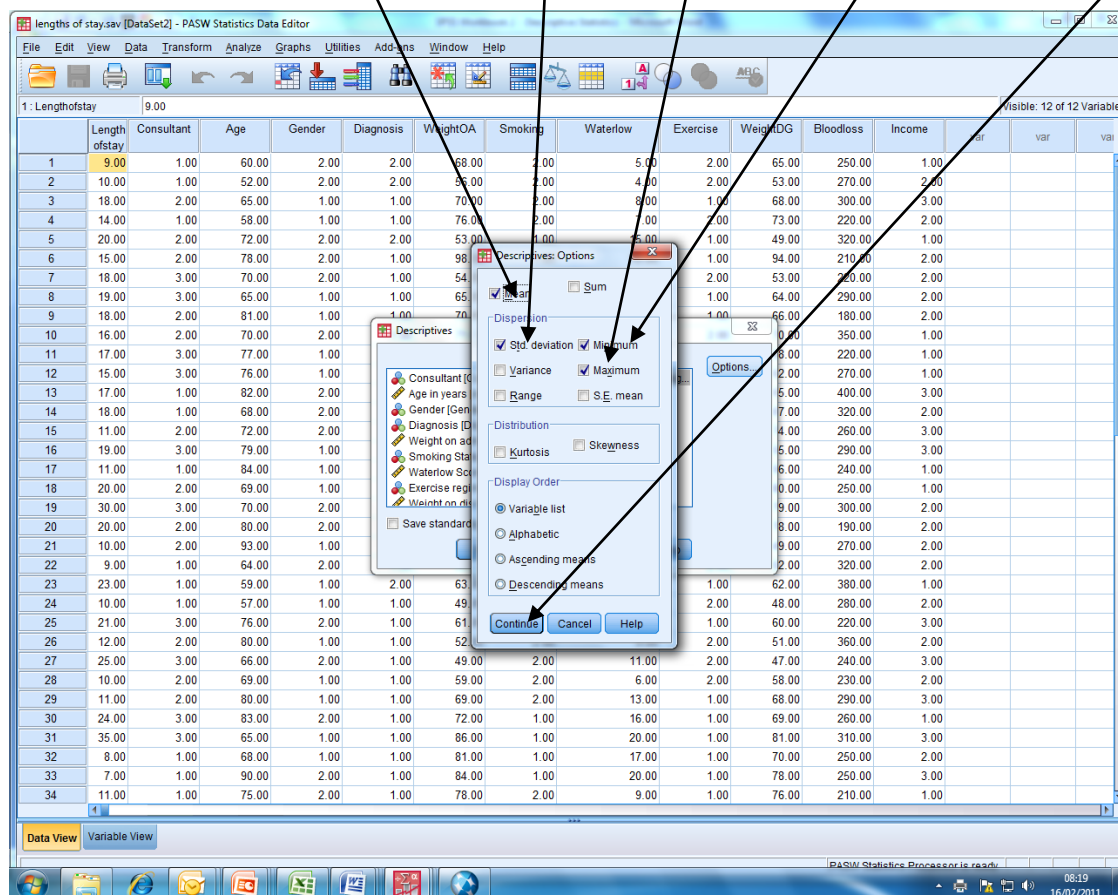
We are now going to obtain the following descriptive statistics:

Mean (measure of central tendency) and Standard deviation (SD = measure of dispersion)

This can be done in a number of ways but we will do this by selecting :

Analyse – Descriptive Statistics – Descriptives

Move the 4 variables into the right hand column and click on **Options**. Ensure that the following are selected **Mean, Standard deviation, Maximum, Minimum**. Now click **Continue** and then **OK** to finish.



Record the descriptive statistics for each variable:

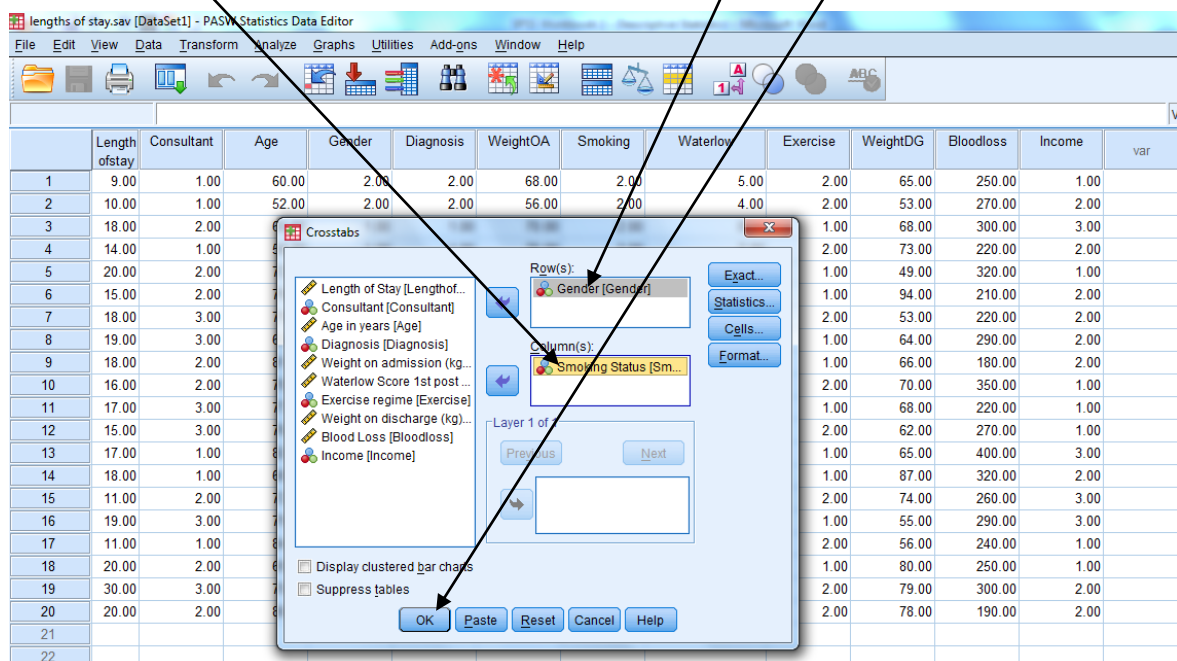
Table 1 – Descriptive Statistics

	Mean	SD	Min	Max
LengthofStay				
Age				
WeightOA				
BloodLoss				

So far we have looked at variables individually. This is often referred to to **univariate** analysis. We are now going to look at some **bivariate** analysis – ie looking at the interactions between 2 variables (remember we are still looking at descriptive statistics so we are not yet looking at cause and effect).

We want to know if there is a relationship between Gender and Smoking status. As both variables are **nominal** (what does this mean?) we will carry out a **Crosstabulation** by:

Selecting – **Analyse – Descriptives – Crosstabs**. Select **Gender** and move this to the row box and select **Smoking** and move this to the column box. Click on **OK** to finish.



Now complete the table below:

Table 2 : Crosstabulation Gender: Smoking Status

		Smoking Status	
		Smoker	Non-smoker
Gender	Male		
	Female		

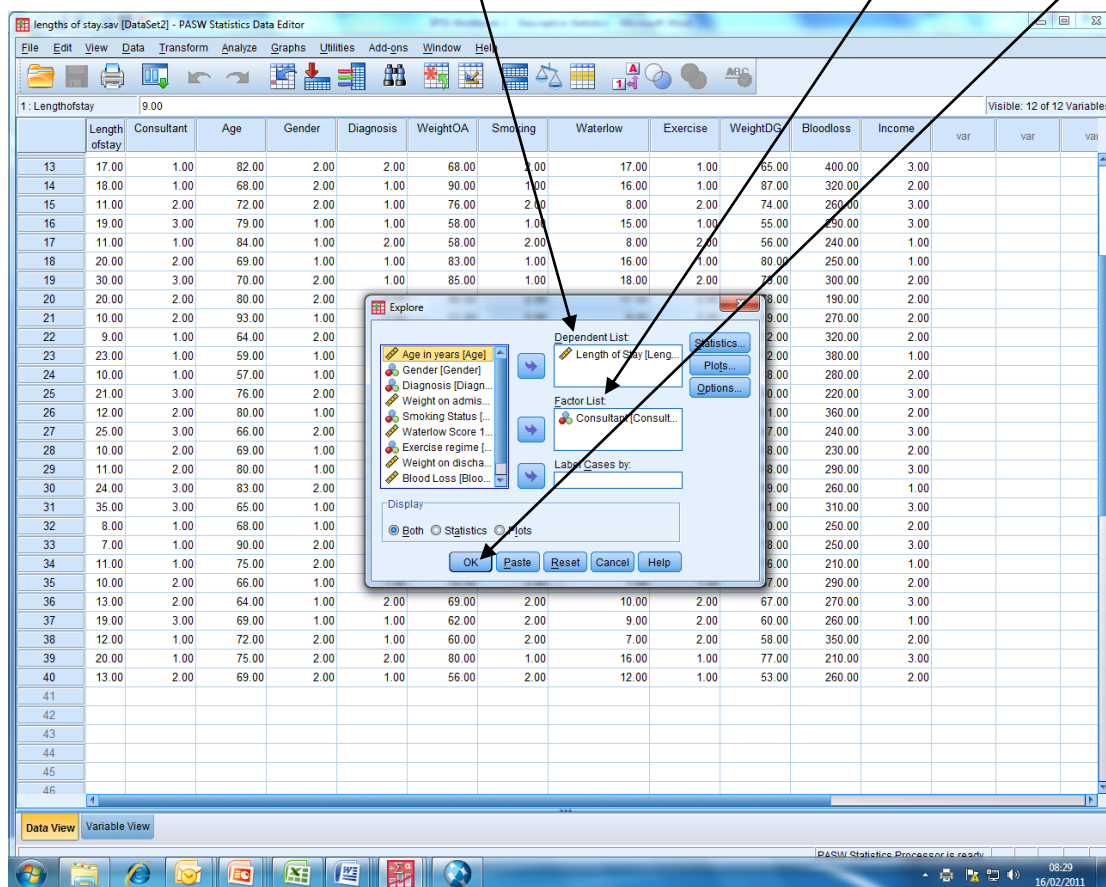
Q2. What do the results suggest?

The Crosstabulation we have just carried out looked at the relationship between 2 nominal variables (remember: **nominal** data is categorical data). We now want to see if there is a difference in Length of stay for patients under the care of different consultants. Length of stay is **ratio** level data (remember: this is continuous data).

We earlier checked to see what the mean “Lengthofstay” was for our sample and this was 16.75 days. We now want to look more closely and see if that was the same for all consultants.

Select - **Analyse – Descriptives – Explore.**

Move **Lengthofstay** into the **Dependent List** and **Consultant** into the **Factor list**. Click **OK** to continue.



Now complete the table below:

Table 3

		Length of Stay (Days)	
		Mean	SD
Consultant	Mr Smith		
	Mr Jones		
	Mr Wilder		

Q3. What do the results suggest?

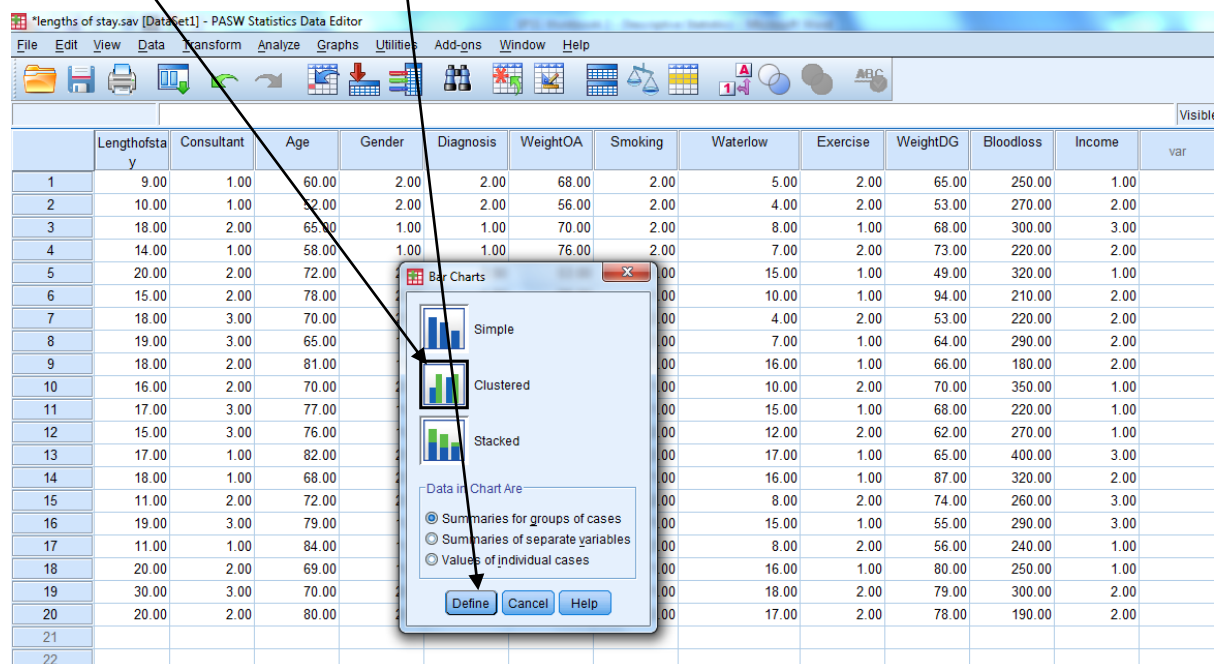
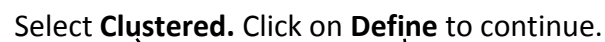
.....

