Exploring the 2012 AuSSA Data

In this workshop we will:

1. Explore the 2012 AuSSA data
2. Learn about missing values
3. How to construct a new variable by using the Recode and Compute command
4. How to compute descriptive statistics.

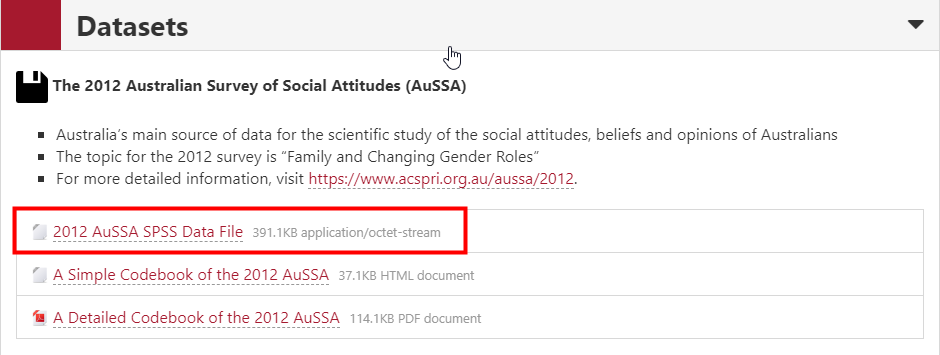
The third workshop introduces the 2012 Australian Survey of Social Attitudes (AuSSA). AuSSA is a longitudinal survey about social attitudes, beliefs and opinions of Australians. It is a biennial survey that began in 2003. It has been the main source of data for studies about how Australians think and feel about their lives as well as about how they change over time.

AuSSA is also the Australian component of the International Social Survey Programme (ISSP). The ISSP is a cross-national collaboration on surveys covering important topics for social science research. In addition, the ISSP chooses a special topic each year, which is called as module, repeating that topic from time to time. The module topic for 2012 was “Family and Changing Gender Roles”. This is the most recent ISSP data on this topic, which was surveyed four times in 1988, 1994, 2002, and 2012.

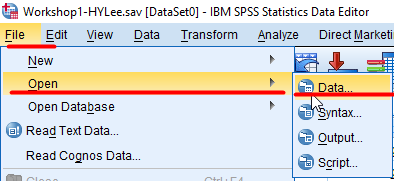
We are using this dataset in this workshop. The dataset is extracted from the 2012 ISSP data but includes only Australian respondents. This 2012 AuSSA is one of the datasets that you can use for your final research project. If you want to know more about the 2012 AuSSA, visit <https://www.acspri.org.au/aussa/2012>.

# Opening the 2012 AuSSA

Let’s get this data. Go to the course iLearn page and find “2012 AuSSA SPSS Data File” under the tab of Datasets. Download this file. Also, look at “A Simple Codebook of the 2012 AuSSA” and “A Detailed Codebook of the 2012 AuSSA” which provide useful information on variables and their values.

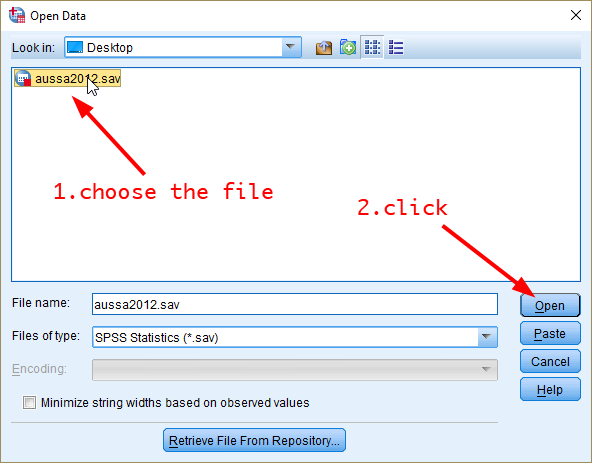
<Figure 1>

After the download, **go to File > Open > Data**.



