

# **IBM SPSS Statistics Set 2**

## **Section 2 Some Basic Statistical Procedures in SPSS**

**2.1 PRODUCING FREQUENCIES AND DESCRIPTIVE STATISTICS**

**2.2 EXAMINING THE DATA IN MORE DETAIL USING EXPLORE**

**2.3 COMPARING MEANS**

**2.4 WORKING WITH CATEGORICAL VARIABLES**



<b>2. 1 PRODUCING FREQUENCIES AND DESCRIPTIVE STATISTICS .....</b>	<b>5</b>
<b>2.1.1. Introduction</b>	<b>5</b>
<b>2.1.2. The basics of the FREQUENCIES procedure</b>	<b>5</b>
<b>2.1.3. Producing barcharts using the FREQUENCIES procedure</b>	<b>6</b>
<b>2.1.4. Summary statistics using the FREQUENCIES procedure</b>	<b>8</b>
<b>2.1.5. Producing histograms using the FREQUENCIES procedure</b>	<b>10</b>
<b>2.1.6. FREQUENCIES syntax (Optional)</b>	<b>12</b>
<b>2.1.7. Basics of the DESCRIPTIVES procedure</b>	<b>16</b>
<b>2.1.8. DESCRIPTIVES Options</b>	<b>18</b>
<b>2.1.9. The REPORT procedure</b>	<b>22</b>
<b>2.1.10 One-click descriptive statistics.</b>	<b>28</b>



## 2. 1 PRODUCING FREQUENCIES AND DESCRIPTIVE STATISTICS

### 2.1.1. Introduction

In this section you will be introduced to the two most commonly used procedures for obtaining univariate descriptive statistics: the FREQUENCIES procedure and the DESCRIPTIVES procedure.

The FREQUENCIES procedure is used to present data as frequency distributions where the data consists of a relatively small number of categories. The level of measurement for the data may be *nominal (categorical)*, *ordinal* or *interval*, but in the latter the data may need to be grouped into a small number of classes before using FREQUENCIES to produce a useful table that does not contain too many categories. This method of grouping is called *recoding* the data which will be covered later.

The FREQUENCIES procedure can also be used for computing a number of statistics for interval level data such as the mean, standard error, and skewness. You can also use this procedure to obtain graphs such as barcharts and histograms.

The DESCRIPTIVES procedure is similar to the FREQUENCIES procedure in many ways, although it cannot be used to produce graphs. You can use DESCRIPTIVES to obtain means, standard errors, inter-quartile ranges and so on about a variable. You cannot use this procedure to obtain FREQUENCY distributions. However it is useful for calculating standardised (*Z*) scores for interval variables.

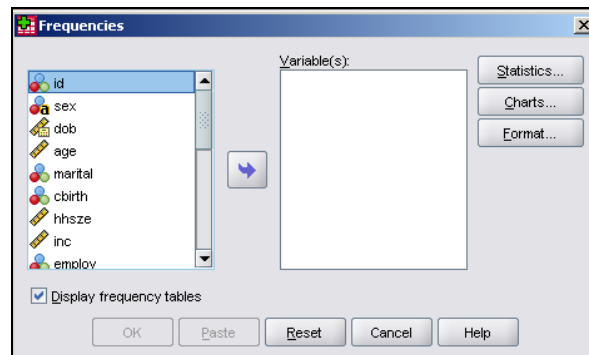
**Note:** *In the computer exercises which follow you may use the file from your disk QUESTSS.SAV used in section 1.3.*

### 2.1.2. The basics of the FREQUENCIES procedure

In its simplest form the FREQUENCIES procedure produces tables as shown in Computer exercise 1. You have already done some work with Frequencies, so if you are happy with the basics, you might like to move onto Section 2.1.4

#### ***Computer exercise 1: Some Basic Frequencies***

(a) *With the file QUESTSS.SAV open, obtain the Frequencies dialog box as below. Then obtain frequency tables for the variable MARITAL and two other categorical variables.*



The frequency table for the variable MARITAL is shown below.

		Marital status			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Never married	21	35.0	38.2	38.2
	Married	23	38.3	41.8	80.0
	Divorced	7	11.7	12.7	92.7
	Other	4	6.7	7.3	100.0
	Total	55	91.7	100.0	
Missing	Missing	5	8.3		
Total		60	100.0		

(a) Explain the difference between the Percent and Valid Percent values.

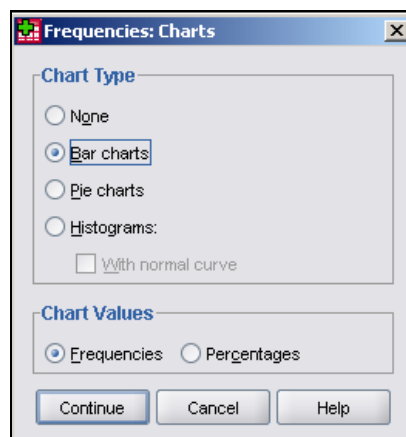
### 2.1.3. Producing barcharts using the FREQUENCIES procedure

As you have already seen, basic output from FREQUENCIES can be obtained by naming the variables in the dialog box. Extra features are available in this procedure by clicking on one or more of the following:



Any features which are selected apply to all variables specified in the **Frequencies** dialog box. If you want different forms of output for different variables then separate FREQUENCIES jobs must be run. For example, for categorical variables such as marital and country of birth you may want to obtain the modes and barcharts while for continuous data such as age and income you may wish to obtain means and histograms.

Using the **Charts** option, the dialog box below is obtained.



This gives you the choice of no graphs, barcharts, pie charts or histograms. Your selection will depend on the data you are analysing. You can display frequencies or percentages on Barcharts and Pie charts and a normal curve with Histograms.

### Computer exercise 2: Some basic frequencies and barcharts

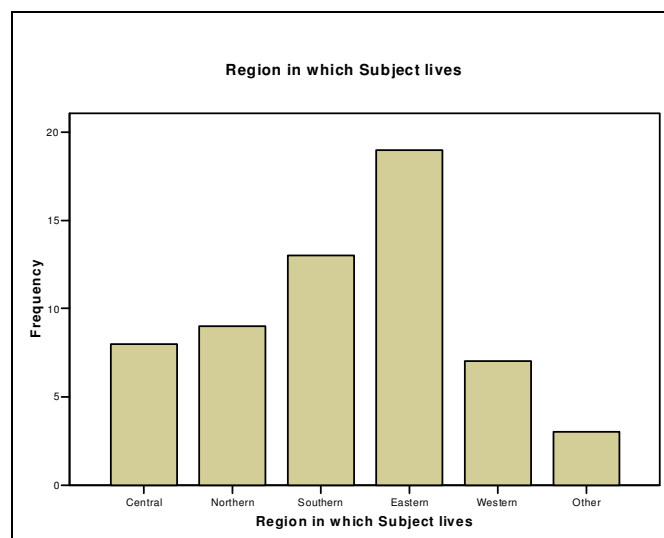
With the *QUESTSS.SAV* file open do the following:

(a) Click on **Analyze, Descriptive Statistics, Frequencies**

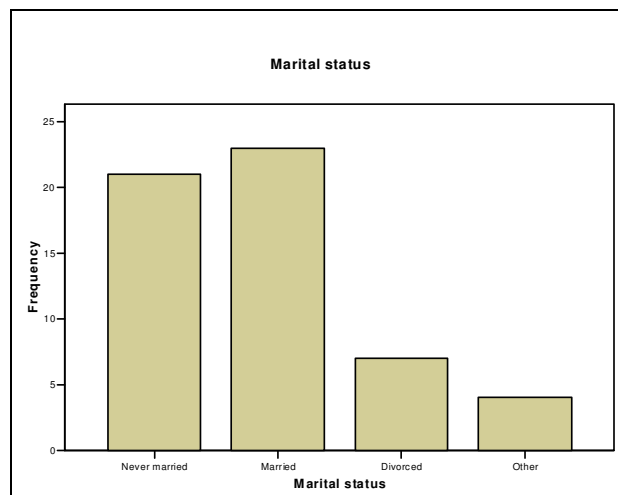
Select the variables *REGION* and *MARITAL* then click on **OK**. Have a look at the output. Then clear the output window.

(b) Repeat (a) but this time also click on **Charts** and choose a **Bar chart**, press **Continue** then **OK**. You should have two barcharts, one for *MARITAL* and another for *REGION* which will look like this.

#### Bar chart for REGION



#### Bar chart for MARITAL



Clear the output window

(c) Repeat (b) but this time obtain percentages rather than frequencies on the vertical axis.

(d) *RESET* the frequencies then repeat (a) for the variables *MARITAL* and *REGION* with no statistics or charts but select **Formats** and choose **Ascending counts**, then **Continue** and **OK**.

*Inspect the output to see the effect of these commands.*

#### 2.1.4. Summary statistics using the FREQUENCIES procedure

There are a number of statistics which can be obtained using the FREQUENCIES procedure. These include, among others, the mean, median, mode, quartiles, standard deviation and skewness. It is important to know in what situations it is appropriate to use these measures. For example a mode would be appropriate for categorical data, whereas a mean or standard deviation would be better suited for an interval level variable.

**Note:** For continuous scaled data it is usually sensible to suppress the frequency tables since too much output is usually generated.

#### *Computer exercise 3: Using Statistics with frequencies*

(a) Select **Analyze, Descriptive Statistics, Frequencies** again. *RESET* then select the continuous variable *AGE*. Check that the frequencies are not to be displayed by clicking in the  **Display frequency tables** box so that the little square box at the left is blank.

Click on **Statistics** and look at what statistics are available. Select the mean, median, standard deviation and skewness and continue.

Click **OK**. The results should look like this.

Statistics		
Age of Subject		
N	Valid	58
	Missing	2
Mean		32.43
Median		31.50
Std. Deviation		8.046
Skewness		.478
Std. Error of Skewness		.314

*Comment briefly on the results.*

**Note:** *If skewness is greater than two standard errors of skewness, the skewness it can be regarded as significant.*

*Clear the output window.*



(b) *RESET the Frequencies dialog box and obtain only the maximum, minimum and kurtosis for AGE.*

**Statistics**

Age of Subject

N	Valid	58
	Missing	2
Kurtosis		-.249
Std. Error of Kurtosis		.618
Minimum		19
Maximum		53

To see what the measures of Skewness and Kurtosis are you can look at websites like:  
<http://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm>

(c) *RESET the Frequencies dialog box again and obtain the 25<sup>th</sup>, 50th and 75th percentile for AGE. You should obtain output that looks like this.*

**Statistics**

Age of Subject

N	Valid	58
	Missing	2
Percentiles	35	28.65
	50	31.50
	75	37.25

*Explain what the 35th percentile tells you.*

---

(d) *RESET the Frequencies dialog box again and divide the distribution of AGE into 10 equal sized groups.*

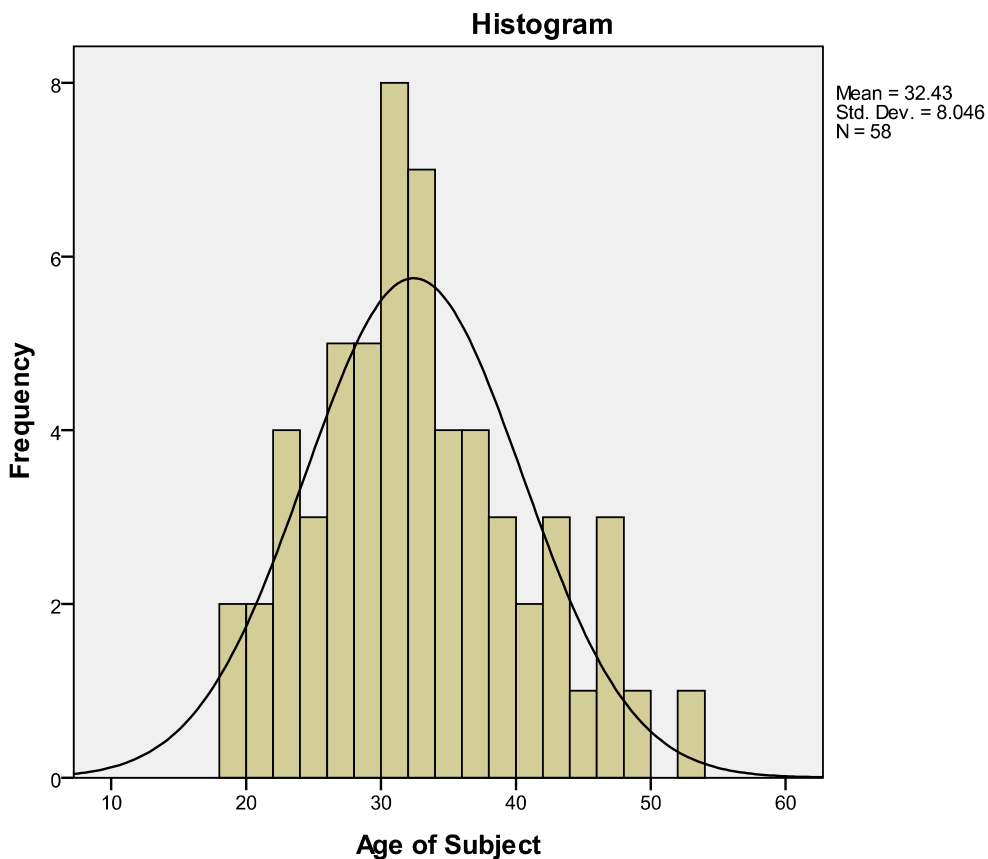
\* (e) Use the file GSS91T.SAV to investigate the Skew and Kurtosis of the variables Age and Educ

### 2.1.5. Producing histograms using the FREQUENCIES procedure

When the variable being analysed is of a *continuous* nature rather than categorical, it is more appropriate to graph it using a *histogram* rather than a barchart. The option to draw a curve on the histogram enables you to get a visual impression of the normality of the data.

