

Using IBM-SPSS

IBM-SPSS is a computer package design to cater for your every statistical need. It carries out analyses that take hours by hand in a matter of seconds. Here's a list of what IBM-SPSS can and cannot do:

IBM-SPSS will:

- Do complex statistical analysis in a few seconds.
- Produce graphs.
- Save you time.
- Do anything statistical that you ask it to do.

IBM-SPSS won't:

- Decide what test needs to be done in a given circumstance.
- Pass your exam for you.
- Make you breakfast.

So, although IBM-SPSS is a very powerful tool, it works only within the limits of your own statistical knowledge. In many respects it is rather stupid because if you provide it with data and ask it to run a completely meaningless analysis, it will happily oblige. Therefore, you still need to use your brain to initiate the correct analysis.

SPSS, PASW, or IBM-SPSS?

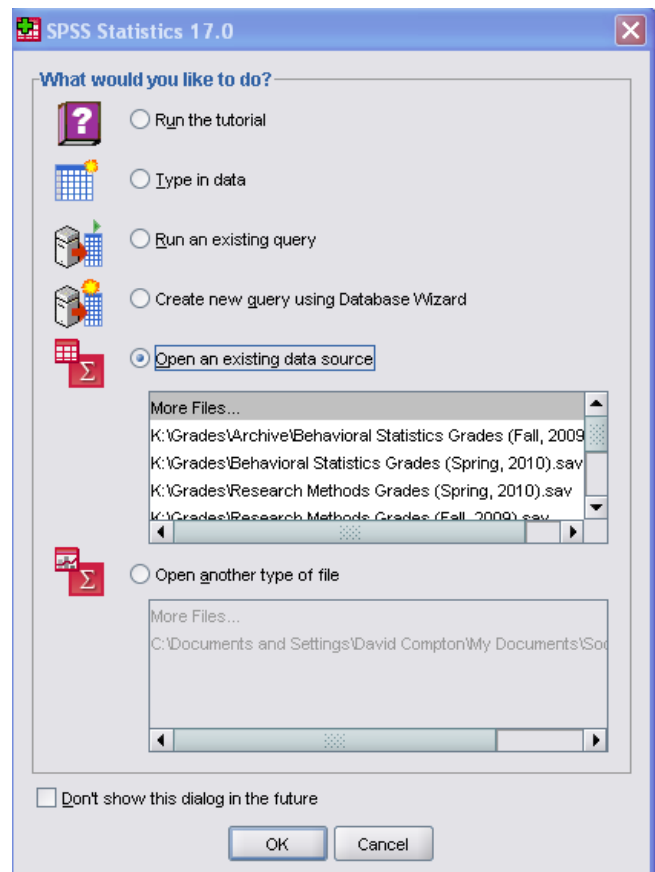
This year, purely to annoy data analysts everywhere, SPSS (the company) decided to change the name of their software from SPSS to PASW (Predictive Analysis SoftWare). This means that if you use versions before 17.0.2 you'll find that your software is called SPSS Statistics but if you use version 17.0.2 or later it will be called PASW Statistics (and now IBM-SPSS). Confused? Don't worry, we all are.

Getting started with IBM-SPSS

There are several excellent texts that give introductions to the general environment within which IBM-SPSS operates.

Once IBM-SPSS has been activated, a start-up window will appear which allows you to select various options. If you already have a data file on disk that you would like to open then select *Open an existing data source* by clicking on the so that it looks like : this is the default option. In the space underneath this option there will be a list of recently used data files that you can select with the mouse. If you want to open a data file that isn't in the list then simply select *More Files...* with the mouse and click on .

