Introduction to SPSS

Presented by : edgar

For Statistical and financial accounting class

Outline

- Review of Concepts (stats and scales)
- Data entry (the workspace and labels)
  - By hand
  - Import Excel
- Running an analysis- frequency, central tendency, correlation

Types of Variables

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

  - Brand (Trek, Raleigh,Avon)
  - Type (road, mountain, racer)
  - Components (Shimano, no name)
  - Age
  - Condition (Excellent, good, poor)
  - Price
  - Frame size
  - Number of gears

Types of Scales

- **Nominal**- objects or people are categorized according to some criterion (gender, job category)
- **Ordinal-** Categories which are ranked according to characteristics (income- low, moderate, high)
- **Interval-** contain equal distance between units of measure- but no zero (calendar years, temperature)
- **Ratio-** has an absolute zero and consistent intervals (distance, weight)

Parametric vs Non-parametric

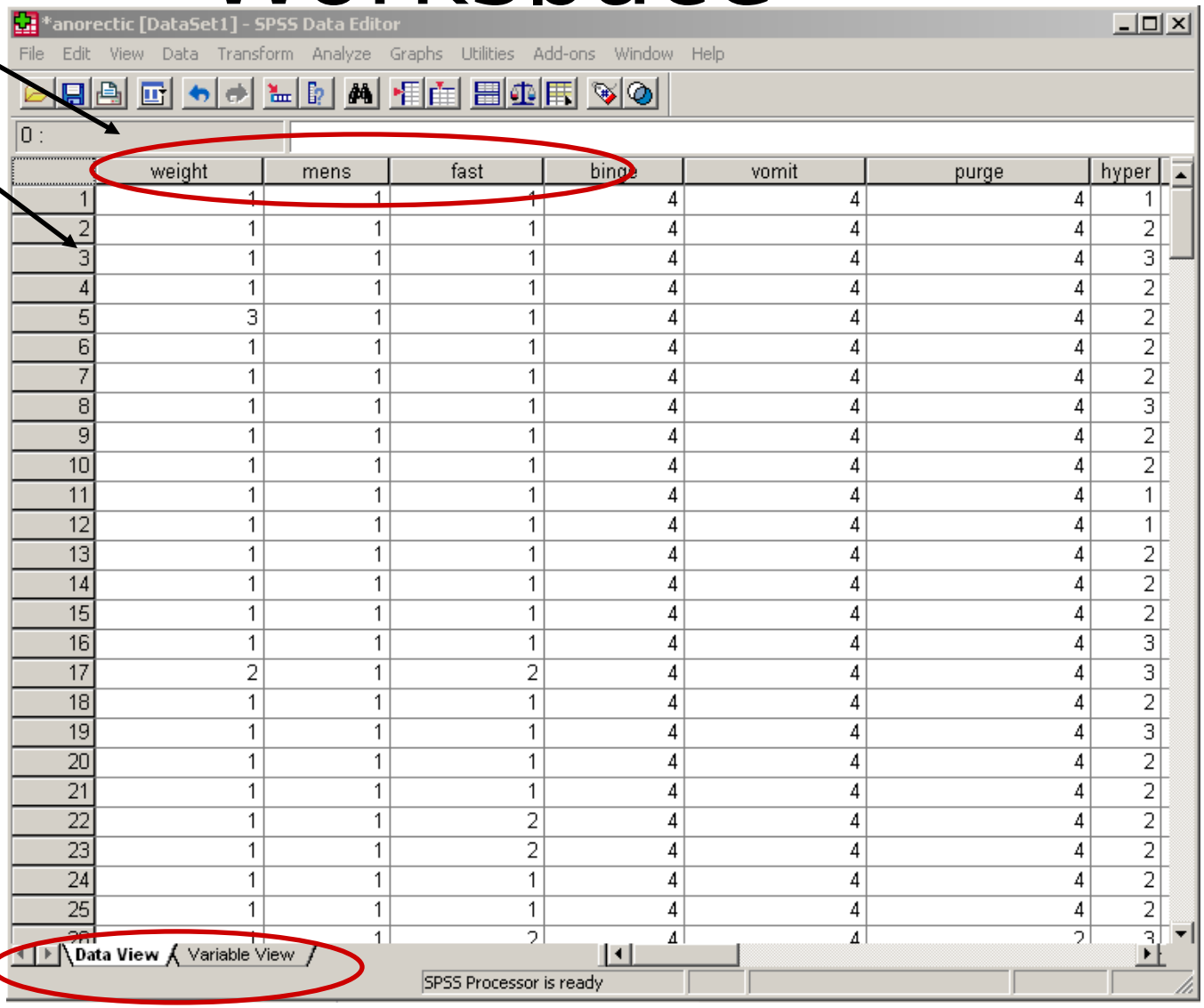
- **Parametric** stats are more powerful than non-parametric stats- for real numbers- T test
- **Non-parametric** stats are not as powerful but good for category variables - Mann-Whitney U (likert)

The SPSS Workspace

Variables

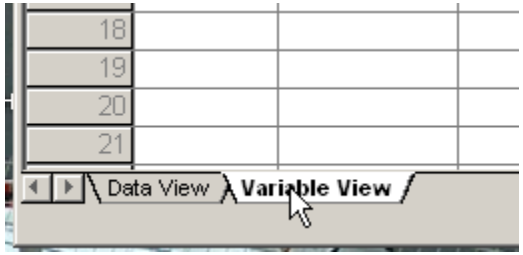
Cases

Toggle between Data and Variable Views

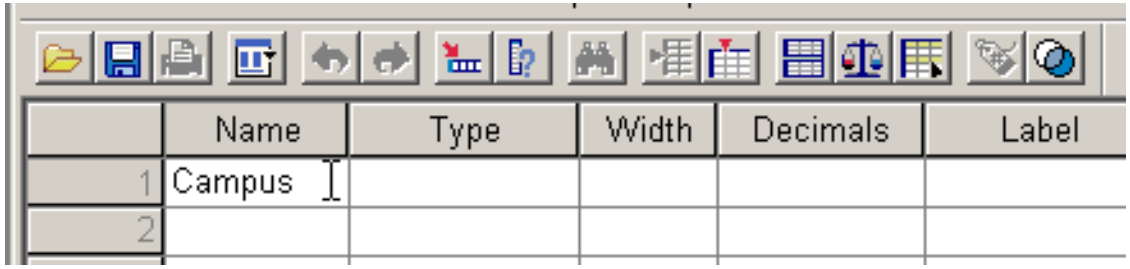


Data Entry (by hand)