#### *Computer exercise 4: Histograms with frequencies*

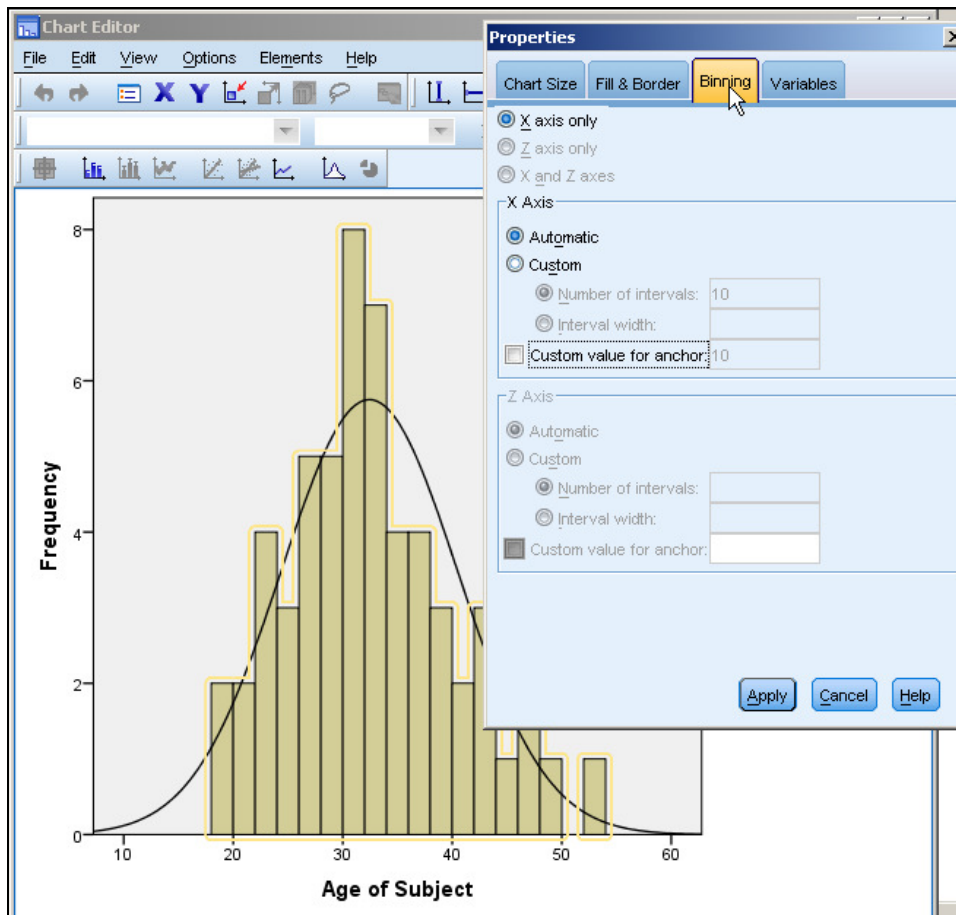
(a) *RESET* and obtain a histogram with a normal curve for AGE. Do not display the frequency table. View the graph as before. It should look similar to the chart below.



*Clear the output window.*

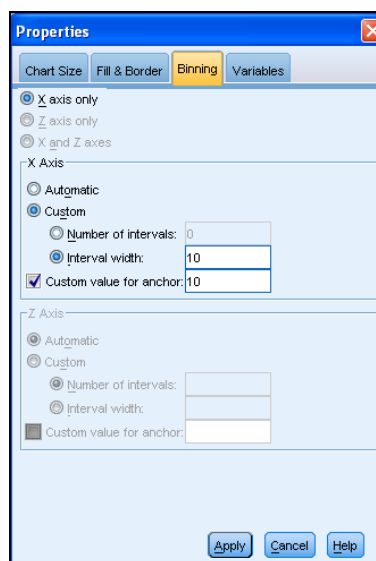
#### **Changing the width of the columns:**

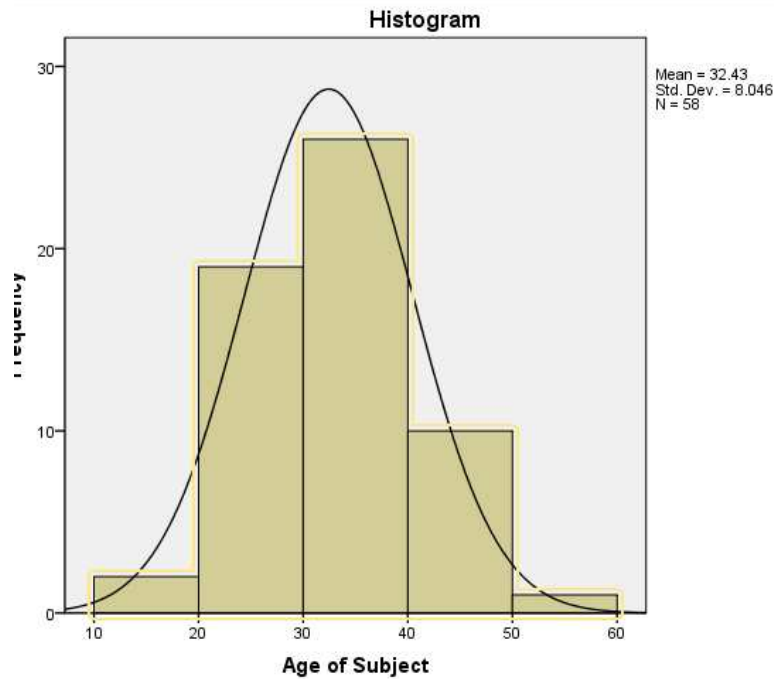
The width of the columns can be changed in the **Chart Editor** by right clicking in any column, selecting the Properties Window, eg using the right side mouse button, then with the right mouse button choose the **Binning Tab**.



Then select the required axis values. For example choose a beginning point, Anchor, and or an Interval width or number of intervals.

(b) Try this with an interval width of 10, and anchor at 10.





(c) Repeat (b) using income as the variable and choose appropriate intervals.

(d) Investigate some of the other features found in the Properties.