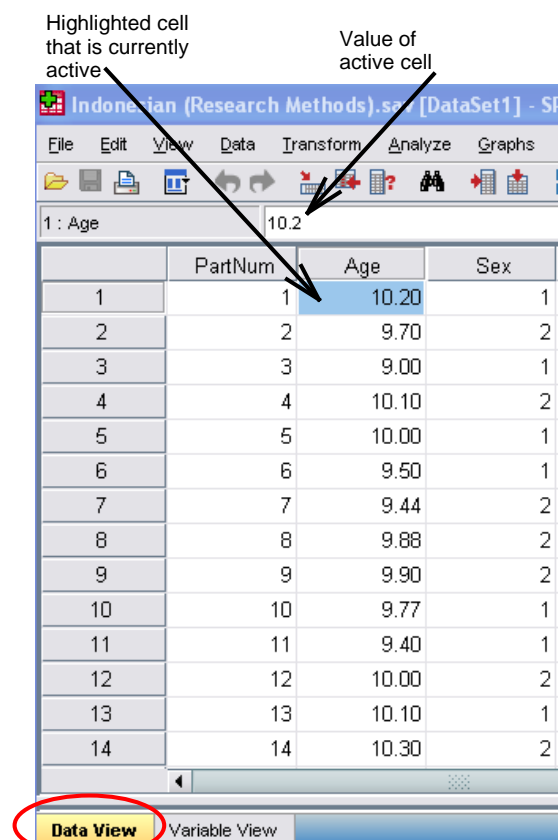
The main IBM-SPSS window includes a data editor for entering data. This window is where most of the action happens. The data editor has two views: the **data view** and the **variable view**. The data view is for entering data into the data editor, and the variable view allows us



to define various characteristics of the variables within the data editor. At the bottom of the data editor, you should notice that there are two tabs labeled 'Data View' and 'Variable View'

(**Data View** Variable View) and all we do to switch between these two views is click on these tabs (the highlighted tab tells you which view you're in, although it will be obvious).

Figure 2: The IBM-SPSS Data Editor ('Data View')



Below is a brief reference guide to each of the menus and some of the options that they contain. This is merely a summary and we will discover the wonders of each menu as we progress through our class:

File This menu allows you to do general things such as saving data, graphs or output. Likewise, you can open previously saved files and print graphs, data or output.

Edit This menu contains edit functions for the data editor. In IBM-SPSS it is possible to cut and paste blocks of numbers from one part of the data editor to another (which can be very handy when you realize that you've entered lots of numbers in the wrong place). You can also use the

Options...

to select various preferences such as the font that is used for the output.

Data This menu allows you to make changes to the data editor. The important features are **Insert Variable**, which is used to insert a new variable into the data editor (i.e. add a column);

Insert Cases, which is used to add a new row of data between two existing rows of data;

Split File..., which is used to split the file by a grouping variable; and **Select Cases...**, which is used to run analyses on only a selected sample of cases.

Transform We would use this menu if we want to manipulate one of our variables in some way. The compute function is also useful for transforming data (e.g. you can create a new variable that is the average of two existing variables). This function allows you to carry out any number of calculations on your variables.

Analyze The fun begins here, because the statistical procedures are in this menu. Below is a brief guide to the options in the statistics menu that will be used during our class (this is only a small portion of what is available):

Descriptive Statistics This menu is for conducting descriptive statistics (mean, mode, median, etc.), frequencies and general data exploration. There is also a command called

crosstabs that is useful for exploring frequency data and performing tests such as chi-square, Fisher's exact test and Cohen's kappa.

Compare Means ▶ This is where you can find t-tests and one-way independent ANOVA.

General Linear Model ▶ This menu is for ANOVAs such as two-way (unrelated, related or mixed), and any ANOVA with repeated measures (if available).

Correlate ▶ You can do bivariate correlations such as Pearson's R, Spearman's rho (ρ) and Kendall's tau (τ) as well as partial correlations.

Regression ▶ There are a variety of regression techniques available in IBM-SPSS. You can do simple linear regression and multiple linear regression.

Data Reduction ▶ You'll find factor analysis here.

Scale ▶ Here you'll find reliability analysis.

Nonparametric Tests ▶ There are a variety of non-parametric statistics available such the chi-square goodness-of-fit statistic, the binomial test, the Mann-Whitney test, the Kruskal-Wallis test, Wilcoxon's test and Friedman's ANOVA.

Graphs IBM-SPSS has some graphing facilities and this menu is used to access the Chart Builder (remember Behavioral Statistics). The types of graphs you can do include: bar charts, histograms, scatterplots, box-whisker plots, pie charts and error bar graphs (plus others).

Help This is an invaluable menu because it offers you online help on both the system itself and the statistical tests. The statistics help files are fairly incomprehensible at times (the program is not designed to teach you statistics) and are certainly no substitute for our textbooks.

As well as the menus there are also a set of *icons* at the top of the data window (see Figure 1) that are shortcuts to specific, frequently used, facilities. All of these facilities can be accessed via the menu system but using the icons will save you time. Below is a brief list of some of the most popular icons and their function:



This icon gives you the option to open a previously saved file (if you are in the data editor IBM-SPSS assumes you want to open a data file; if you are in the output viewer, it will offer to open a viewer file).