1. Click Variable View

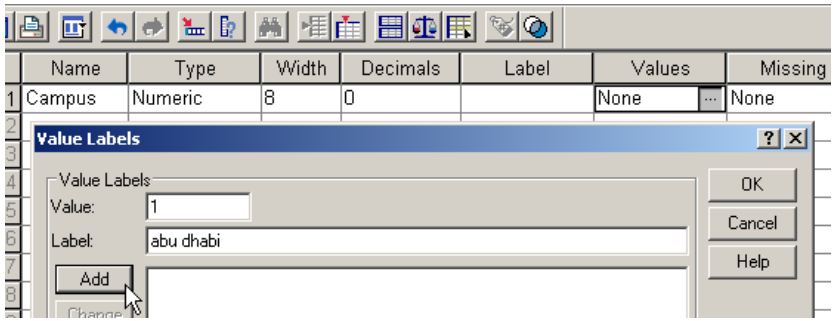


2. Click the Row 1, Name cell and type Campus (no spaces allowed in name)



Data Entry (by hand)

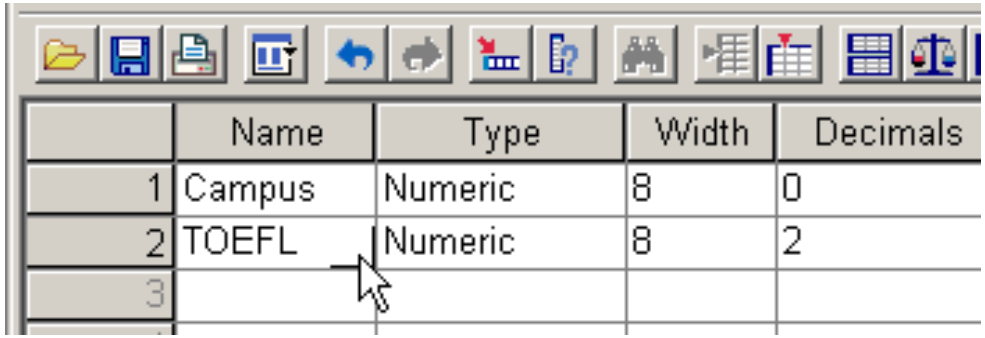
3. Click the Row 1, Values cell and type 1 for the value and abu dhabi for the label- click Add



4. Type 2 for the value and dubai for the label- click Add and then OK

Data Entry (by hand)

5. Click the Row 2, Name cell and type TOEFL



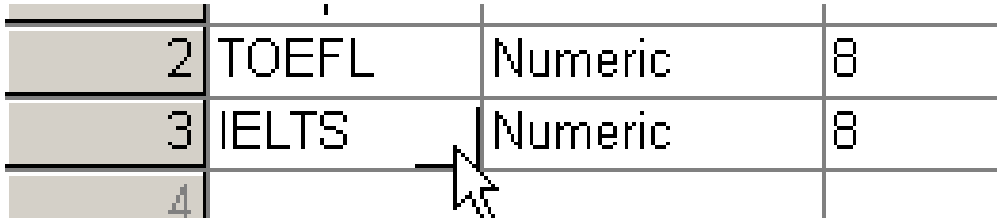
	Name	Type	Width	Decimals
1	Campus	Numeric	8	0
2	TOEFL	Numeric	8	2
3				

6. Click the Row 2, Label cell and type Paper based TOEFL Scores

Decimals	Label	Values	Missing
0		{1, abu dhabi}..	None
2	Paper based TOEFL Scores	None	None

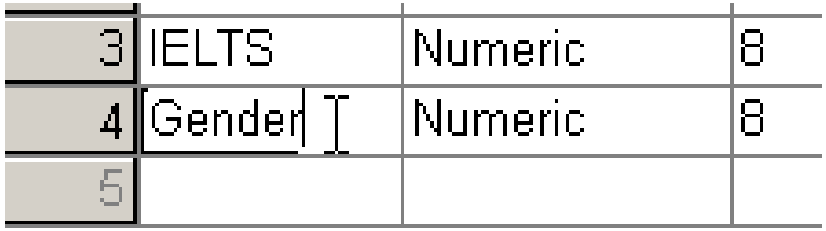
Data Entry (by hand)

7. Click the Row 3, Name cell and type IELTS



2	TOEFL	Numeric	8
3	IELTS	Numeric	8
4			

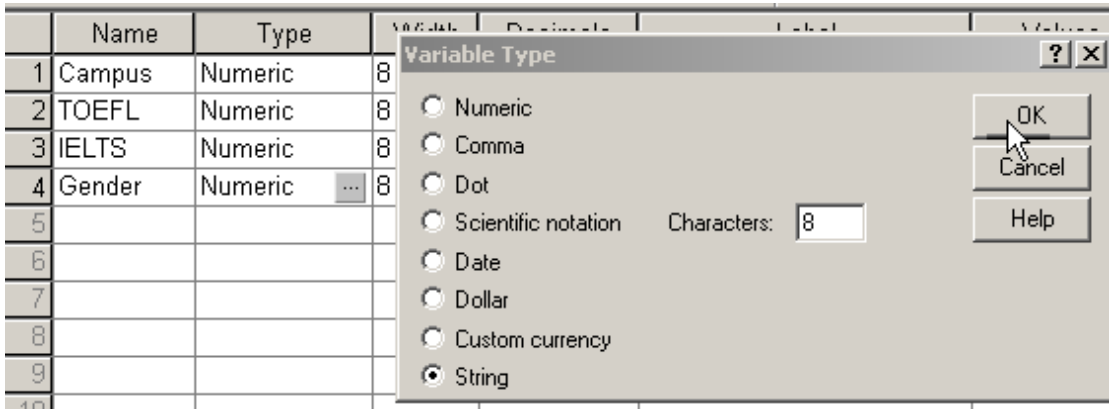
8. Click the Row 4, Name cell and type Gender