### 2.1.6. FREQUENCIES syntax (Optional)

Syntax rather than dialog boxes can also be used to perform the functions of the FREQUENCIES command. For example, rather than using the X axis properties menu in the Chart Editor to specify your minimum and maximum ranges of the histogram, syntax can be written to perform exactly the same process.

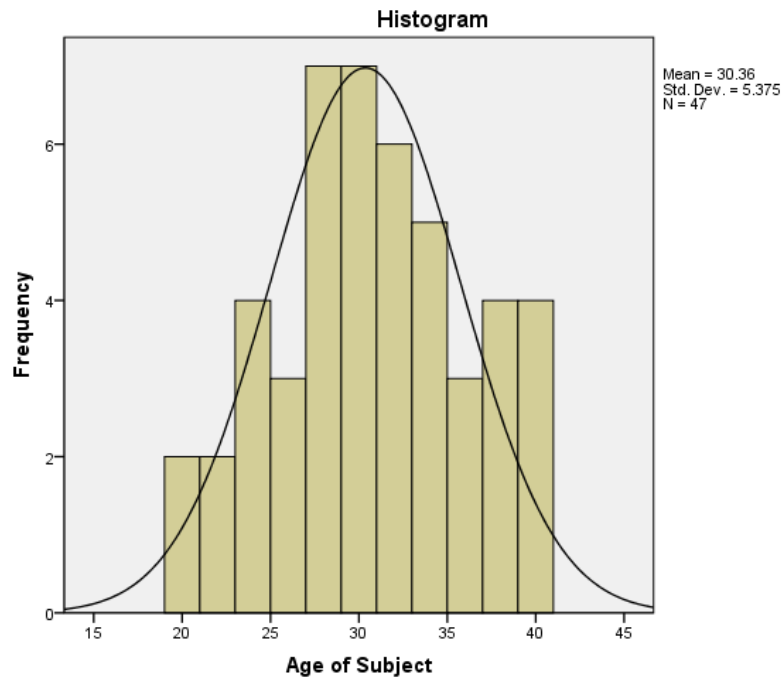
Firstly, if you **paste** the syntax used to obtain the first histogram in Computer exercise 4 you will see

```
FREQUENCIES VARIABLES=age
  /FORMAT=NOTABLE
  /HISTOGRAM NORMAL
  /ORDER=ANALYSIS.
```

To limit the range of the histogram we could modify the syntax so that only ages from 20 to 40 are graphed.

```
FREQUENCIES
  /FORMAT=NOTABLE
  /VARIABLES=age
  /HISTOGRAM MIN(20) MAX(40) NORMAL
  /ORDER ANALYSIS.
```

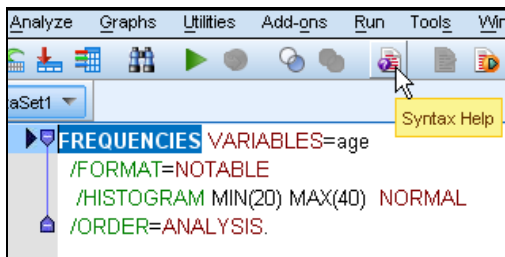
Highlight the syntax then run the job by clicking the RUN key,  to graph as below.



Note: The graph extends from just under 20 to just over 40 due the way the bins are formed. You would need to use the Columns Binning menu to show exactly.

### FREQUENCIES Syntax help (Optional)

By highlighting the word Frequencies in the Syntax window, and clicking on the Syntax help button, you see all the options available in Frequencies syntax.



```

FREQUENCIES VARIABLES=varlist [varlist...]

  [/FORMAT= [{NOTABLE }] [{AVALUE**}]
             {LIMIT(n)}  {DVALUE  }
                       {AFREQ   }
                       {DFREQ   }

  [/MISSING=INCLUDE]
  [/BARChart=[MINIMUM(n)] [MAXIMUM(n)] [{FREQ(n)  }]]
                                     {PERCENT(n) }
  [/PIEChart=[MINIMUM(n)] [MAXIMUM(n)] [{FREQ  }]] [{MISSING }]]
                                     {PERCENT}  {NOMISSING}
  [/HISTOGRAM=[MINIMUM(n)] [MAXIMUM(n)] [{FREQ(n)  }]] [{NONORMAL}]]
                                                     {NORMAL  }

  [/GROUPED=varlist [{(width)          }]]
                  {(boundary list)}

  [/NTILES=n]
  [/PERCENTILES=value list]
  [/STATISTICS=[DEFAULT] [MEAN] [STDDEV] [MINIMUM] [MAXIMUM]
               [SEMEAN] [VARIANCE] [SKEWNESS] [SESKEW] [RANGE]
               [MODE] [KURTOSIS] [SEKURT] [MEDIAN] [SUM] [ALL]
               [NONE]]
  [/ORDER=[{ANALYSIS}] [{VARIABLE}]]

```

How this works is best seen in the examples given.

**Computer exercise 5: Use of syntax with FREQUENCIES**

(a) Carry out the task above.

(b) Repeat (a) using INCOME as the variable and choose a restricted range to graph.

**Further examples using the Japanese files (Optional)**

**Files in Japan folder**

**Computer exercise 6:**

With the PERSBGC.SAV file open do the following:

(a) Click on **Analyze, Descriptive Statistics, Frequencies**

Select the variables XJPNEXP and XJPNREAS then click on **OK**. Have a look at the output. Then clear the output window.

(b) Repeat (a) but this time also click on **Charts** and choose a **Bar chart**, press **Continue** then **OK**. You should have two barcharts

Clear the output window.

(c) Repeat (b) but this time obtain percentages rather than frequencies on the vertical axis.

(d) **RESET** the frequencies then repeat (a) for the variables XJPNEXP and XJPNREAS with no statistics or charts but select **Formats** and choose **Ascending counts** then **OK**.

Inspect the output to see the effect of these commands.

**Computer exercise 7: Using Statistics with frequencies**

(a) With the PERSBGC.SAV file open select **Analyze, Descriptive Statistics, Frequencies**, again. **RESET** then select the continuous variable XBTHYR. Check that the frequencies are not to be displayed

Obtain the mean, median, standard deviation and minimum and maximum and continue.

Click **OK**. The results should look like this.