This icon allows you to save files. It will save the file you are currently working on (be it data or output). If the file hasn't already been saved it will produce the Save Data As dialog box.



This icon activates a dialog box for printing whatever you are currently working on (either the data editor or the output). The exact print options will depend on the printer you use. One useful tip is to select parts of the output by clicking on branches in the viewer window. When the

print dialog box appears remember to click on the option to print only the selected text. Selecting parts of the output will save a lot of trees because by default IBM-SPSS will print everything in the output window.



This icon looks a bit like your data have had a few too many beers and have collapsed in the gutter by the side of the road. The truth is considerably less exciting: it enables you to go directly to a case (a case is a row in the data editor and represents something like a participant, an organism or a company). This icon can be used to skip directly to a case. Clicking on this icon activates a dialog box in which you type the case number required (for example, participant number 222).



This icon doesn't allow you to spy on your neighbors (unfortunately), but it does enable you to search for words or numbers in your data file and output window. In the data editor it will search within the variable (column) that is currently active. This option is useful if, for example, you realize from a graph of your data that you have typed 66.6 instead of 6.6, you can simply search for 66.6 within that variable and replace that value with 6.6.


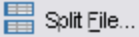


Clicking on this icon inserts a new case in the data editor (so it creates a blank row at the point that is currently highlighted in the data editor). This function is very useful if you need to add new data at a particular point in the data editor.


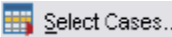


Clicking on this icon creates a new variable to the left of the variable that is currently active (to activate a variable simply click once on the name at the top of the column).



Clicking on this icon is a shortcut to the   function. There are often situations in which you might want to analyze groups of cases separately. In IBM-SPSS we differentiate groups of cases by using a coding variable and this function lets us divide our output by such a variable. For example, we might test males and females on their statistical ability. We can code each participant with a number that represents their gender (e.g. 1 = female, 2 = male). If we then want to know the mean statistical ability of each gender we simply ask the computer to split the file by the variable Gender. Any subsequent analyses will be performed on the men and women separately.



This icon is a shortcut to the   function. If you want to analyze only a portion of your data, this is the option for you! This function allows you to specify what cases you want to include in the analysis.



Clicking on this icon will either display or hide the value labels of any coding variables. We often group people together and use a coding variable to let the computer know that a certain participant belongs to a certain group. For example, if we coded gender as 1 = female, 2 = male then the computer knows that every time it comes across the value 1 in the Gender column, that person is a female. If you press this icon, the coding will appear on the data editor rather than the numerical values; so, you will see the words male and female in the Gender column rather than a series of numbers. This idea will become clear when we enter data (and do other things).

Entering Data Using the Data Editor

Overview

When you first load IBM-SPSS it will provide a blank data editor (showing the 'data view') with the title *New Data*. When inputting a new set of data, you must input your data in a logical way. The IBM-SPSS data editor is arranged such that *each row represents data from one participant while each column represents a variable*. There is no discrimination between independent and dependent variables, both types should be placed in a separate column. It follows from this arrangement that any variable measured with the same participants (a repeated measure) should be represented by several columns (each column representing one level of the repeated measures variable). However, when a between-group design was used (e.g. different participants were assigned to each level of the independent variable) the data will be represented by two columns: one that has the values of the dependent variable and one that is a coding variable indicating to which group the participant belonged. This idea will become clearer as you learn about how to carry out specific procedures.

The data editor is made up of lots of **cells**, which are just boxes in which data values can be placed. When a cell is active it becomes highlighted in light blue (usually with a black surrounding box (see above, Figure 1). You can move around the data editor, from cell to cell, using the arrow keys ← ↑ ↓ → (found on the right of your keyboard) or by clicking the mouse on the cell that you wish to activate. To enter a number into the data editor simply move to the cell in which you want to place the data value, type in the value, then press the arrow button appropriate to the direction in which you wish to move. So, to enter a row of data, move to the far left of the row, type the first value and then press → (this inputs the value and then moves you into the next cell on the right).

The first step in entering your data is to create some variables using the 'variable view' of the data editor, and then to input your data using the 'data view' of the data editor. We'll go through these two steps by working through an example.

Creating a Variable: The 'Variable view'

Before we input any data into the data editor, we need to create the variables. To create variables we use the 'Variable View' of the data editor. To access this view click on the 'Variable View' tab at the bottom of the data editor (**Data View** **Variable View**); the contents of the window will change (see Figure 3).

