



A Student Guide to SPSS

- Getting Started
- Data Operations
- Univariate Analysis
- Bivariate Analysis
- Multiple Linear Regression
- Exploratory Factor Analysis

Luke Sloan
School of Social Science
Cardiff University
SloanLS@cardiff.ac.uk

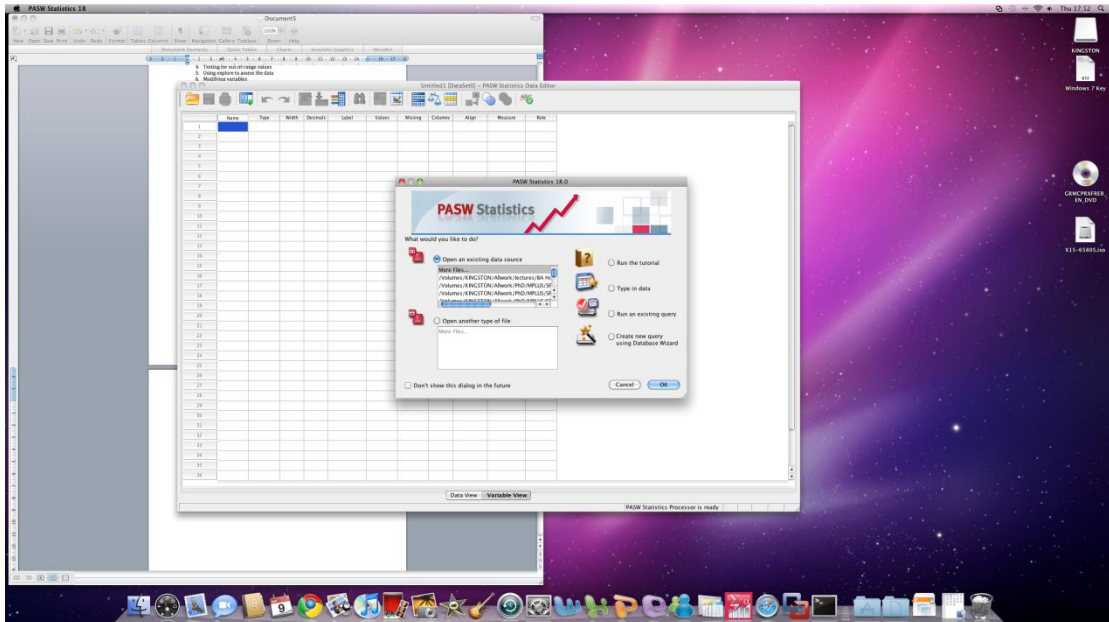
Version 4 – updated 11/01/2012

Contents

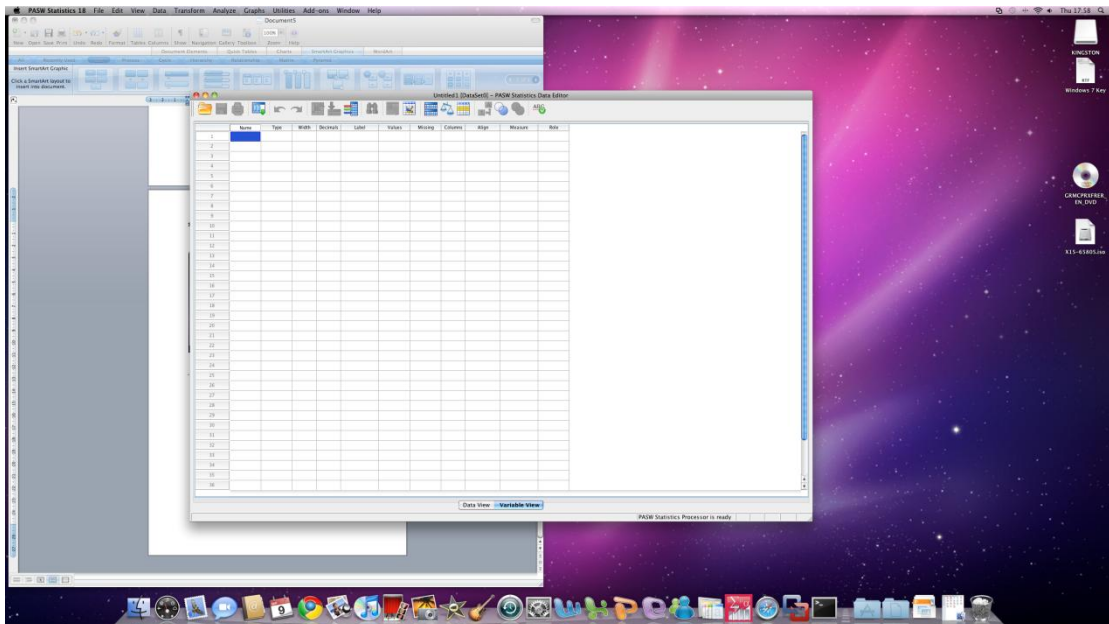
1) Opening an SPSS File.....	3
2) Creating a Database.....	4
3) Entering Data into SPSS	5
4) Using Explore for Testing Input Quality.....	6
5) Using Explore to Assess Measures of Central Tendency (Univariate Analysis).....	7
6) Using Frequencies to Analyse Response Patterns.....	7
7) Modifying Variables	8
8) Bivariate Analysis 1: Cross-Tabulations and Chi-Square for 2 Categorical Variables	11
9) Bivariate Analysis 2: Independent Samples T-Test for Comparing Mean Scores Between 2 Groups (1 Categorical, 1 Continuous).....	13
10) Bivariate Analysis 3: Paired Samples T-Test for Comparing Mean Scores for One Respondent on Two Occasions (1 Categorical, 1 Continuous)	15
11) Bivariate Analysis 4: Correlation for Assessing Relationships Between 2 Continuous Variables	16
12) Linear Regression: for Two Continuous Independent Variables	17
13) Multiple Linear Regression: for One Continuous Dependent Variable and Numerous Independent Variables.....	19
14) Exploratory Factor Analysis (EFA)	21

1) Opening an SPSS File

- Open SPSS (or PASW). Press cancel where it asks what you would like to do. This screen is for loading established databases.

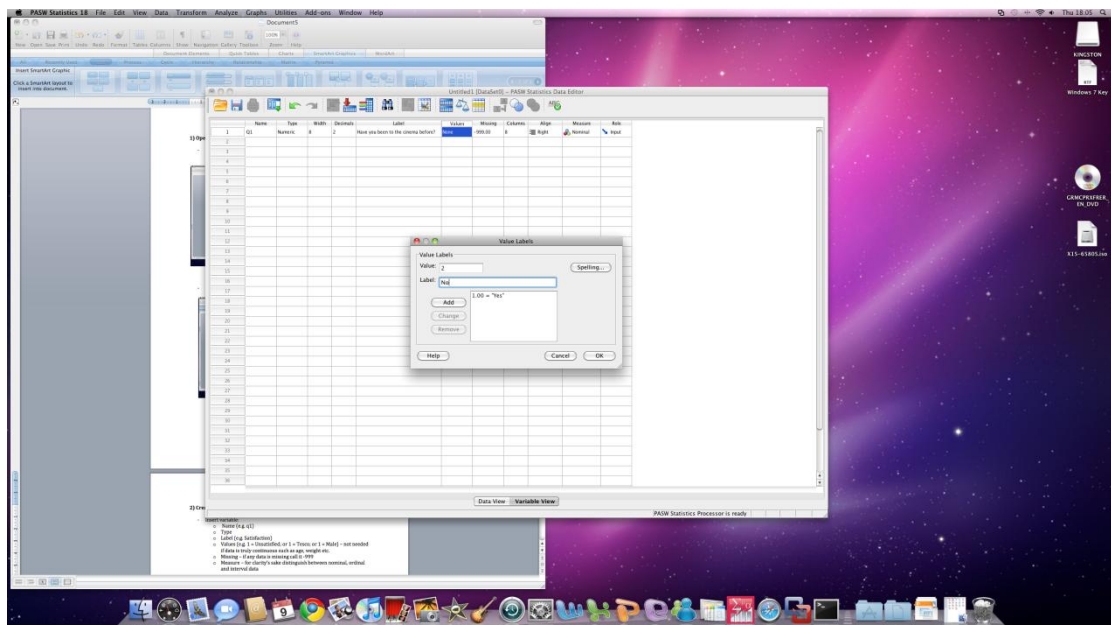


- Make sure that you are in **variable view** (not data view)



2) Creating a Database

- Insert variable:
 - o Name (e.g. q1)
 - o Type
 - o Label (e.g. Satisfaction)
 - o Values (e.g. 1 = Unsatisfied; or 1 = Tesco; or 1 = Male) – not needed if data is truly continuous such as age, weight etc.
 - o Missing – if any data is missing call it -999
 - o Measure – for clarity's sake distinguish between nominal, ordinal and interval data