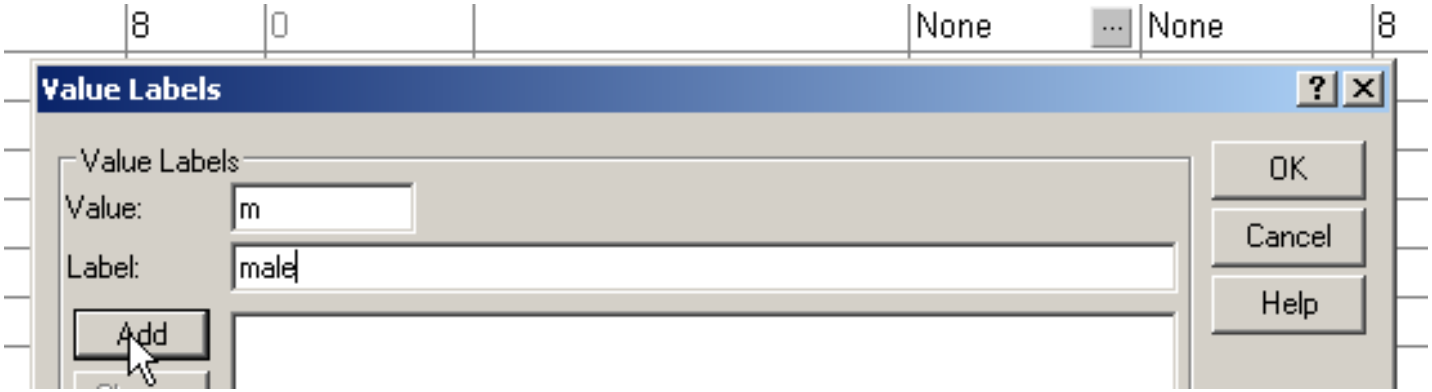
3	IELTS	Numeric	8
4	Gender	Numeric	8
5			

Data Entry (by hand)

9. Click the Row 4, Type cell and click String and click OK

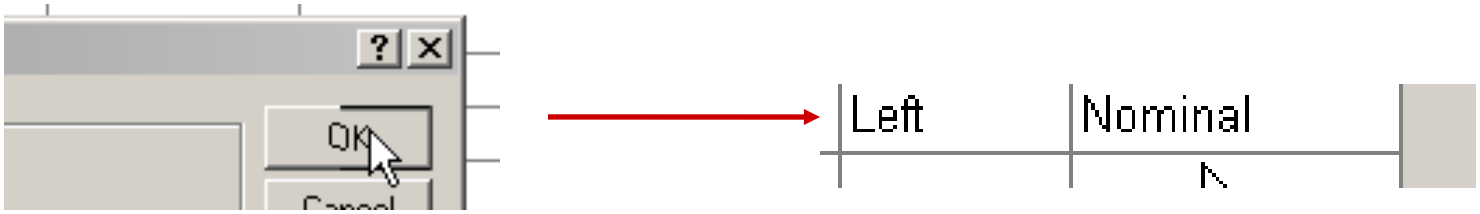


10. Click the Row 4, Values cell and type m for the value and male for the label- click Add

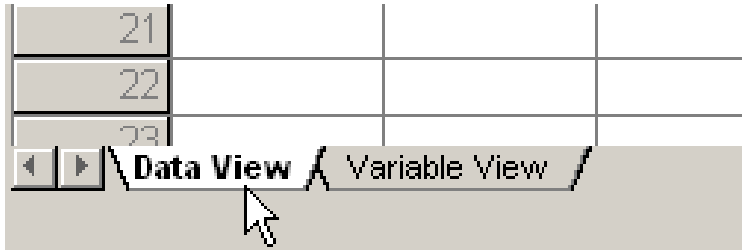


Data Entry (by hand)

11. Type f for the value and female for the label- click Add and then OK (notice the measure is now nominal)



12. Click Data View in the bottom left corner to start entering the data



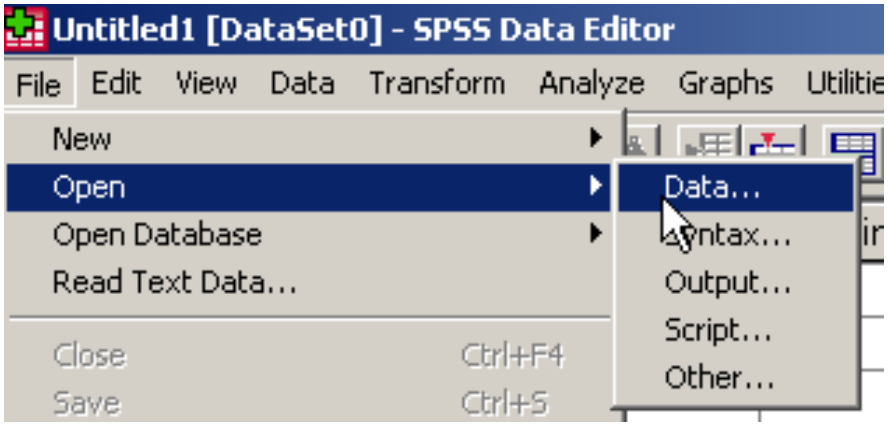
Data Entry (by hand)

13. Click on the cells and enter the data (either type numbers of select from the dropdown menu)

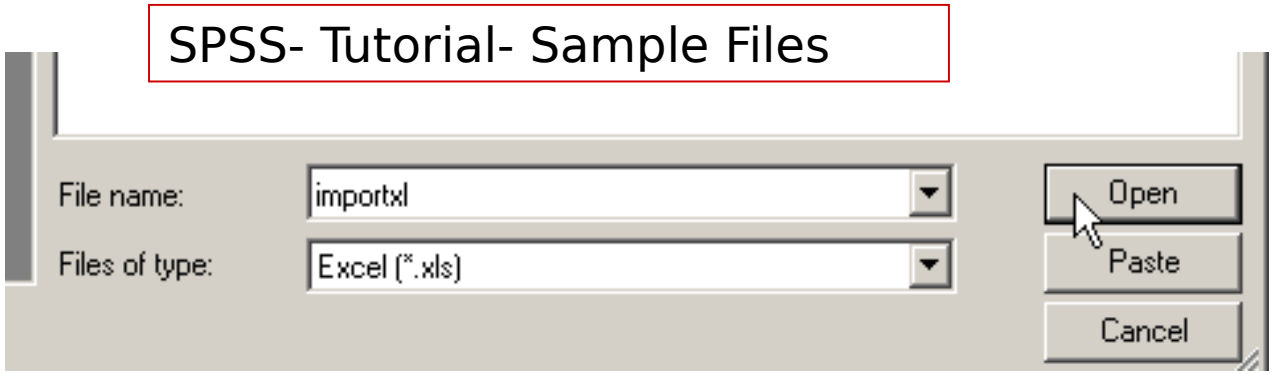
	Campus	TOEFL	IELTS	Gender
1	abu dhabi	502.00	5.00	male
2	dubai	500.00	5.00	male
3	dubai	388.00	4.00	male
4	abu dhabi	433.00	4.50	female
5	dubai	433.00	4.50	male
6	dubai	567.00	7.00	female
7	abu dhabi	600.00	7.00	female

Data Entry (import from Excel)

