Basic Statistical Analysis & SPSS

By
Dr. Rajeev Pandey
Professor & Head
Department of Statistics
University of Lucknow, India

Web: http://www.dacg.in

What is Statistics?

The purpose of statistics is to develop and apply methodology for extracting useful knowledge from both experiments and data. In addition to its fundamental role in data analysis, statistical reasoning is also extremely useful in data collection (design of experiments and surveys) and also in guiding proper scientific inference (Fisher, 1990).

What is Statistics?

- Statistics is neither really a science nor a branch of mathematics. It is perhaps best considered as a meta-science (or meta-language) for dealing with data collection, analysis, and interpretation. As such its scope is enormous and it provides much guiding insight in many branches of science, business.

My Opinion

Brief History of Statistics

- The Word statistics have been derived from Latin word "**Status**" or the Italian word "**Statista**", meaning of these words is "**Political State**" or a Government.
- Shakespeare used a word Statist is his drama Hamlet (1602). In the past, the statistics was used by rulers.
- The application of statistics was very limited but rulers and kings needed information about lands, agriculture, commerce, population of their states to assess their military potential, their wealth, taxation and other aspects of government.

Brief History of Statistics

- Basic forms of statistics have been used since the beginning of civilization. Early empires often collated censuses of the population or recorded the trade in various commodities.
- The Roman Empire was one of the first states to extensively gather data on the size of the empire's population, geographical area and wealth.

Brief History of Statistics

- During the 20th century several statistician are active in developing new methods, theories and application of statistics.
- Now these days the availability of electronics computers is certainly a major factor in the modern development of statistics.

HISTORY OF STATISTICS TIME LINE

TIME	CONTRIBUTOR	CONTRIBUTION
ANCIENT GRECE	PHILOSPHERS	IDEAS – NO QUANTITATIVE ANALYSIS
17th century	Graunt, Pascal, Petty & Bernoulli	Study affairs of States, Vital Statistics of Population Study Probability through games of chance & gambling
18th century	Laplace & Gauss	Normal Curve, Regression through Astronomy
19th century	Quetelet Galton	Astronomer who first applied statistical analyses to human biology. Studied Genetic Variation in Human (Used Correlation & Regression)
20th century	Pearson, Gossett (Student) & Fisher	Studied natural selection using correlation. Formed first Academic Department of Statistics, Biometrika Journal, helped developed Chi-Square Analysis, studied process of brewing. Altered the Statistics Community about problems with small sample sizes, developed Student's t-test. Evolutionary biologist developed ANOVA, stressed the importance of Experimental design

	Wilcoxon	Biochemist studied Pesticides, Non-Parametric Equivalent of two sample test
	Kruskal, Wallis, Spearman	Economists who developed the , Non-Parametric Equivalent of the ANOVA
	Spearman	Psychologist who developed the , Non-Parametric Equivalent of the correlation coefficient
20th Century (later)	Kendall	Statistician who developed the another Non-Parametric Equivalent of the correlation coefficient
	Tukey	Statistician who developed multiple comparisons procedure
	Dunnett	Biochemist who studied pesticides and developed multiple comparisons procedure for control groups
	Keuls	Agronomist who developed multiple comparisons procedures
	COMPUTER TECHNOLOGY	Provided many advantages over calculations by hands or calculators, stimulated the growth of investigation into new techniques

Statisticians

Carl Friedrich Gauss,
-mathematician who developed the method of least squares in 1809.

Sir William Petty
-a 17th-century economist who used early statistical methods to analyze demographic data.

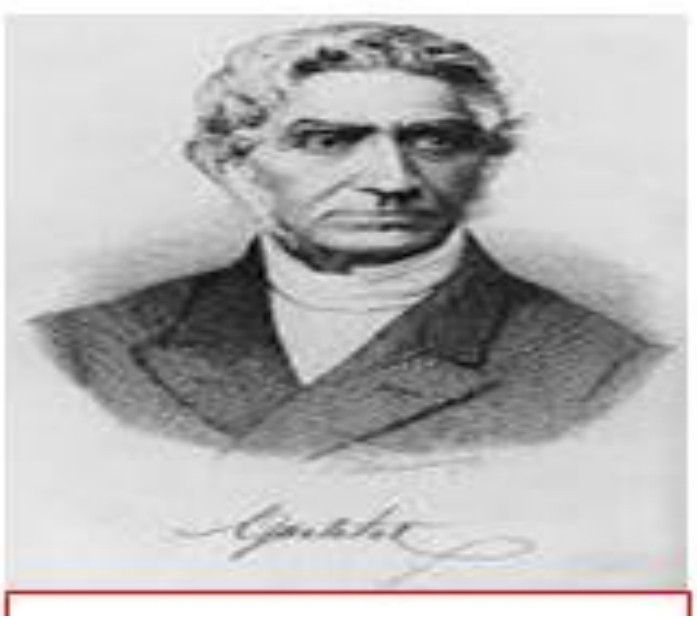
Pierre-Simon
-marquis de Laplace, one of the main early developers of Bayesian statistics.

Thomas Bayes
(c. 1702 – April 17, 1761)

Karl Pearson, the founder of mathematical statistics.

James Lind
-carried out the first ever clinical trial in 1747, in an effort to find a treatment for scurvy.

Ronald Fisher
"A genius who almost single-handedly created the foundations for modern statistical science",



Adolphe Quetelet, in full **Lambert Adolphe Jacques Quetelet**, (born February 22, 1796, Ghent, Belgium—died February 17, 1874, Brussels), Belgian mathematician, astronomer, statistician, and sociologist known for his application of statistics and probability theory to social phenomena.

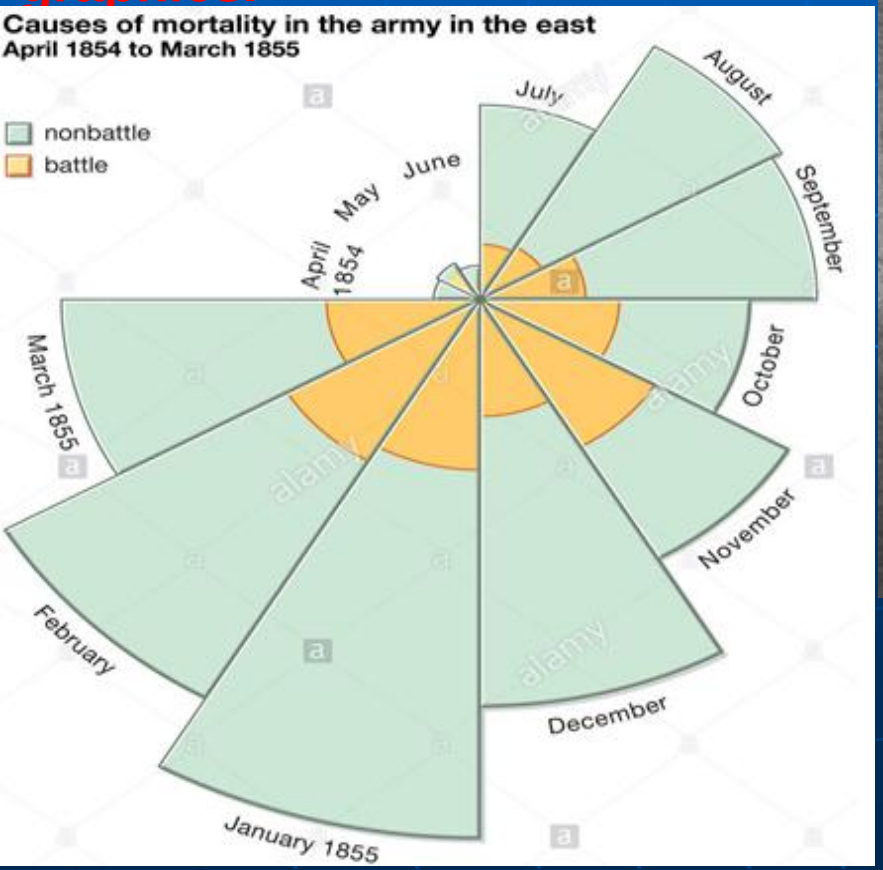
FATHER OF STATISTICS

No recognition
Yet
My personal belief

"A father of Statistics"

TESTING OF HYPOTHESIS

~~Florence Nightingale~~ exhibited a gift for mathematics from an early age and excelled in the subject under the tutelage of her father. Later, ~~Nightingale~~ became the first pioneer in the visual presentation of information and statistical graphics.



Florence Nightingale
(12 May 1820 – 13 August 1910)

Why Study Statistics?

- ***Communication***
  - Understanding the language of statistician who facilitates communication and improves problem solving.

- ***Computer Skills***
  - The use of spreadsheets for data analysis and word processors or presentation software for reports improves upon your existing skills.

What is Statistics? (OLD)

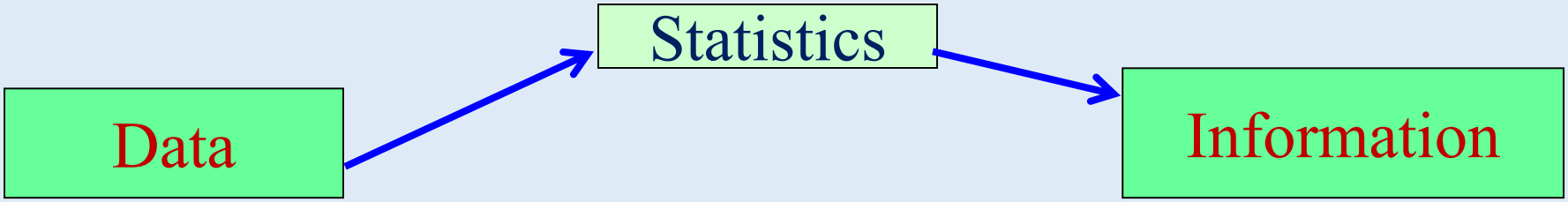
- *Statistics* is the science of collecting, organizing, analyzing, interpreting, and presenting data.

- A *statistic* is a single measure (number) used to summarize a sample data set. For example, the average height of students in this class.

- A *statistician* is an expert with at least a master's degree in mathematics or statistics or a trained professional in a related field.

What is Statistics?

- "Statistics is a way to get information from data"



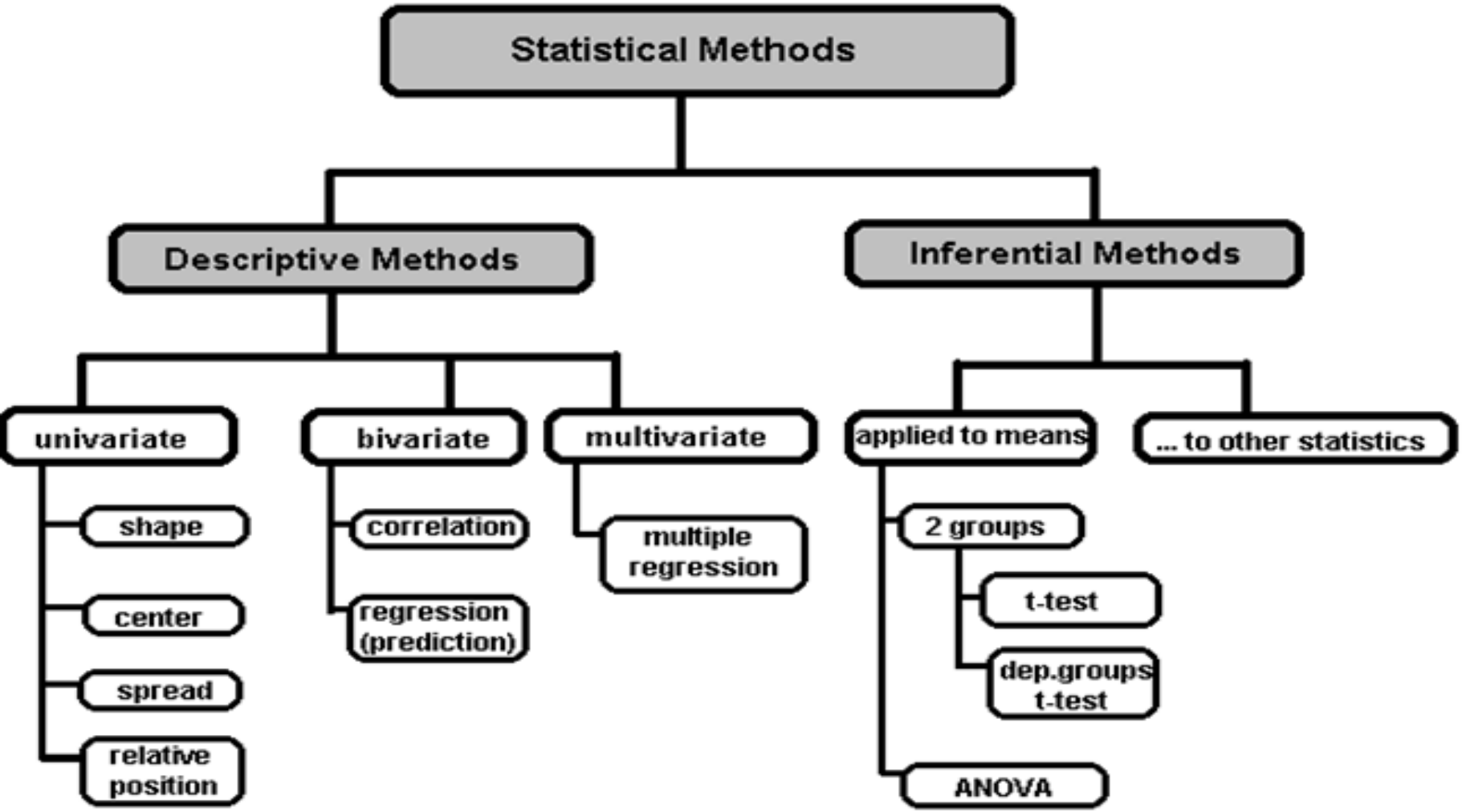
Data: **Facts, especially numerical facts, collected together for reference or information.**

Information: Knowledge communicated concerning some particular fact.

Statistics is a *tool* for creating *new understanding* from a set of numbers.

Statistics is a science of getting informed decisions.

A Taxonomy of Statistics



Statistical Description of Data

- Statistics describes a numeric set of data by its
  - Center
  - Variability
  - Shape
- Statistics describes a categorical set of data by
  - Frequency, percentage or proportion of each category

Data Presentation

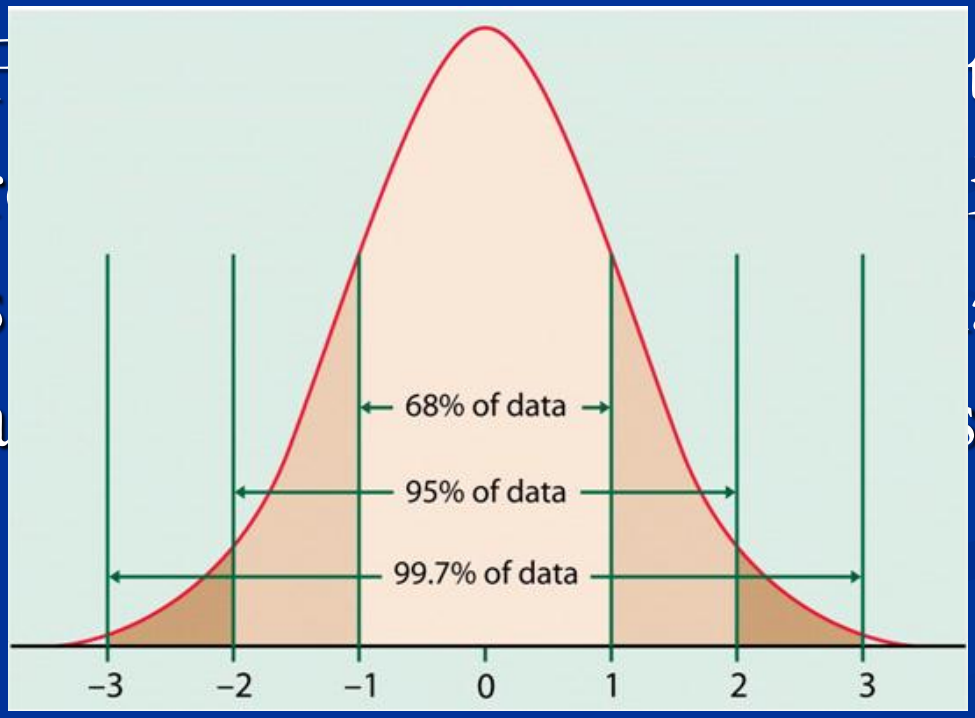
Two types of statistical presentation of data - graphical and numerical.

Graphical Presentation: **We look for the overall pattern and for striking deviations from that pattern. Over all pattern usually described by shape, center, and spread of the data. An individual value that falls outside the overall pattern is called an *outlier.***

- **Bar diagram and Pie charts are used for categorical variables.**
- **Histogram, stem and leaf and Box-plot are used for numerical variable.**

Role of Normality

- Many statistical methods require that the numeric variables we are working with have an approximate **normal distribution**.
- F ts, and
  r ire in some
  s ables are
  a tributed.