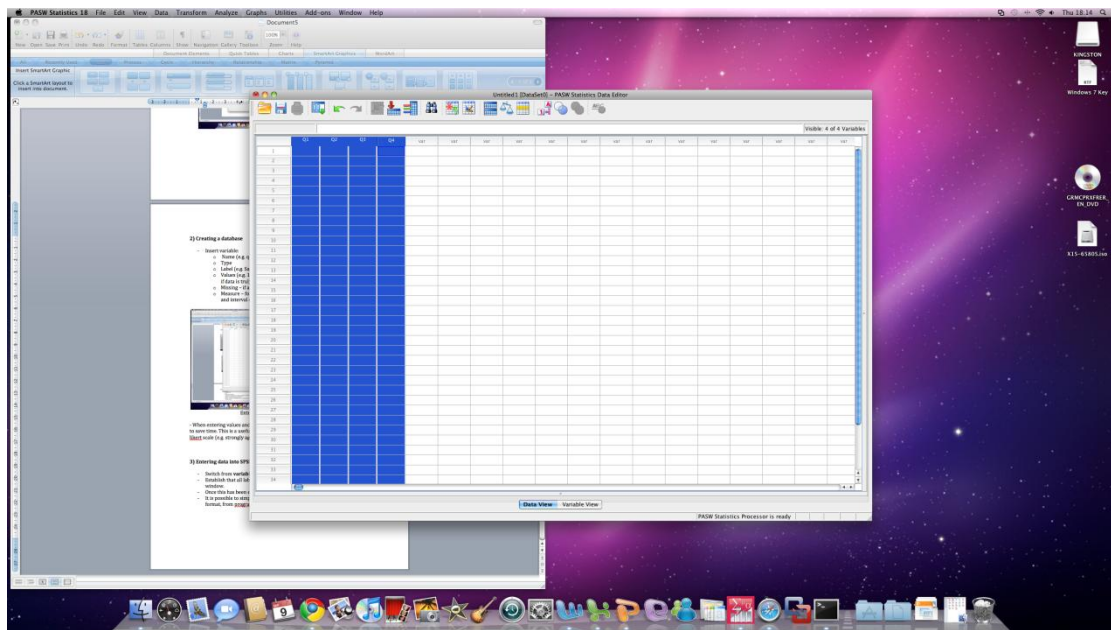


Entering values for variables in SPSS

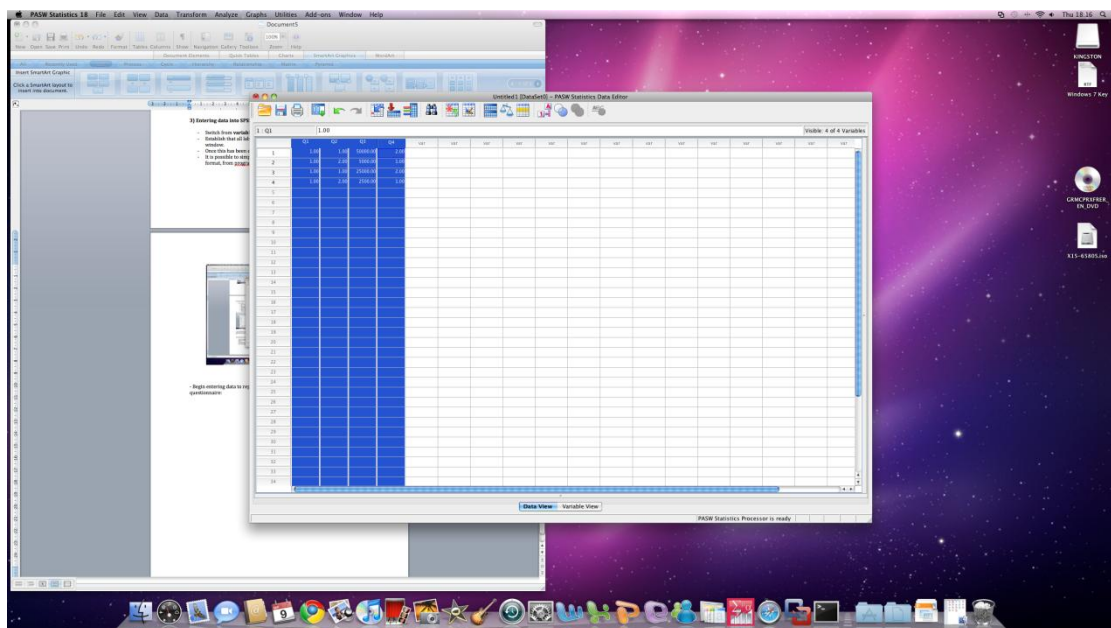
- When entering values and missing data values, it is possible to copy and paste to save time. This is a useful trick when you have 30 items measured on the same likert scale (e.g. strongly agree to strongly disagree).

3) Entering Data into SPSS

- Switch from **variable view** to **data view**
- Establish that all labels are evident across the top row of the data view window.
- Once this has been established it is possible to begin inputting data.
- It is possible to simply copy and paste data, providing it is in the same format, from programmes such as MS Excel (amongst others).



- Begin entering data to represent the answers given by respondents in the questionnaire:

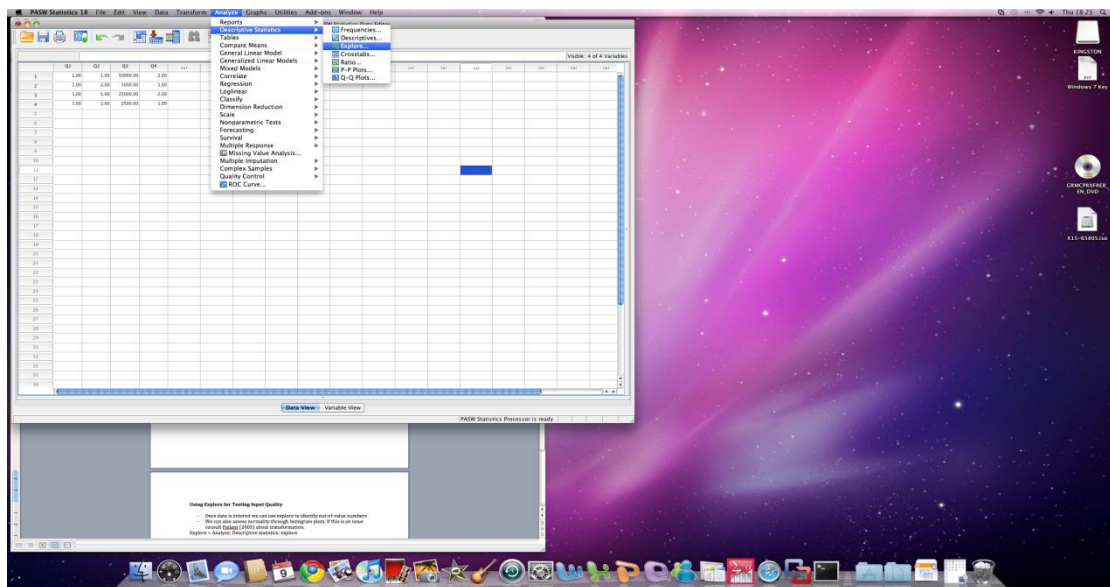


- Scan the database for any mistakes and unintentional errors. Often we double impute numbers meaning we get 55 instead of 5. Nonetheless, we can test for this using the explore function.

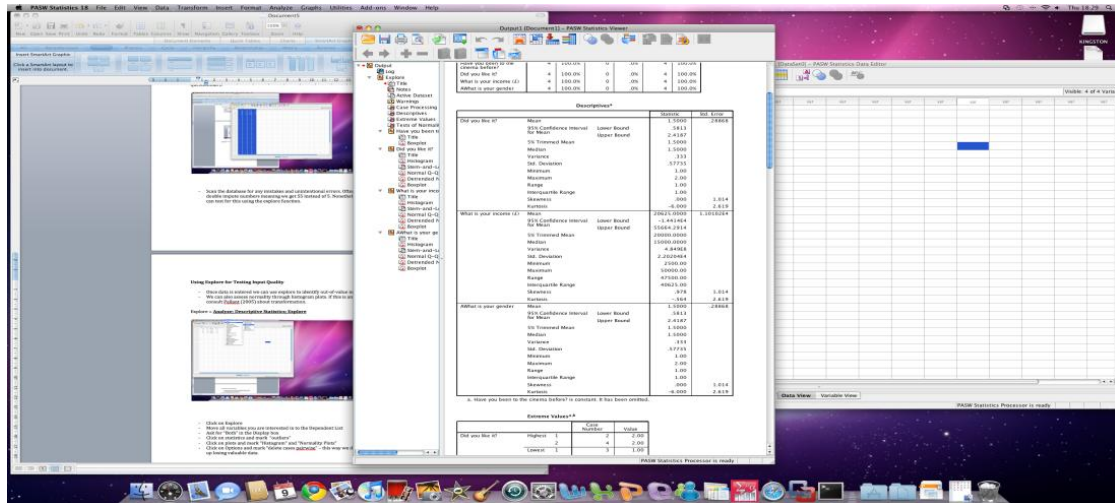
4) Using Explore for Testing Input Quality

- Once data is entered we can use explore to identify out-of-value numbers
- We can also assess normality through histogram plots. If this is an issue you should consult Pallant (2005) about transformation.

Explore = **Analyze; Descriptive Statistics; Explore**



- Click on Explore
 - Move all variables you are interested in to the Dependent List
 - Ask for “Both” in the Display box
 - Click on statistics and mark “outliers”, “minimum” and “maximum”
 - Click on plots and mark “Histogram” and “Normality Plots”
 - Click on Options and mark “delete cases pairwise” – this way we don’t end up losing valuable data.
- Study the output
 - Identify the minimum and maximum values. Are any of these bigger than they should be? For example, I have had a 500 year old respondent identified here due to a data entry mistake.
 - If yes, study the Extreme Values – which case number is it? Change.
 - Study the histograms and identify whether normality is an issue!

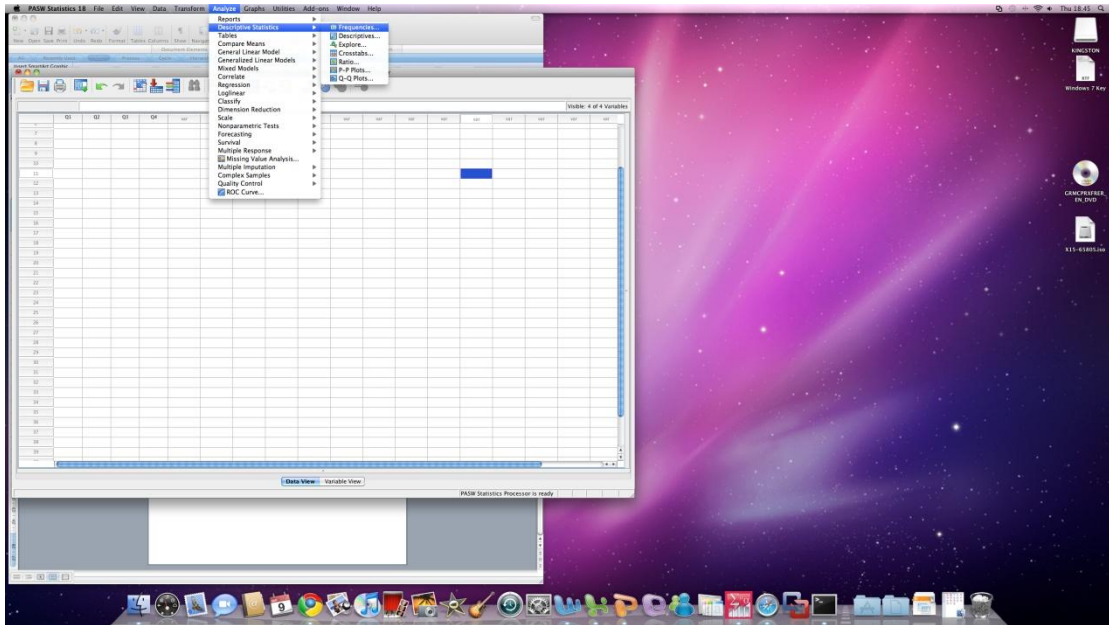


5) Using Explore to Assess Measures of Central Tendency (Univariate Analysis)