Statistics		
QE.1 birth year		
N	Valid	282
	Missing	202
Mean		75.38
Median		77.00
Std. Deviation		5.76
Minimum		9
Maximum		79

*Comment briefly on the results.*

---

*Note the minimum value of 9 is an unlikely birth year. We will assume it is a missing value so set it as a missing value.*

*Clear the output window.*

*(b) RESET the Frequencies dialog box and obtain only the maximum, minimum and kurtosis for AGE.*

*(c) RESET the Frequencies dialog box again and obtain the quartiles, the 35th and 85th percentile for AGE. You should obtain output that looks like this.*

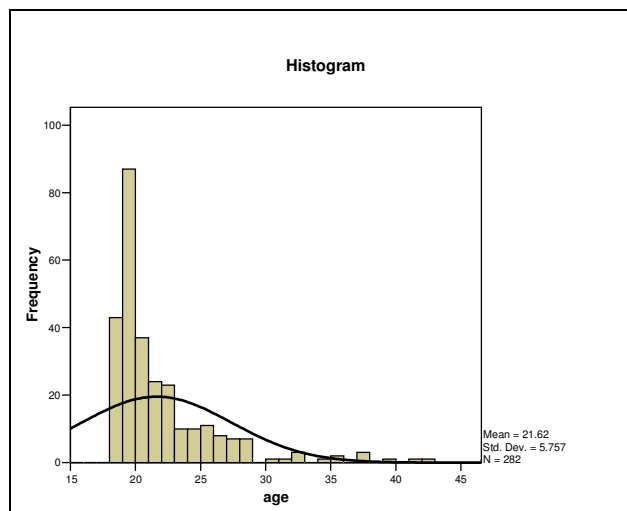
**Statistics**

AGE		
N	Valid	282
	Missing	202
Percentiles	25	19,00
	35	19,00
	50	20,00
	75	22,00
	85	25,00

*Explain what the 35th percentile tells you*

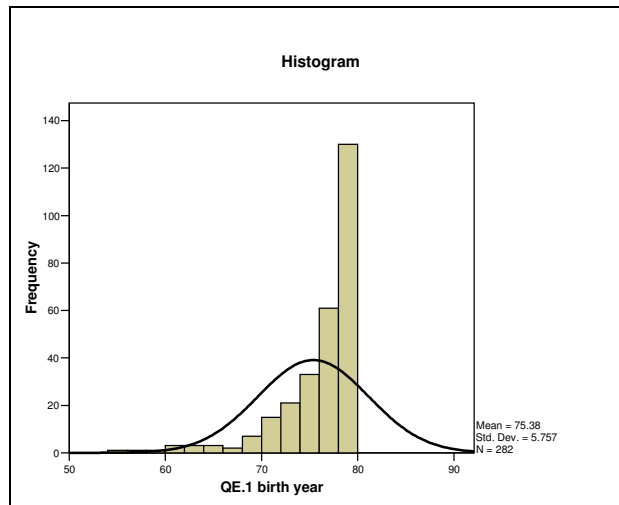
---

*(d) Obtain a histogram with interval width of 1, and the range displayed from 15 to 45.*



*Comment on what you expect the direction of the skewness and kurtosis would be. Check this.*

(e) Produce a histogram for birth year from 50 to 90 with intervals of 2 years as shown below.



Comment on what you expect the direction of the skewness and kurtosis would be. Check this.

### 2.1.7. Basics of the DESCRIPTIVES procedure

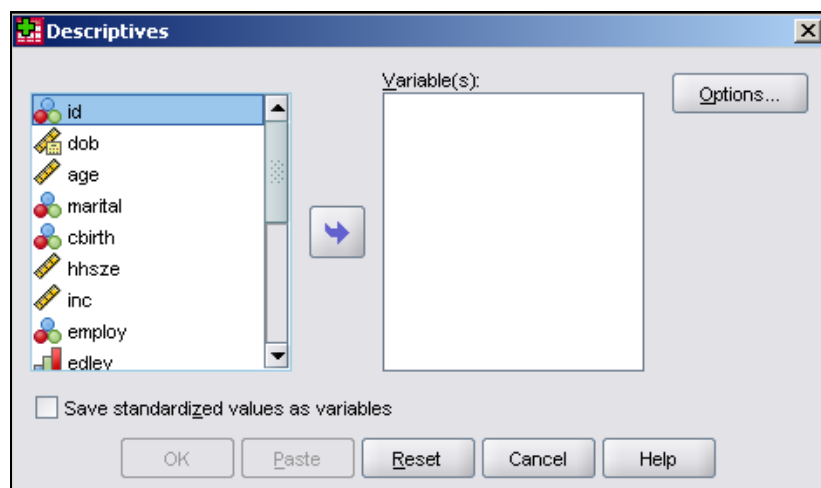
This procedure provides a number of descriptive statistics such as the mean and standard deviation for variables which are measured on a metric scale. These may be interval or ratio level data.

When you are using the DESCRIPTIVES procedure through the dialog boxes non-numeric variables will not be included on the list. Hence SPSS will not determine the descriptive statistics for any variable which has a non-numeric type.

To access DESCRIPTIVES click on the following:

**Analyze,**  
**Descriptive Statistics,**  
**Descriptives**

The DESCRIPTIVES dialog box is shown below. Select the variables required. The example below is based on the QUESTSS.SAV dataset.





**Computer exercise 8: Obtaining basic descriptive statistics**

- (a) (i) Use the *QUESTSS.SAV* file try to obtain descriptive statistics for the variables *SEX*, *AGE*, *MARITAL* and *INC*.

Why is there a problem obtaining statistics for some of the variables?

---

- (ii) Use the questionnaire file to obtain descriptive statistics for *ALL* the variables which appear in the dialog box.

What does the mean for Rugby tell you?

---

- (b) Obtain the same *DESCRIPTIVES* for *AGE* and *INC*.

Click on **Save standardized values as variables** to obtain Z scores for *AGE* and *INC*.

A **Z score** indicates how many standard deviations the score is above or below the mean. It is calculated as follows:

**Z score** = (score-mean)/standard deviation.

For the variable *AGE*, write down the

mean \_\_\_\_\_

and the *S.D.* \_\_\_\_\_

Sort by *ID*. Now have a look at the last variables in the **Data Editor window** to see the newly created variables which were added to your file. *Zage* corresponds to the standardised variable for *AGE* and *Zinc* corresponds to the standardised variable for *INC*. You can rename these variables in the *Data Editor* if you wish.

<b>Zage</b>	<b>Zinc</b>
<b>-0.03136</b>	<b>-1.13476</b>
<b>-0.27386</b>	<b>-1.04143</b>
<b>2.02990</b>	<b>2.83173</b>
<b>1.78740</b>	<b>2.59841</b>
.	<b>-1.27475</b>
.	<b>-0.66811</b>
<b>2.51490</b>	.
<b>0.81739</b>	.
<b>-0.51636</b>	<b>-0.10814</b>
<b>-0.27386</b>	<b>-1.06476</b>

**Note:** If these Z variables already exist SPSS will allocate different names for the new saved variables.

Use an appropriate procedure (**Analyze, Reports, Case summaries**) to look at all the values of AGE and ZAGE for the first 10 cases. Some of the output is shown below:

**Case Summaries<sup>a</sup>**

	Respondents identification number	Age of Subject	Zscore: Age of Subject
1	1	32	-.03136
2	2	30	-.27386
3	3	49	2.02990
4	4	47	1.78740
5	5	-1	.
6	6	-1	.
7	7	53	2.51490
8	8	39	.81739
9	9	28	-.51636
10	10	30	-.27386
Total	N	10	8

a. Limited to first 10 cases.