Standardized normal distribution with empirical rule percentages.

Tools for Assessing Normality

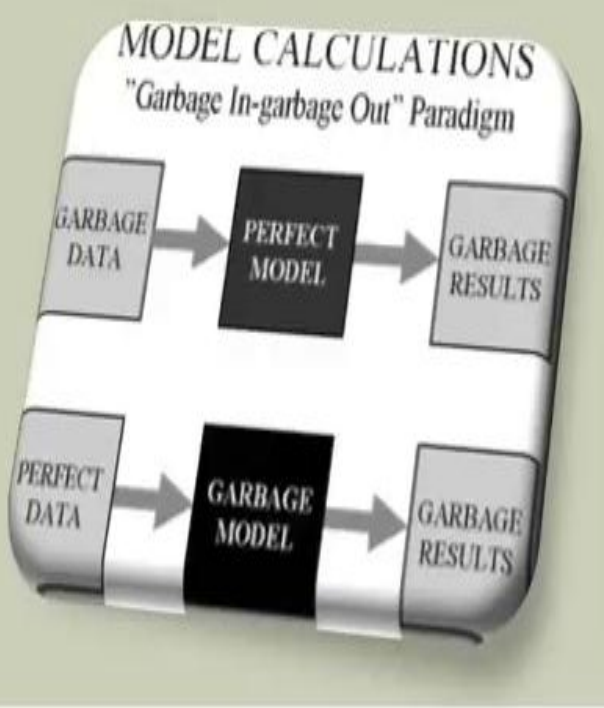
- Histogram and Boxplot
- Normal Quantile Plot
  (also called Normal Probability Plot)
- Goodness of Fit Tests

  Shapiro-Wilk Test (JMP)
  Kolmogorov-Smirnov Test (SPSS)
  Anderson-Darling Test (MINITAB)

LOOK AT YOUR DATA GRAPHICALLY FIRST

...*Before* starting all the fun, cool, whiz-bang analysis.

Get to know the data. Look for patterns, potential problems, initial relationships, etc.

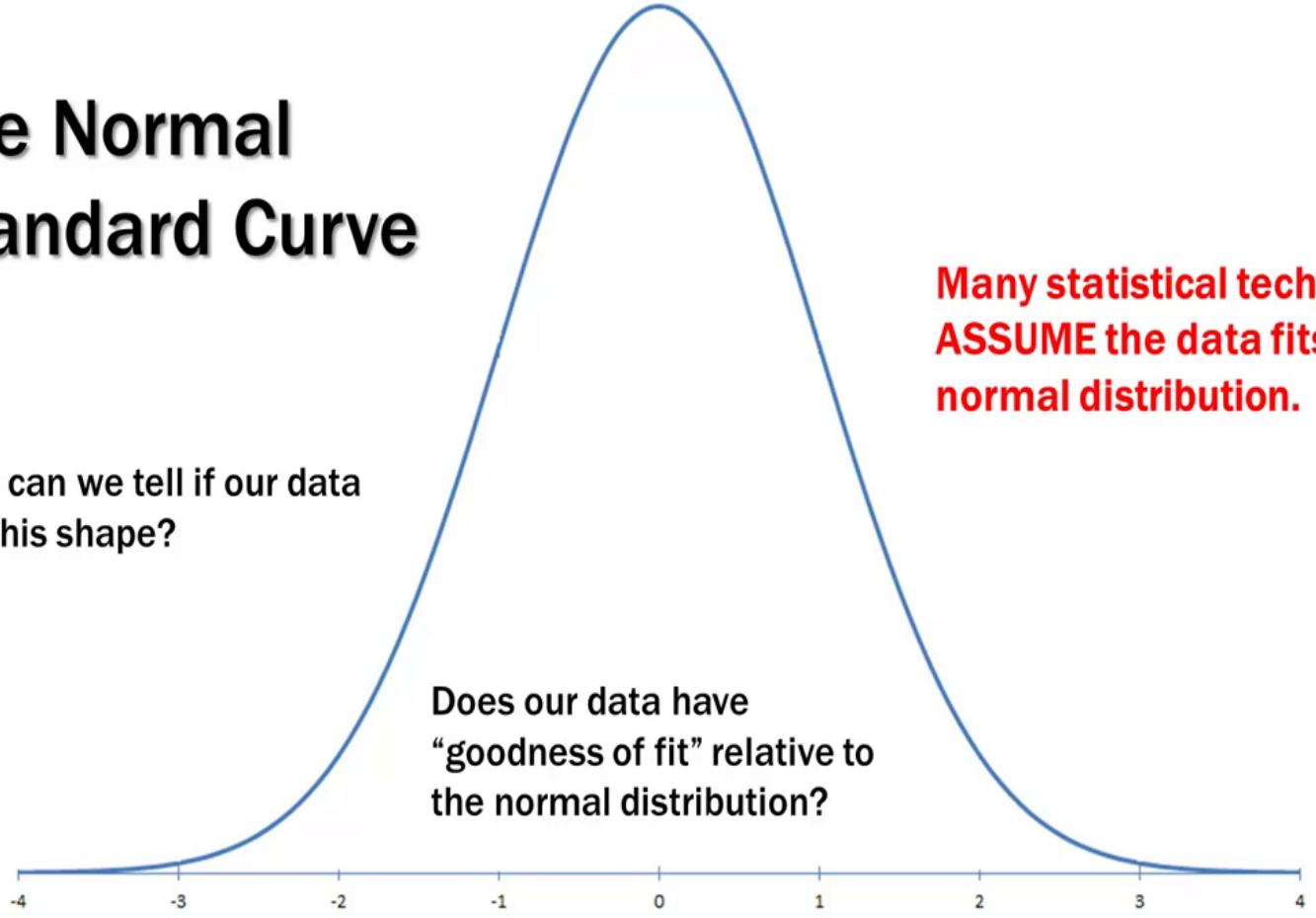


GARBAGE IN, GARBAGE OUT.

GRAPHICAL DATA EXPLORATION

- By using a few simple visual tools, we can learn a tremendous amount of information about our data
- Our data may have excess skew (lopsided), kurtosis (very fat tails), be bi-modal (two humps like a camel), or follow a distribution other than the normal distribution
- In this presentation we will briefly discuss the following tools to determine if our data is “normal”:
  - Histograms
  - Stem and Leaf Plots
  - Box Plots (Box and Whisker Plots)
  - P-P Plots
  - Q-Q Plots

The Normal Standard Curve



Many statistical techniques ASSUME the data fits a normal distribution.

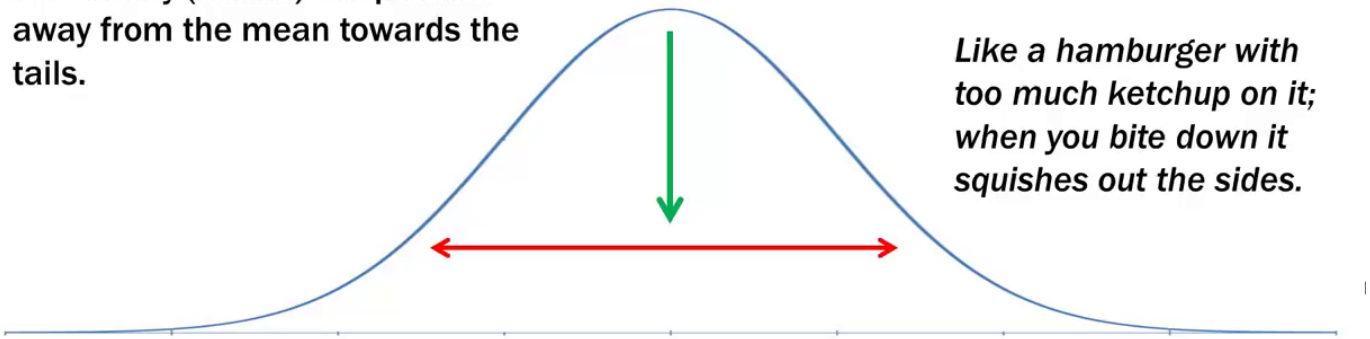
How can we tell if our data fits this shape?

Does our data have "goodness of fit" relative to the normal distribution?

Excess Kurtosis

More probability than expected in the tails of the distribution due to extreme values away from the mean.

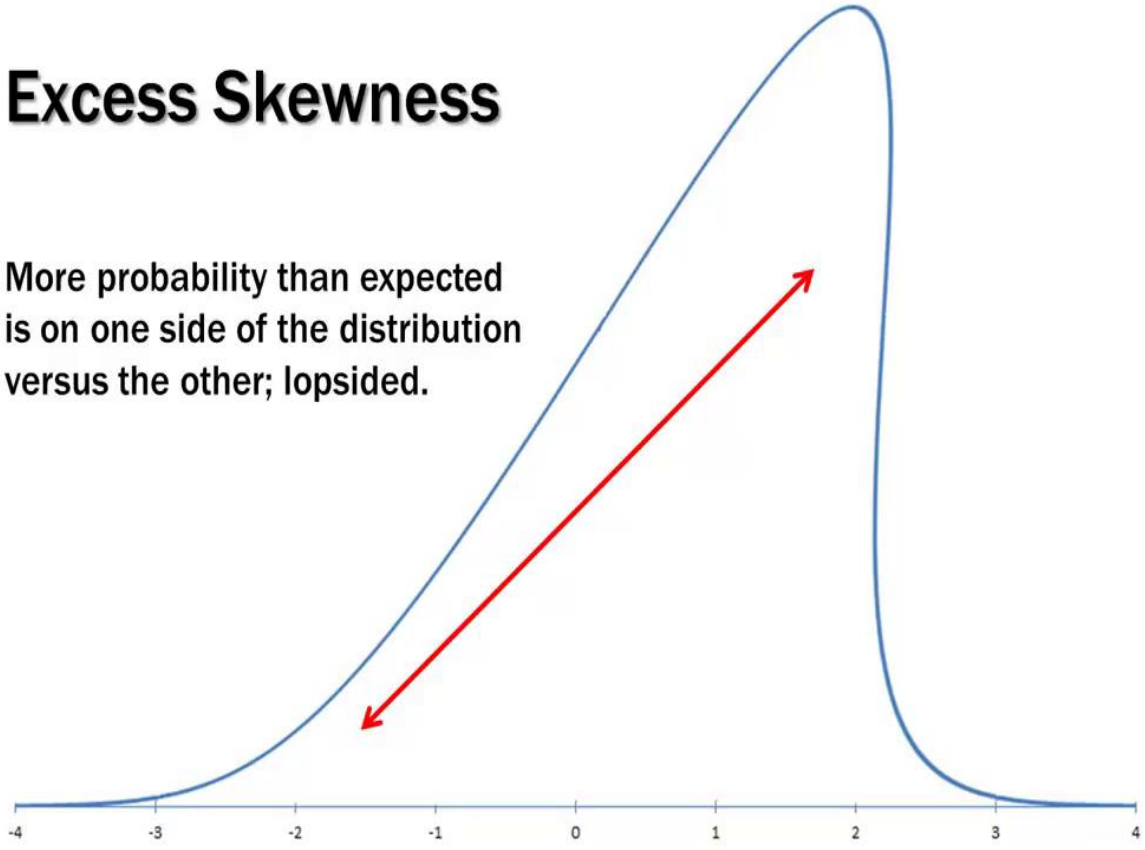
Probability (values) are pushed away from the mean towards the tails.



Like a hamburger with too much ketchup on it; when you bite down it squishes out the sides.

Excess Skewness

More probability than expected is on one side of the distribution versus the other; lopsided.



OTHER PROBABILITY DISTRIBUTIONS

Oftentimes data fits another type of distribution much better:

Exponential

Lognormal

Among others....

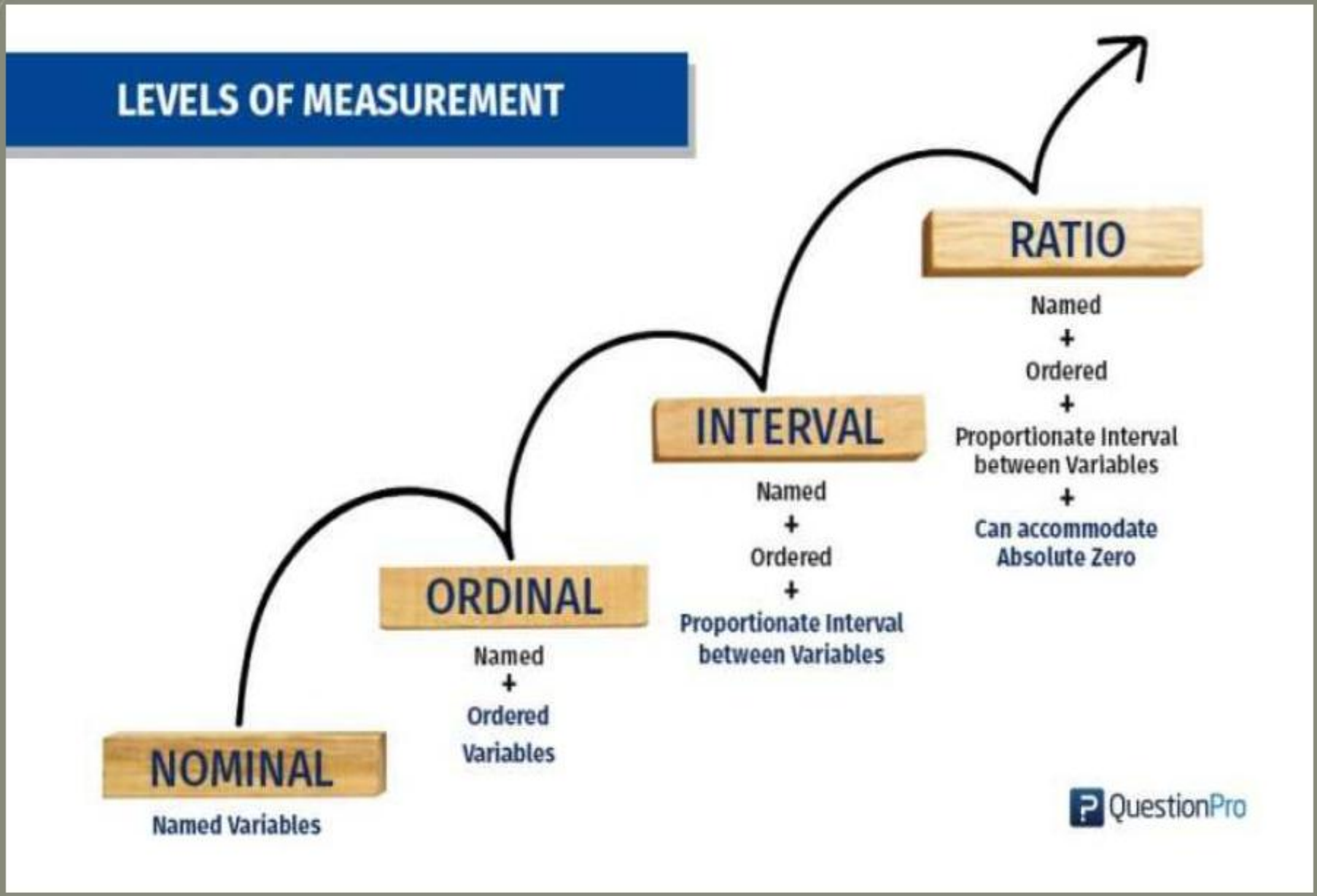
Weibull

Uniform

Some Definitions

Variable – any characteristic of an individual or entity. A variable can differ values for different individuals. *Variables can be categorical or quantitative*.

- **Nominal** - Categorical variables with no inherent order or ranking sequence such as names or classes (e.g., gender). Value may be a numerical, but without numerical value (e.g., I, II, III). The only operation that can be applied to Nominal variables is enumeration.
- **Ordinal** - Variables with an inherent rank or order, e.g. mild, moderate, severe. Can be compared for equality, or greater or less, but not *how much* greater or less.
- **Interval** - Values of the variable are ordered as in Ordinal, and additionally, differences between values are meaningful, however, the scale is not absolutely anchored. Calendar dates and temperatures on the **Fahrenheit scale are examples.** Addition and subtraction, but not multiplication and division are meaningful operations.
- **Ratio** - Variables with all properties of Interval plus an absolute, non-arbitrary zero point, e.g. age, weight, temperature (Kelvin). Addition, subtraction, multiplication, and division are all meaningful operations.