<Figure 2>

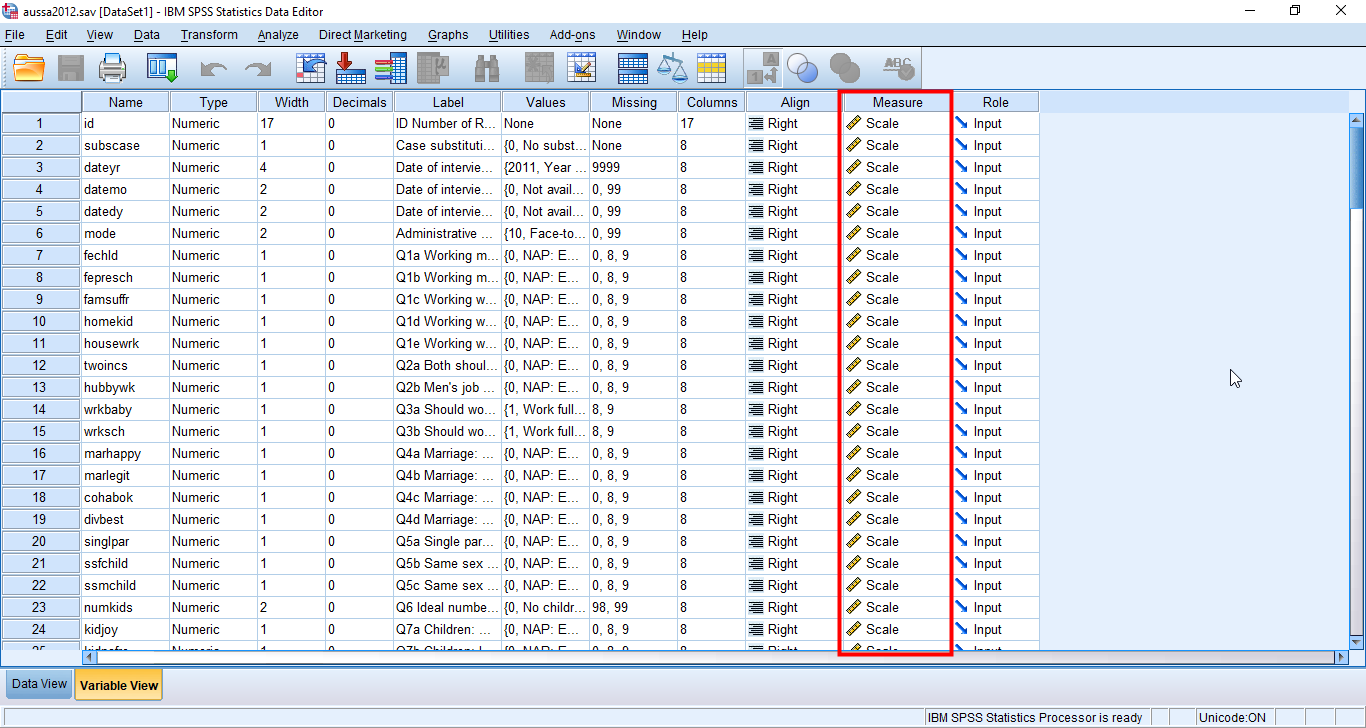
And **choose the downloaded file (“aussa2012.sav”)**. Then, **click Open**. You will see the dataset imported on your SPSS.



<Figure 3>

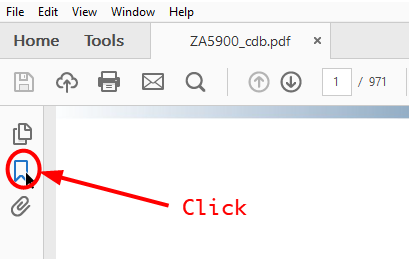
# Looking around the 2012 AuSSA

**Click the Variable View tab** at the bottom of the Data Editor window. You will see all the variable information. Note that the levels of measurement for all variables are intentionally set to “Scale”. Thus, it is necessary to assign a relevant level of measurement for variables you are analysing before you start an analysis.



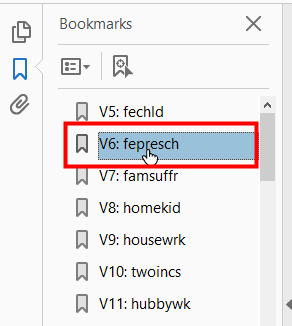
<Figure 4>

For instance, suppose that we are analysing a variable, *fepresch*. For detailed information on this variable, open “A Detailed Codebook of the 2012 AuSSA” using Adobe Acrobat Reader. Click the icon of Bookmark in the left pane.



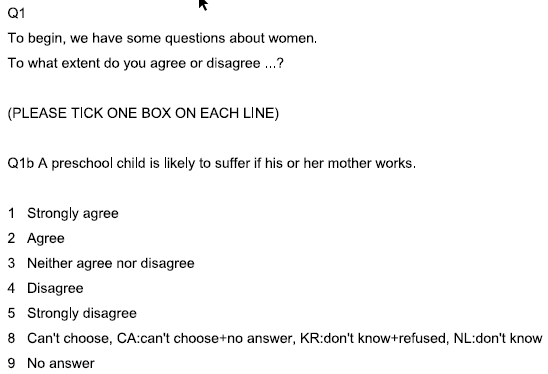
<Figure 5>

Then, you will see the variable list in the left pane. Click the variable you want to study. In this example the variable is *fepresch*. Then, the pdf will show the page which explains the variable of your choice.