.....

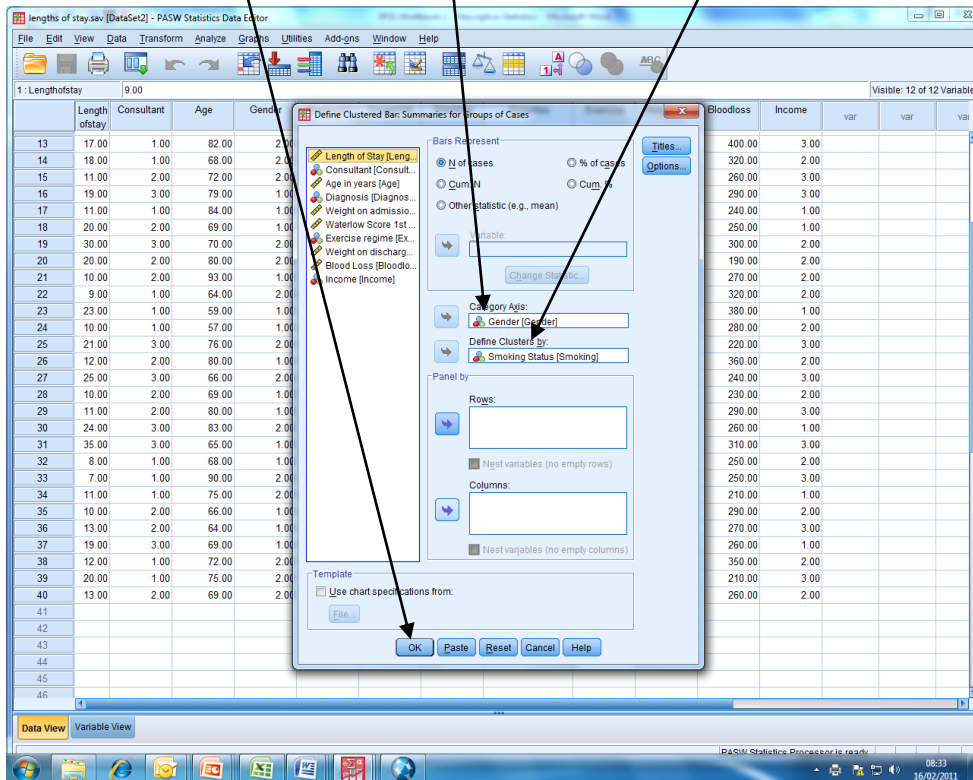
.....

Now think about how you would present each set of data in a graph – try to draw these below.