LEVELS OF MEASUREMENT
NOMINAL
Named Variables
ORDINAL
Named
+
Ordered
Variables
INTERVAL
Named
+
Ordered
+
Proportionate Interval between Variables
RATIO
Named
+
Ordered
+
Proportionate Interval between Variables
+
Can accommodate Absolute Zero
QuestionPro

Statistical Packages

Command Base (Programmable)

- SPLUS
- R and RPLUS
- Win Bugs
- Matlab

Window Base

- STATISTICA
- NCSS
- STATA
- SPSS (Syntax base)

History & Development

- The original statistical software packages - written for IBM mainframes. Its development started in 1957, at UCLA Health Computing Facility. SPSS arrived second, developed by social scientists at the University of Chicago, starting around 1968.
- The program, originally called Statistical Package for the Social Sciences, was released in 1968 and quickly became one of the most widely used statistics programs in the social sciences, including in healthcare, government, market research and surveying.

What are Variables?

- Variables are things that we measure, control, or manipulate in research.
- They differ in many respects, most notably in the role they are given in our research and in the type of measures that can be applied to them.
  - e.g. predictor, criterion, moderator, etc.

What is a variable?

- A variable is any characteristic or attribute of an object under investigation that takes on numerical values.
- For example, variables associated with employees may be their talent, work-ethic, wage, gender, age, productivity level, etc.

What are variables you would consider in buying a second hand bike?

- Brand (Atlas, Hero, etc.)
- Type (road, mountain, racer)
- Age
- Condition (Excellent, good, poor)
- Price
- Frame size
- Number of gears, etc.

Latent vs. Manifest Variables

- A *manifest variable* can be observed.

  e.g. age, gender, productivity-level, tenure and wage

- A *latent variable* is not observed and can only be measured indirectly.

  e.g. talent, work ethic

Dependent vs. Independent Variables

- An *independent variable* has an antecedent or causal role.

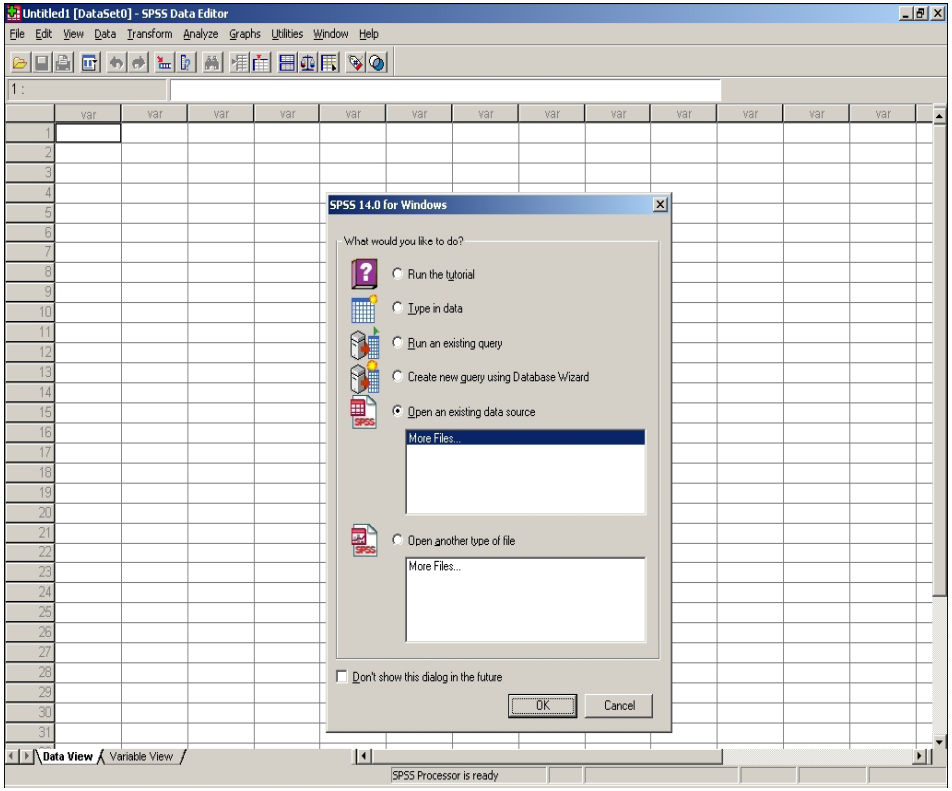
  e.g. talent, work-ethic, age, tenure
- A *dependent variable* plays a consequent, or affected, role in relation to the independent variable.

  e.g. productivity

Starting SPSS

- From the **Start menu go to All Programs / SPSS for Windows and select SPSS 22/25**
- **SPSS Data Editor window opens up with a queries window**
- **superimposed (Figure).**

- Select **Cancel or Type in data to this query window. The window then closes.**



Example

2.1 Imagine that you interviewed some people on their smoking habits using the questions shown below:

Reference number ☐☐☐

University of BRISTOL

Smoking Questionnaire

1. How old were you on your last birthday? ☐☐

2. Indicate your sex — Male ☐ Female ☐

3. Do you smoke at all? — Yes ☐$_1$ No ☐$_2$
   If No, go to question 9

4. Do you smoke cigarettes? — Yes ☐$_1$ No ☐$_2$
   If No, go to question 6

5. On average how many cigarettes a day would you say that you smoked? ☐☐☐

6. Do you smoke a pipe? — Yes ☐$_1$ No ☐$_2$

7. Do you smoke cigars? — Yes ☐$_1$ No ☐$_2$

8. Have you ever tried to give up smoking? — Yes ☐$_1$ No ☐$_2$

9. Tell me what you think on each of the following three statements:-

   Tax on tobacco is too high
   strongly disagree ☐$_1$ disagree ☐$_2$ agree ☐$_3$ strongly agree ☐$_4$

   Smoking is dangerous to your health
   strongly disagree ☐$_1$ disagree ☐$_2$ agree ☐$_3$ strongly agree ☐$_4$

   Smoking should not be allowed in cinemas
   strongly disagree ☐$_1$ disagree ☐$_2$ agree ☐$_3$ strongly agree ☐$_4$

Response of Data Set

Ref Number	Age	Sex	Smoke	Smoke Cigs	How many	Pipe	Cigars	Give Up	Tax	Health	Cinema
1	27	F	1	1	10	2	2	1	3	3	3
2	31	M	2						4	2	1
3	35	M	2						4	1	1
4	58	M	2						3	1	2
5	56	M	2						4	3	2
6	25	F	1	1	20	2	2	2	3	4	4
7	41	F	1	1	30	2	1	1	3	1	3
8	38	F	1	1	999	2	2	1	4	4	4
9	43	F	1	2		2	1	1	4	2	2
10	29	M	1	1	40	2	2	2	2	4	4

Label

ref_no	Reference number
age	Age last birthday
sex	Sex of respondent
smoker	Do you smoke?
cigs	Do you smoke cigarettes?
num_cigs	How many cigarettes per day?
pipe	Do you smoke a pipe?
cigars	Do you smoke cigars?
give_up	Have you tried to give up smoking?
tax	Do you think tax on tobacco is too high?
danger	Do you think smoking is dangerous to your health?
cinemas	Do you think smoking should be allowed in cinemas?

Data Editor

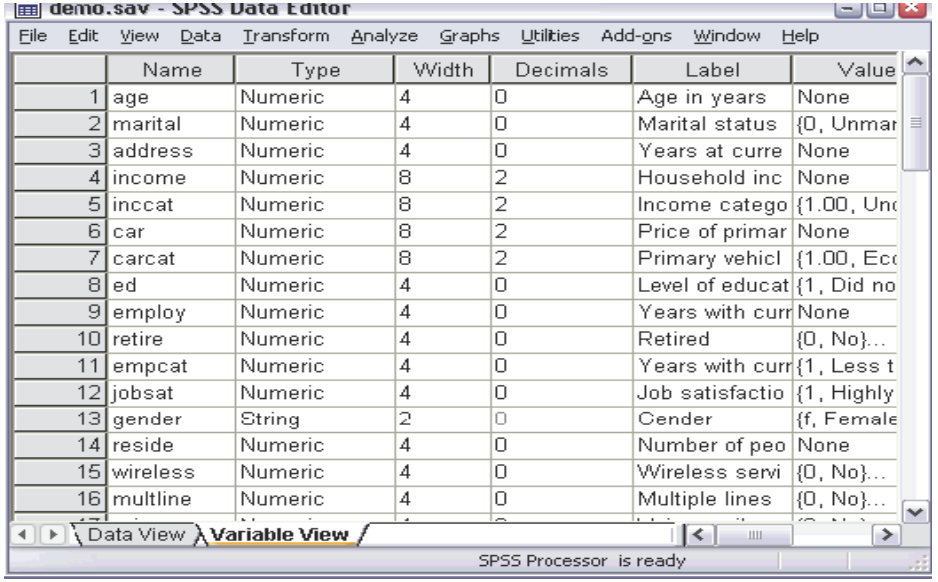
provides two views of your data:

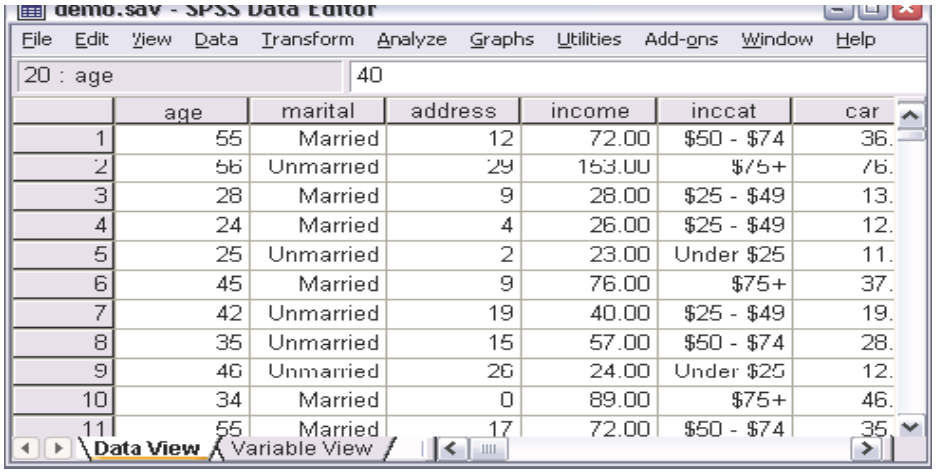
The Data Editor

***Data View.** This view displays the actual data values or defined value labels.

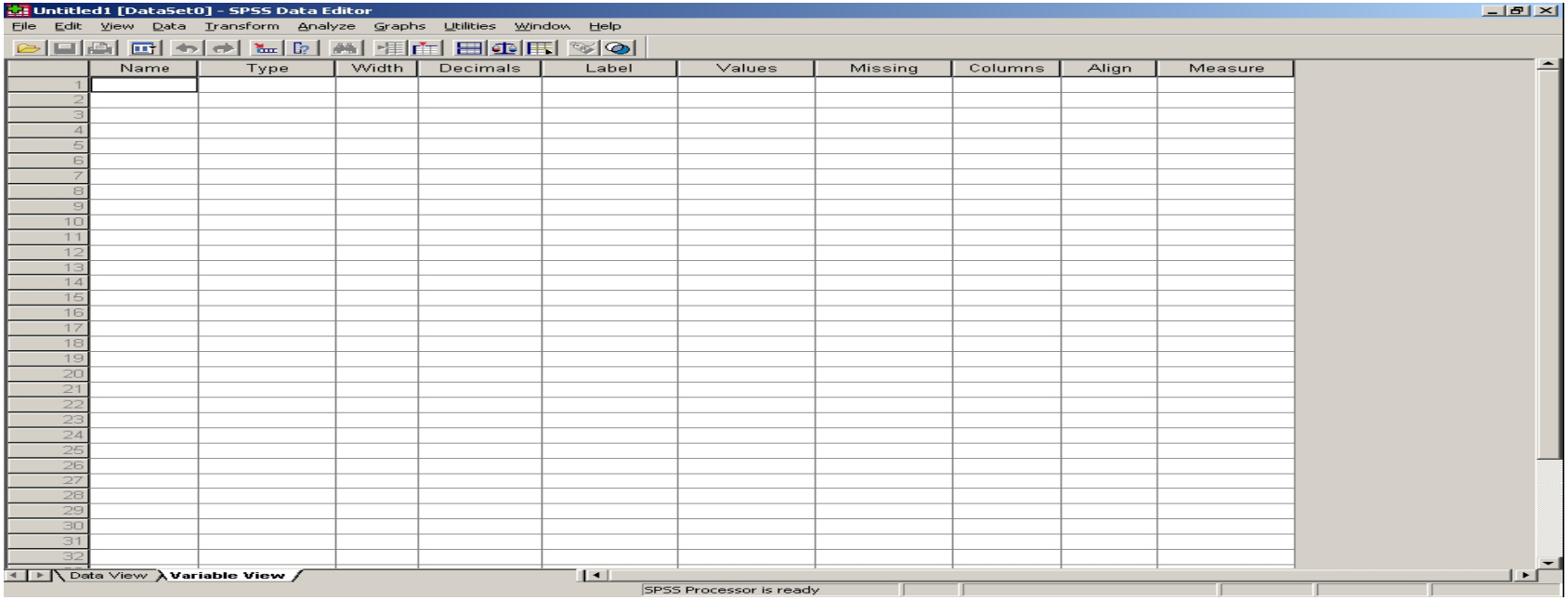
* **Variable View. This view displays variable definition information, including defined variable and value labels, data type (for example, string, date, and numeric), measurement level (nominal, ordinal, or scale), and user-defined missing values.**

(In both views, you can add, change, and delete information that is contained in the data file).





Variable View of the Data Editor window:
Click on the **Variable View tab in the bottom left hand corner of the screen**

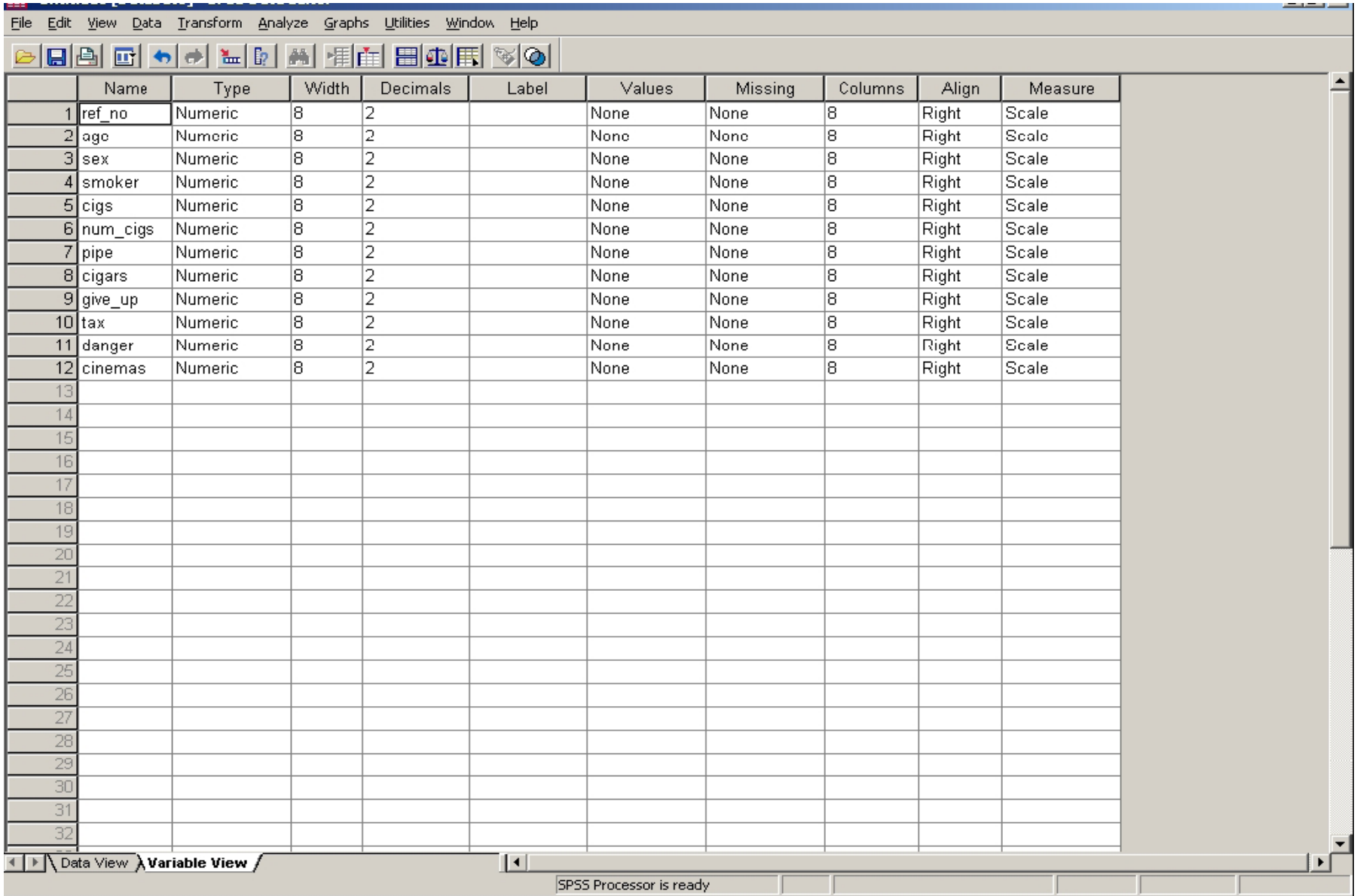


window already has a defined structure.
There are **ten** columns headed: -
Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align and Measure