14. Click Open- Data…

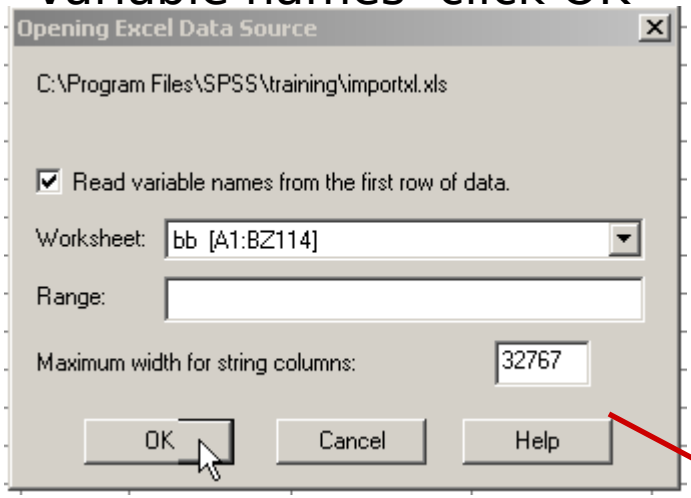


15. Change Files of type to Excel, then browse and open the file.

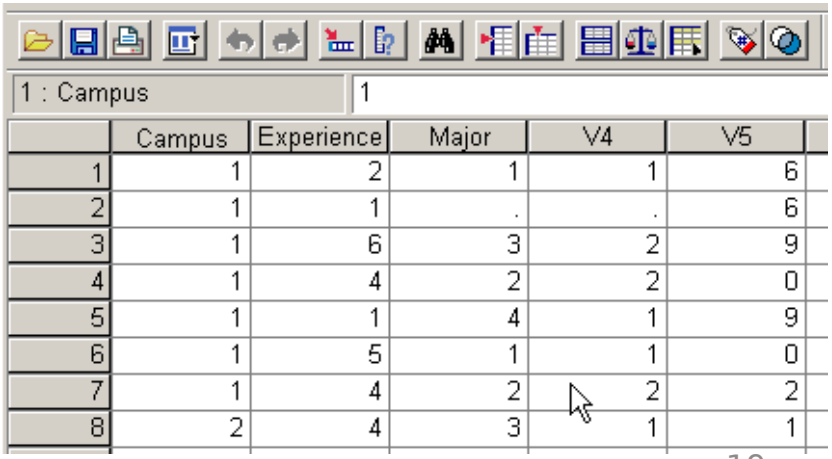


Data Entry (import from Excel)

16. Select the worksheet, the range (if desired), and if to read variable names- click OK



The data and variable names will appear

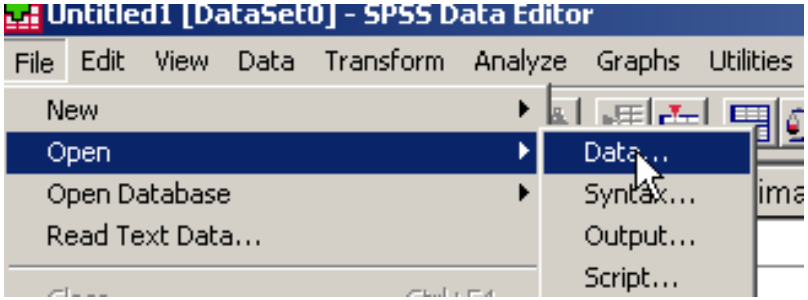


1 : Campus | 1

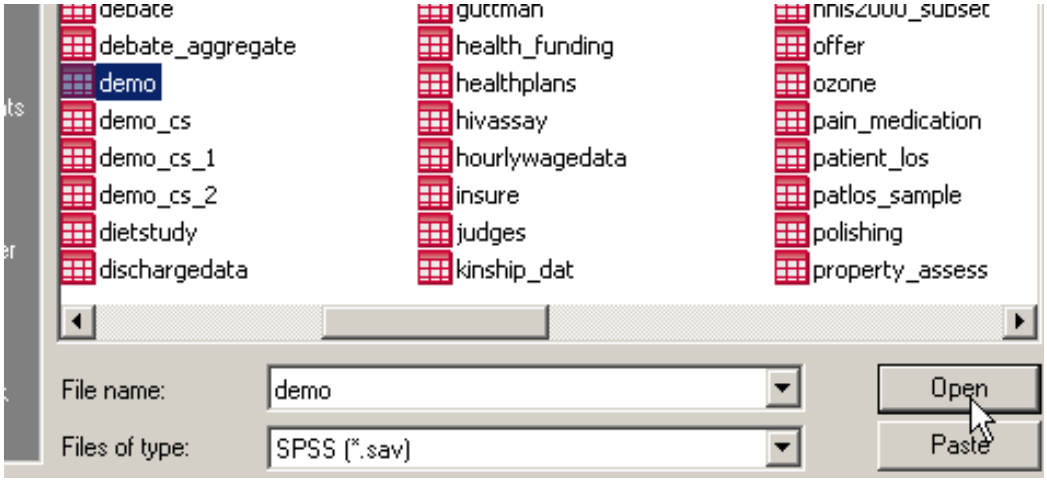
	Campus	Experience	Major	V4	V5
1	1	2	1	1	6
2	1	1	.	.	6
3	1	6	3	2	9
4	1	4	2	2	0
5	1	1	4	1	9
6	1	5	1	1	0
7	1	4	2	2	2
8	2	4	3	1	1