<Figure 6>

The codebook shows the questionnaire and response options, which will allow you to figure out the levels of measurement.



<Figure 7>

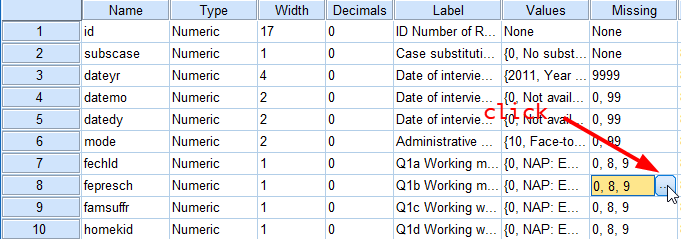
<Figure 7> shows information on *fepresch*. It asks respondents the extent to which they agree or disagree with the statement that a preschool child is likely to suffer if his or her mother works. Although you can see nine response options, we will take into account only five options (Strongly agree; Agree; Neither agree nor disagree; Disagree’ Strongly disagree) as valid options. The other two options will be counted as system missing values, and I will explain missing values in the next section. Since the five response options have internal ranks but do not have the same distance between them, *fepresch* is an ordinal variable. You learned how to change the level of measurement in the Variable View tab (if you cannot recall it, see p. 8 of Workshop 1).

I recommend you spend some time in looking around variables in the 2012 AuSSA. The dataset includes so many variables relating to gender and family issues. You may find many variables that inspire your intellectual curiosity.

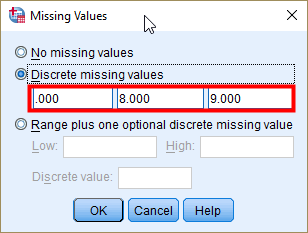
# Understanding Missing Values

Missing or invalid values are generally too common to ignore. Survey respondents may refuse to answer certain questions, may not know the answer, or may answer in an unexpected format. If you do not filter or identify these data, your analysis may not provide accurate results.

For numeric data, empty data fields or fields containing invalid entries are converted to system-missing, which is identifiable by a single period(.). But more common practice is that some numerical values are treated as missing values. For example, in <Figure 7> a numeric value eight denotes “Can’t choose”, and a numeric value of nine represents “No answer”. These responses are typical types of missing values. In the Variable View tab, if you **click a small square in the column of Missing for *fepresch*** *(see* <Figure 8>*)*, you will see that the value 8 and 9 are treated as missing values in the popped-up window (see <Figure 9>). **To go back to the previous Variable View tab, click *OK*** in the popped-up window.



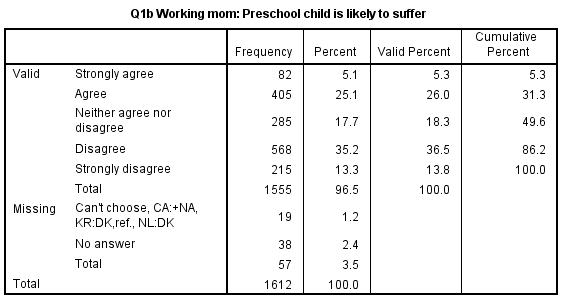
<Figure 8>



<Figure 9>

The reason value is missing may be important to your analysis. For example, you may find it useful to distinguish between those respondents who refused to answer a question and those respondents who didn’t answer a question because it was not applicable.

When analysing variables, SPSS always take into account of missing values. For example, look at the frequency table of *fepresch* in <Figure 10>. You will see the frequency of two missing categories (8 = “Can’t choose”; 9 = “No answer”) in the bottom. Note that the “Percent” in the table is based on all responses including missing categories, while the “Valid Percent” is based on only valid responses which exclude missing categories. Then, the question is which percent (more broadly, which statistics) we should report. The norm is to report the valid percent unless the non-response rates (the percent of missing values) are too much high. This because a high level of non-response rate could provide an inaccurate estimation. In <Figure 10>, the non-response rate is 3.5%, which is very small.