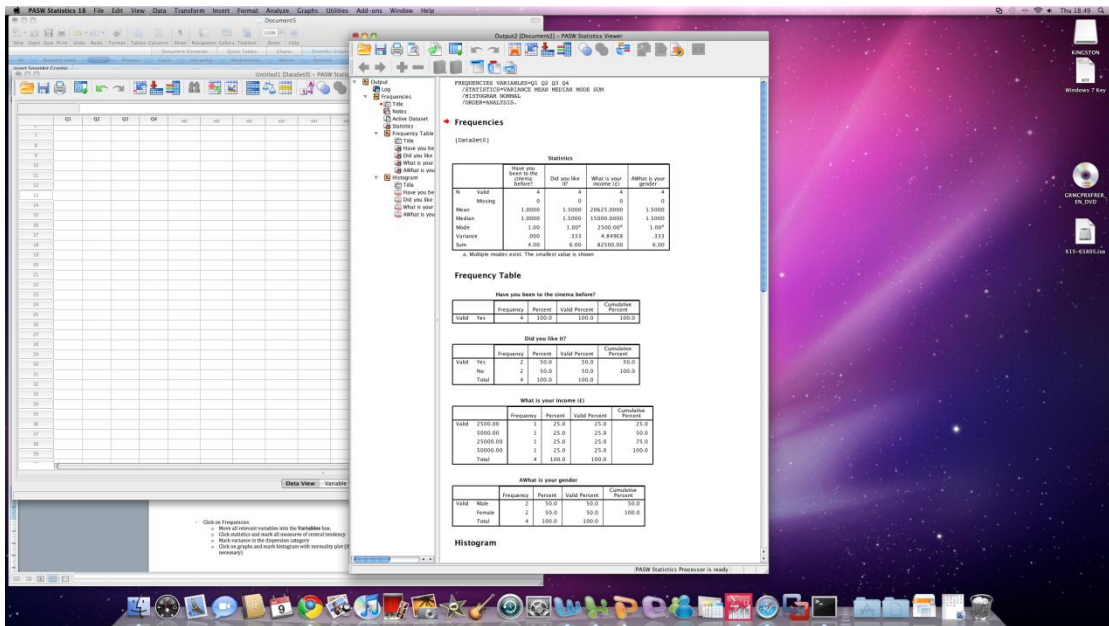
- The explore function also provides estimates for the mean, median, mode, range, standard deviation etc.
- This should be interpreted with caution. Mean values are not applicable or usable for categorical data. Therefore **NEVER** say something like “THE MEAN GENDER SCORE WAS 1.2”

6) Using Frequencies to Analyse Response Patterns

- It is also possible to use the FREQUENCIES function to assess how many people have responded to a certain question (e.g. male = 35; female = 25)
- As before, click **Analyze; Descriptive Statistics; Frequencies**
- Click on Frequencies
- Move all relevant variables into the **Variables** box.
- Click statistics and mark all measures of central tendency
- Mark variance in the dispersion category
- Click on graphs and mark histogram with normality plot (if necessary)



- Evaluate the output. This is very useful for assessing who responded to the questionnaire. It also provides percentages.



7) Modifying Variables

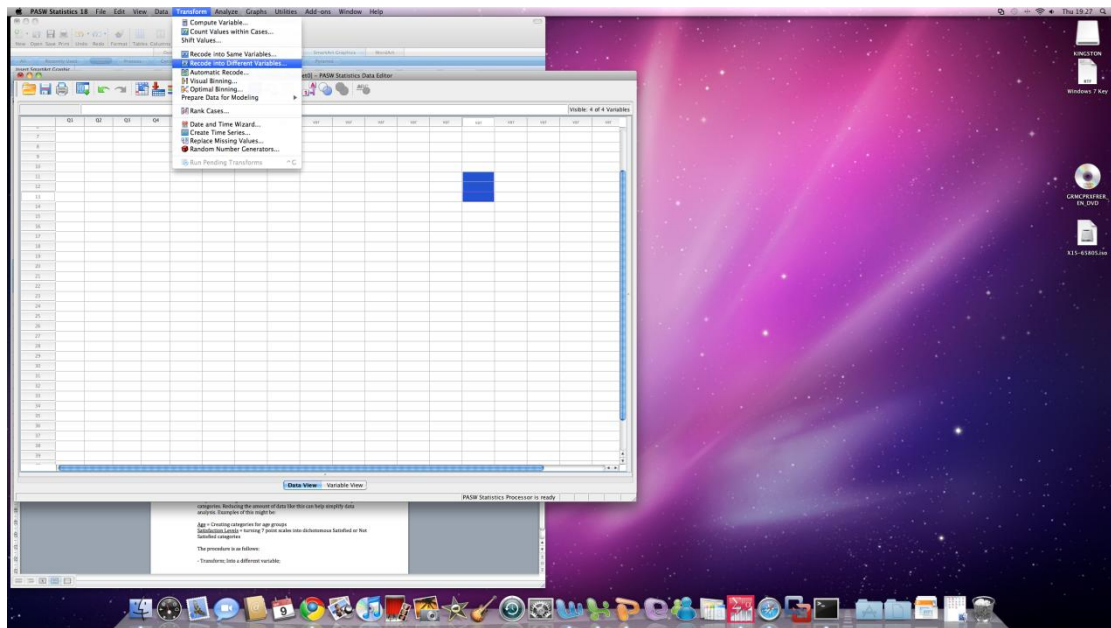
Very often we might find that we wish to break down our variable response categories. Reducing the amount of data like this can help simplify data analysis. Examples of this might be:

Age = Creating categories for age groups

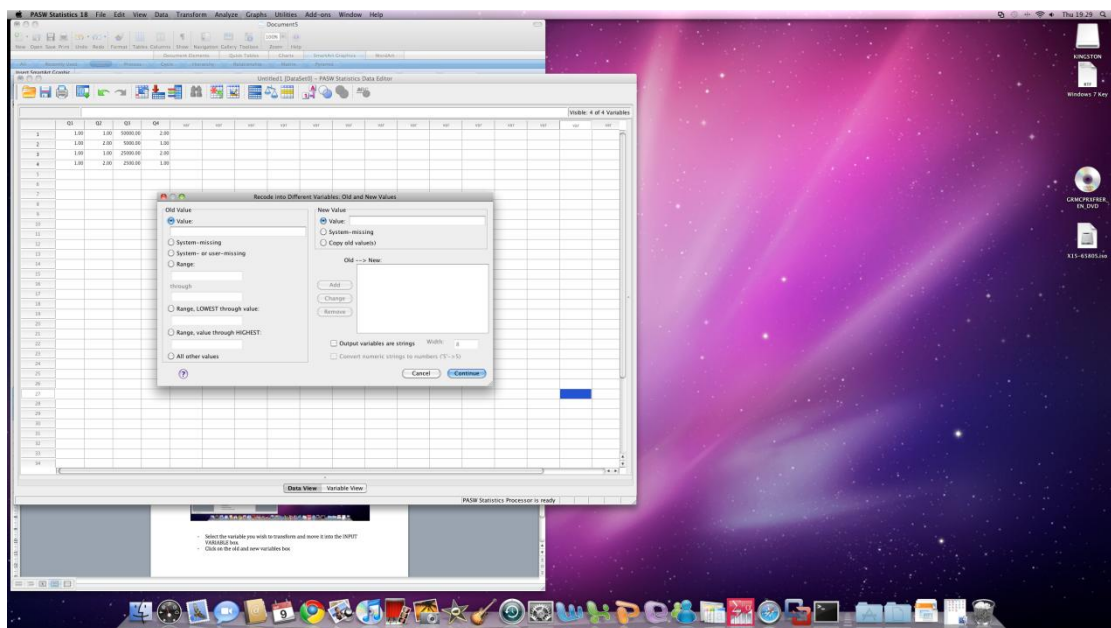
Satisfaction Levels = turning 7 point scales into dichotomous Satisfied or Not Satisfied categories

The procedure is as follows:

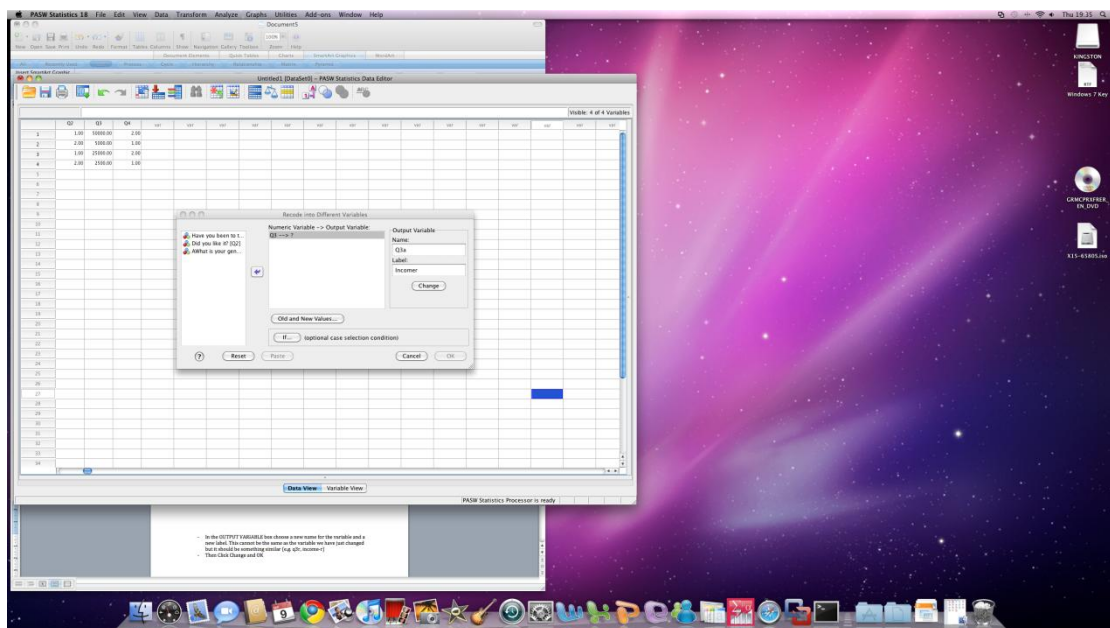
- Transform; Transform into a different variable;