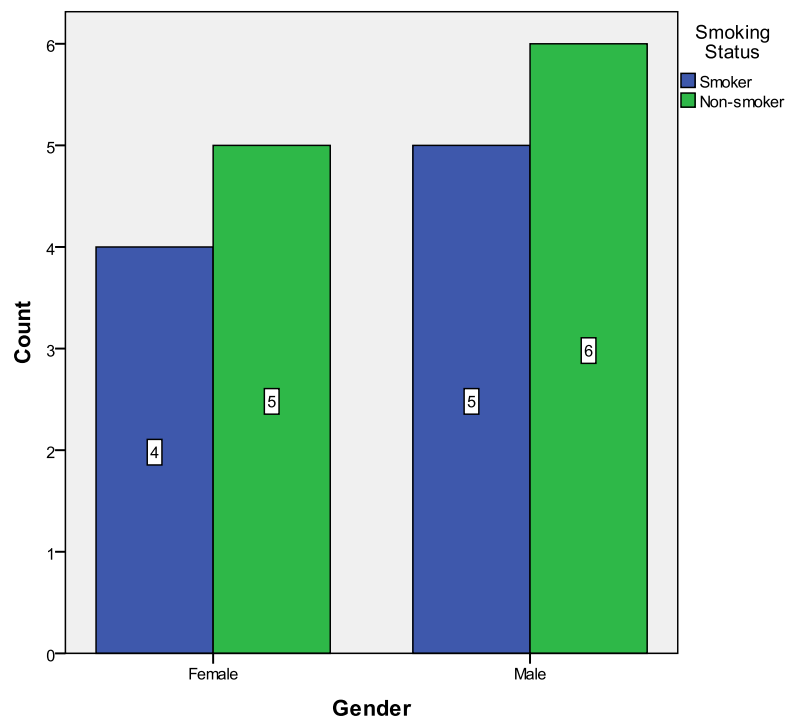
Clustered Bar Chart (Gender/Smoking Status) - ~~Graphs~~ – Legacy Dialogs – Bar



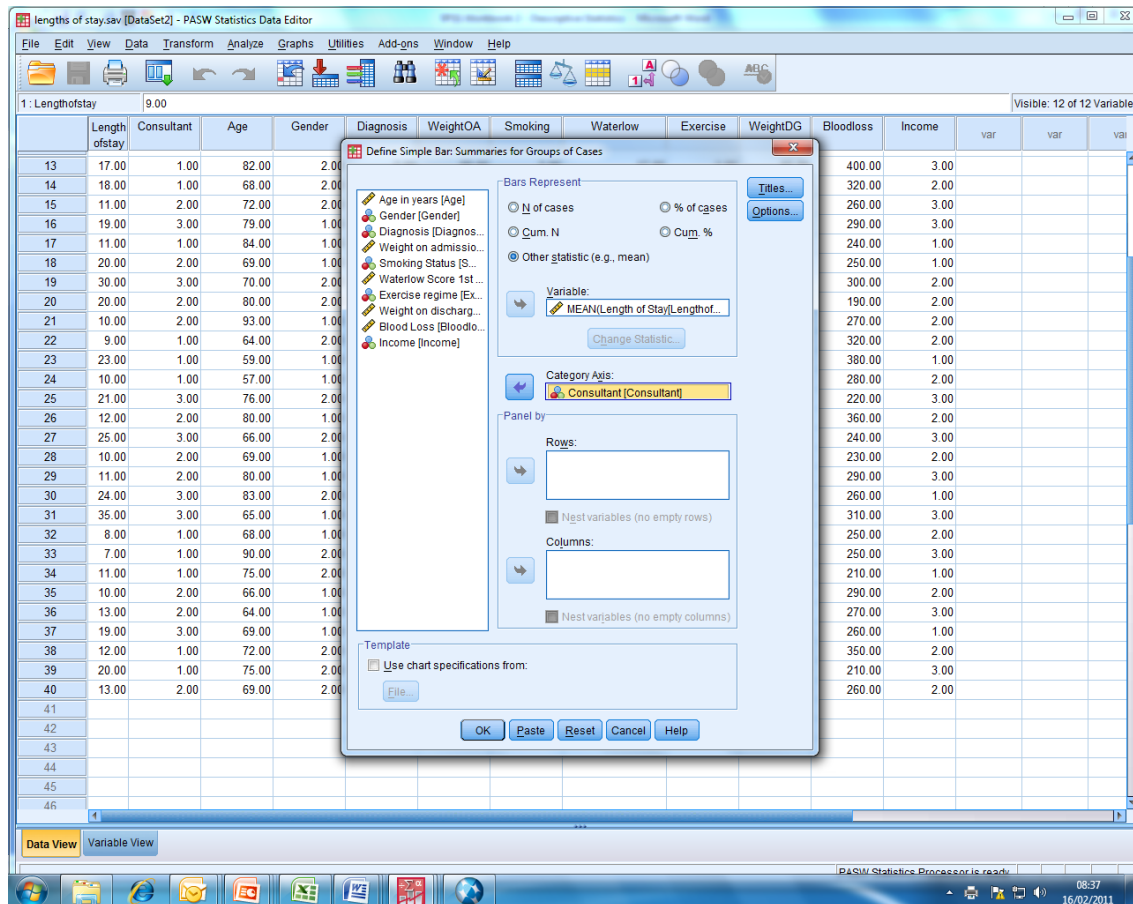
Now move **Gender** into the **Category Axis** box and **Smoking status** into the **Define Clusters by** box and click **OK** to continue, (See below)