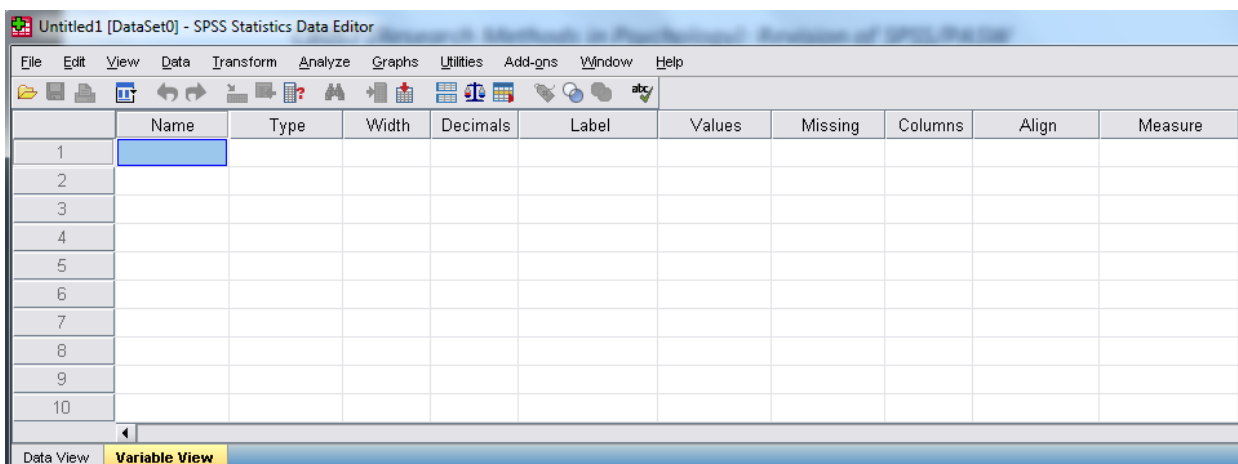


Figure 3: The 'Variable View' of the IBM-SPSS Data Editor

Every row of the variable view represents a variable, and you set characteristics of a particular variable by entering information into the labeled columns. You can change various characteristics of the variable by entering information into the following columns (play around and you'll get the hang of it):

Name

You can enter a name in this column for each variable. This name will appear at the top of the corresponding column in the data view, and helps you to identify variables in the data view. There are certain symbols you can't use (mainly symbols that have other uses in IBM-SPSS such as +, -, \$, &), and you can't use spaces. Underscore does work here though. If you use a character that IBM-SPSS doesn't like you'll get an error message saying that the variable name is invalid when you click on a different cell, or try to move off the cell using the arrow keys.

Type

You can have different types of data. Mostly we'll use numeric variables (which just mean that the variable contains numbers - IBM-SPSS assumes this data type). You will come across string variables, which consist of strings of letters. If you wanted to type in people's names, for example, you would need to change the variable type to be string rather than numeric. You can also have currency variables (i.e., \$) and date variables.

Label

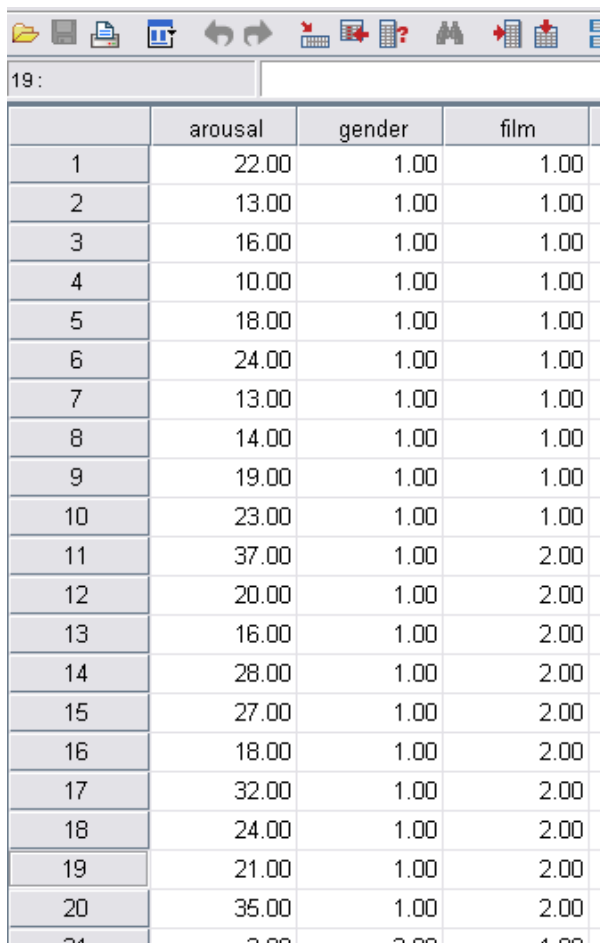
The name of the variable (see above) has some restrictions on characters, and you also wouldn't want to use huge long names at the top of your columns (they become hard to read). Therefore, you can write a longer variable description in this column. Anything goes, even [Supercalifragilisticexpialidocious](#). This may seem pointless, but is actually one of the best habits you can develop.