Running Analyses

17. With SPSS open, select file- Open- Data

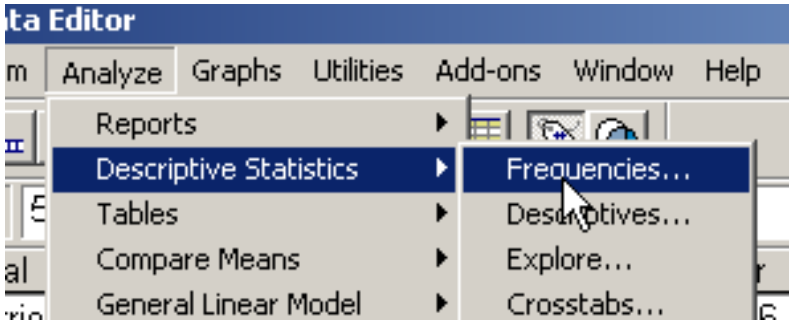


18. Navigate to SPSS- Tutorial- sample_files- select demo, click Open

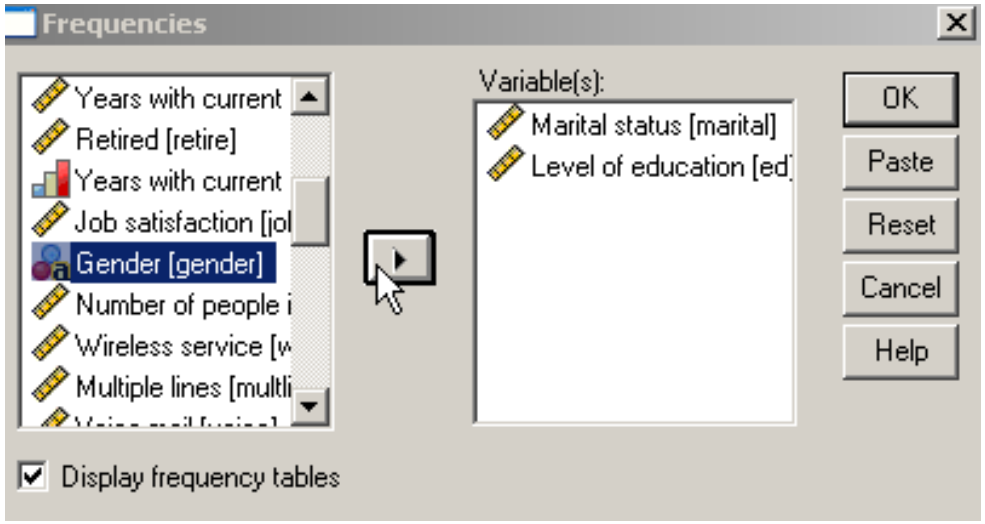


Running Analyses (Frequency)

19. Select Analyze- Descriptive Stats- Frequencies

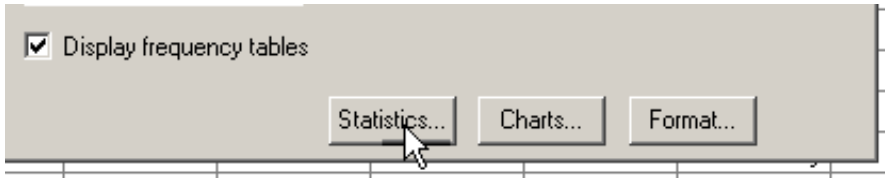


20. Select the desired variables and click the arrow to move them to the right side

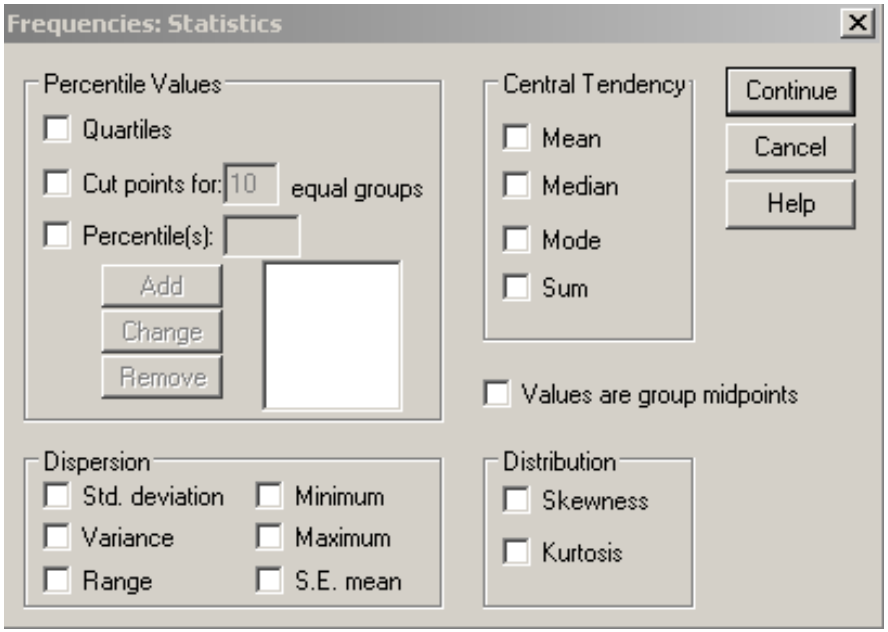


Running Analyses (Frequency)

21. Click Statistics



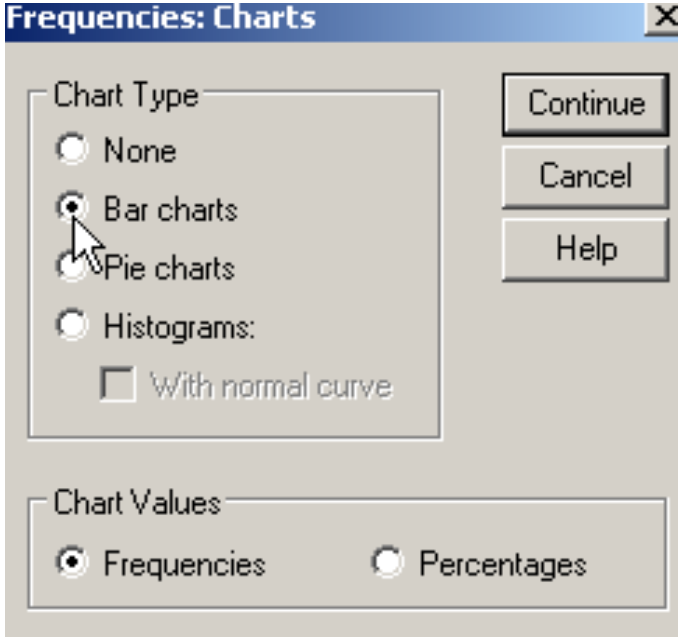
22. Select any stats that you want to see, click Continue



Running Analyses (Frequency)