Note that age of 47 has a Z score of 1.78740 This tells us that this age is about 1.8 standard deviations above the mean.

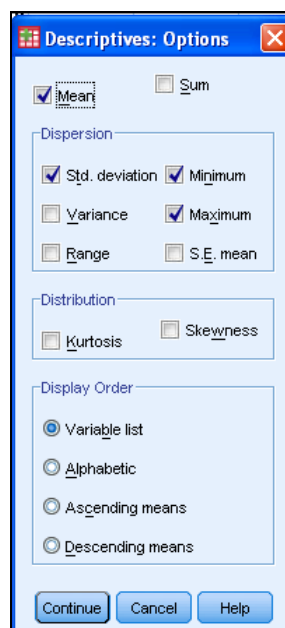
*What can you say about the person aged*

(i) 30? \_\_\_\_\_

(ii) 53? \_\_\_\_\_

### 2.1.8. DESCRIPTIVES Options

The dialogue box for **Descriptives: Options** is shown below.



As you can see a range of statistics to describe the dispersions of a variable and its distribution are available. You can also choose whether you want the statistics for a group of variables to be displayed in ascending or descending order according to the value of the mean of the variable, or whether you want them displayed in alphabetical or variable list order of the selected variable names. In the computer exercise below, you learn more about the options.

**Computer exercise 9: Using the DESCRIPTIVES options**

- (a) Use the **Options** to obtain the default statistics which include the mean, standard deviation, minimum, maximum and optional statistics which include the **sum, SE mean and skewness** for AGE and INC.

The output should look like this:

Descriptive Statistics									
	N	Minimum	Maximum	Sum	Mean		Std. Deviation	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Std. Error
Age of Subject	58	18	53	1871	32.26	1.083	8.247	.414	.314
Weekly income	52	\$40	\$1,000	\$20,445	\$393.17	\$29.717	\$214.296	.909	.330
Valid N (listwise)	50								

**Note:** If none of the STATISTICS options are selected, only the **default statistics, mean, SD, minimum and maximum** are given.

- (b) Select six of the variables from the list which are appropriate for DESCRIPTIVES . Obtain the **default** statistics only and use the **Display Order** option to display all the statistics in alphabetic order of name. Notice the different style of output to that above.
- (c) Obtain the mean, standard deviation, skewness, range, and sum for the variables AGE, INC, EXPF, EXPT and EXPL. Compute Z scores and use the Options to display in order of descending means.
- (d) Use **Data, Sort Cases** to sort the cases in order of age. Then use **Analyze, Reports, Case summaries** to list AGE and ZAGE. Compare the age corresponding to the ZAGE closest to 0 with the mean and median age obtained in Computer exercise 3 above using Frequencies (Mean was 32.3, Median was 31.5). What age from this list would you expect to be closest to the median age?

Some output is shown below:

	Zscore:
Age of Subject	Age of Subject
_____	_____
30	-.27386
31	-.15261
31	-.15261
32	-.03136
32	-.03136
32	-.03136
32	-.03136
32	-.03136
33	.08989
33	.08989
33	.08989

34	.21114
34	.21114
35	.33239

**Note:** There were 2 missing cases, case#'s 1 and 2, for Age in the 60 cases. Hence the 58 valid cases started at Case# 3 and the median is the 29.5th case which represents the age of 31.5. The mean corresponds to the ZAGE = 0, just over 32.

**Note:** The median is not available in the DESCRIPTIVES procedure.

What procedure can you use to obtain it? \_\_\_\_\_

**Computer exercise 10: Obtaining basic descriptive statistics using the Japanese files –See Japan Folder. (Optional)**

(a) Open the SPSS file STUD97C.SAV file

A variable which is the sum of QBI to QB39 has been obtained using a COMPUTE command (covered in detail in a later section)

Try to obtain descriptive statistics for the variables UNI, LEVEL, ID and STUDSC

Why is there a problem obtaining statistics for some of the variables?

(b) Obtain descriptive statistics for the variables STUDSC and click on **Save standardized values** to obtain Z scores for STUDSC

For the variable STUDSC, write down the

mean \_\_\_\_\_ and the S.D. \_\_\_\_\_

If you followed the above procedures, your output should look like this:

**Descriptive Statistics**

	N	Minimum	Maximum	Mean	Std. Deviation
STUDSC	399	43	166	95.84	35.02
Valid N (listwise)	399				

Note that ZSTUDSC1 corresponds to the standardised variable STUDSC. Now have a look at the last variable in the **Data Editor window** to see this newly created variable which was added to your file. You can rename this variable in the Data Editor if you wish.

**Note:** If the Z variable already exists SPSS will allocate a different name for the new saved variable.

Use **Analyze, Reports, Case summaries** or the **List procedure** to look at the values of **STUDSC** and **ZSTUDSC**. Some of the output is shown below:

STUDSC	Zscore (STUDSC)
149	1.51816
.	.
53	-1.22342
58	-1.08063
64	-.90928
65	-.88072
57	-1.10919
81	-.42379
.	.
.	.
60	-1.02351
90	-.16677
122	.74709

Note that a score of 149 has a Z score of 1.51816. This tells us that this age is about 1.5 standard deviations above the mean.

What can you say about a score of

(i) 53? \_\_\_\_\_

(ii) 131? \_\_\_\_\_

**Computer exercise 11: Using the DESCRIPTIVES options with the Japanese data**

- (a) Once again using SPSS file **STUD97C.SAV** use the **Options** to obtain the default statistics which include the mean, standard deviation, minimum, maximum and optional statistics which include the **SE mean, the variance, the skewness and the range**.

The output should look like this:

**Descriptive Statistics**

	N	Range	Minimum	Maximum	Mean		Std.	Variance	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error
STUDSC	399	123	43	166	95.84	1.75	35.02	226.140	.292	.122
Valid N (listwise)	399									

**Note:** If none of the STATISTICS options are selected, only the **default statistics, mean, SD, minimum and maximum** are given.

Select several of the variables from the list which are appropriate for **DESCRIPTIVES**. Obtain the **default** statistics only and use the **Display Order** option to display all the statistics in alphabetic order of name. Notice the different style of output to that above.