Rules apply to variable name

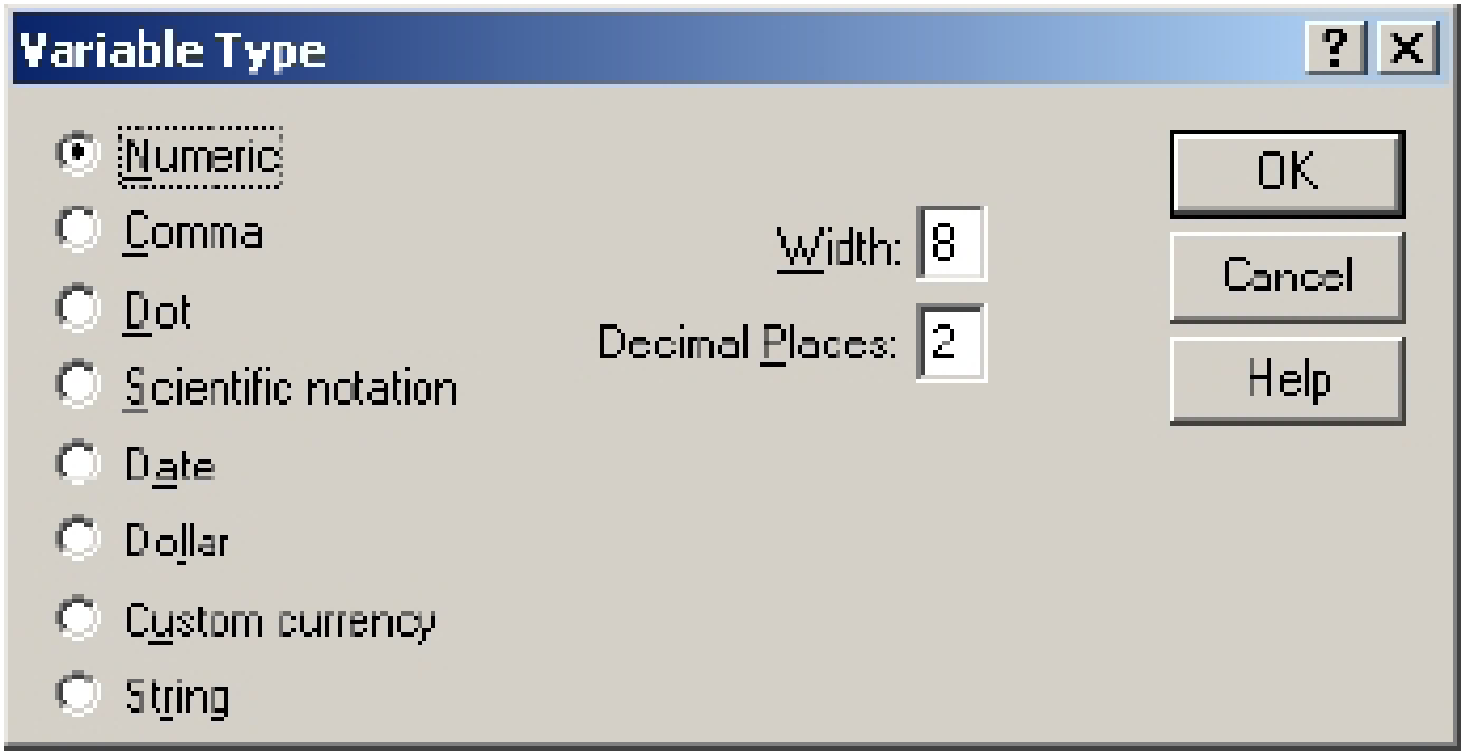
- **The name must begin with a letter**. The remaining characters can be any letter, any digit, a period, or the symbols **@, #, _, or $**.
- **Variable names cannot** end with a period.
- **Ending variable names** with an underscore should be avoided (to avoid conflict with variables that are automatically created by some procedures).
- **The length of the name** cannot exceed 64 bytes. Typically, 64 bytes means 64 characters in single-byte languages (for example, English, French, German).
- **Blanks and special** characters (for example, !, ?, ‘, and *) **cannot** be used.
- **Reserved keywords** cannot be used as variable names. Reserved keywords are: **ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH.(Predefined)**

SPSS WORKSHEET



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	ref_no	Numeric	8	2		None	None	8	Right	Scale
2	age	Numeric	8	2		None	None	8	Right	Scale
3	sex	Numeric	8	2		None	None	8	Right	Scale
4	smoker	Numeric	8	2		None	None	8	Right	Scale
5	cigs	Numeric	8	2		None	None	8	Right	Scale
6	num_cigs	Numeric	8	2		None	None	8	Right	Scale
7	pipe	Numeric	8	2		None	None	8	Right	Scale
8	cigars	Numeric	8	2		None	None	8	Right	Scale
9	give_up	Numeric	8	2		None	None	8	Right	Scale
10	tax	Numeric	8	2		None	None	8	Right	Scale
11	danger	Numeric	8	2		None	None	8	Right	Scale
12	cinemas	Numeric	8	2		None	None	8	Right	Scale
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										

Variable Type



Variable Type options box

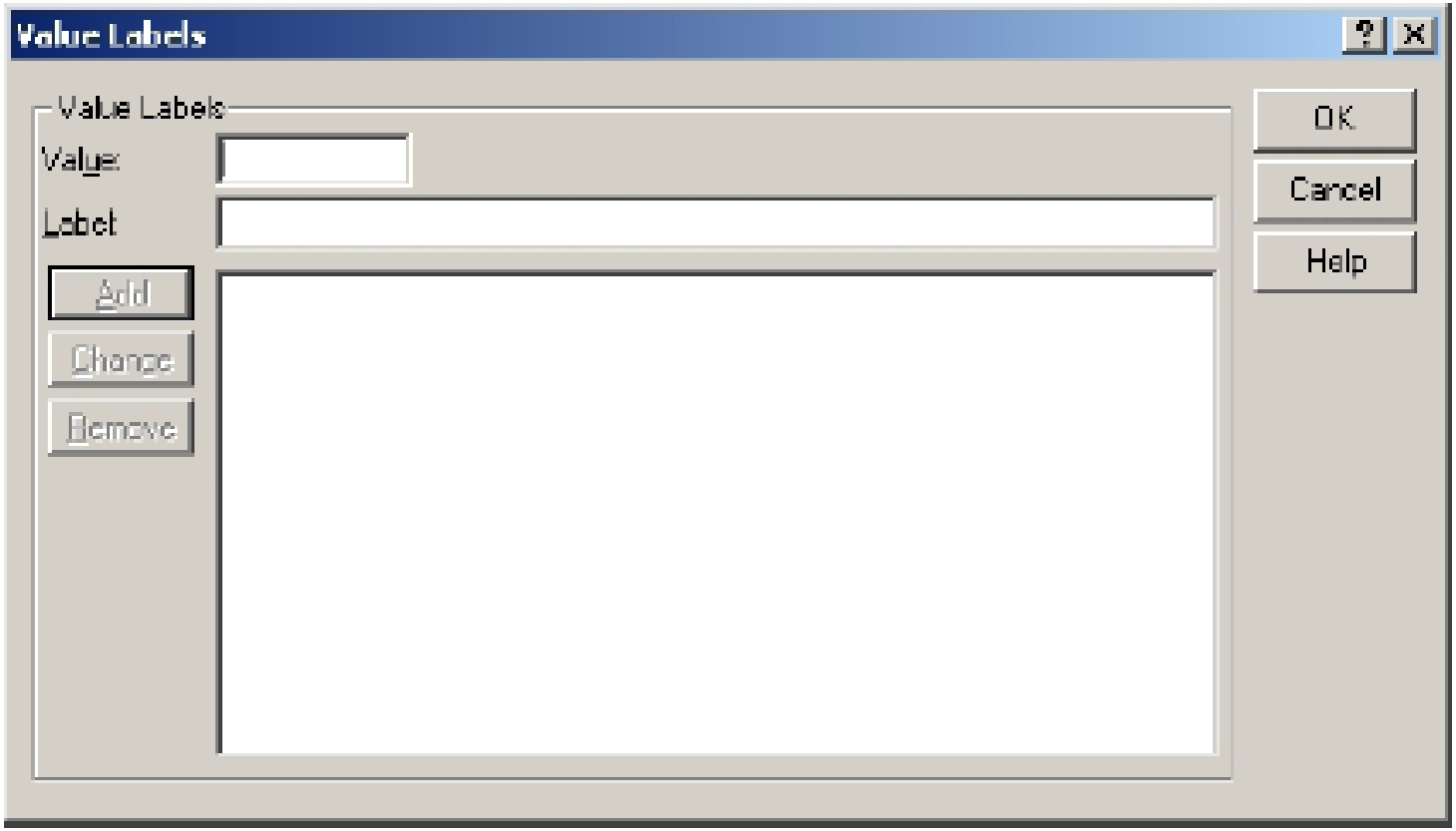
Variable Width and Decimals

- **Width *1, Decimals 0 for variables smoker, cigs, pipe, cigars, give_up, tax, danger***

and cinemas.

- **Width *2, Decimals 0 for variable age.***
- **Width *3, Decimals 0 for variable num_cigs.***
- **Width *4, Decimals 0 for variable ref_no.***
- **Width *1 for string variable sex***

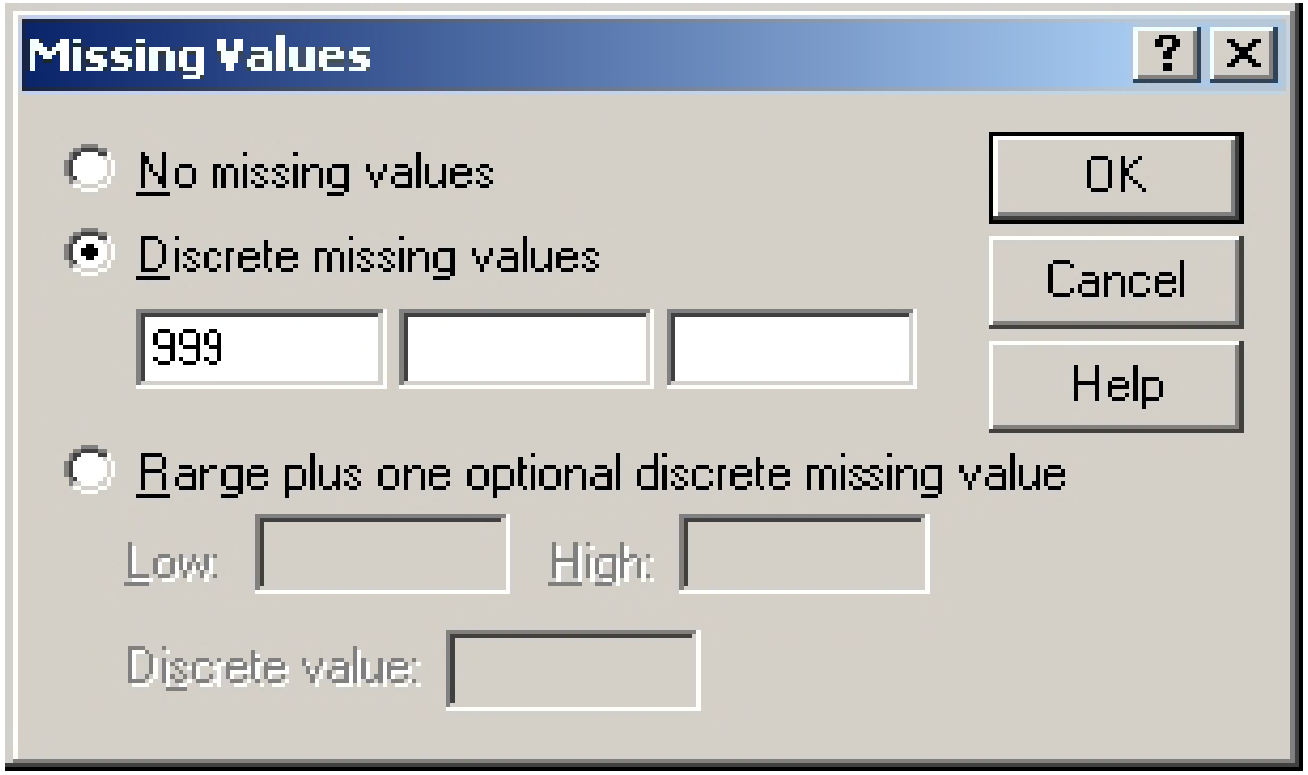
Values


Value Labels
Value Labels
Value:
Label:
Add
Change
Remove
OK
Cancel
Help

- Move to row 3 column 6 and click in the cell. A dropdown menu appears so you can
- In the box by the word **Value type *F. In the box by the word Label type Female.***
- Click on **Add and watch the value and its label move to the bottom box.**
- In the box by the word **Value now type *M and the word Male in the Label box.***
- Click on **Add. Now that all the Value Labels for this variable are complete** click on **OK to return to the Variable View**

Missing Values

The next column of the Variable View sheet is Missing Values.



Column, Align, Measure

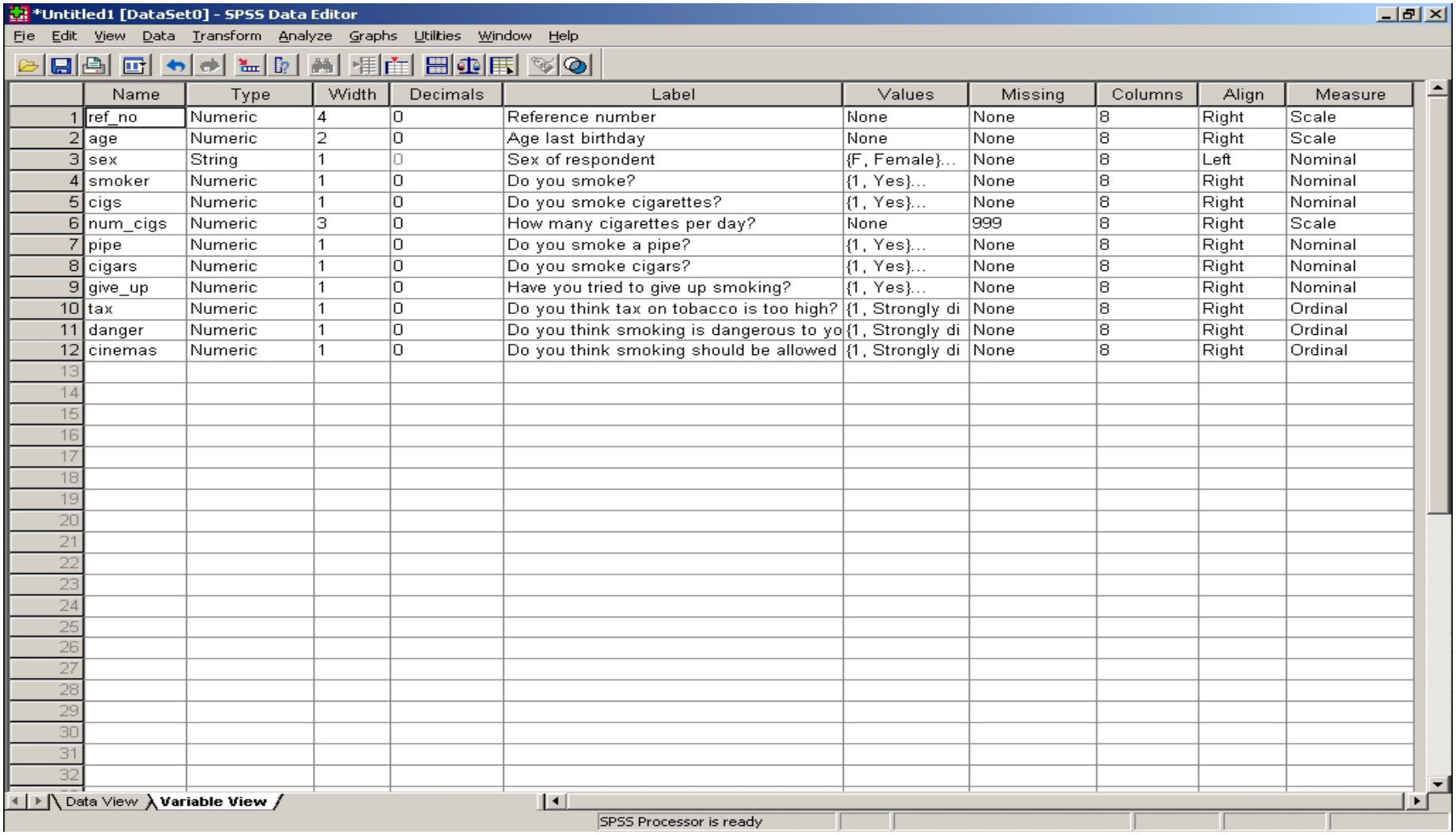
- ***Column indicates the width of the Particular Variable column***
- ***Align is as usual of formatting in word or Excel alignment i.e. 1. Left 2. centre 3. Right***
- **The three measures are used for variables**