Values

This column is for assigning numbers to represent groups of people (more later).

Measure

This is where you define the level at which a variable was measured (Nominal, Ordinal or Scale; Scale equals interval & ratio). This is important, especially if you want to draw graphs!



	arousal	gender	film
1	22.00	1.00	1.00
2	13.00	1.00	1.00
3	16.00	1.00	1.00
4	10.00	1.00	1.00
5	18.00	1.00	1.00
6	24.00	1.00	1.00
7	13.00	1.00	1.00
8	14.00	1.00	1.00
9	19.00	1.00	1.00
10	23.00	1.00	1.00
11	37.00	1.00	2.00
12	20.00	1.00	2.00
13	16.00	1.00	2.00
14	28.00	1.00	2.00
15	27.00	1.00	2.00
16	18.00	1.00	2.00
17	32.00	1.00	2.00
18	24.00	1.00	2.00
19	21.00	1.00	2.00
20	35.00	1.00	2.00
21	2.00	2.00	1.00

first white cell in the column labeled Name.

Setting Up a Between-Groups Design



To begin with, imagine that a film company director was interested in whether there was really such a thing as a 'chick flick' (a film that typically appeals to women more than men). He took 20 men and 20 women and showed half of each sample a film that was supposed to be a 'chick flick' (*Bridget Jones' Diary*), and the other half of each sample a film that didn't fall into the category of 'chick flick' (*Avatar*). In all cases he measured their physiological arousal as an indicator of how much they enjoyed the film.

We need to pay attention to the rule of the data editor. Remember that each score above comes from a different person; therefore we will need 40 rows (not 10 as above). To [enter the data](#) into the IBM-SPSS data editor we need to create several variables. If we begin with the variable **arousal**, we should follow these steps:

1. Move the on-screen arrow (using the mouse) to the

2. Type the word Arousal.
3. Move off of this cell using the arrow keys on the keyboard (you can also just click on a different cell, but this is actually a very slow way of doing it).

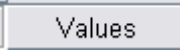


You've just created your first variable! You should notice that once you've typed a name, IBM-SPSS creates default settings for the variable (such as assuming its numeric and assigning 2 decimal places). Now because I want you to get into good habits, move to the cell in the column labeled label and type 'Average Physiological Arousal During the Film'.

Once the variable has been created, you can return to the data view by clicking on the 'data view' tab at the bottom of the data editor (). The contents of the window will change, and you'll notice that the first column now has the label arousal. To enter the data, click on the white cell at the top of the column labeled arousal and type the first value (22). To register this value in this cell, we have to move to a different cell and because we are entering data down a column, the most sensible way to do this is to press the  key on the keyboard. This action moves you down to the next cell, and the number 22.00 should appear in the cell above. Enter the next number (13) and then press down arrow key to move down to the next cell, and so on.

Creating Coding Variables

A coding variable (also known as a grouping variable) is a type of variable that you will use on numerous occasions: it is a variable consisting of a series of numbers that represent levels of an independent variable. In experiments, coding variables are used to represent independent variables that have been measured between-groups (i.e., different participants were assigned to different groups). So, if you were to run an experiment with one group of participants in an experimental condition and a different group of participants in a control group, you might assign the experimental group a code of 2, and the control group a code of 1. When you come to put the data into the data editor, then you would create a variable (which you might call group) and type in the value 2 for any participants in the experimental group and a 1 for any participant in the control group. This tells the computer that all of the cases that have been assigned the value 1 should be treated as belonging to the same group, and likewise for the participants assigned the value 2.