**Note:** The median is not available in the **DESCRIPTIVES** procedure.

What procedure can you use to obtain it? \_\_\_\_\_

### 2.1.9. The REPORT procedure

We have seen how the REPORT procedure can be used to list values of variables. You can also perform other tasks such as list value labels, break the file into groups and obtain summary statistics. The following exercises demonstrate some of the features of this procedure.

You can choose Summaries in Columns or Summaries in rows, depending on the statistics you want to display.

**NB.** You can obtain the percentage of cases within a certain range, which cannot be done in other SPSS procedures.

#### A. Summaries in Rows

##### Computer exercise 12: The REPORT command

(a) Using the QUESTSS.SAV file and **Analyze, Reports, Report summaries in Rows...** for the variables ID, SEX, AGE, CBIRTH and INC.

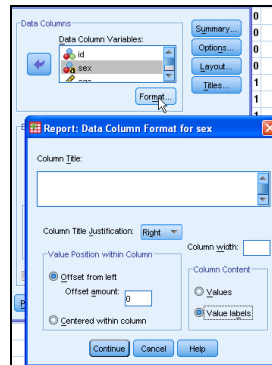


Make sure Display cases is ticked.

You should see a listing in the Output window similar to:

Respondents					
identification number	Sex of subject	Age of Subject	Country of Birth	Weekly income	
1	F	32	1	\$150	
2	F	30	1	\$170	
3	M	49	2	\$1,000	
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
Grand Total					
N	60	58	58	57	52

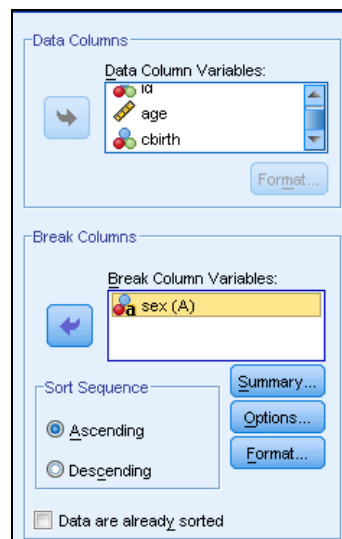
(b) To obtain value labels click on the relevant variable, say SEX, then **Format** and **Value Labels**, as in the example below:



Do this for SEX and CBIRTH and obtain an output like that below:

Respondents identification number	Sex of subject	Age of Subject	Country of Birth	Weekly income
1	Female	32	Australia	\$150
2	Female	30	Australia	\$170
3	Male	49	Other	\$1,000
4	Female	47	Other	\$950
5	Male	.	Australia	\$120
6	Female	.	Australia	\$250
7	Female	53	Other	.
8	Female	39	Other	.
9	Female	28	Other	\$370
10	Female	30	Other	\$165

(c) The report can be broken down into groups such as males and females using the **BREAK** feature.



**Note:** It has not been assumed that the data has already been sorted, so we have not selected the *Data already sorted* box.

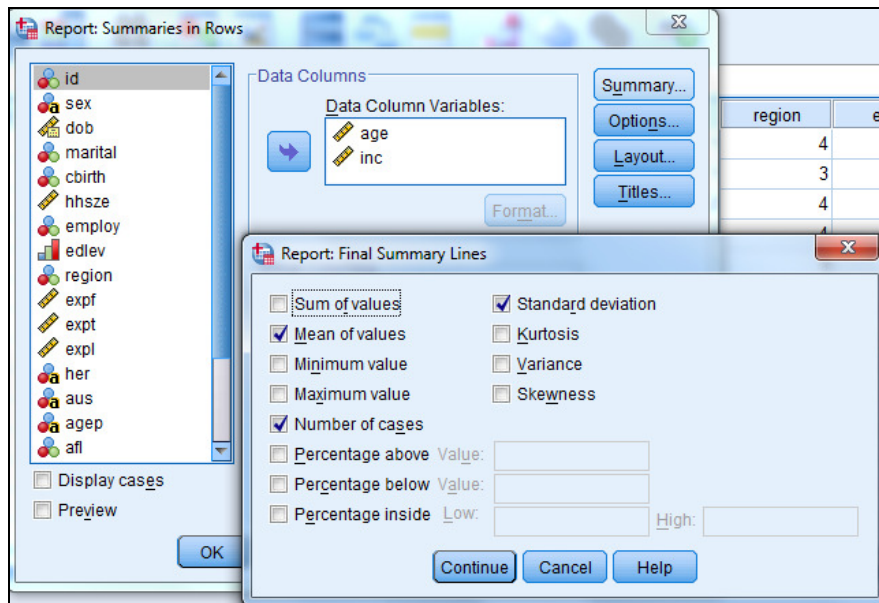
**Some output:**

Sex of subject	Respondents identification number	Age of Subject	Country of Birth	Weekly income
Female	1	32	Australia	\$150
	2	30	Australia	\$170
	4	47	Other	\$950
	6	.	Australia	\$250
	7	53	Other	.
	.	.	.	.
	.	.	.	.
Male	3	49	Other	\$1,000
	5	.	Australia	\$120
	11	24	.	\$480
	14	35	Australia	\$390
	.	.	.	.
	.	.	.	.

**(d) Statistics from Report**

For relevant variables, descriptive statistics can be obtained.

For example for Age and Income, under **Summary** Select the **Mean of values**, **Standard deviation** and **Number of cases** then **Continue** and **OK** to get an output as below: (Best turn off Display cases)