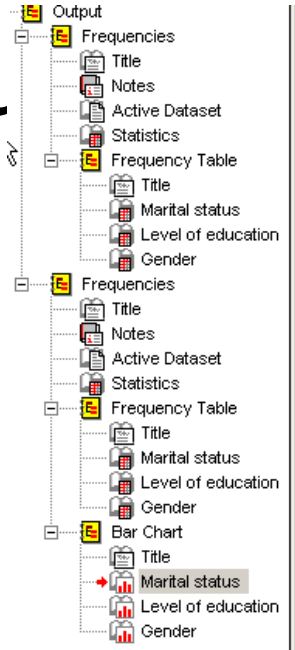
23. Click Charts

24. Select the type of chart you want, click Continue, then OK



Result Tables and Graphs will appear

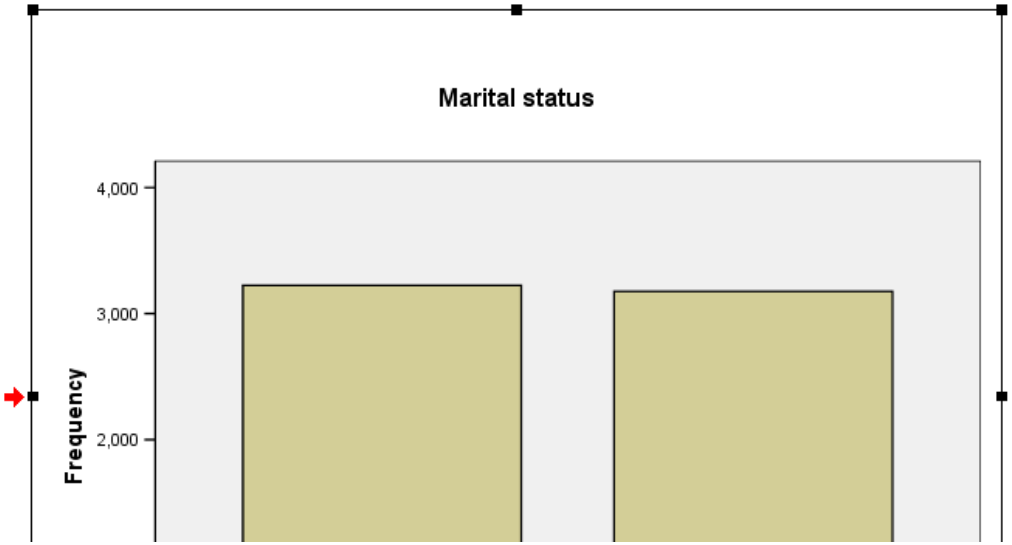
Ru y)



Gender

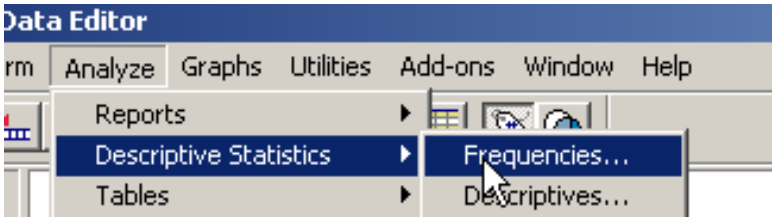
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	3179	49.7	49.7	49.7
	Male	3221	50.3	50.3	100.0
	Total	6400	100.0	100.0	

Bar Chart

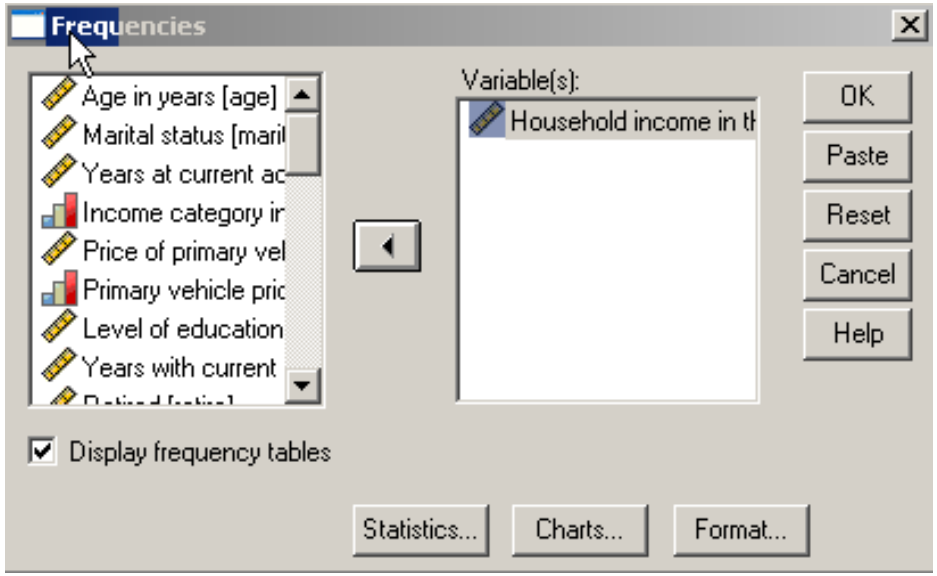


Running Analyses (Central Tendency)

25. Select Analyze- Descriptive Stats- Frequencies

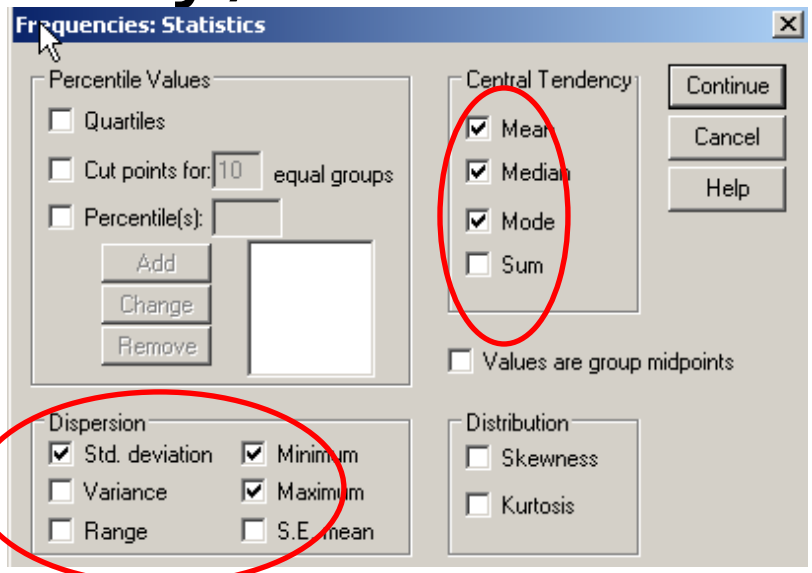


26. Select the desired variables (household income) and click the arrow to move them to the right side



Running Analyses (Central Tendency)