➢ **Scale: to represent a numeric variable that can take discrete or continuous values along a range**

➢ **Ordinal: to represent values that, although numeric, only represent an ordered listing of such values**

➢ **Nominal: to represent values that are simply names**

Variable View screen with defined information



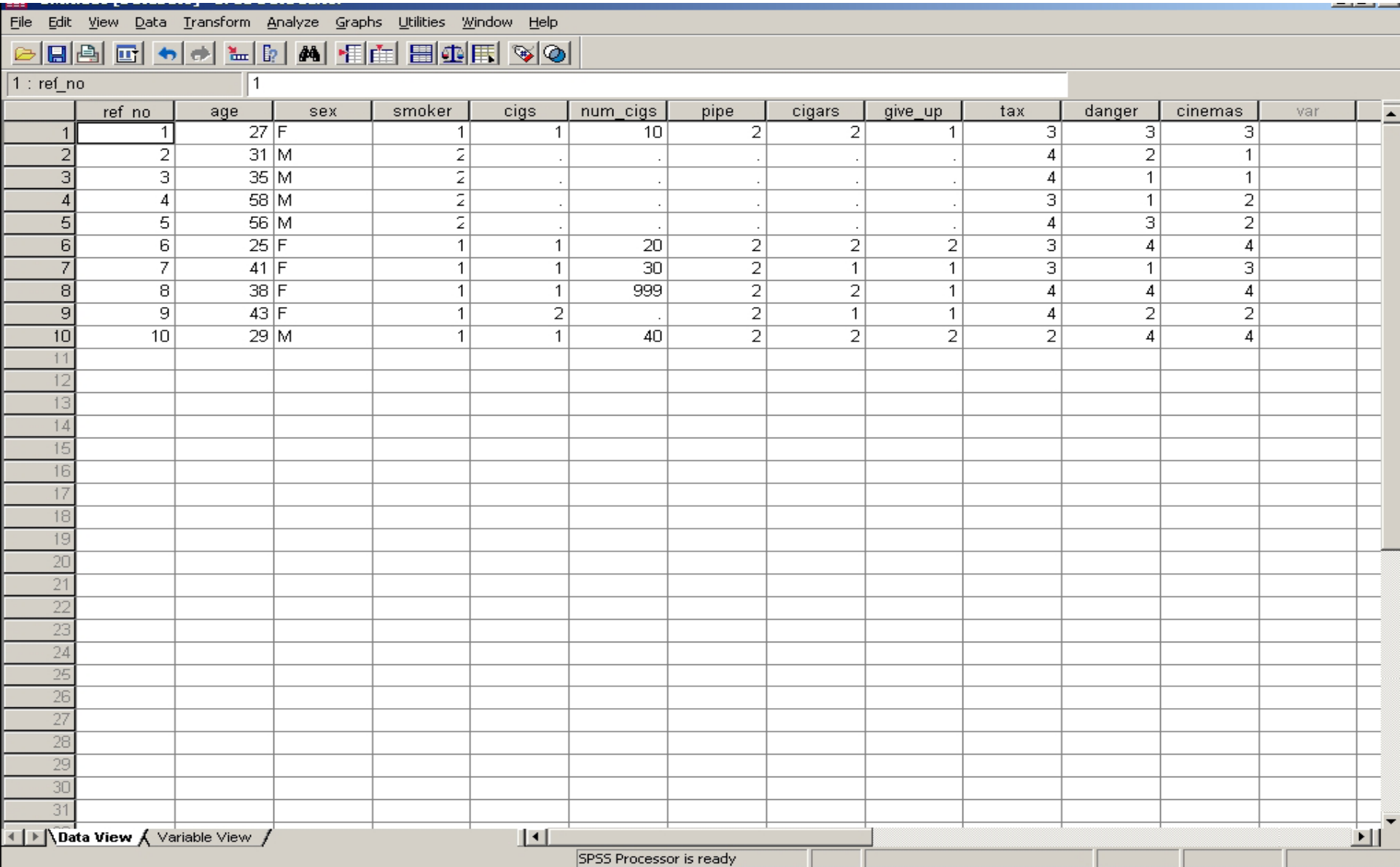
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	ref_no	Numeric	4	0	Reference number	None	None	8	Right	Scale
2	age	Numeric	2	0	Age last birthday	None	None	8	Right	Scale
3	sex	String	1	0	Sex of respondent	{F, Female}...	None	8	Left	Nominal
4	smoker	Numeric	1	0	Do you smoke?	{1, Yes}...	None	8	Right	Nominal
5	cigs	Numeric	1	0	Do you smoke cigarettes?	{1, Yes}...	None	8	Right	Nominal
6	num_cigs	Numeric	3	0	How many cigarettes per day?	None	999	8	Right	Scale
7	pipe	Numeric	1	0	Do you smoke a pipe?	{1, Yes}...	None	8	Right	Nominal
8	cigars	Numeric	1	0	Do you smoke cigars?	{1, Yes}...	None	8	Right	Nominal
9	give_up	Numeric	1	0	Have you tried to give up smoking?	{1, Yes}...	None	8	Right	Nominal
10	tax	Numeric	1	0	Do you think tax on tobacco is too high?	{1, Strongly di	None	8	Right	Ordinal
11	danger	Numeric	1	0	Do you think smoking is dangerous to yo	{1, Strongly di	None	8	Right	Ordinal
12	cinemas	Numeric	1	0	Do you think smoking should be allowed	{1, Strongly di	None	8	Right	Ordinal
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										
28										
29										
30										
31										
32										

You may return to the **Variable View window at any time if further changes are** needed.

Entering Data

- Click on the **Data View tab of the Data Editor Window**
- To enter the first person's data, click the first cell of **ref_no.**
- Type ***1.***
- Press the **<Tab> key or right arrow once and the heavy outline moves to the next** column.
- Type in ***27 and press the <Tab> key.***
- Type in ***F and press the <Tab> key.***
- Type in ***1 and press the <Tab> key.***
- Follow the same procedure along the first row until all twelve data values are entered.
- Move back to row 2, column 1 and start to enter the values for interview 2. Press the **<Tab> key twice to skip over a column. Notice that a dot appears in the cell.** This is the system-missing value

Data editor window with all interview data entered



File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : ref_no | 1

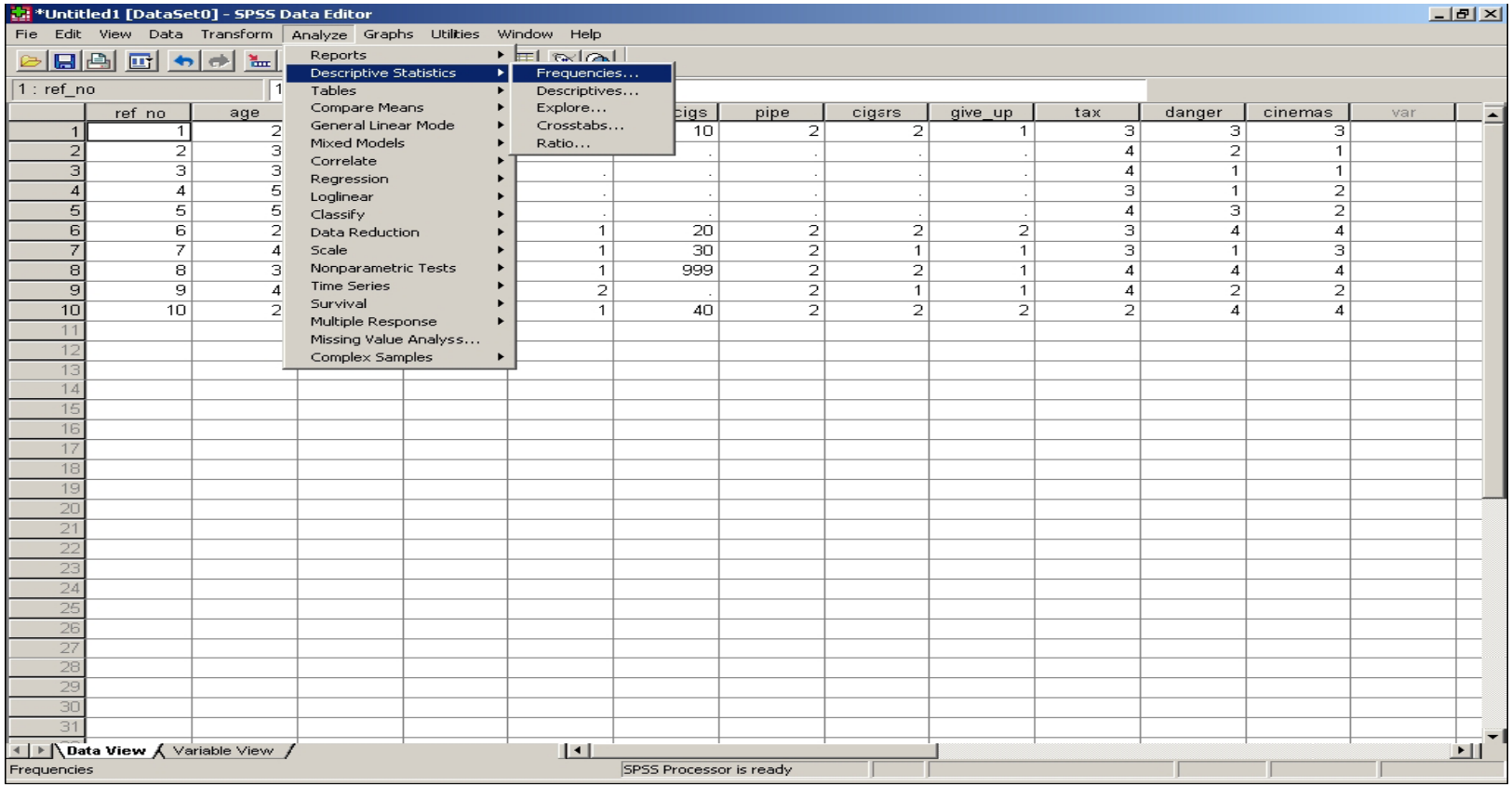
	ref_no	age	sex	smoker	cigs	num_cigs	pipe	cigars	give_up	tax	danger	cinemas	var
1	1	27	F	1	1	10	2	2	1	3	3	3	
2	2	31	M	2	.	.	.	.	.	4	2	1	
3	3	35	M	2	.	.	.	.	.	4	1	1	
4	4	58	M	2	.	.	.	.	.	3	1	2	
5	5	56	M	2	.	.	.	.	.	4	3	2	
6	6	25	F	1	1	20	2	2	2	3	4	4	
7	7	41	F	1	1	30	2	1	1	3	1	3	
8	8	38	F	1	1	999	2	2	1	4	4	4	
9	9	43	F	1	2	.	2	1	1	4	2	2	
10	10	29	M	1	1	40	2	2	2	2	4	4	
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
31													

Data View | Variable View

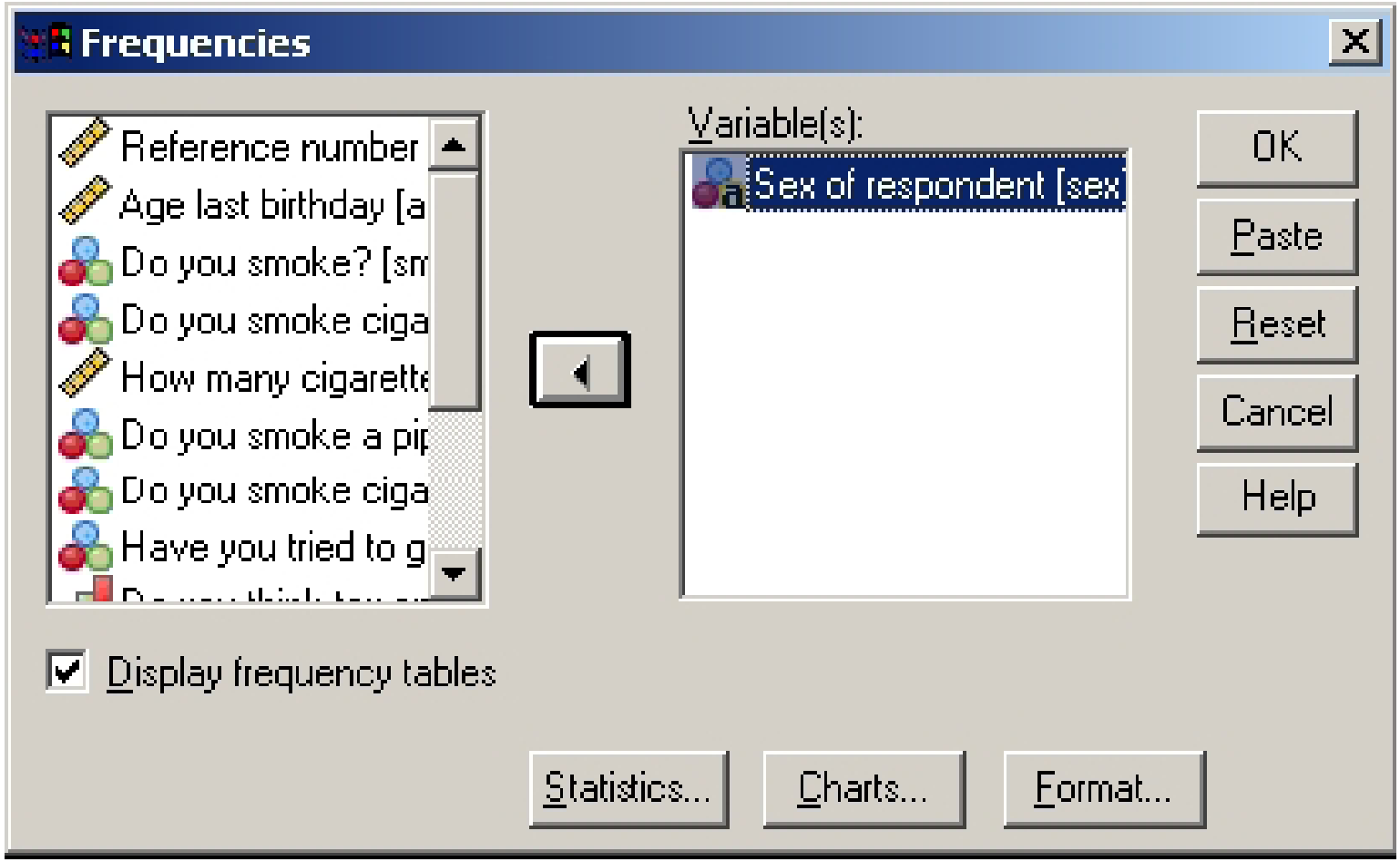
SPSS Processor is ready

Frequency Tables - the frequencies procedure

- In the **Data Editor window select Analyze.**
- From the **Analyze menu select Descriptive Statistics.**