**Output**

	Age of Subject	Weekly income
Grand Total		
Mean	32	\$393
N	58	52
StdDev	8	\$214

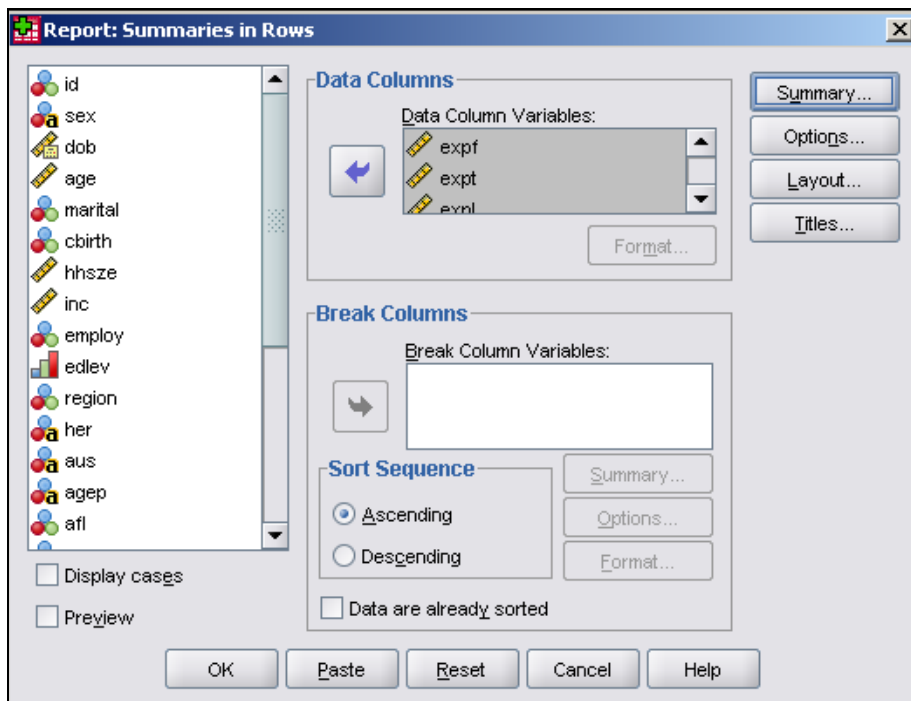


Notice any statistics requested apply to all the variables chosen. So make sure suitable variables are included if you want statistics printed. For example if we had the three expenditure variables, EXPF, EXPT and EXPL selected, we could ask for the Mean, SD and N as below.

	Weekly food expenditure	Weekly travel to work expenditure	Weekly leisure expenditure
Grand Total			
Mean	\$137	\$23	\$49
N	52	54	54
StdDev	\$108	\$33	\$27

A number of other features are available with REPORT which you can investigate if you are interested. In particular are some of the summary statistics which you can obtain which are not available elsewhere in SPSS. For example we may want to find the percentage of people who spent less than \$50 on the three expenditure variables and the percentage of people who spent between 100 and \$150 on these variables.

To obtain these percentages the **Summary** dialog box needs to be completed as follows:



These statistics apply to all variable in the Data columns box, so this is only appropriate if all the variables are on the same scale, eg the expenditure variables

	Weekly travel to work expenditure	Weekly leisure expenditure
Grand Total		
< 50	11.5%	40.7%
In 100 to 150	36.5%	7.4%

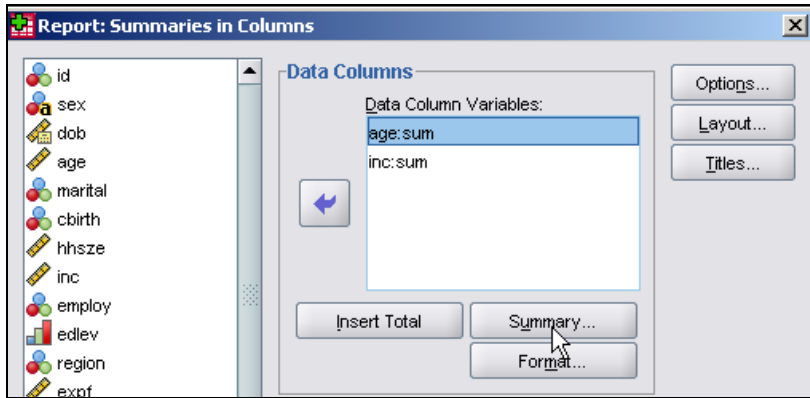
Therefore the summary dialog box above indicates the statistics available to the user.

## B. Summaries in Columns

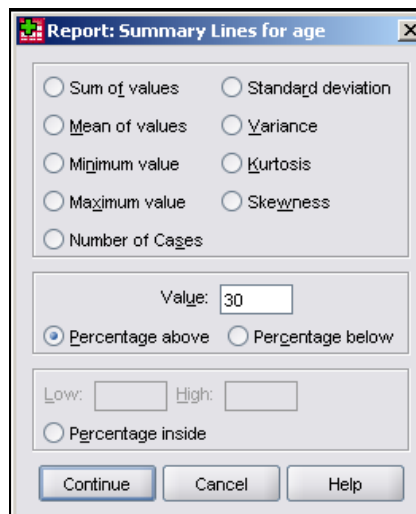
Sometimes it is not appropriate to request the same statistics for a number of variables. For example if we wanted the percent of people aged over 30 and the percentage earning over \$500, we could not do this in a single **Summaries in Rows command**.

If you did want to report different summary statistics for different variables in a single command, you would choose **Analyze, Reports, Report Summaries in Columns...**

For example for the variables AGE and INC. If you were interested in the percentage of people over 30 and the percentage of people earning less than \$500 you could use:



Change the **Summary** for each variable as required:



For example obtain an output like:

Age of	Weekly
Subject	income
> 30	< 500
<hr/>	
Grand Total	
53.4%	71.2%

### ***Computer exercise 13: Other features of REPORT***

Investigate any of the other features available with the REPORT procedure which you may need to use.

### 2.1.10 One-click descriptive statistics

In Version 21 a One-click descriptive statistics feature was introduced which gives summary descriptive statistics (for example, mean, median, frequency counts) for the selected variables. Appropriate statistics are automatically determined based on the measurement level of the variable.

To obtain descriptive statistics for selected variables:

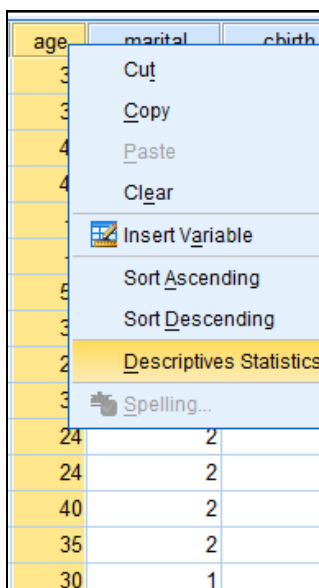
1. Right-click the selected variables in either Data View or Variable View.
2. From the pop-up menu select Descriptive Statistics.

By default, frequency tables (tables of counts) are displayed for all variables with 24 or fewer unique values. Summary statistics are determined by variable measurement level and data type (numeric or string):

- String. No summary statistics are calculated for string variables.
- Numeric, nominal or unknown measurement level. Range, minimum, maximum, mode.
- Numeric, ordinal measurement level. Range, minimum, maximum, mode, mean, median.
- Numeric, continuous (scale) measurement level. Range, minimum, maximum, mode, mean, median, standard deviation.

*For example use the file QUESTSS.SAV.*

*For example using the right mouse button, click on Age*



age	marital	chirth
3		
3		
4		
4		
5		
3		
2		
3		
24	2	
24	2	
40	2	
35	2	
30	1	

You should see some descriptive statistics and a frequency table.

**Statistics**

Age of Subject

N	Valid	58
	Missing	2
Mean		32.43
Median		31.50
Std. Deviation		8.046
Range		34
Minimum		19
Maximum		53

*When you do it on Sex, you just see a frequency table.*

*What is produced for (a) Region? \_\_\_\_\_*

*(a) Household size? \_\_\_\_\_*

*You can select several variables at the same time. Try this.*