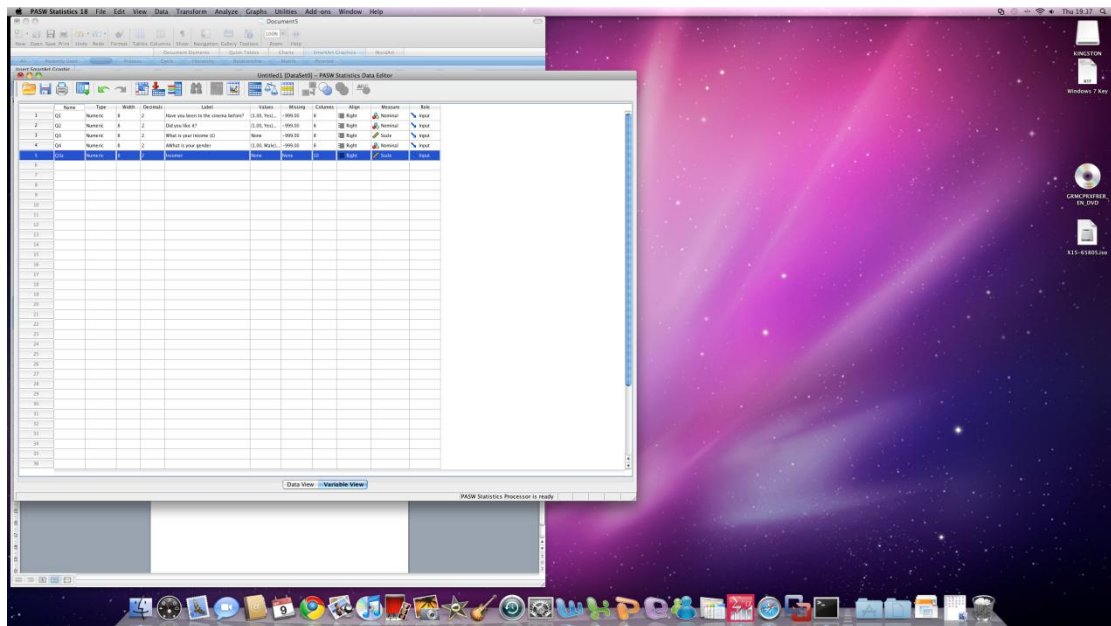
- Select the variable you wish to transform and move it into the INPUT VARIABLE box.
- Click on the old and new variables box



- Depending on the variable you wish to change, select new values for it. For example if we use income £0-15,000 will become 1, £15,001-30,000 will become 2, etc.
- Select a new value for each code and press add...this will be entered into the box
- Once the variable has been separated into the correct number of categories desired by the researcher press CONTINUE
- In the OUTPUT VARIABLE box choose a new name for the variable and a new label. This cannot be the same as the variable we have just changed but it should be something similar (e.g. q3r, income-r)
- Then Click Change and OK



- In the VARIABLE VIEW mode you will see that a new variable has been added to the dataset with values as specified.
- In the Values column, we should label what our new coding system means (1 = £0-15,000, 2 = £15,001-30,000, etc.). This is the same as when we initially set up the dataset.
- It is good practice to run a frequency analysis (as described above) to test that the new variable has worked.



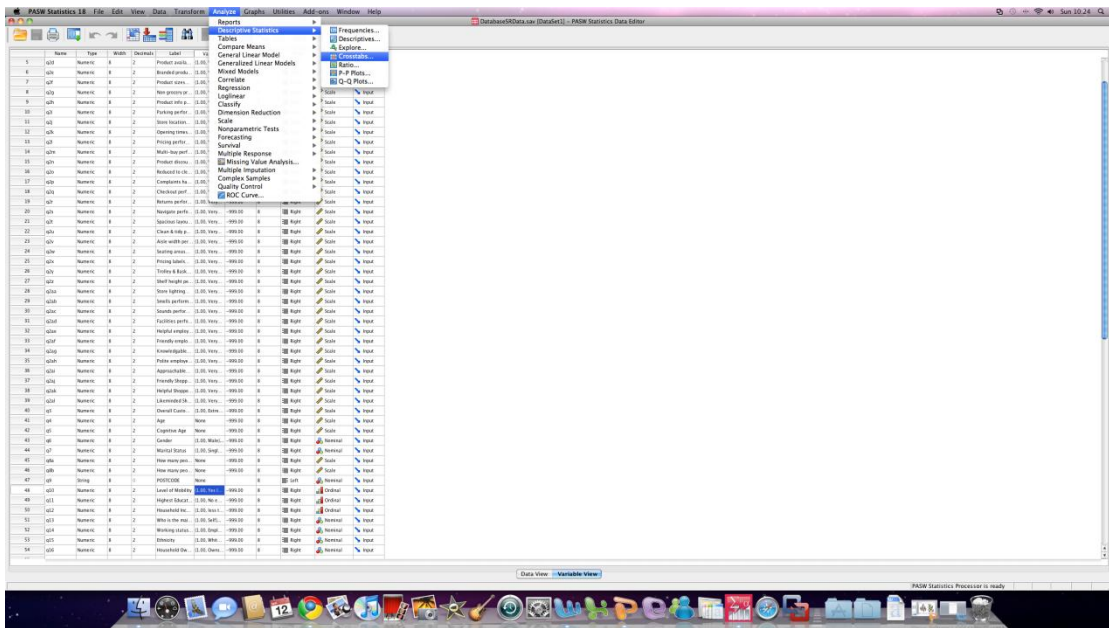
8) Bivariate Analysis 1: Cross-Tabulations and Chi-Square for 2 Categorical Variables

TEST: Analyses whether the count (i.e. no. responses) in one category (e.g. male) is higher than the alternative (e.g. female). Chi-square statistically assesses whether a difference truly exists.

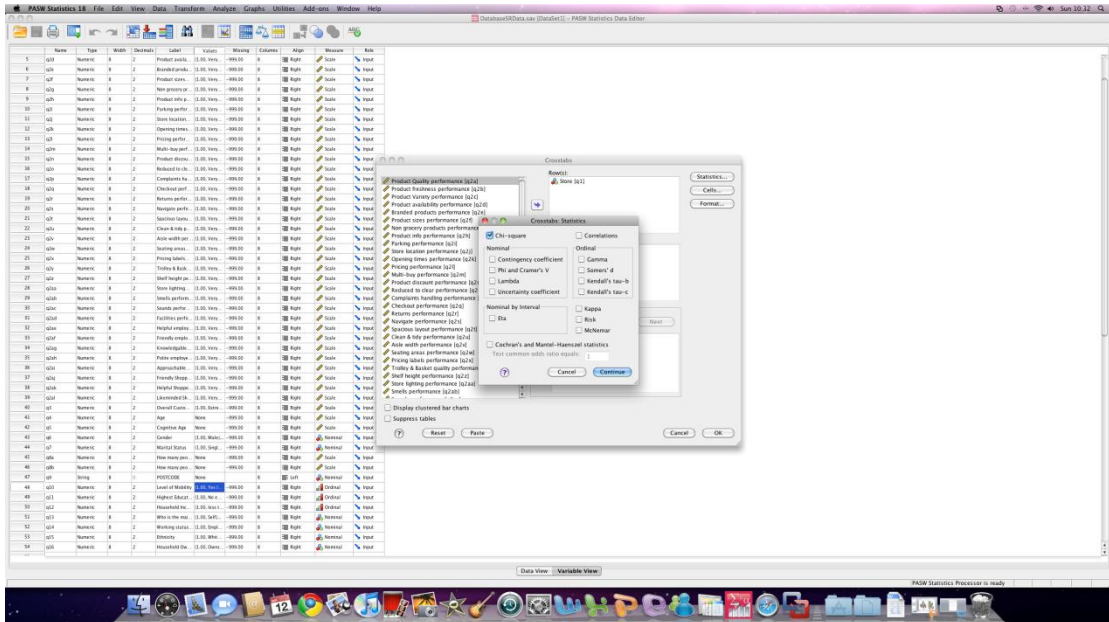
NEEDS: Two categorical variables with 2 or more independent category options.

ASSUMPTIONS: Requires at least 5 counts of data in each category (i.e. in a 2X2 matrix would need a minimum of 20 responses).

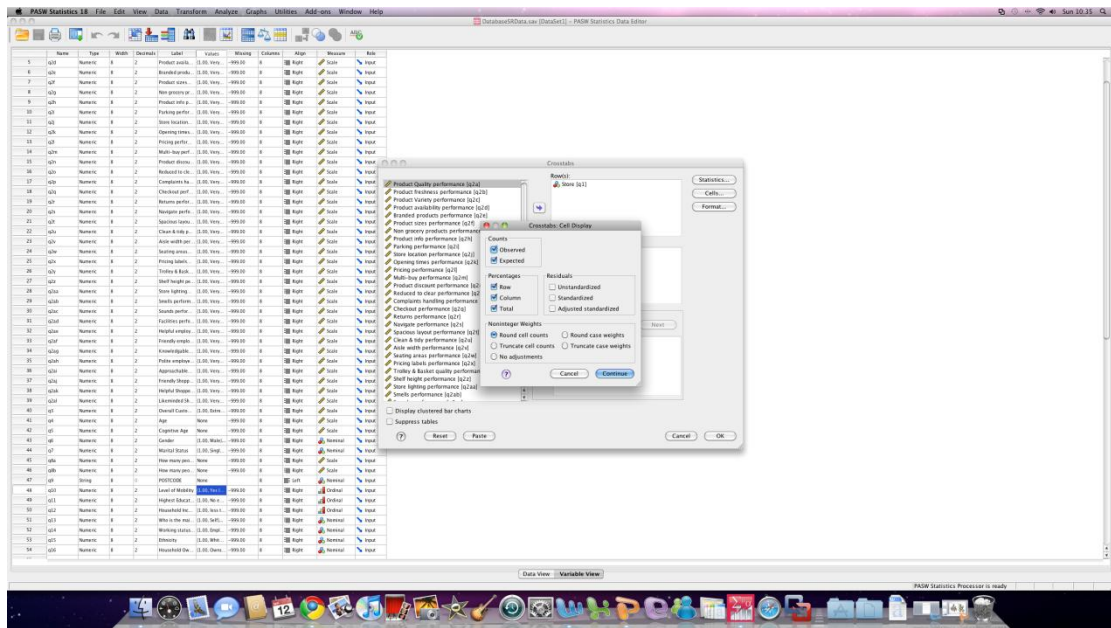
- Open up the relevant SPSS (PASW) file. By this stage we should have completed the basic univariate analyses such as those described with the measures of central tendency – descriptive, frequencies etc.
- In the icon bar select:
 - Analyze: Descriptives: Cross-tabs



- Click on the cross-tabs icon. From the list of variables move one variable into the row section and the other into the column. This shows how the two variables will be displayed in the cross-tabulation.
- Click on the Statistics box and ensure that the chi-square option is clicked.
- Click continue and return.



- Click on the cells option. Ensure that both “observed” and “expected” counts are ticked in the counts section.
- In the percentages section make sure that “row”, “column” and “total” are clicked.
- Click on the continue option.



- Click on the OK button to run the analysis and receive the output.