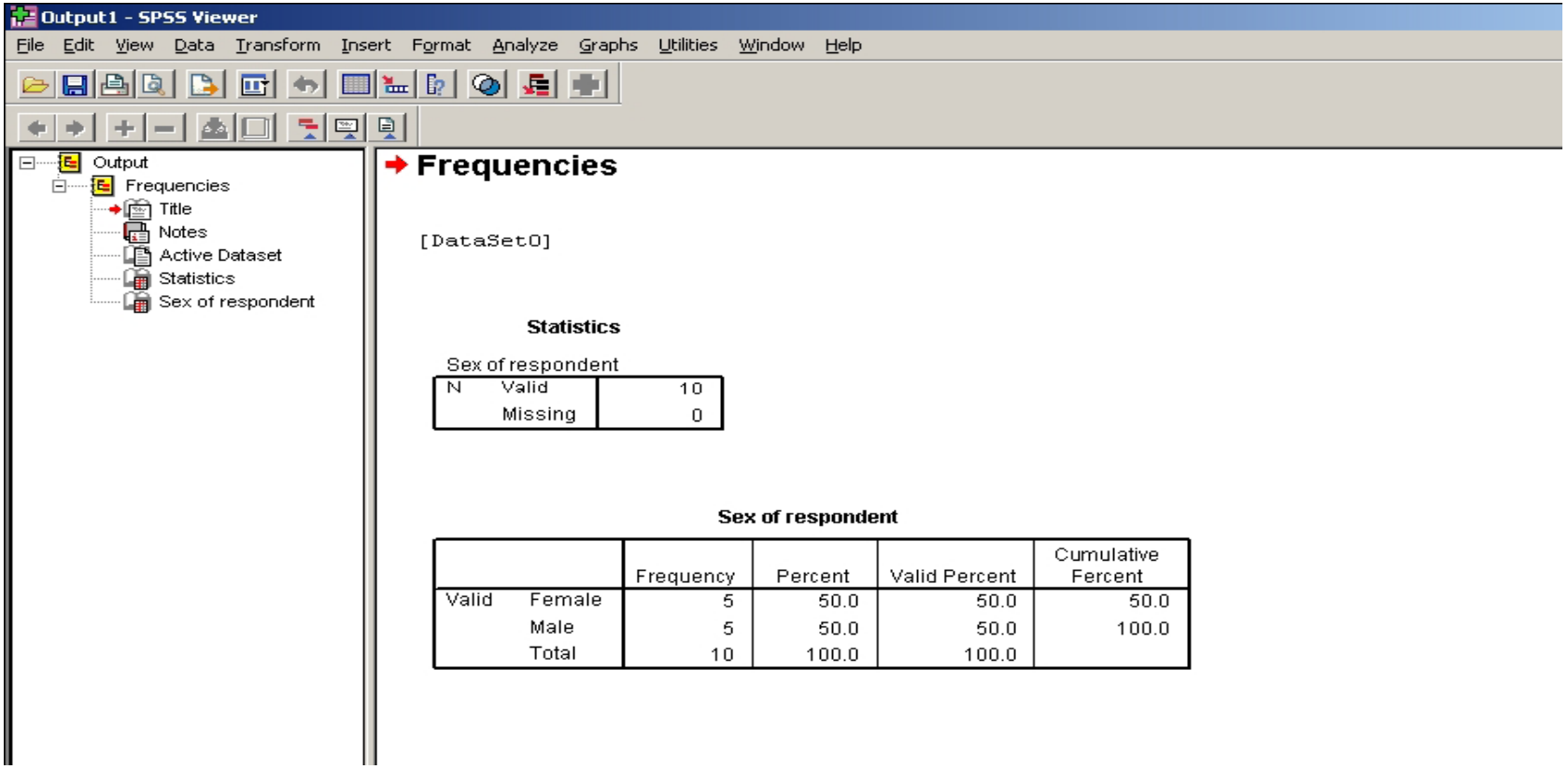


- From the **Descriptive Statistics** submenu, select **Frequencies**



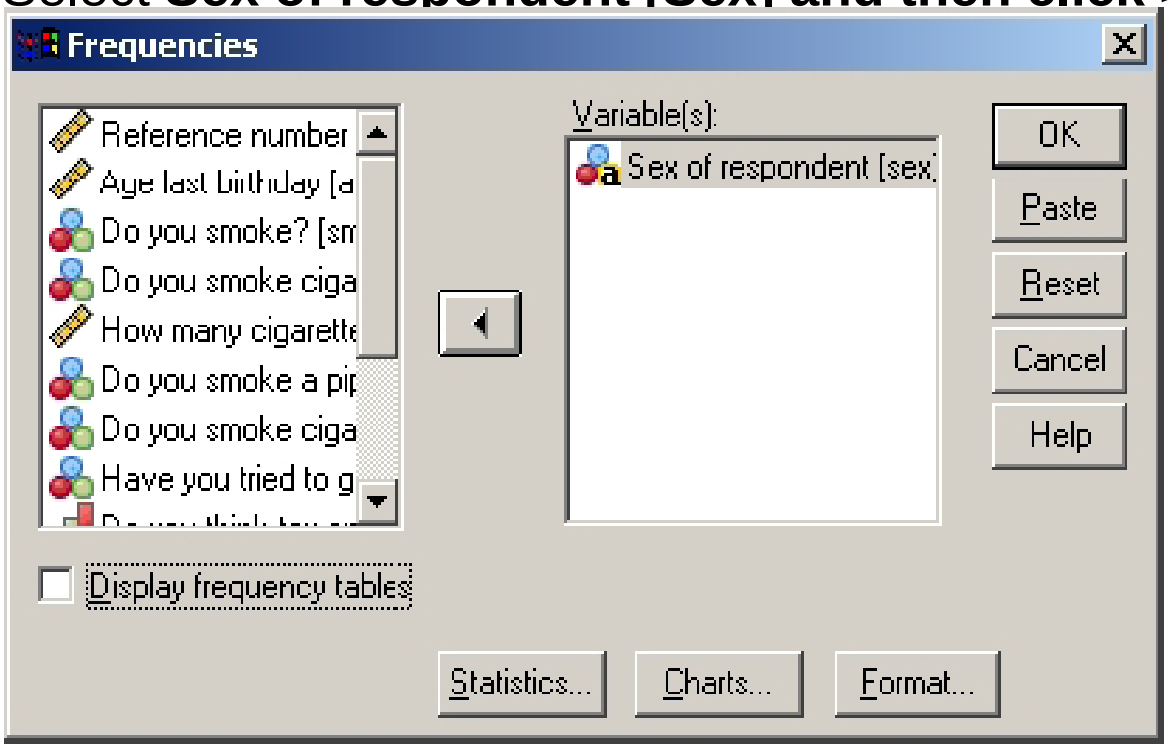
- Select **Sex of respondent.**
- Click the right pointing arrow head (**>) to move sex into the Variables box.**
- Click **OK.**

A frequency table is produced. Note that tables, statistics and charts are displayed in the **SPSS Viewer window – a completely different window from the Data Editor**

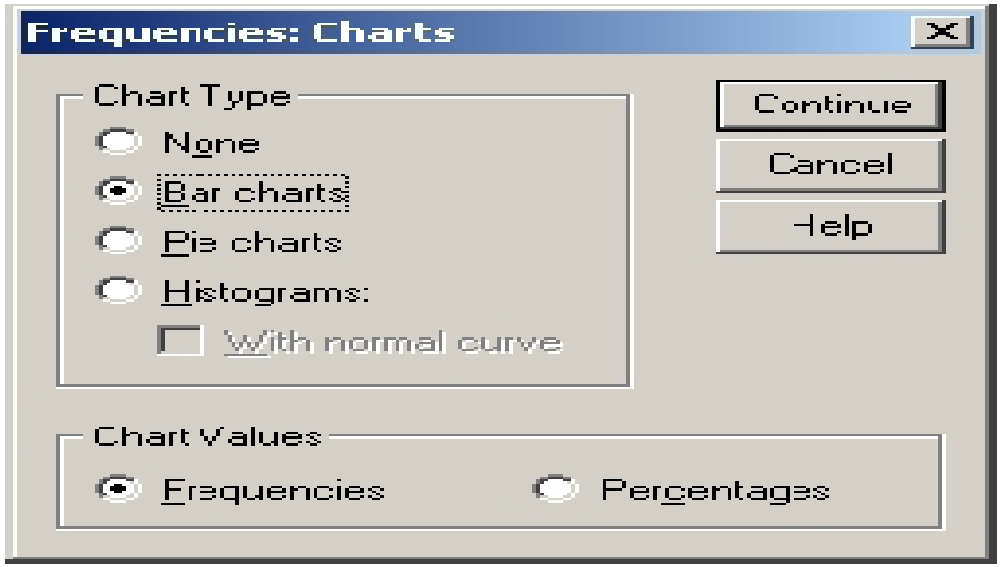


Producing a bar chart from frequencies

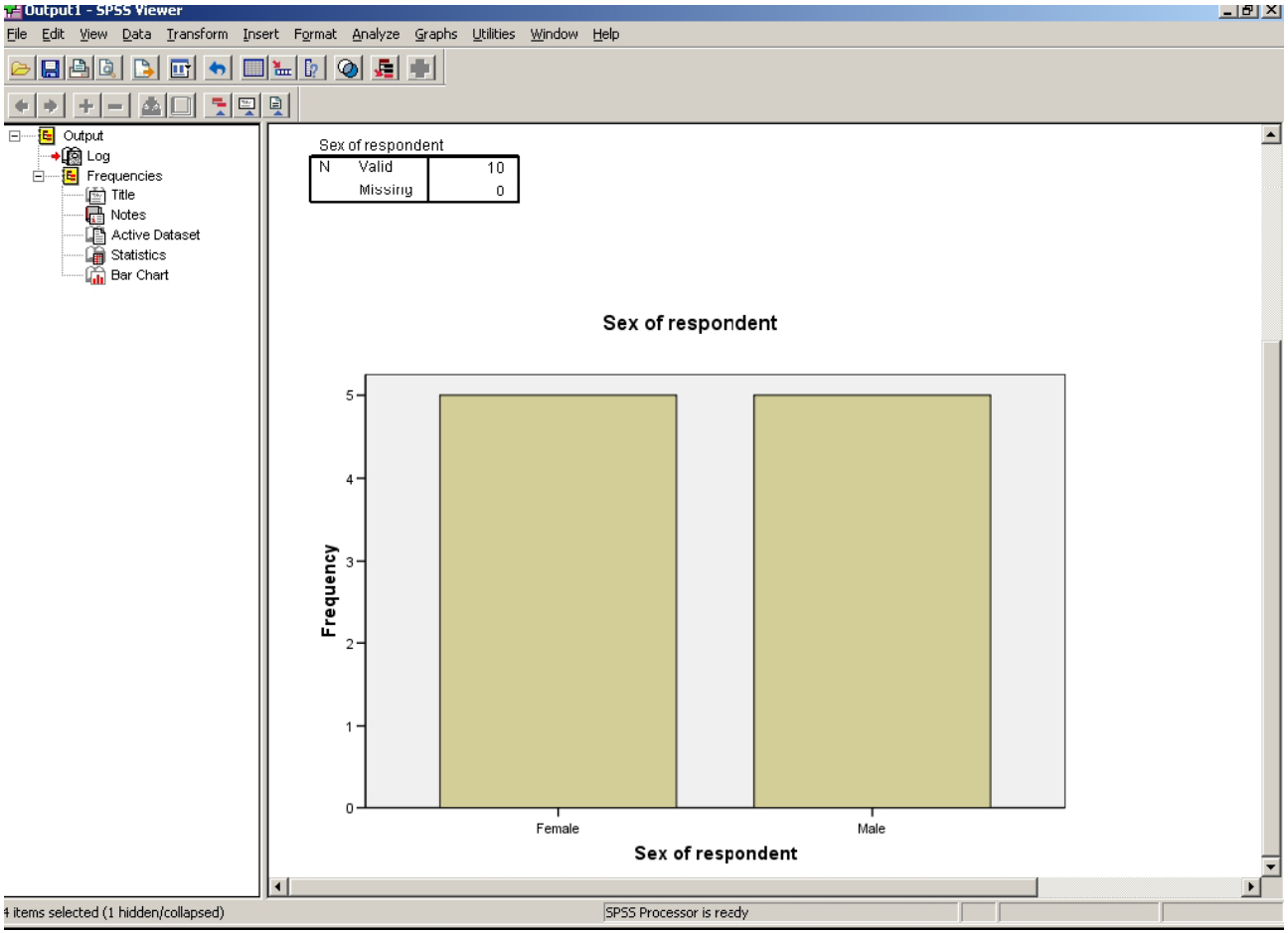
- In the **Viewer or Data Editor window click Analyze.**
- From the **Analyze menu, click Descriptive Statistics.**
- From the **Descriptive Statistics submenu, click Frequencies.**
- Click **Reset.**
- Select **Sex of respondent [Sex] and then click >**



- Click **Charts**



- Click the **Bar chart(s) option and click on Continue.**
- Click **Display Frequency Tables to suppress the display of the frequency table.**
- Click **OK.**



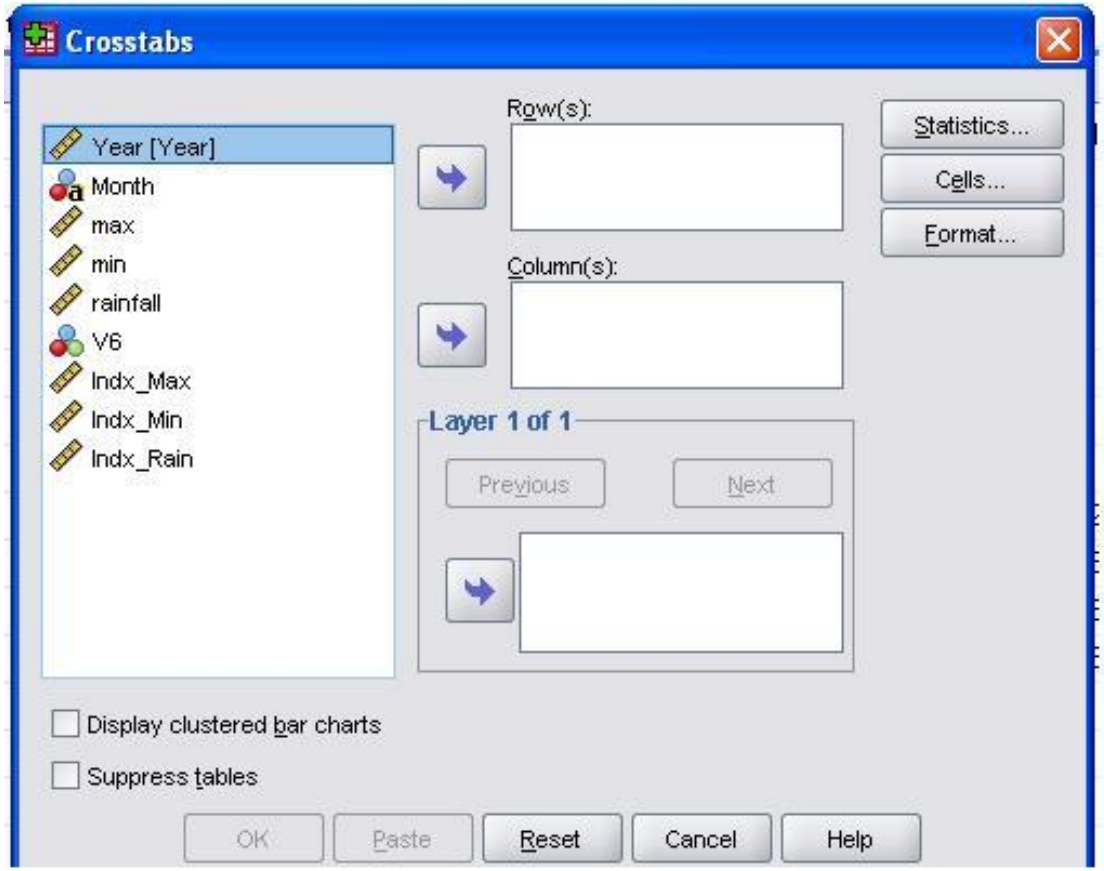
SPSS viewer output produced using the Bar Chart option from the Frequencies procedure

Crosstabulation

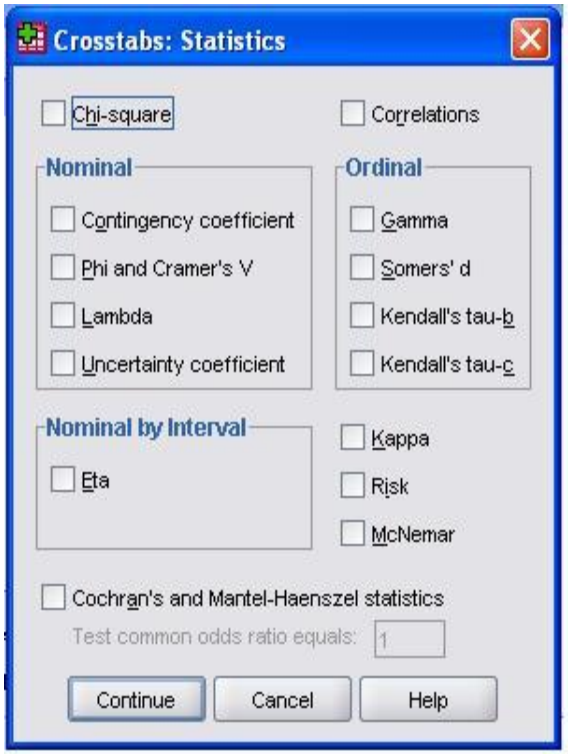
To get a crosstabulation:

- Select **Analyze.**
- Select **Descriptive Statistics.**
- Select **Crosstabs.**
- Select **Do you smoke? [smoker] from the source variable list.**
- Click **> adjacent to the Row(s) text box.**
- From the source variable list select **Sex of respondent [sex].**
- Click **> adjacent to the Column(s) text box.**
- To see the crosstabulation click **OK.**
- SPSS produces a crosstabulation of **smoker by sex.**

The cells of the table show the Counts.



Click Statistics



Adding cell percents and the chi-square statistic

Select Analyze.