27. Select some measures of central tendency and dispersion- click Continue then OK



Results will appear

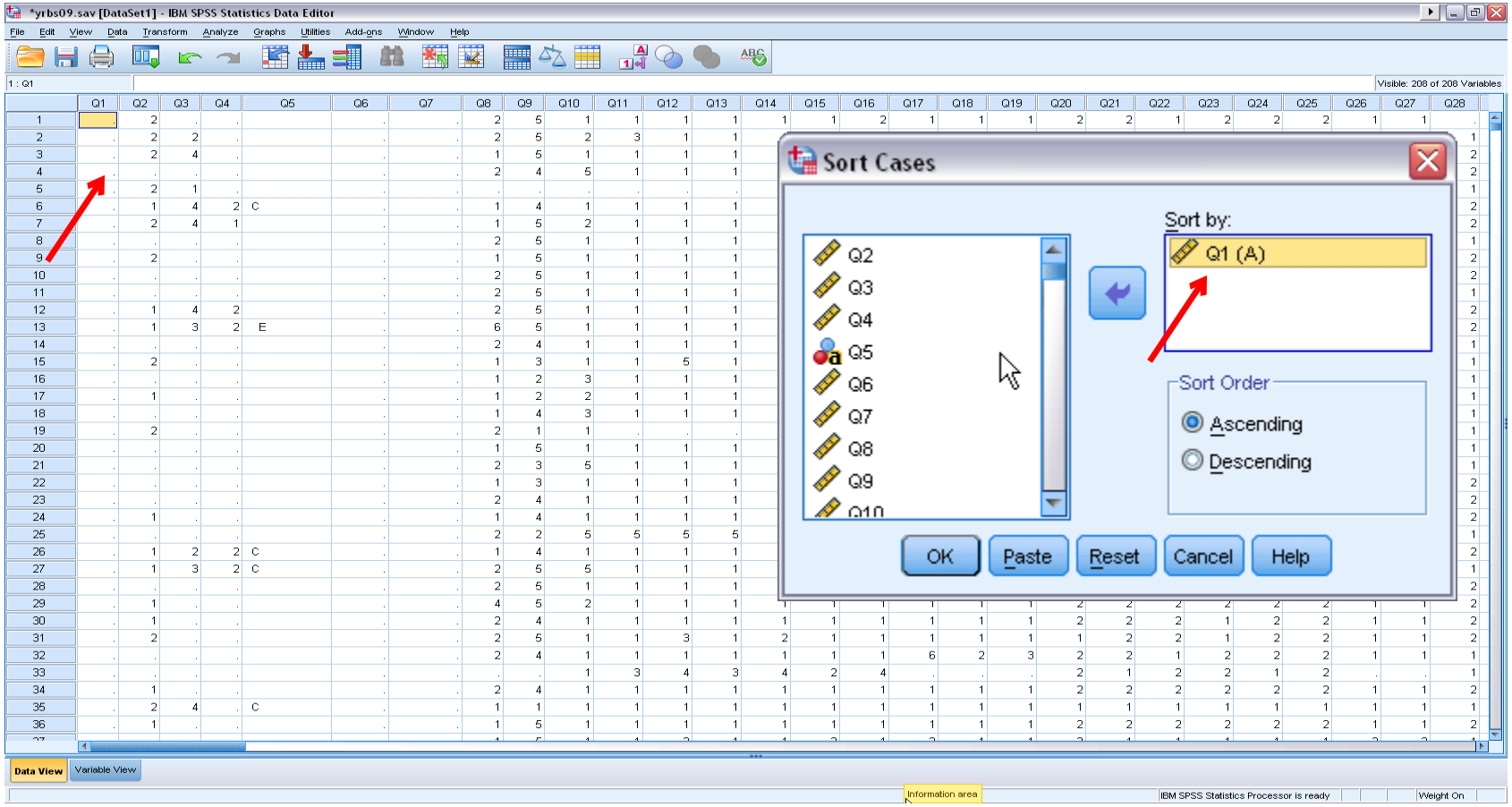
Statistics

Household income in thousands

N	Valid	6400
	Missing	0
Mean		69.4748
Median		45.0000
Mode		25.00
Std. Deviation		78.71856
Minimum		9.00
Maximum		1116.00

Sort and select cases

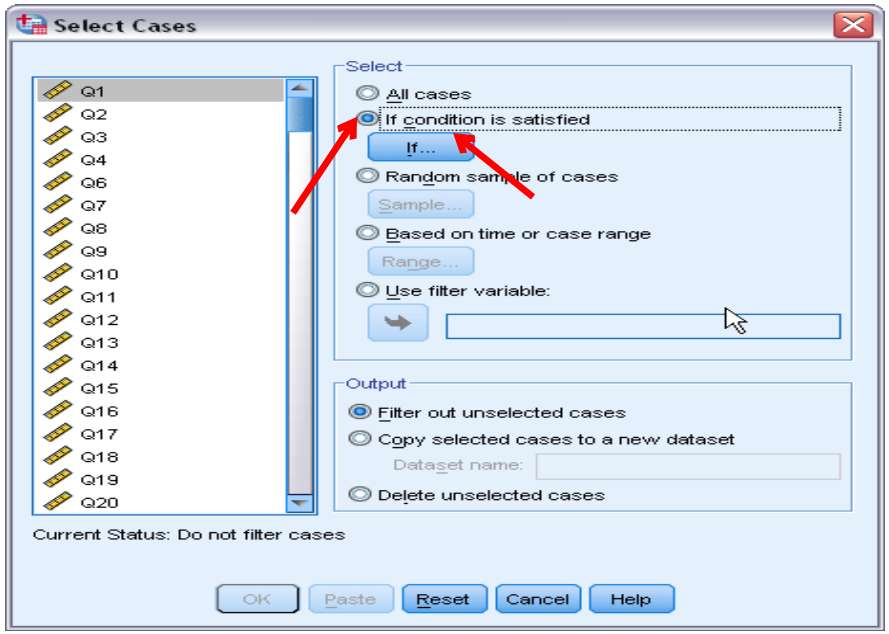
- Sort cases by variables: ~~Data~~ → Sort Cases
- You can use Sort Cases to find missing.



Sort and select cases

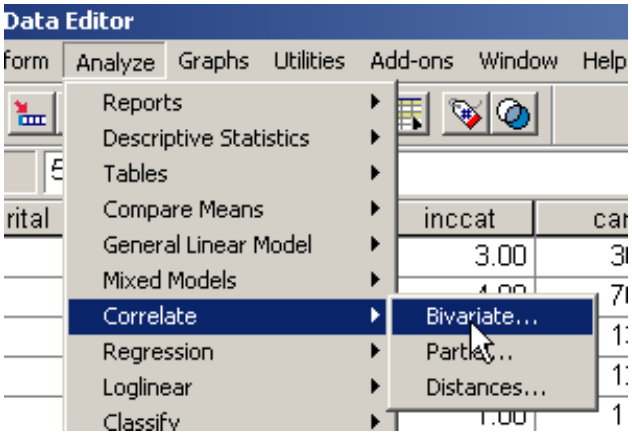
- Select cases
  - Example 1. Select Females for analysis.