There is a simple rule for how variables should be placed in the IBM-SPSS data editor: levels of the between-group variables go down the data editor whereas levels of within-subject (repeated measures) variables go across the data editor. We will practice this rule over the coming weeks.

We have two coding variable in our data: the one describing whether a person was male or female, and a second describing whether the person watched 'Bridget Jones' diary' or 'Memento'. Let's create the variable gender first. To create this coding variable, we follow the steps for creating a normal variable, but we also have to tell the computer which numeric codes have been assigned to which groups. So, first of all, move to the cell in the second row in the column labeled Name type a name (let's call it **sex**). I'm still trying to push you to develop good habits, so move along the third row to the column called Label and give the variable a full description such as 'Participant's Sex'. Then to define the group codes, move along the row to the column labeled  and into this cell: . Click on  to access the Value Labels dialog box (see Figure 3). Click on the appropriate cell in the column labeled Values

Click on the appropriate cell in the column labeled Values

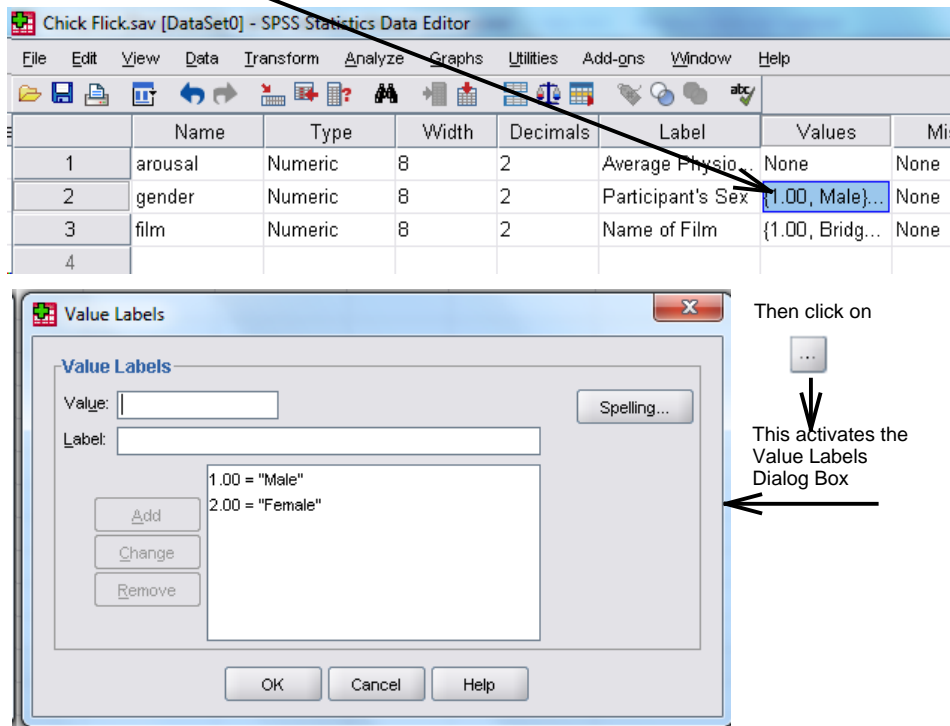
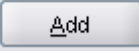
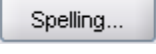








Figure 5: Defining coding values in IBM-SPSS

The Value Labels dialog box is used to specify group codes. This can be done in three easy steps. First, click with the mouse in the white space next to where it says Value (or press Alt and u at the same time) and type in a code (e.g., 1). These codes are completely arbitrary; for the sake of convention people typically use 0, 1, 2, 3, etc, but in practice you could have a code of 561 if you were fond of our area code. The second step is to click the mouse in the white space below, next to where it says Value Label (or press Tab, or Alt and e at the same time) and type in an appropriate label for that group. In Figure 5 I have already defined a code of 1 for Males, and then I have typed in 2 as my code and given this a label of Female. The third step is to add this coding to the list by clicking on . When you have defined all of your coding values you can click on  and IBM-SPSS will check your variable labels for spelling

errors (always a good idea). To finish, click on ; if you click on  and have forgotten to add your final coding to the list, IBM-SPSS will display a message warning you that any pending changes will be lost. In plain English this simply tells you to go back and click on  before continuing.

Finally, coding variables always represent categories and so the level at which they are measured is nominal  Nominal (or ordinal  Ordinal if the categories have a meaningful order). Therefore, you should specify the level at which the variable was measured by going to the column labeled **Measure** and selecting (or if the groups have a meaningful order) from the drop-down list.