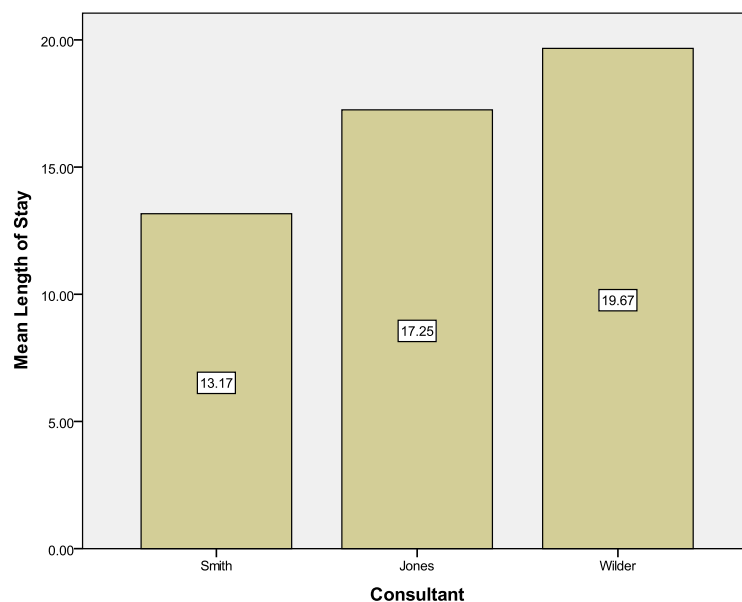
Does the graph below agree with the one you drew earlier?



Select - **Graphs – Legacy Dialogs – Bar Chart** and this time select **Simple** instead of **Clustered**.. Move variables as detailed below and select **OK** to continue.