<Figure 10>

# Recoding Variables

## How to Recode to Make a Dichotomous Variable

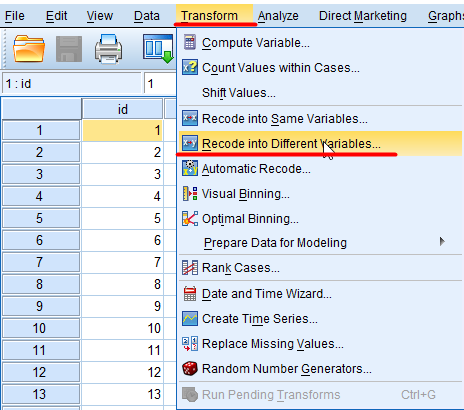
Suppose we want a variable where people are categorised into dichotomous views about the statement of whether a preschool child is likely to suffer if his or her mother works. We will make a new variable (*dichfepresch*) in which those who agree or strongly agree to that statement are coded as 1, and those who do not agree are coded as 0. <Table 1> shows the recoding scheme of this new variable.

|  |  |  |  |
| --- | --- | --- | --- |
| **Old Variable(*fepresch*)** | | **New Variable (*dichfepresch*)** | |
| **Value** | **Label** | **Value** | **Label** |
| 1 | Strongly agree | 1 | Agree |
| 2 | Agree |
| 3 | Neither agree nor disagree | 0 | Don’t agree |
| 4 | Disagree |
| 5 | Strongly disagree |
| 8 | Can’t choose | . | System-missing |
| 9 | No answer |

<Table 1>

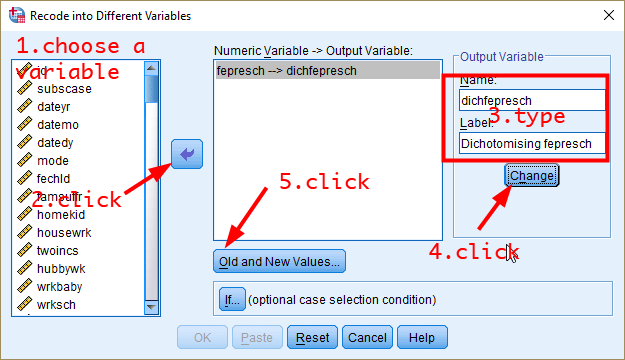
Do you remember how we made an age group variable by recoding the age variable in the first workshop? We will follow the same steps (See pp.17—21 in the Workshop 1).

**Go to Transform > Recode into Different Variables**.



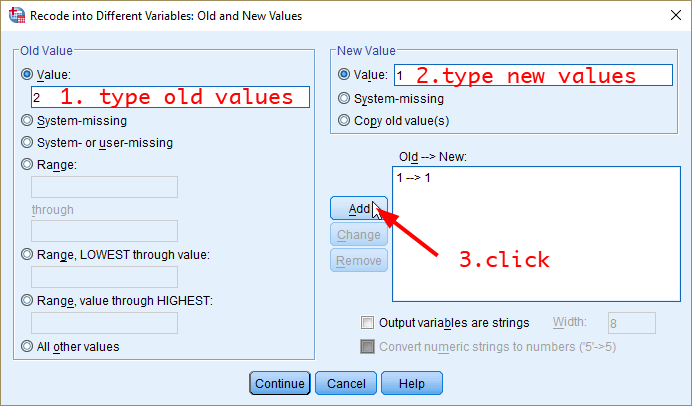
<Figure 11>

In the popped-up dialog box, **select *fepresch* in the left variable box and then click the right arrow**. Then, you will see “fepresch → ?”. **In the right box titled *Output Variable*, type a new variable name (*dichfepresch*) and its label (Dichotomising fepresch)**. Then, **click Change.** “fepresch → ?” will be changed into “fepresch → dichfepresch”. Then, **click *Old and New Values***.



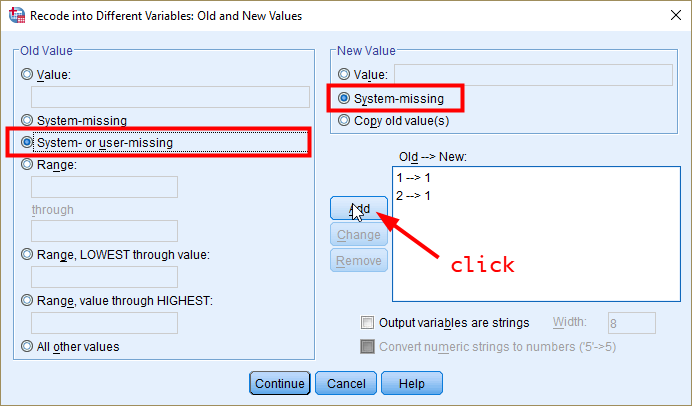
<Figure 12>

In the popped-up window, select Value and type “1” in the *Old Value* section. Type “1” in the *New Value* section. Click *Add*. Type “2” in the *Old Value* section and “1” in the *New Value* section. Click *Add*.