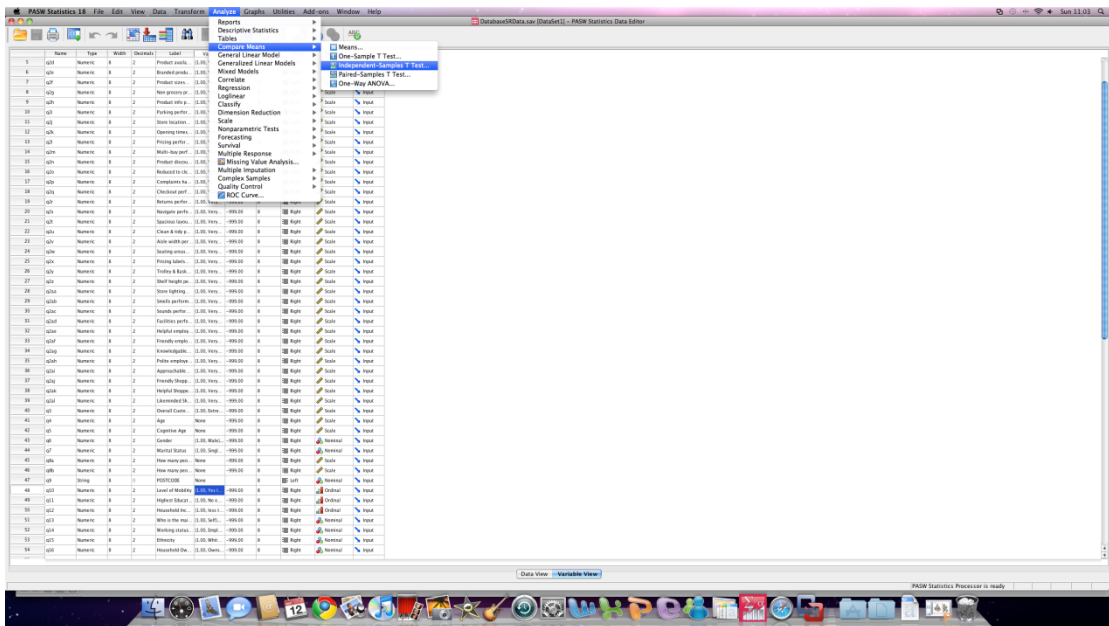
9) Bivariate Analysis 2: Independent Samples T-Test for Comparing Mean Scores Between 2 Groups (1 Categorical, 1 Continuous)

TEST: Analyses whether the mean score on one continuous variable (e.g. age, satisfaction, income) differs between two groups (e.g. male/female, young/old, educated/non-educated).

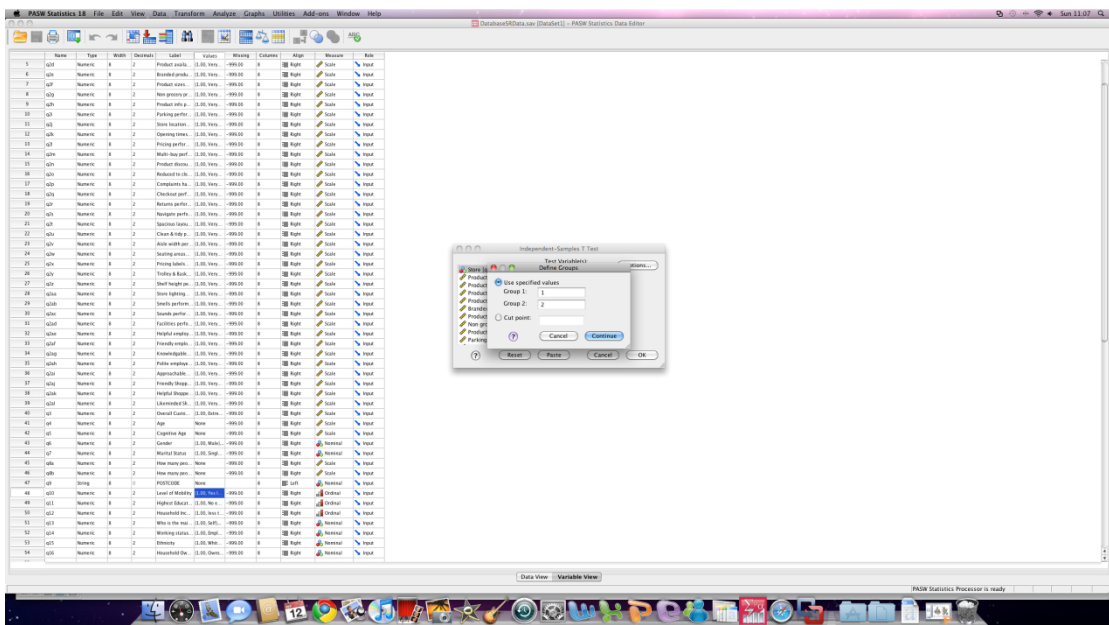
NEEDS: One categorical variable which defines the two groups, and one continuous variable.

ASSUMPTIONS: Requires the continuous variable to be normally distributed – check histogram, requires the variances to be homogenous – nonetheless SPSS provides an alternative test if this assumption isn't met. Refer to the lecture slides and Pallant (2004) for more of a discussion of this.

- In your SPSS file follow the instructions below:
 - o Analyze: Compare Means: Independent Samples T-Test



- After clicking on this test, move the continuous variable you will be using (e.g. income, satisfaction etc.) into the Test Variable box.
- Move your categorical variable, e.g. gender, into the Grouping Variables Box.
- To qualify that the coding matches, click on the Define Groups icon. For both groups provide the number which you use to describe them in the database. For example group 1 represents men so mark it as '1'. Group 2 is female so mark it as '2'.
- Click on Options to ensure that cases are deleted PAIRWISE only.



- Click on continue to run the analysis.
- REMEMBER – look at the Levene’s test P-Value before selecting the t-test statistic.

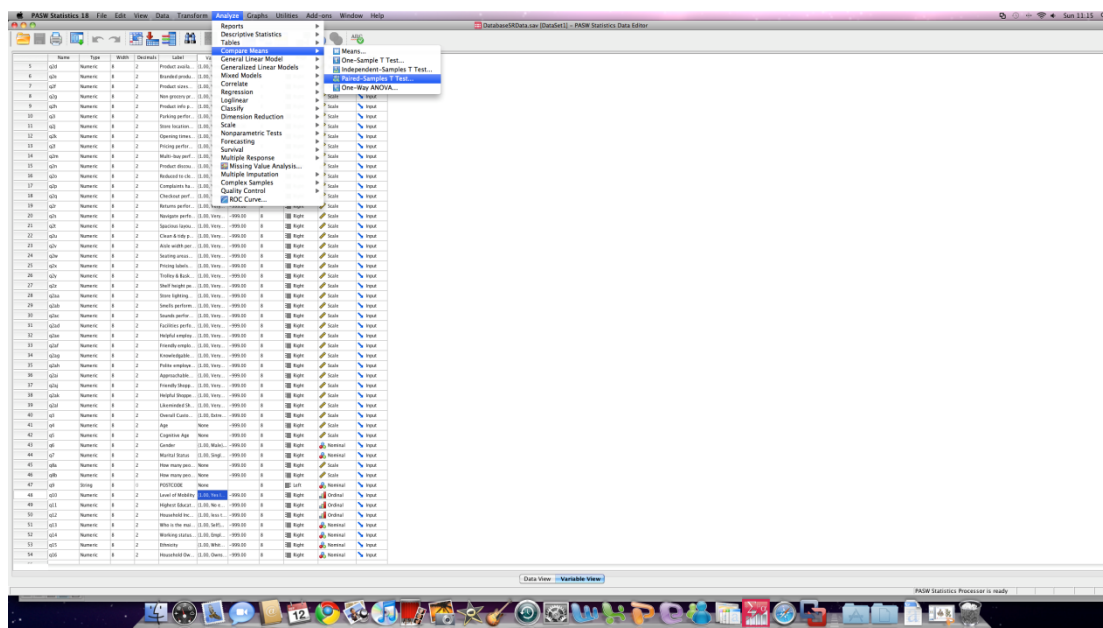
10) Bivariate Analysis 3: Paired Samples T-Test for Comparing Mean Scores for One Respondent on Two Occasions (1 Categorical, 1 Continuous)

TEST: Analyses whether the mean score on a single continuous variable differ before and after an intervention.

NEEDS: 2 identical continuous variables taken at different times

ASSUMPTIONS: Requires the continuous variable to be normally distributed – check histogram. Refer to the lecture slides and Pallant (2004) for more of a discussion of this.

- In your SPSS file follow the instructions below:
 - o Analyze: Compare Means: Paired Samples T-Test



- Once you have clicked on the test, move your two identical continuous variables into the Paired variables Box.
- Click on Continue to run the test.

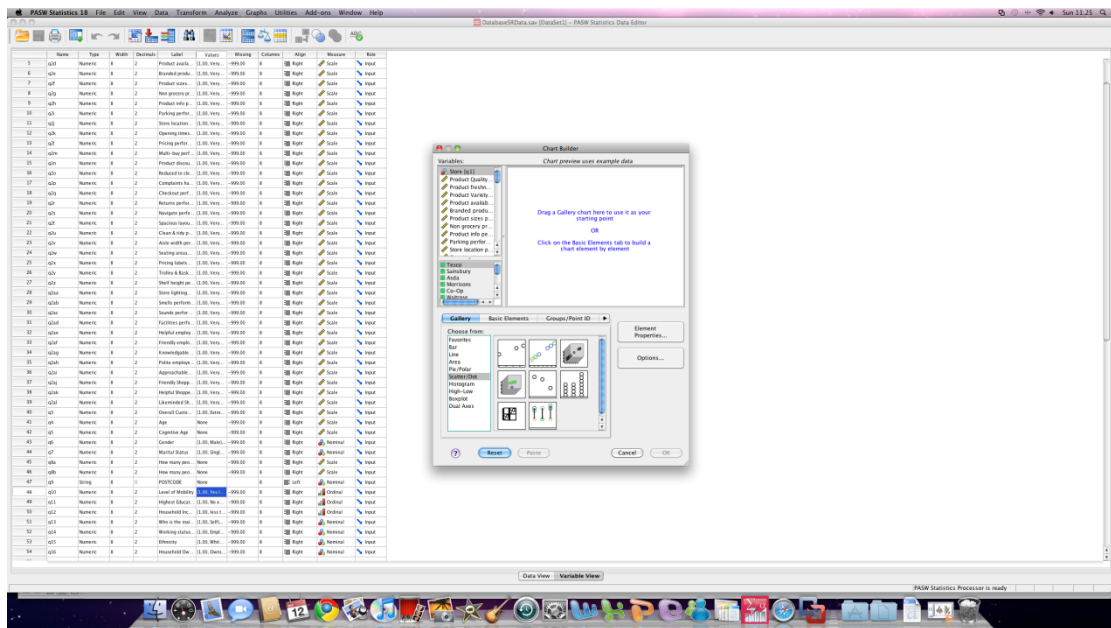
11) Bivariate Analysis 4: Correlation for Assessing Relationships Between 2 Continuous Variables

TEST: Tests the degree and direction (e.g. negative, positive) of the relationship between two continuous variables.