Having defined your codes, you can then go to the data view (by clicking the tab at the bottom of the window) and type these numerical values into the appropriate column (so if a person was a male type a value of 1, if they were a female type the value 2). What is really cool is that you can get the computer to display the numerical codes, or the value labels that you gave them by clicking on .

Having created these variables arousal and gender, try to create the variable film and enter the rest of the data yourself. Save these data in a file called ChickFlick.sav.

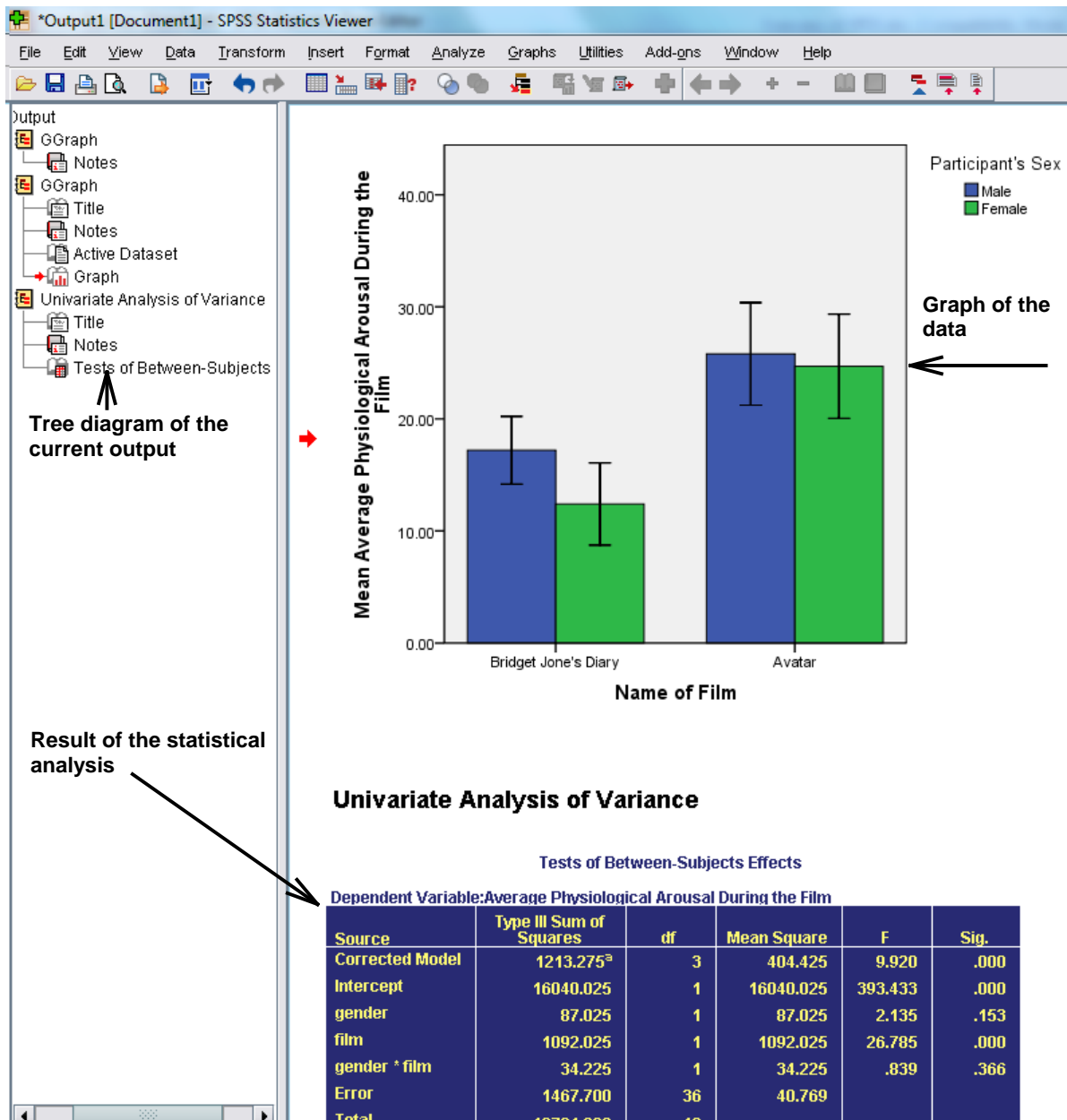


Figure 6: The Output Navigator