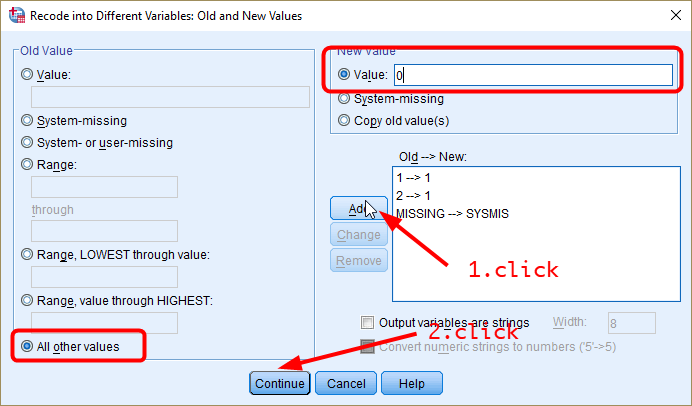
<Figure 13>

Then, select “System- or user-missing” in the *Old Value* section and “System-missing” in the *New Value* section. Click *Add*.



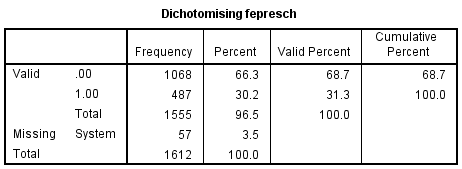
<Figure 14>

After then, select “All other values” in the *Old Value* section. Select *Value* in the *New Value* section and type “0”. Then, click *Add*. Complete the process by clicking Continue and then OK. Then you will see the message in the output window that the new variable is created.



<Figure 15>

To check whether the new variable is created correctly, produce a frequency table of *dichfepresch*. And compare it with that of *fepresch*(see <Figure 10>). First, compare the number of missing cases. It should be the same. Second, the frequency of “1” should equal the sum of “Strongly agree” and “Agree” in <Figure 10>. <Figure 16> meets all the requirements.



<Figure 16>

It would also be great if you assign labels for each value in *dichfepresch*. You should know how to do this. If you cannot recall it, see pp.10—11 in the workshop 1.

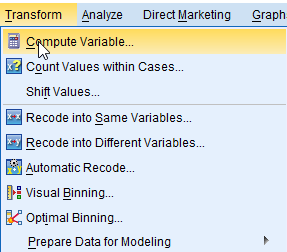
# Computing Variables

## Creating a Variable of Birth Years

Suppose that we want to make a new variable of birth years using a variable of *age*. Given that the survey was conducted in 2012, the relationship between birth years and age is:

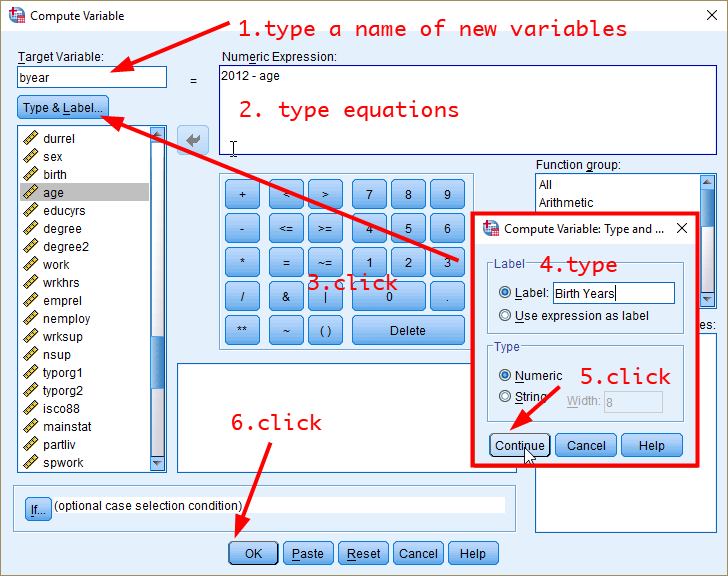
Let’s make a variable of birth years using the *Compute* command. The Compute command will let you construct new variables by using functions like arithmetic, statistical, random number, string and date/time functions. This time we will use an arithmetic function (i.e., subtraction).