NEEDS: Two continuous variables (e.g. age, income, satisfaction)

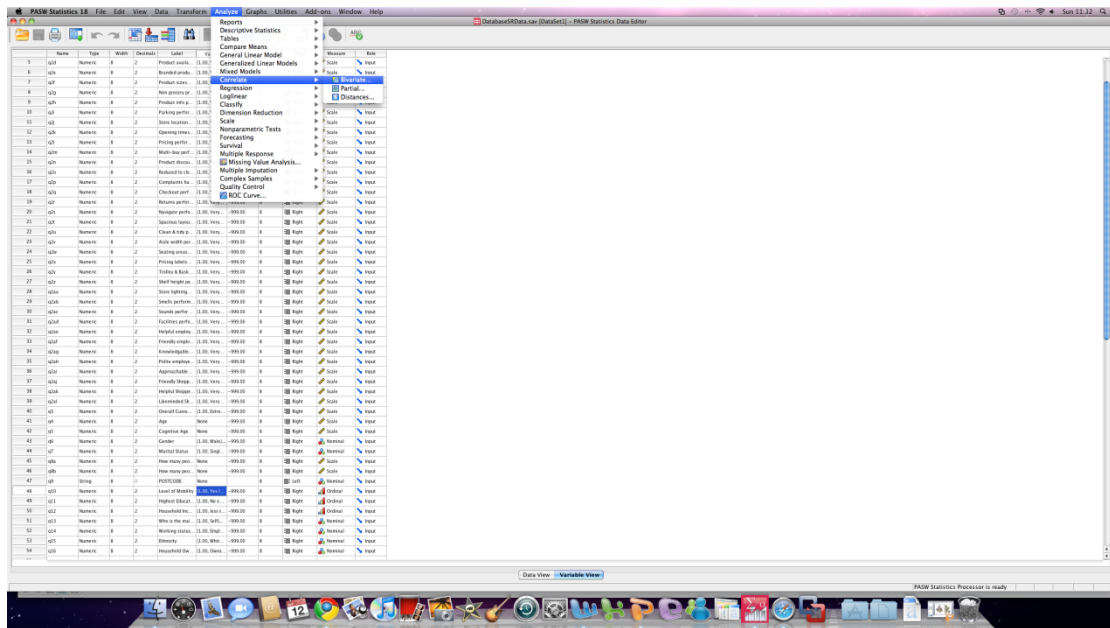
ASSUMPTIONS: Requires the continuous variable to be normally distributed – check histogram. Refer to the lecture slides and Pallant (2004) for more of a discussion of this.

- Before running a correlation analysis, it is always worthwhile creating a scatter plot. This is fairly simple to do. But requires a fair bit of clicking.
- Click on the following icons:
 - o Graphs: Chart Builder.
- Click on OK if provided with an option to define variables.
- From the screen below, ensure that the Gallery option is highlighted.



- Click on the scatter/dot option and then highlight one of the 8 types of chart offered. I normally use the top left option which is a simple scatter chart. Drag this into the chart builder box using the mouse.
- From the variables box, drag your two continuous variables on to the axis of the chart. For example, one must go on the X-Axis and the other on the Y-Axis.
- You can use the Element Properties box to edit the labels for the charts. This is always a good idea for when you are producing a report/document.
- Click on OK for the graph to be produced in SPSS.

- From the scatter plot it will be clear whether there is any sort of relationship in the data.
- Nonetheless we should formally test this by using Pearson's coefficient (see lecture notes).
- To test correlation, click on the following icons:
 - o Analyze, correlate: bivariate



- Move your two continuous variables into the Test Box.
- Ensure that Pearson's correlation is ticked in the Correlation Coefficient box and that two-tailed significance is marked below this.
- Click on options to ensure that the PAIRWISE deletion box is marked.
- Click on continue.
- Click on OK to run the analysis

12) Linear Regression: for Two Continuous Independent Variables

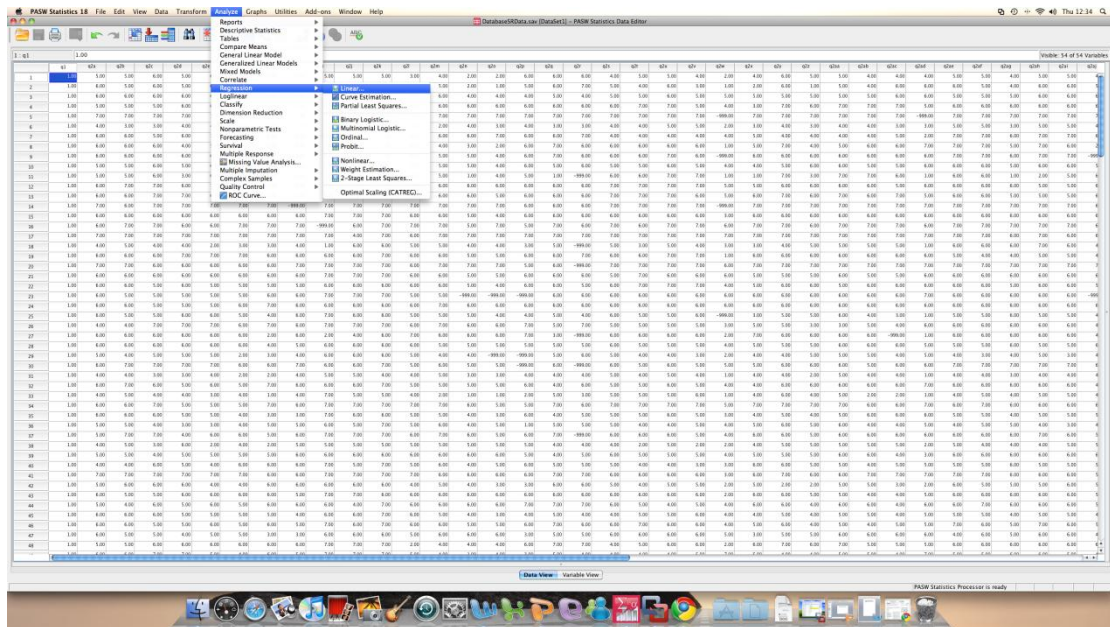
TEST: Tests the degree, direction (e.g. negative, positive) and numeric relationship between two continuous variables.

NEEDS: Two continuous variables (e.g. age, income, satisfaction)

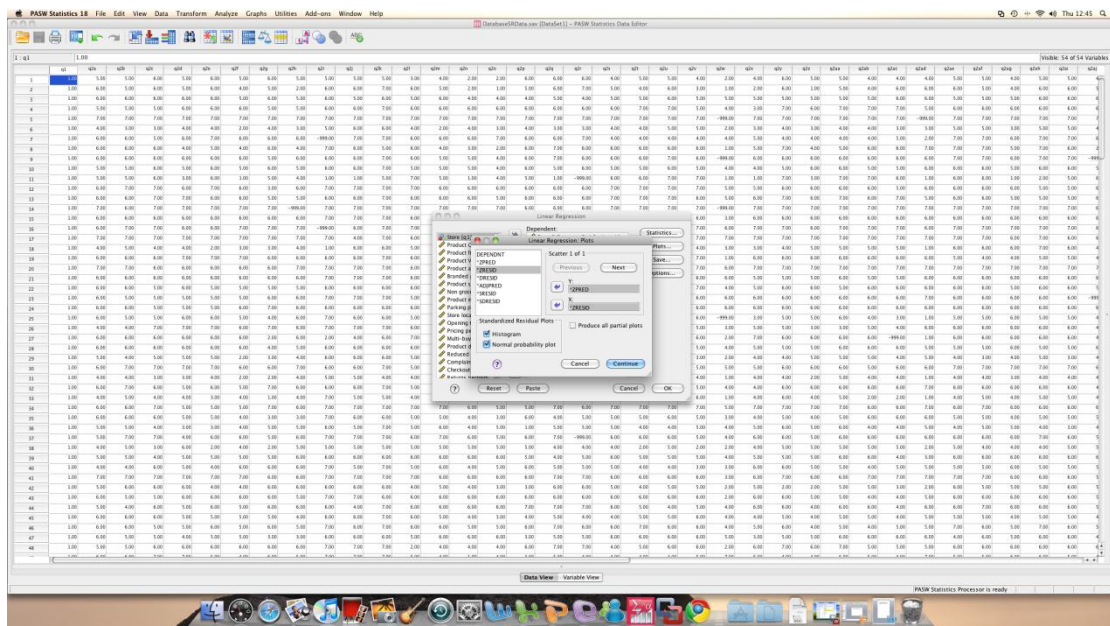
ASSUMPTIONS: Requires both continuous variable to be normally distributed – check histogram. Variables must also share a linear relationship. Homogeneity of

residuals must be observed. Refer to the lecture slides and Pallant (2004) for more of a discussion of this.

- Begin by selecting 'Analyse' from the toolbar.
- Scroll down to 'Regression' and select 'Linear'



- Enter your dependent variable into the 'Dependent' field (the variable that you are trying to explain/predict)
- Enter a single explanatory variable into the 'Independent' field
- Click on the 'Statistics' button and select '*ZRESID' for the 'Y.' field (see screenshot below)
- Then select 'DEPENDNT' for the 'X.' field
- Ensure that the 'Histogram' and 'Normal probability plot' buttons are ticked



- Click 'Continue'
- Then click 'OK' to run the analysis.

CHECKING THE OUTPUT (refer to lecture notes for details):

- What proportion of the variance in 'y' is explained by 'x'? (R^2 ?)
- Does 'x' have a statistically significant effect on 'y'? (p value?)
- What is this effect? (beta coefficient or 'B'?)
- Are standardized residuals normally distributed? (histogram)
- Do the points on the normal p-p plot fall on the line?
- Are the residuals randomly spaced? (i.e. is there any pattern to the scatter?)