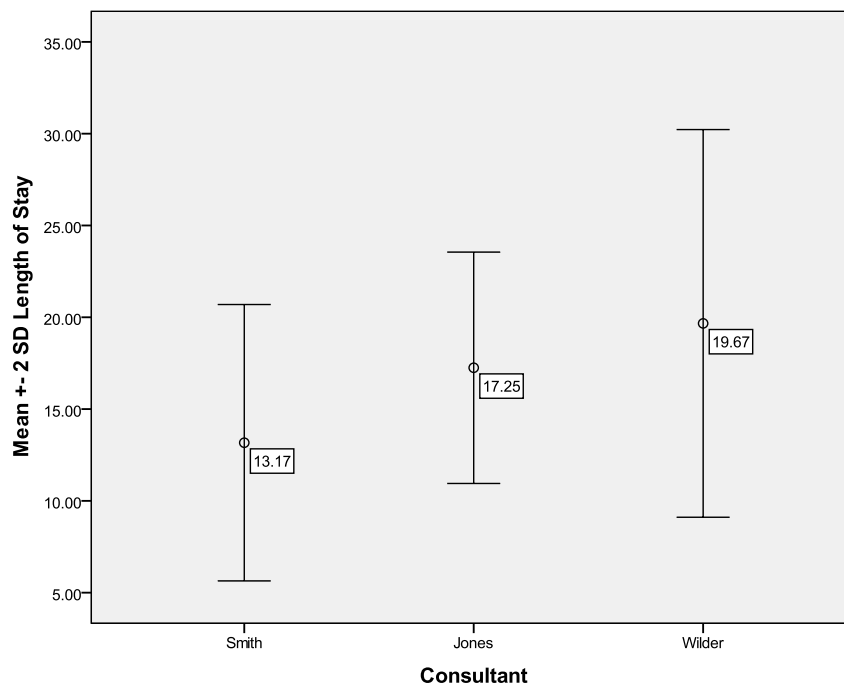
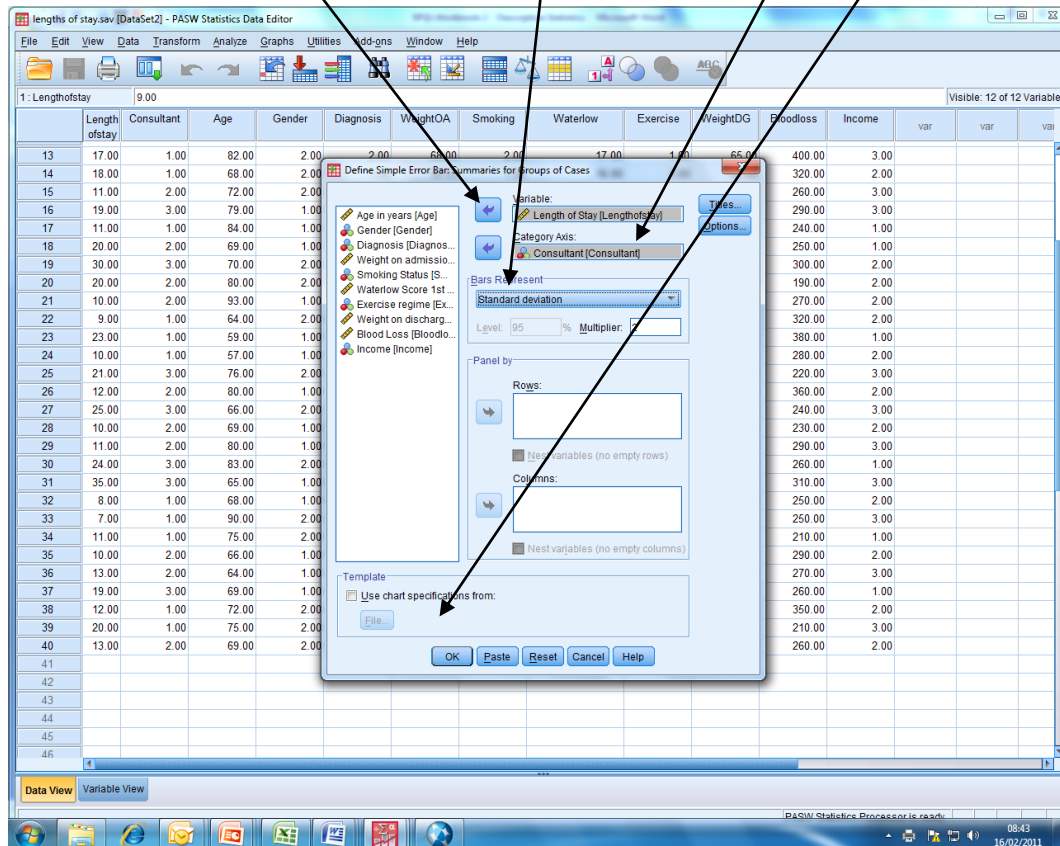


Does the graph below agree with the one you drew earlier?



Now try this last graph for the same data – it's called an Error Bar Chart and contains additional information.

Select – **Graphs – Legacy Dialogs – Error Bar** ensure **Simple** is highlighted and select **Define**. Move **LengthofStay** into the **Variable** box and **Consultant** in to the Category Axis. Change “Bars Represent” to show **Standard Deviation** and select **OK** to continue.



Q4. Compare the last 2 graphs – why is the last graph more suitable?

ANSWERS

Appendix 1 – Answers & Completed tables.

Q1. No – this would only identify mistakes where the input value falls outside of the expected range. If you entered someone's data as male instead of female (ie 1 instead of 2 in the case of our data-file) then you would not know you had done this unless you checked all data carefully. I would suggest that in your SPSS exam you do this to ensure that all your data is input correctly.

Table 1 : Descriptive Statistics

	Mean	SD	Min	Max
LengthofStay	16.75	4.644	9	30
Age	71.4	8.531	52	84
WeightOA	70.75	12.229	53	98
BloodLoss	267.5	55.334	180	400

Table 2 : Crosstabulation Gender: Smoking Status

		Smoking Status	
		Smoker	Non-smoker
Gender	Male	5	6
	Female	4	5

Q2. In terms of table 2 the data suggests that there is no difference in smoking status between men and women.

Table 3

		Length of Stay (Days)	
		Mean	SD
Consultant	Mr Smith	13.1667	3.76386
	Mr Jones	17.25	3.15096
	Mr Wilder	19.6667	5.27889

Q3. The results suggest that patients under the care of Mr Smith typically leave hospital earlier than those under the care of other consultants and that patients under the care of Mr Wilder stay longer than other patients.

Q4. The Error-bar also shows the spread of scores (ie we set this as 2 standard deviations – see this down the y-axis of the graph).