**Go to Transform > Compute Variable…**



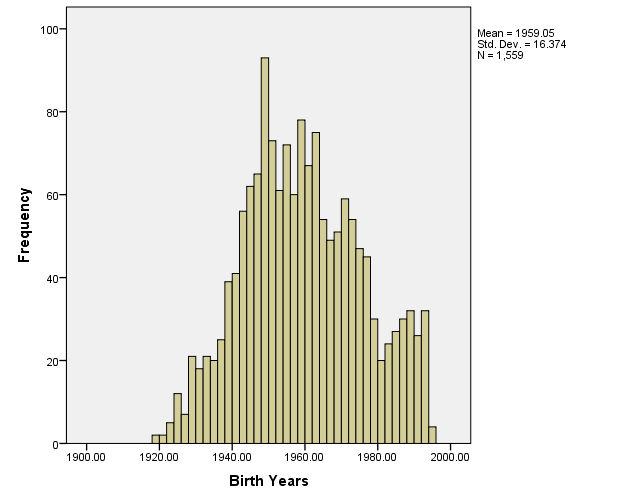
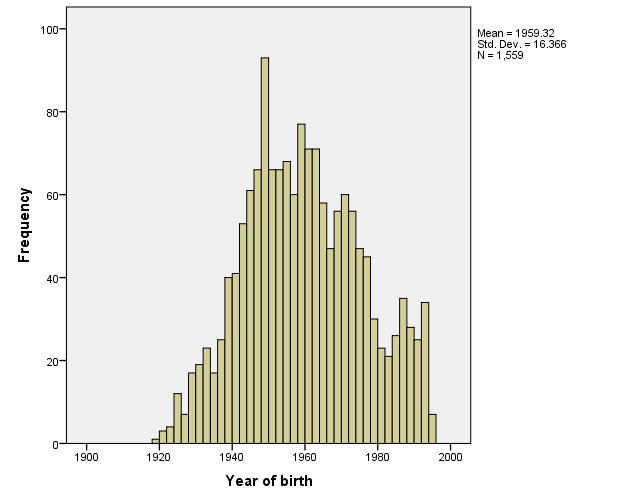
<Figure 17>

In the popped-up window, 1) **type a name of new variables** (in our case, it is *byear*) in the section of *Target Variable:*. 2) **Type equations** that speak the relationship between old and new variables in the section of *Numeric Expression:*(in our case, “2012 – age”). 3) **Click *Type & Label***, and it will show a new window. 4) **Type a label of new variables** and 5) **Click Continue** in a newly popped-up window. 6) **Click OK**. After this, go to the variable view and you will see a newly created variable, *byear* at the bottom.



<Figure 18>

Let’s check whether the new variable of birth years is created correctly. Actually, we already had a variable about birth year, which is *birth*. Comparing the histogram of byear (that we made) and birth (that is included in the 2012 AuSSA) will give the answer. The left histogram is for *byear*, and the right histogram for *birth*. As you see, they are the same, confirming that *byear* is generated correctly.

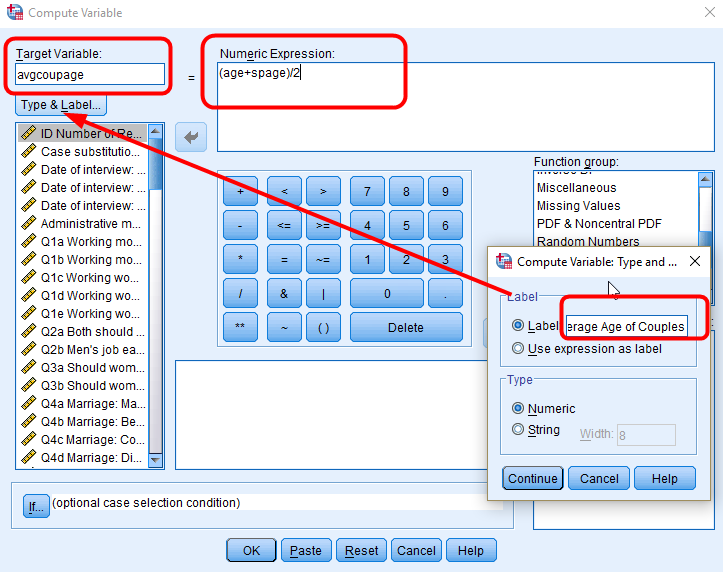
<Figure 19>

## Creating a Variable from a Combination of Multiple variables

This time we will make a new variable that contains the mean age of couples. This new variable(*avgcoupage*) can be generated via:

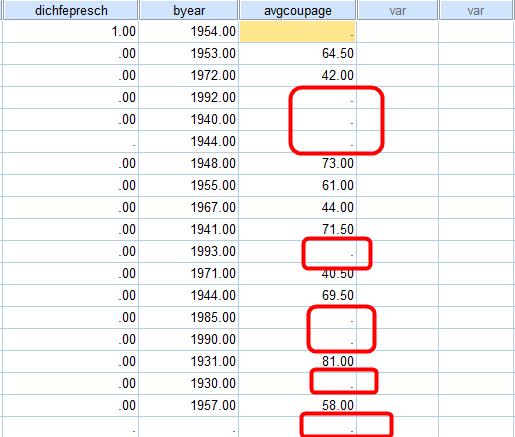
**Go to Transform > Compute Variable…**

In the popped-up window, 1) **type “avgcoupage”** in the section of *Target Variable:*. 2) **Type “(age+spage)/2”** in the section of *Numeric Expression:*. 3) **Click *Type & Label***, and it will show a new window. 4) **Type a label of new variables** and 5) **Click Continue** in the newly popped-up window. 6) **Click OK** at the bottom. Then, you will see a newly created variable, *avgcoupage*.



<Figure 20>

Note that you will see a lot of missing values in a newly made variable (see <Figure 21>). This is because we cannot calculate the average age of couples for people without a partner or spouse.



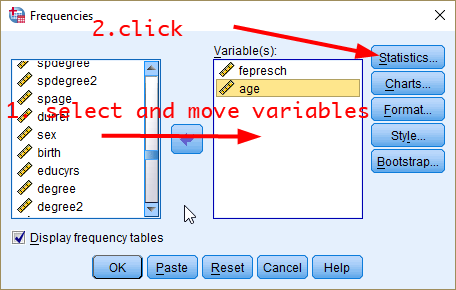
<Figure 21>

# Getting Descriptive Statistics

This section introduces three commands to compute descriptive statistics: Frequencies, Descriptives and Explore. We will compute descriptive statistics for *fepresch* (ordinal variable) and *age* (continuous variable).

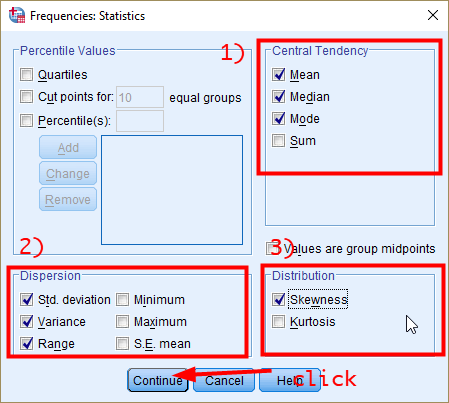
## Compute Descriptive Statistics with the Frequencies Command

**Go to Analyze > Descriptive Statistics > Frequencies**. In the popped-up window, **select variables** for which you want to compute descriptive statistics (in our case, they are *fepresch* and *age*) and **move them into the pane of *Variable(s)***. Then, **click *Statistics***.



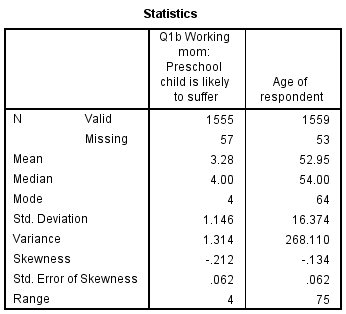
<Figure 22>

In a newly popped-up window, **1) tick *Mean*, *Median* and *Mode* in the section of *Central Tendency*, 2) tick *Std. deviation*, *Variance* and *Range* in the section of *Dispersion* and 3) tick *Skewness* in the section of *Distribution***. Then **click *Continue*** at the bottom. You will be back to the previous window. **Click *OK*** at the bottom.



<Figure 23>

In the output window, you will see the descriptive statistics of both *fepresch* and *age* (see <Figure 24>).



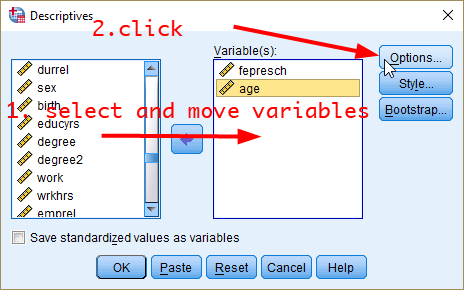
<Figure 24>

**Activity Suggestion**

Discuss which measure of central tendency and variability is most relevant for each variable and how you will interpret the measures.