1. Go to Data → Select Cases
2. Under Select: check the second one
3. Click If button

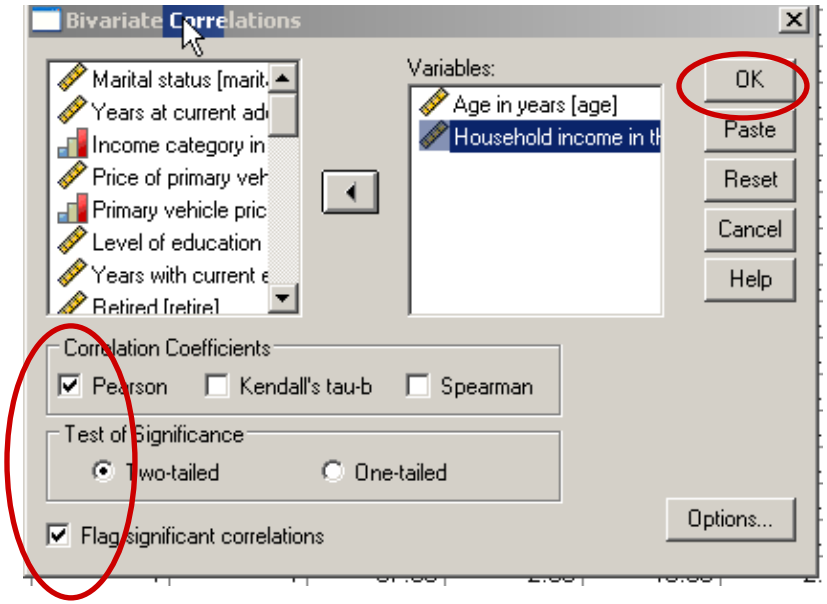


Running Analyses (Correlation)

28. Click Analyze- Correlate- Bivariate



29. Move the two variables of interest to the right side (age & income), click OK



Running Analyses (Correlation)

30. Results appear and tell us that the relationship is weak to moderate and results are not due to chance

Correlations

		Age in years	Household income in thousands
Age in years	Pearson Correlation	1	.335**
	Sig. (2-tailed)		.000
	N	6400	6400
Household income in thousands	Pearson Correlation	.335**	1
	Sig. (2-tailed)	.000	
	N	6400	6400

**. Correlation is significant at the 0.01 level (2-tailed).

- 1. *Skewness*: a measure of the asymmetry of a distribution.The normal distribution is symmetric and has a skewness value of zero.

  Positive skewness: a long right tail.

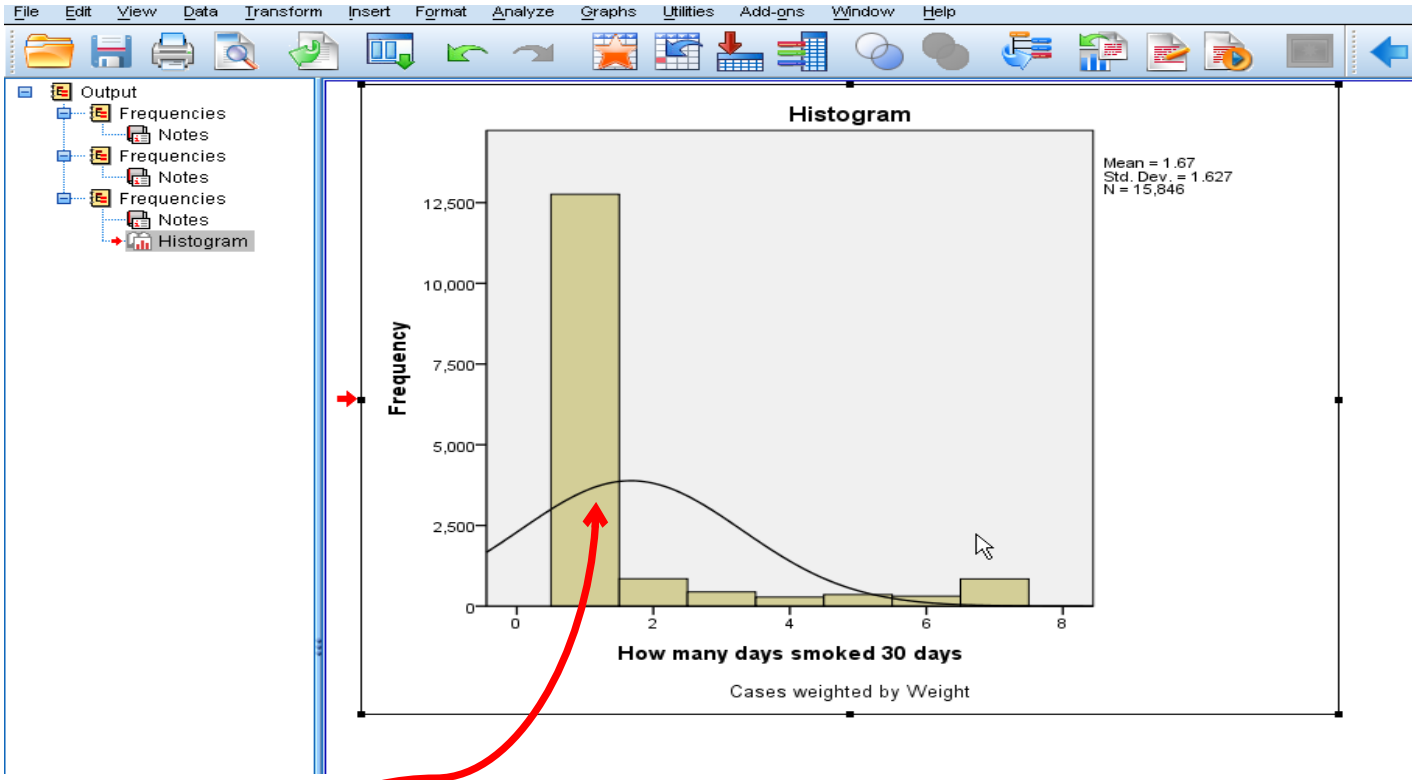
  Negative skewness: a long left tail.

  Departure from symmetry : a skewness value more than twice its standard error.
- 2. *Kurtosis*: A measure of the extent to which observations cluster around a central point. For a normal distribution, the value of the kurtosis statistic is zero.

  Mesokurtic data vales are with a normal curve/distribution

  Leptokurtic data values are more peaked, whereas

  platykurtic data values are flatter and more dispersed along the *X* axis.



Nature of curve

Resources

- Texas A & M- a huge selection of helpful movies http://www.stat.tamu.edu/spss.php
- UCLA- SPSS 17.0 Starter Kit (useful movies, FAQs, etc) http://www.ats.ucla.edu/stat/spss/sk/default.htm
- Indiana University- Getting Started (useful instructions with screenshots) http://www.indiana.edu/~statmath/stat/spss/win/
- University of Toronto- A Brief Tutorial (screenshots, instructions and basic stats) http://www.psych.utoronto.ca/courses/c1/spss/page1.htm
- Central Michigan- Tutorials and Clips (movies, screenshots, instructions- slow loading but good) http://calcnet.mth.cmich.edu/org/spss/toc.htm
- SPSS Statistics Coach and Tutorial (under Help) as well as the ZU library
- Online Statistics Textbook http://www.statsoft.com/textbook/stathome.html