- Select **Descriptive Statistics.**
- Select **Crosstabs.**

The table contents can be changed by clicking on the **Cells button and specifying** options. Some useful options are:

Expected prints expected values

Row includes row percentages

Column includes column percentages

Modify the Crosstabs table to request statistics and include the options Row, Column, Total and Expected as follows:

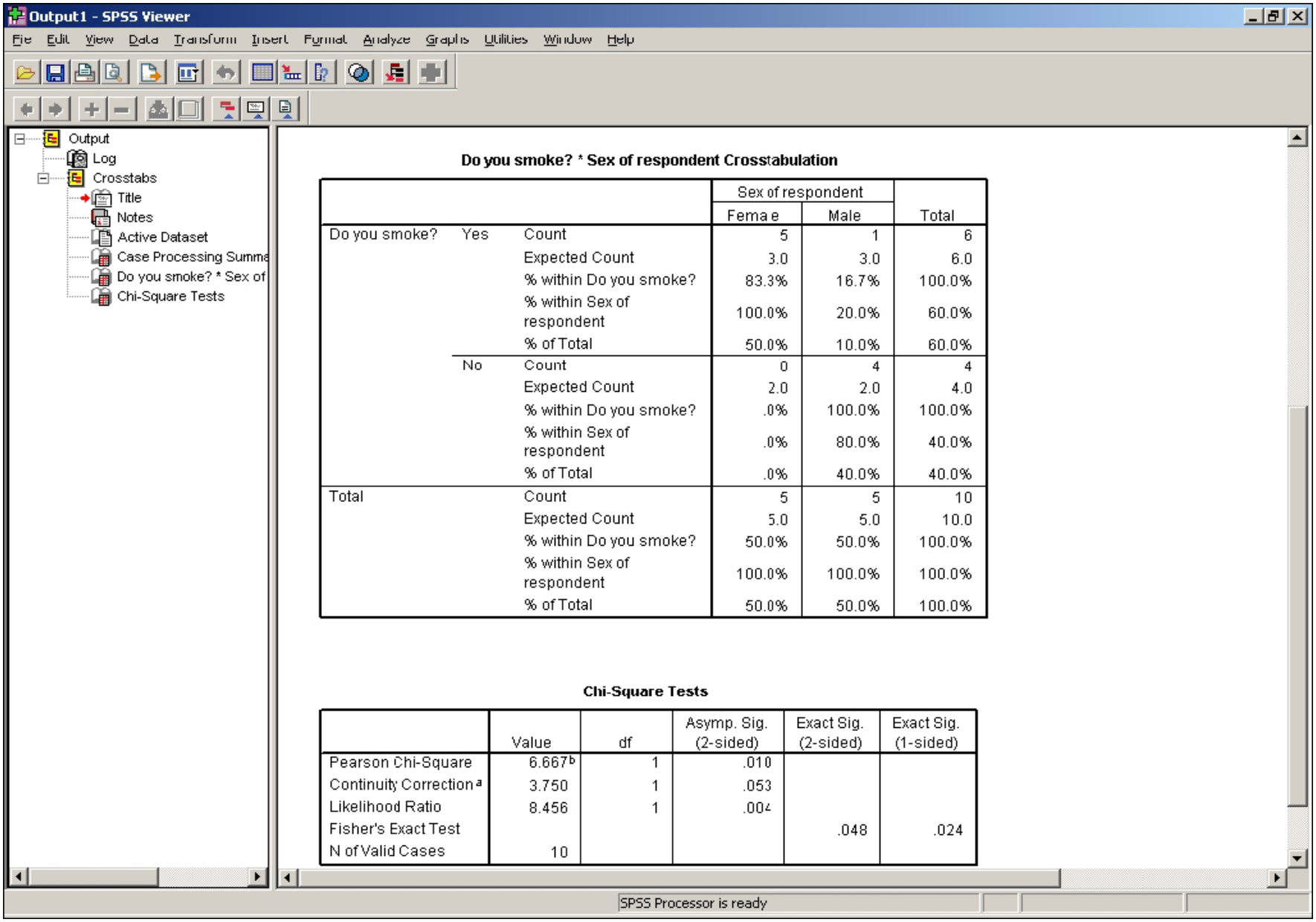
- Click the **Cells button.**
- Select the additional options **Expected, Row Percentages, Column Percentages** and **Total Percentages.**
- Click **Continue.**

The **Statistics button in the Crosstabs window requests statistics.**

- Click the **Statistics button.**

Chi-square requests a Chi-Square (χ 2) test of independence and a Fisher's Exact test when there are fewer than 20 cases in a 2 x 2 table.

- Select the **Chi-square option.**
- Click **Continue.**
- Click **OK**



SPSS viewer output from Crosstabs procedure (with Chi-squared tests)

Correlations

To obtain the Pearson correlation coefficients of **tax, danger and cinemas:**

- Select **Analyze.**
- Select **Correlate.**
- Select **Bivariate**

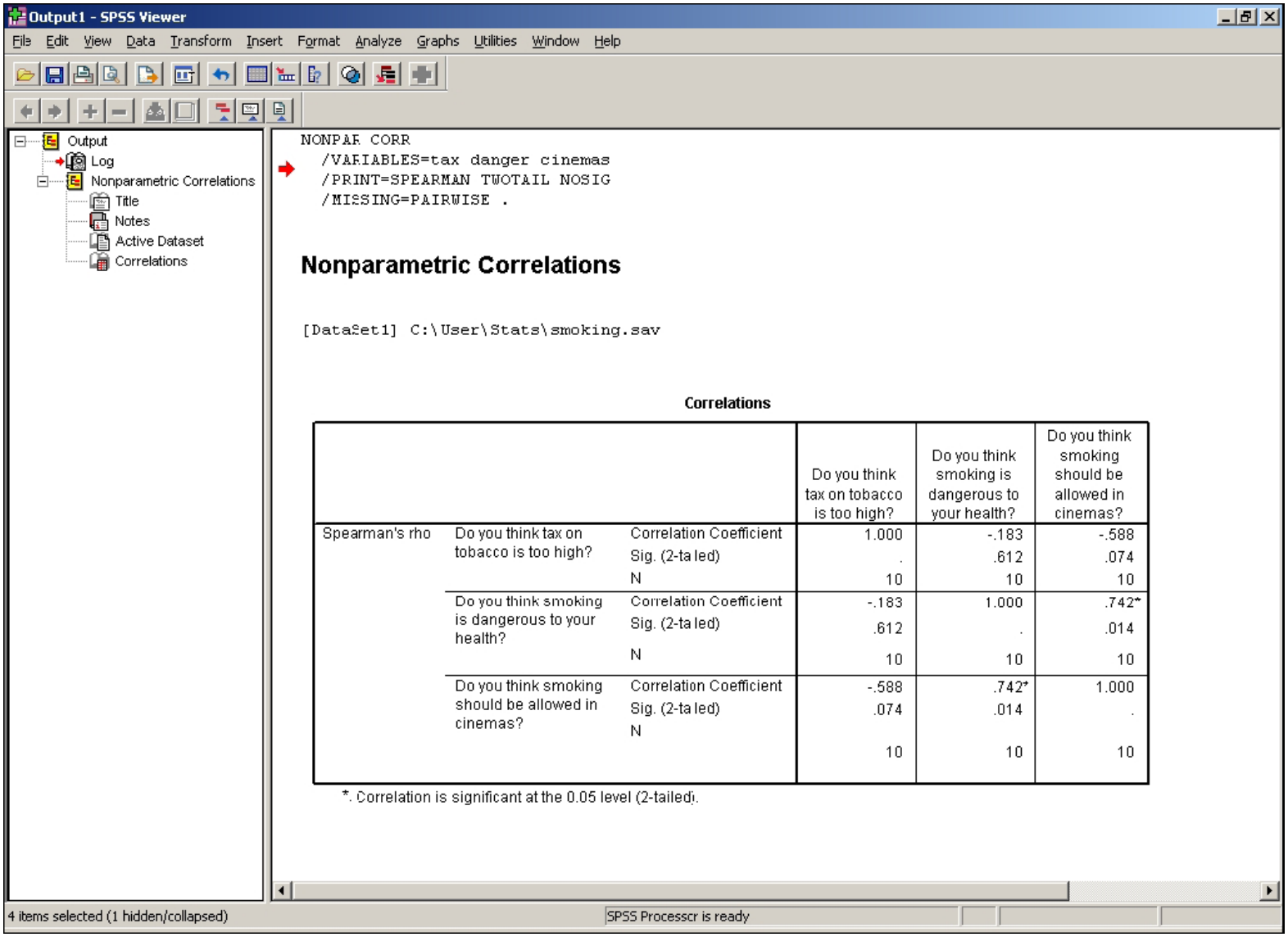
- Move the following to the **Variables box:**

• Do you think smoking should be allowed in cinemas? (cinemas)

• Do you think smoking is dangerous to your health? (danger)

• Do you think tax on tobacco too high? (tax) to the Variables box.

- Click **Flag significant correlations to put a tick in the box.**
- Click **OK.**

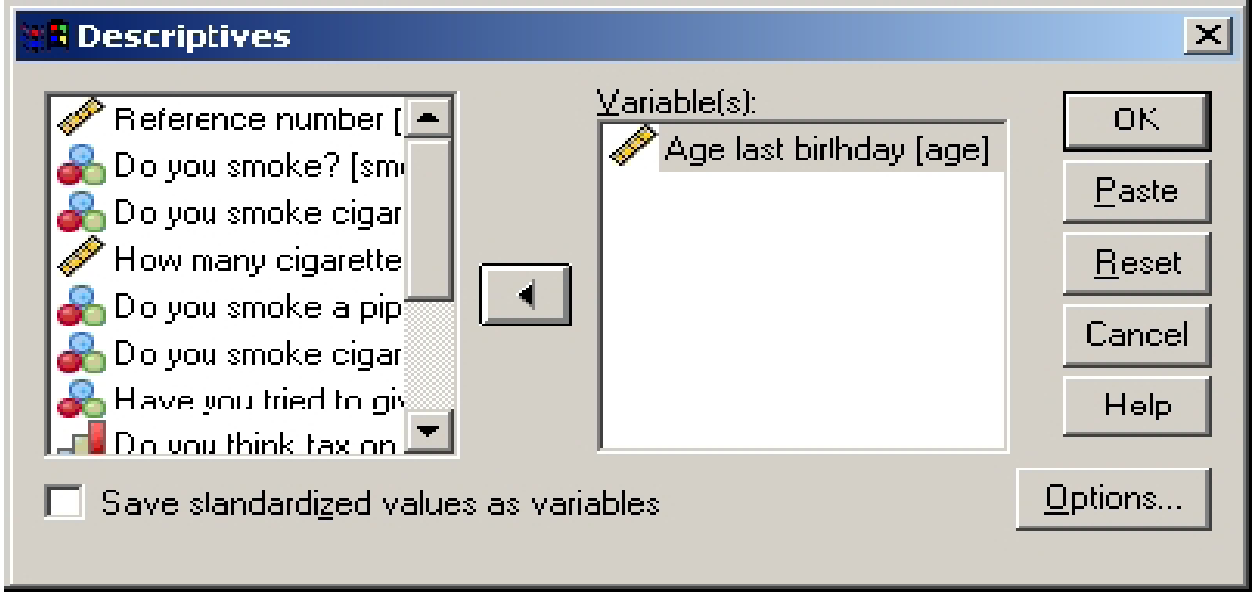


SPSS viewer output from Spearman's Rank correlation

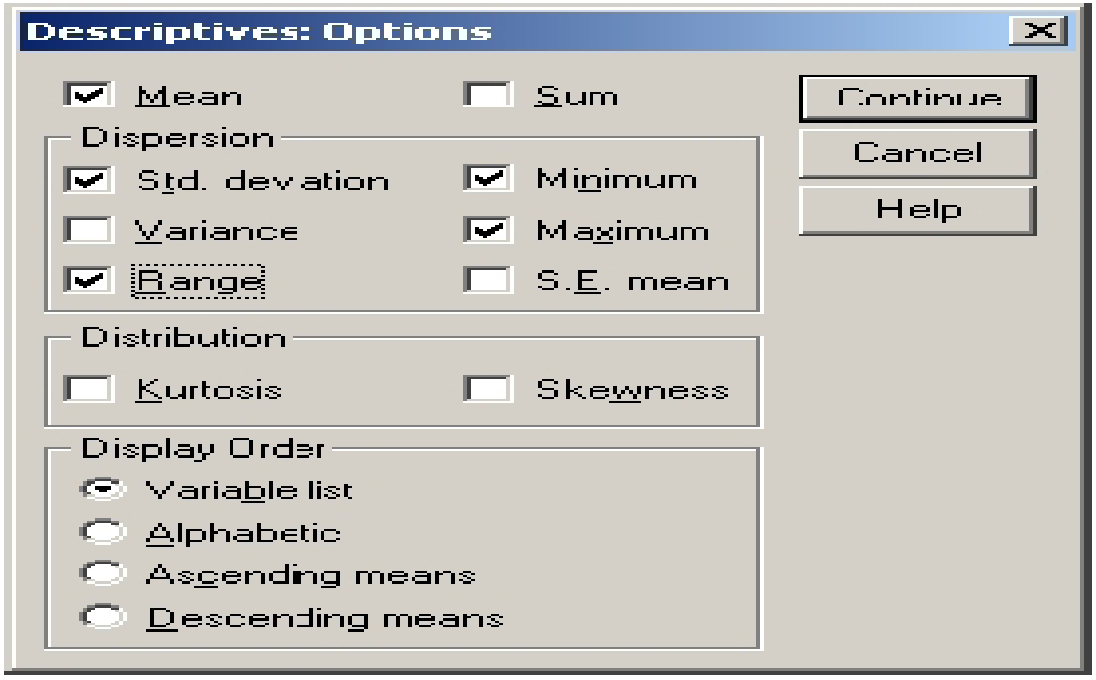
Using summary statistics for continuous variables – the Descriptive procedure

In the **Viewer or Data Editor window select Analyze.**

- From the **Analyze menu, select Descriptive Statistics.**
- From the **Descriptive Statistics submenu, select Descriptives.**
- Select the variable **Age last birthday [age]**



Click the **Options button in the Descriptives window.**



- Click **Continue.**
- Click **OK.**

Getting SPSS to read data from other spreadsheet formats e.g. Excel

Let C:\User\Stats there is a file called Large Smoking Data.xls. This is an Excel spreadsheet that has the same 12 variables as you have used so far in this workbook, but with many more cases than you have entered.

To input the Excel file:

- From the **File menu select Open and then select Data.**
- Ensure the directory in the **Look in box is correct. If you are in one of the Computer**

Centre training rooms change the directory to **C:\User\Stats.**

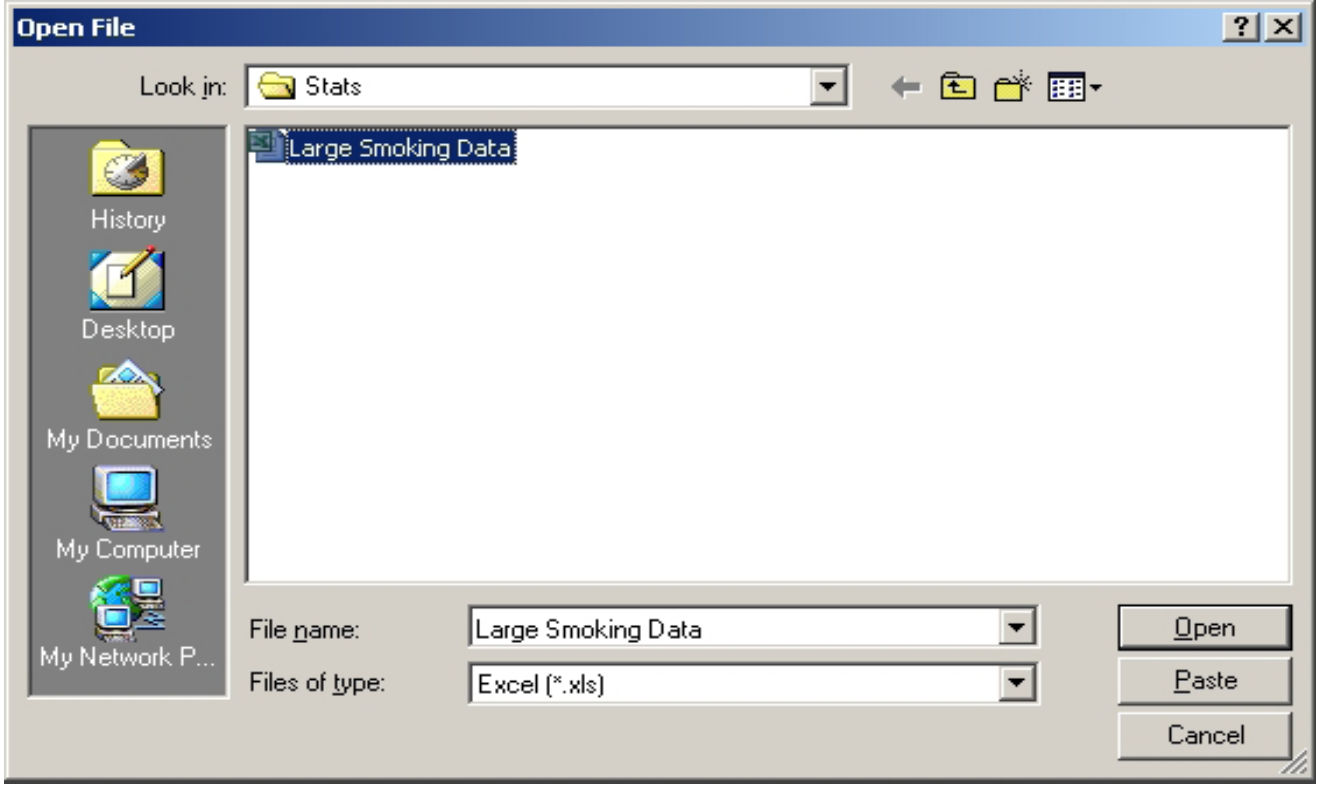
- The **Files of type window will be showing SPSS. This needs to be changed to**