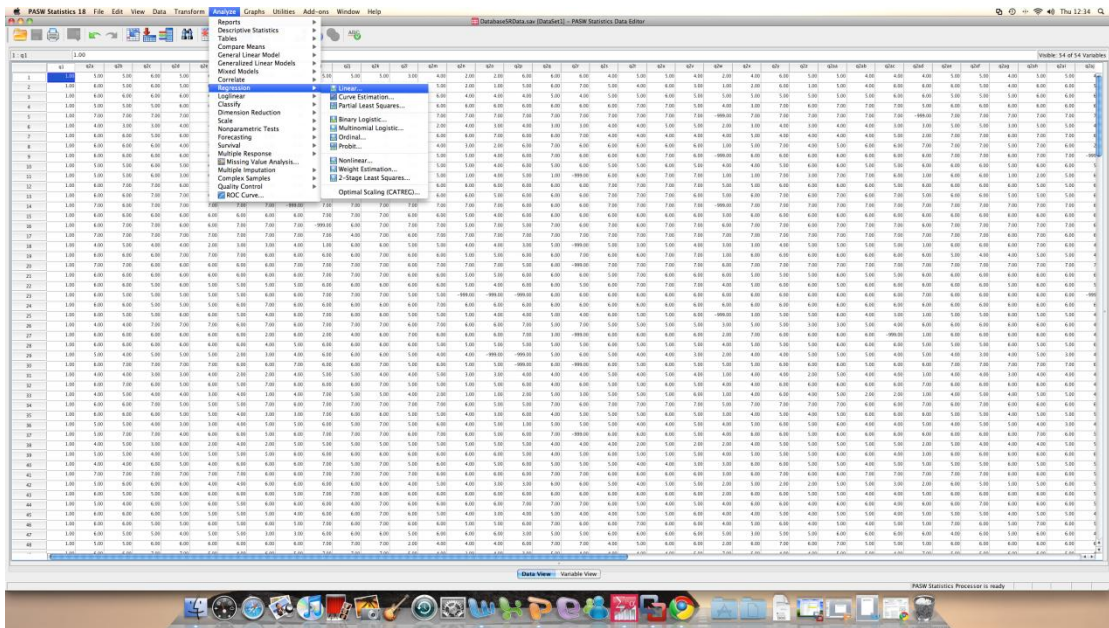
13) Multiple Linear Regression: for One Continuous Dependent Variable and Numerous Independent Variables

TEST: Tests the degree, direction (e.g. negative, positive) and numeric relationship between one continuous dependent variable and several different dependent variables (dependents do not have to all be continuous!).

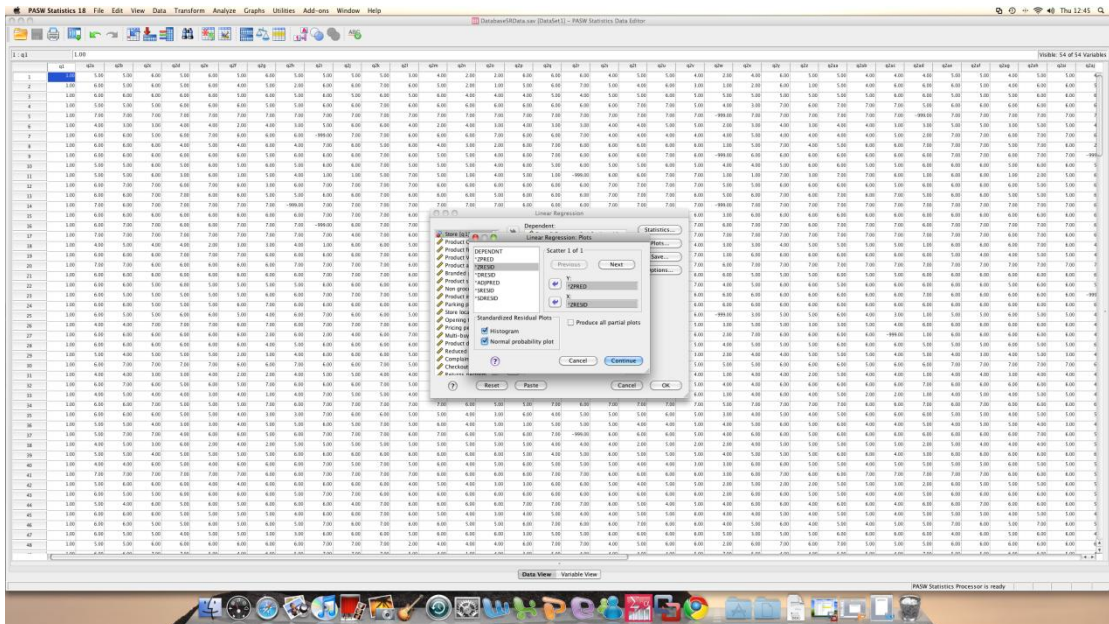
NEEDS: One continuous dependent variable (e.g. income) and several dependent predictor variables (e.g. age – continuous, educational qualification – ordinal, sex – categorical).

ASSUMPTIONS: Requires variables to be normally distributed – check histograms. Continuous variables should share a linear relationship. Independent variables must not be related to each other statistically (multicollinearity). Homogeneity of residuals must be observed. Refer to the lecture slides and Pallant (2004) for more of a discussion of this.

- Begin by selecting 'Analyse' from the toolbar.
- Scroll down to 'Regression' and select 'Linear'



- Enter you dependent variables into the 'Dependent' field (the variables that you are trying to explain/predict)
- Enter a single explanatory variable into the 'Independent' field
- Click on the 'Statistics' button and select '*ZRESID' for the 'Y:' field (see screenshot below)
- Then select '*ZRESID' for the 'X:' field
- Ensure that the 'Histogram' and 'Normal probability plot' buttons are ticked
- Click 'Continue'



a. then click 'OK' to run the analysis.

CHECKING THE OUTPUT (refer to lecture notes for details):

- What proportion of the variance in 'y' is explained by the independent variables (R^2)? Remember that this statistic refers to the overall effect of the whole model (all independent variables at once).
- Is the model a better predictor than simply using means? (sig of 'F')
- Do any of the independent variables have a statistically significant effect on 'y'? (p value?)
- What are these effects? (beta coefficients or 'B'?)
- Are standardized residuals normally distributed? (histogram)
- Do the points on the normal p-p plot fall on the line?

Are the residuals randomly spaced? (i.e. is there any pattern to the scatter?)

14) Exploratory Factor Analysis (EFA)

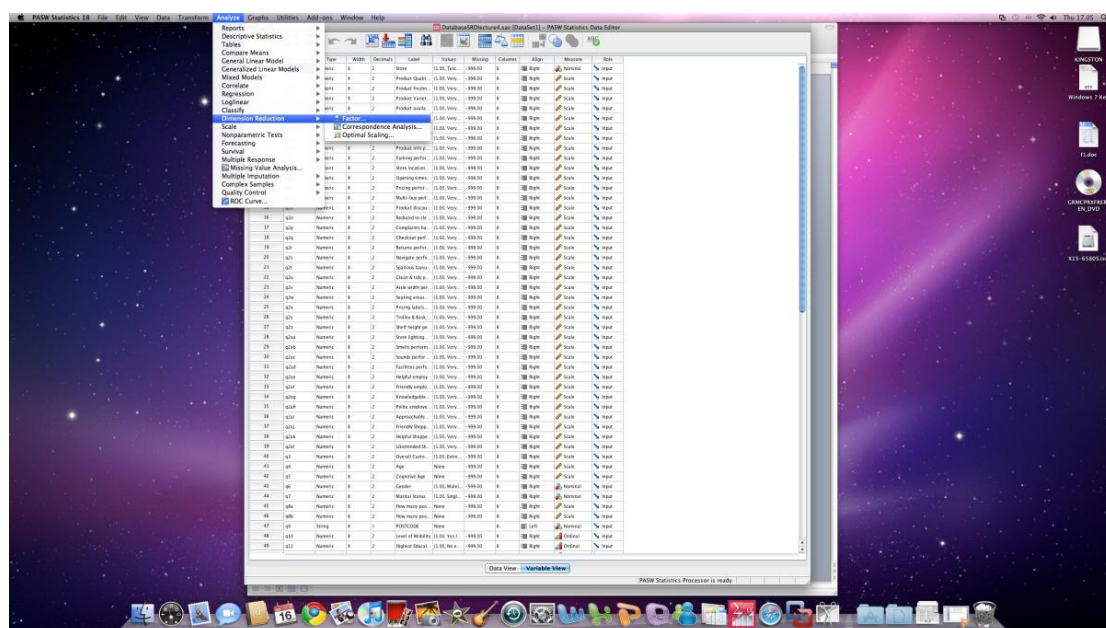
In principle, an EFA is an extremely simple analysis to set up and run. The following instructions outline the process from beginning to end.

WHAT YOU NEED:

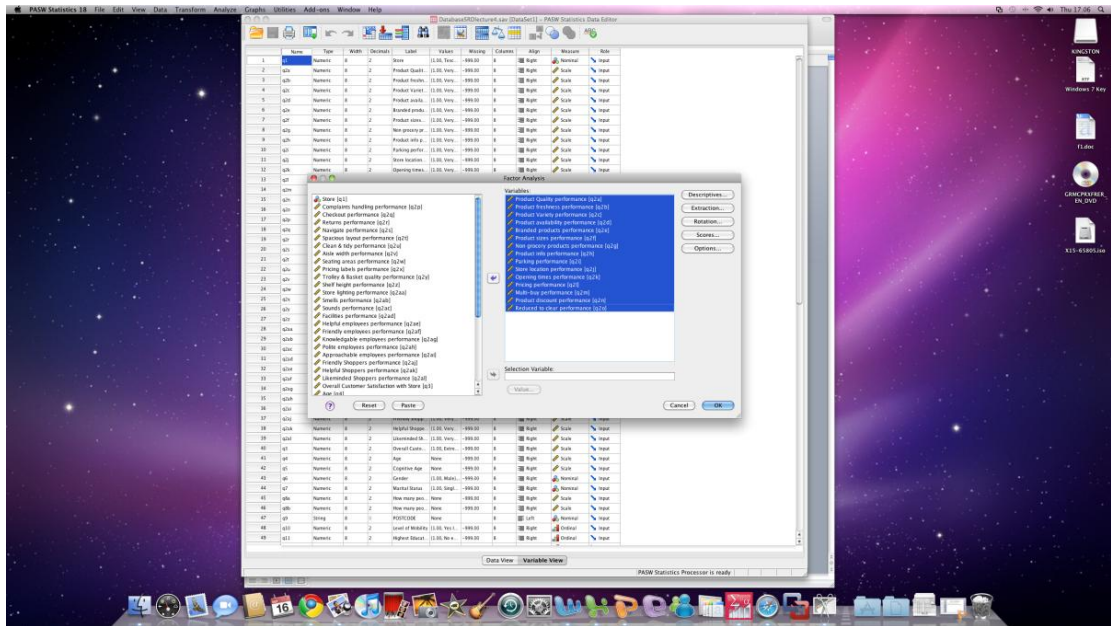
- Two or more continuous-level variables (e.g. ratings scales)
- All variables need to be measured in the same way (e.g. all 7-point scales)

RUNNING THE ANALYSIS:

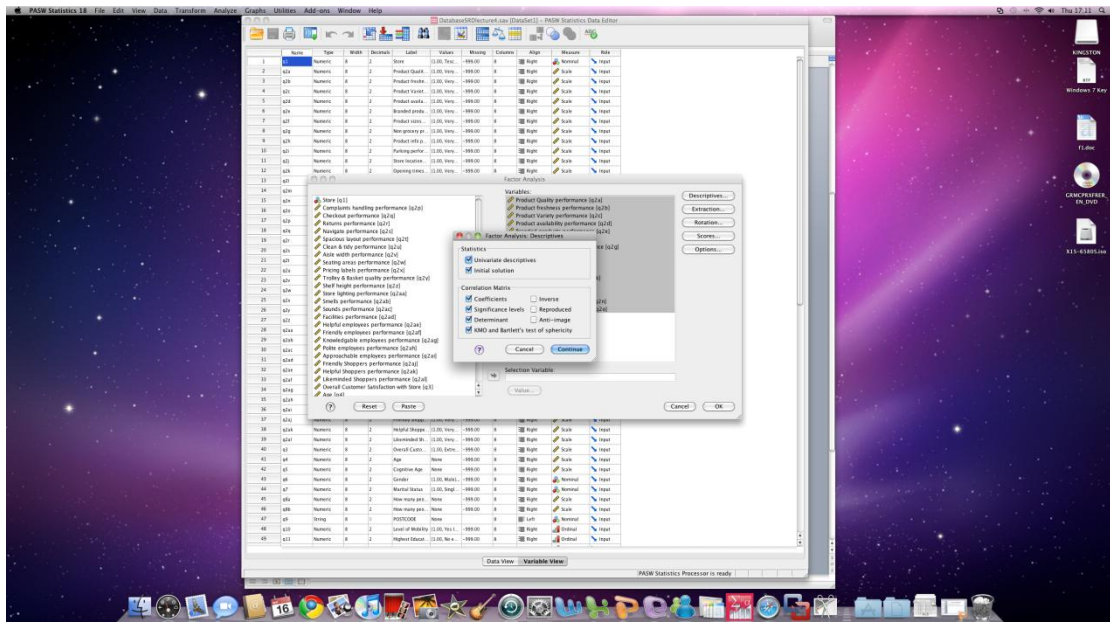
- Open up the relevant SPSS file
- Click on:
 - o 'Analyze'; 'Dimension Reduction'; 'Factor'



- In the analysis box move all of the relevant variables for the test into the 'Variables' box

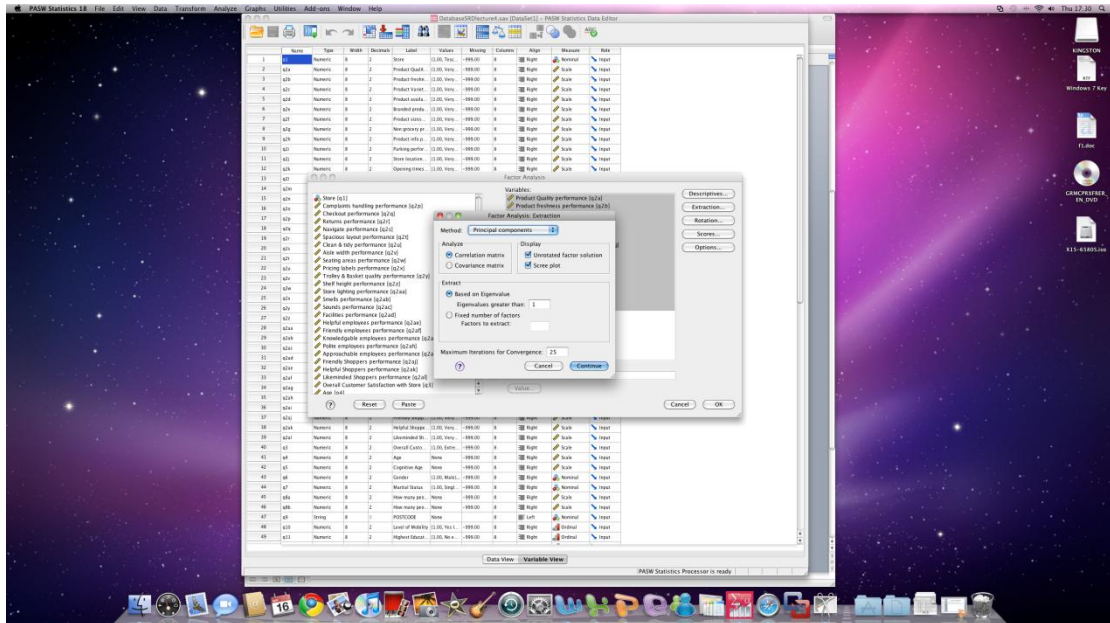


- Once you have moved each of the relevant variables into the 'Variables' box, click on the 'Descriptives' button on the right hand side
- Mark all of the options on the left hand side, namely: 'Univariate Statistics'; 'Initial Solution'; 'Coefficients'; 'Significance Levels'; 'Determinant'; 'KMO & Bartlett's Test of Sphericity'.
- Click on 'Continue'

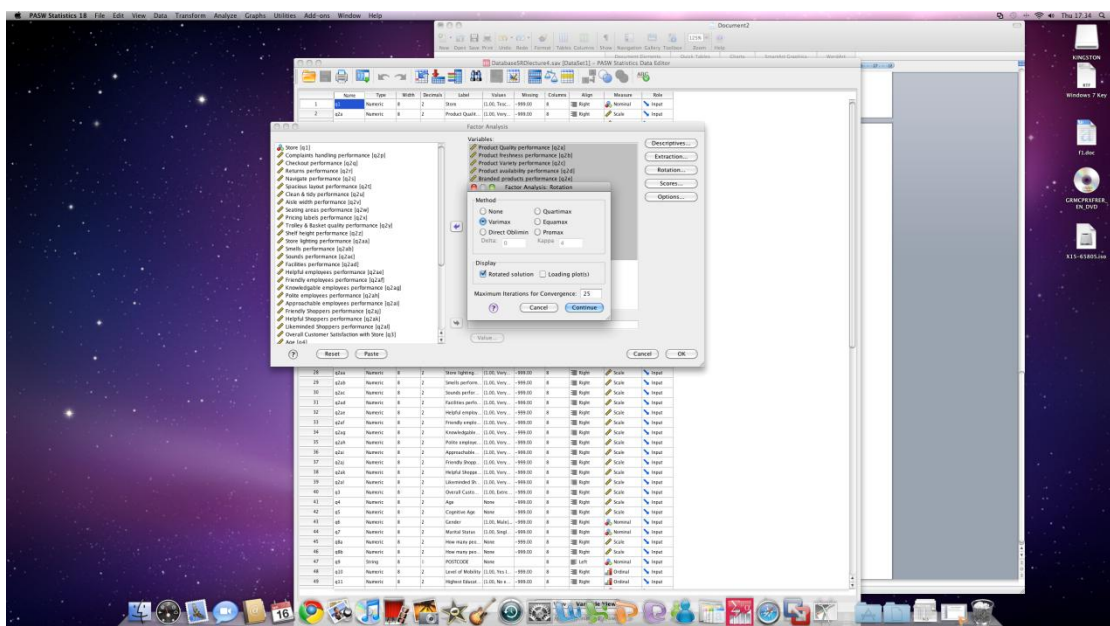


- In the main window click on 'Extraction'
- Ensure that 'Principal Components' is marked as the method
- In the 'Display' box, mark 'Unrotated Factor Solution' and 'Scree Plot'

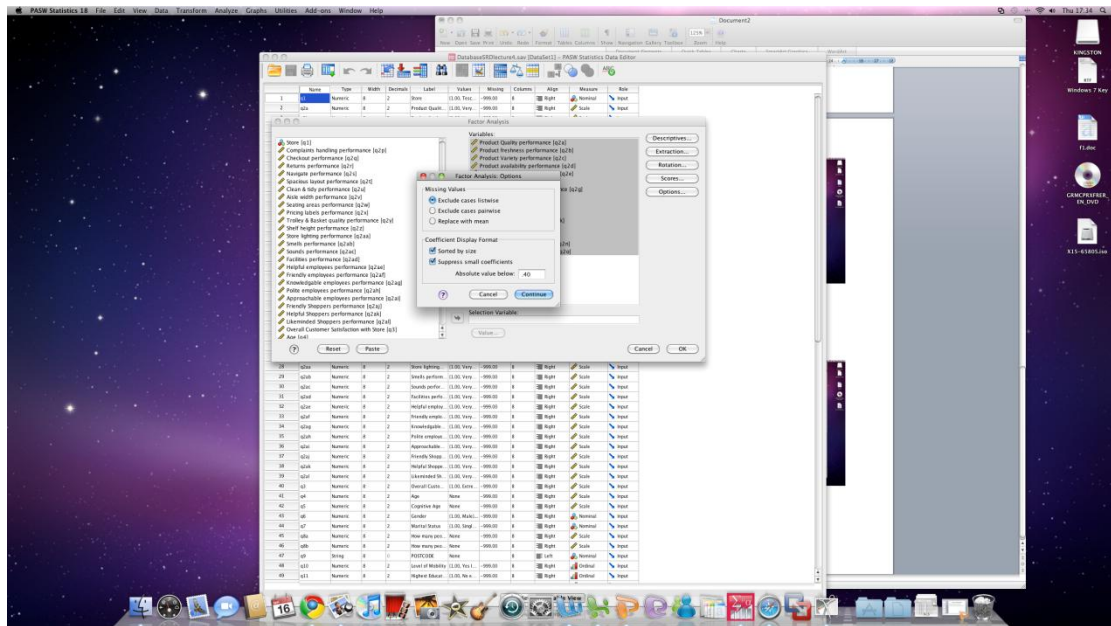
- In the 'Analyze' box, make sure that 'Correlation Matrix' is selected
- The extraction can be completed with Eigenvalues of over 1, or where there is evidence to support a fixed number of factors. For initial runs when there is little evidence of factor structure, make sure that Eigenvalues greater than 1 is marked
- Keep convergence iterations at 25
- Click 'Continue'



- In the main window click on 'Rotation'
- There are a number of methods for rotation available. Varimax is a reliable technique so choose this
- Ensure that the rotated solution is displayed
- Click 'Continue'



- In the main window click on the 'Options' box
- EFA requires that all data in the items is present so ensure that 'Listwise Deletion' is selected
- In the coefficient display format, mark both 'Sorted by Size' and 'Suppress Small Coefficients'
- In the suppress small coefficients box type 0.40. This will make sure that all factor loadings lower than 0.40 are not included in the analysis and will therefore appear blank. This allows for more accurate interpretation.
- Click 'Continue'



- In the main window click 'OK' to run the analysis
- Modify the analysis by removing variables and controlling the number of factors to extract