Excel by clicking on the down arrow at the right-hand end of the Files of Type box

and selecting **Excel (*xls). The file name Large Smoking Data should now be**

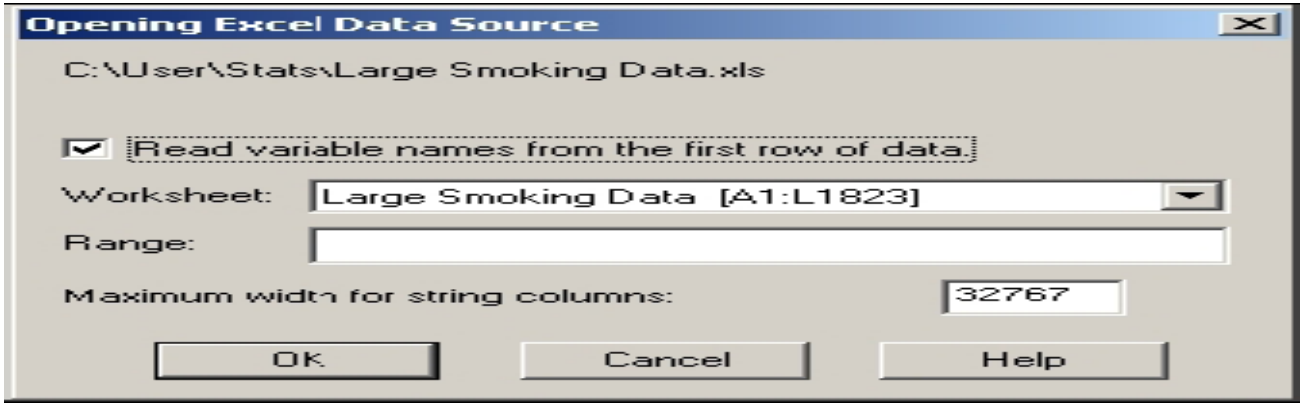
Visible

- Double click the file name.



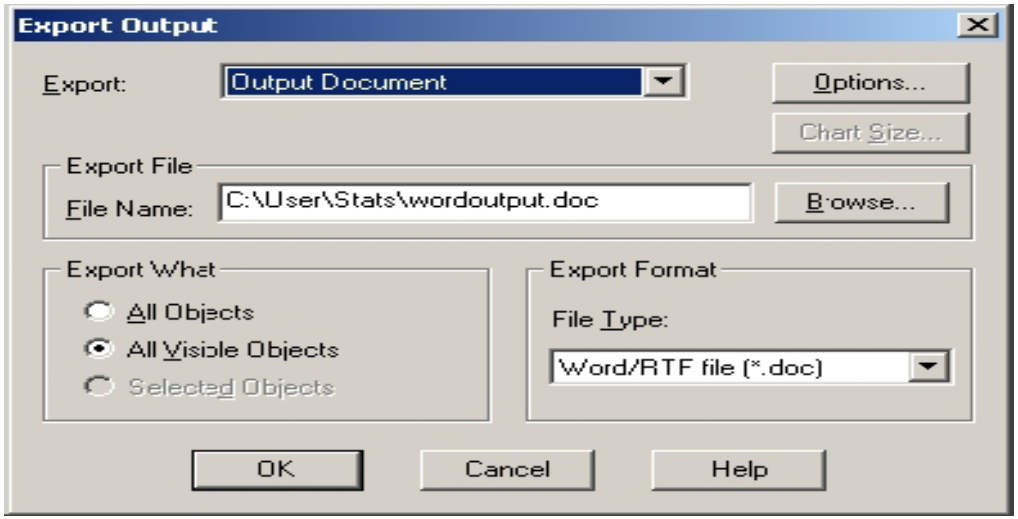
➢The next menu confirms that the Excel file has been recognised. You should check that the box **Read variable names from first row of data is ticked**

➢ Click **OK**



Saving output from SPSS into word processor documents e.g. Microsoft Word

- From the **File menu select Export.**



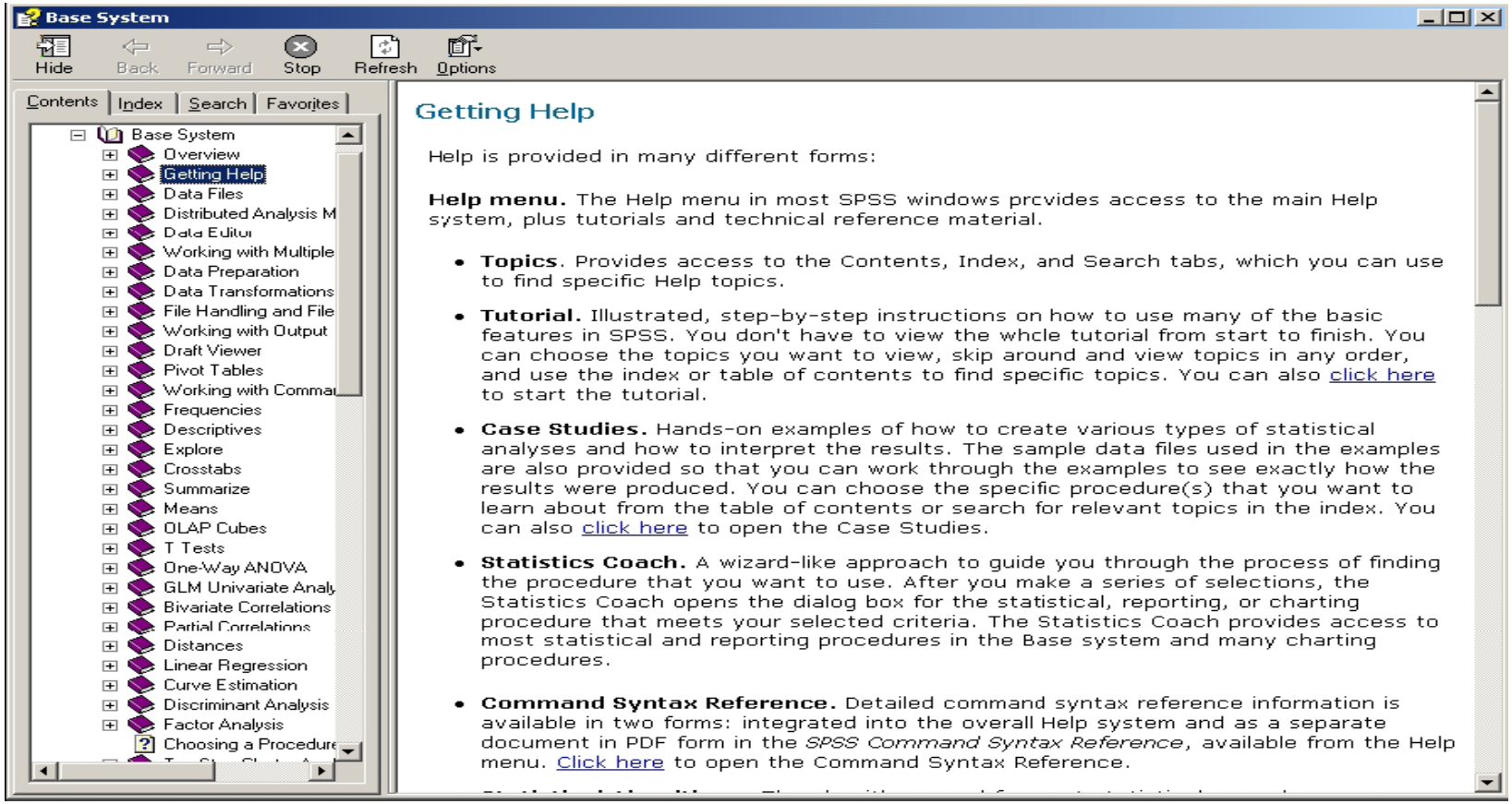
- Under **Export Format, File Type select Word/RTF file (*.doc) from the drop down** menu. **(Make sure you make this choice first).**
- In the **Export File, File Name box type *C:\User\Stats\wordoutput.***
- In the **Export What box, ensure that All Visible objects is selected.**
- Click **OK.**

If Problem Persists

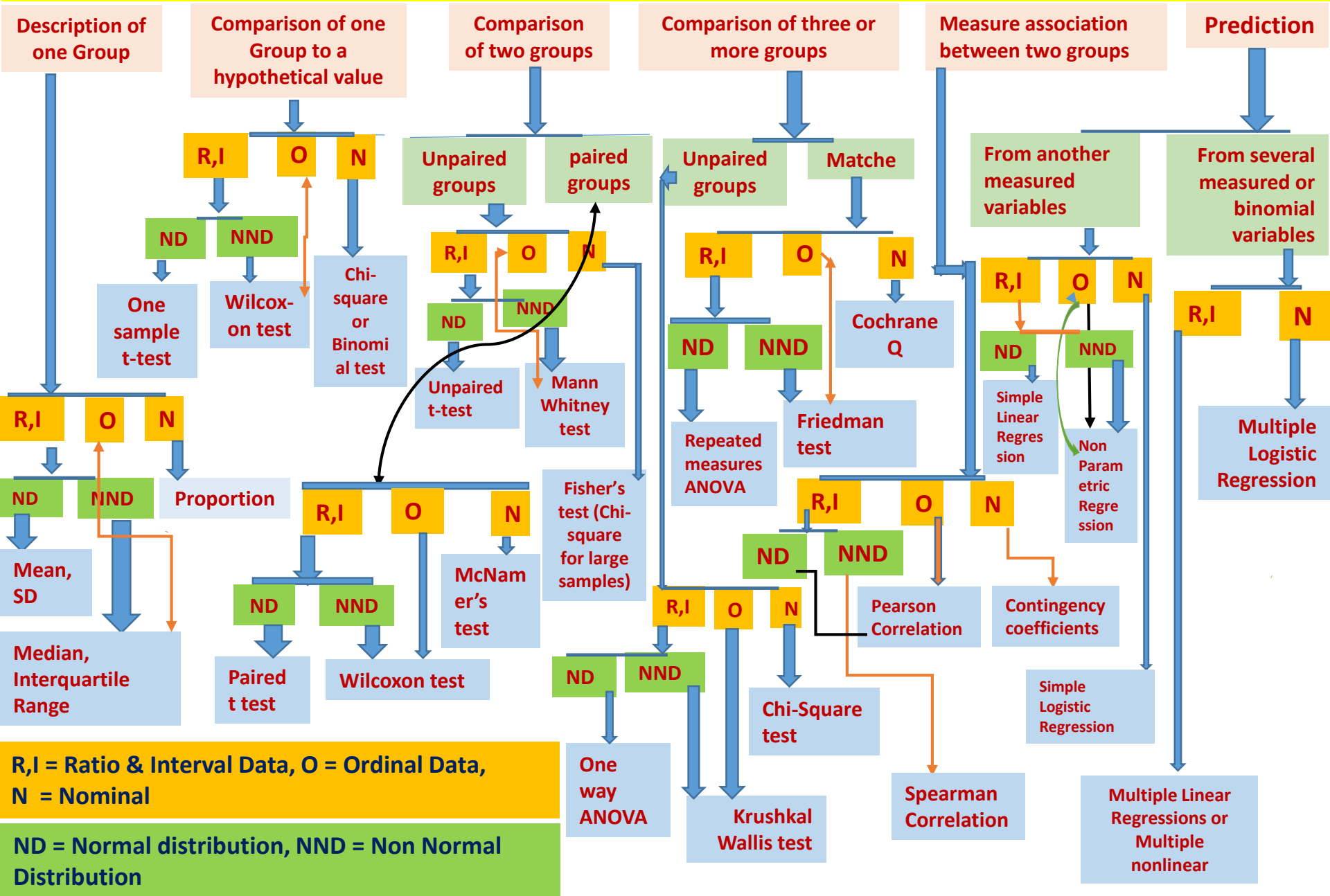
- Click **Help from the main menu.**

A menu appears from which you can choose further topics.

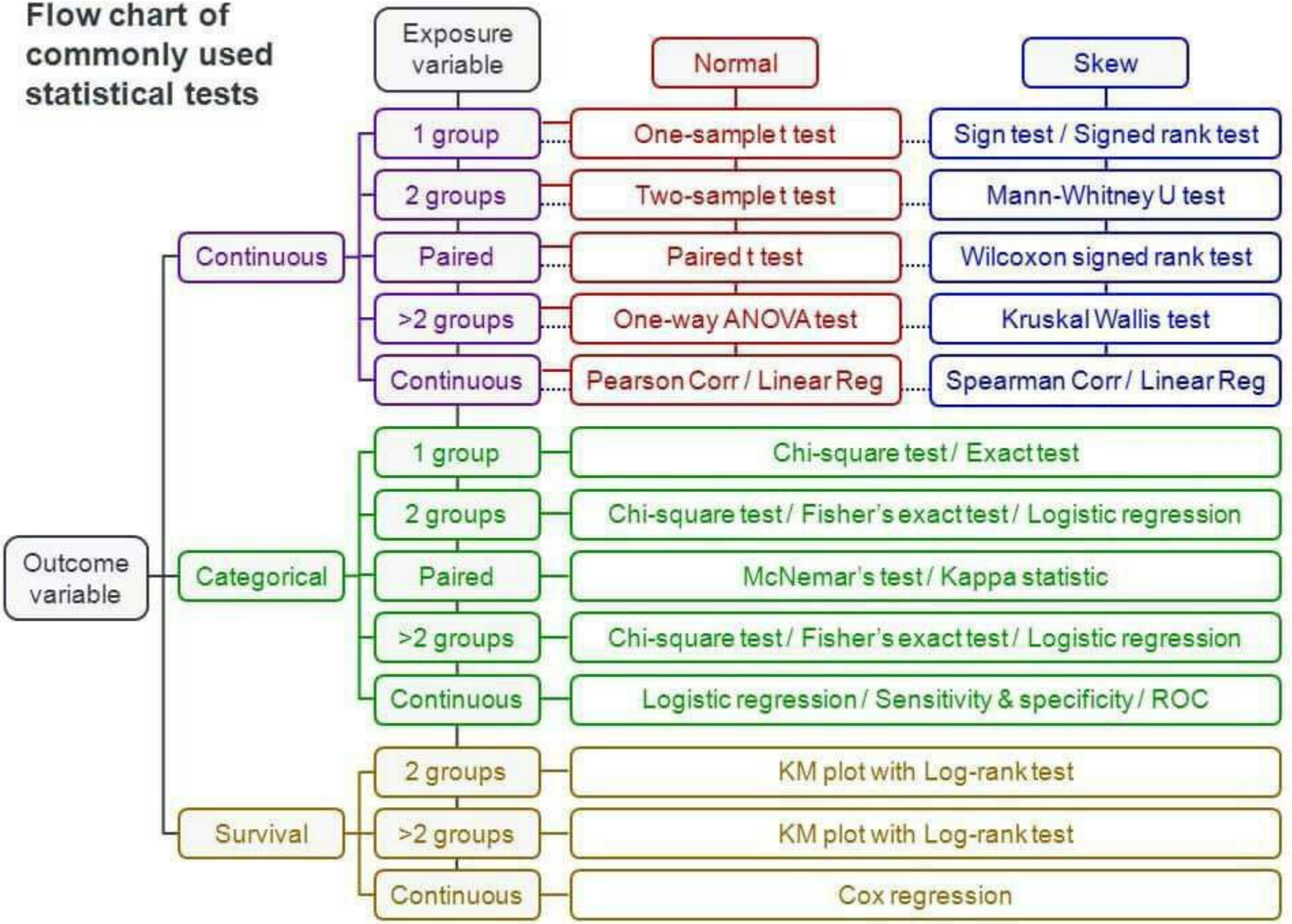
- Click **Topics and an output screen should** open in a new window.



SELECTION OF APPROPRIATE STATISTICAL TESTS



Flow chart of commonly used statistical tests



References

- SPSS for Introductory Statistics Uses and Interpretation by **G.A. Morgan**
- Introduction to SPSS for Windows **Practical Work Book** by University of Bristol Information Services
- **A Handbook of Statistical Analyses using SPSS: *Sabine Landau and Brian S. Everitt* (Chapman & Hall/CRC)**

Normality test using SPSS- How to check whether data are normally distributed.mp4

Thank you

Statistics 101- Is My Data Normal.mp4