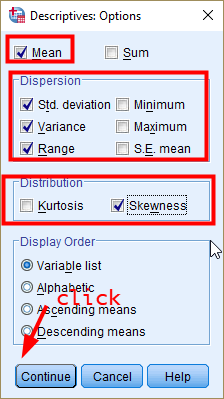
## Compute Descriptive Statistics with the Descriptive Command

**Go to Analyze > Descriptive Statistics > Descriptive**. In the popped-up window, **select variables** for which you want to compute descriptive statistics (in our case, they are *fepresch* and *age*) and **move them into the pane of *Variable(s)***. Then, **click *Options***.



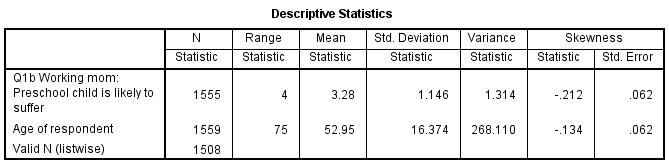
<Figure 25>

In a newly popped-up window, **1) tick *Mean* at the top, 2) tick *Std. deviation*, *Variance* and *Range* in the section of *Dispersion* and 3) tick *Skewness* in the section of *Distribution***. Then **click *Continue*** at the bottom. You will be back to the previous window. **Click *OK*** at the bottom.



<Figure 26>

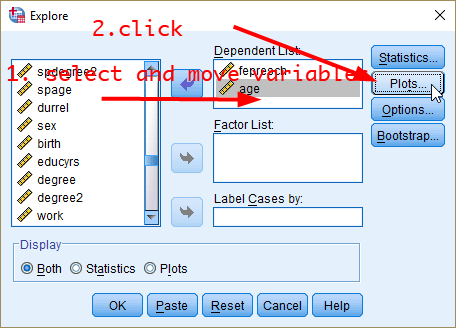
In the output window, you will see the descriptive statistics of both *fepresch* and *age* (see <Figure 27>). As you may notice, the output does not show the median of variables. Thus, I always prefer the Frequency to the Descriptive command.



<Figure 27>

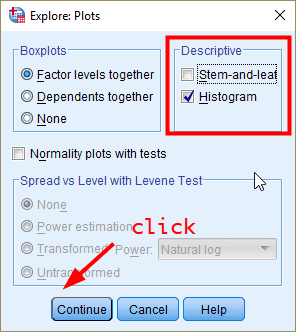
## Compute Descriptive Statistics with the Explore Command

**Go to Analyze > Descriptive Statistics > Explore**. In the popped-up window, **select variables** for which you want to compute descriptive statistics (in our case, they are *fepresch* and *age*) and **move them into the pane of *Dependent List***. Then, **click *Plots***.



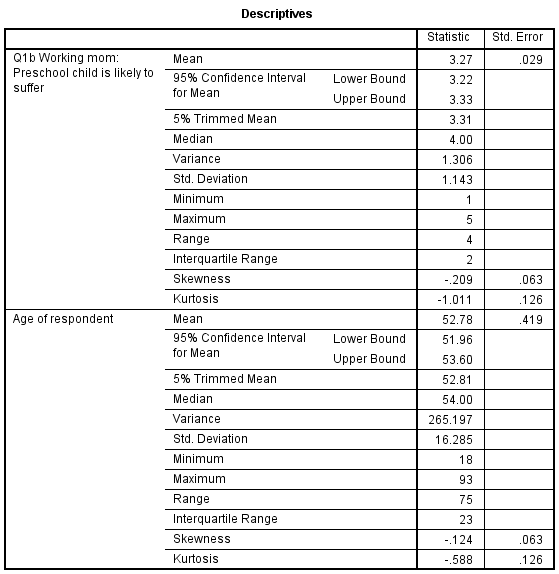
<Figure 28>

In a newly popped-up window, **tick only *Histogram***. Then, **click *Continue***. You will be back to the previous window. **Click *OK*** at the bottom.



<Figure 29>

In the output window, you will see the descriptive statistics of both *fepresch* and *age* (see <Figure 30>). Also, the charts of variables (e.g., histograms and box plots) will be shown as well.

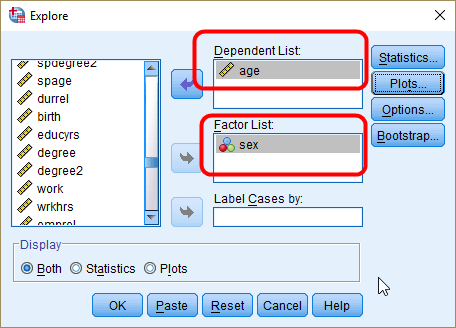


<Figure 30>

## Compare Descriptive Statistics by Group Using the Explore Command

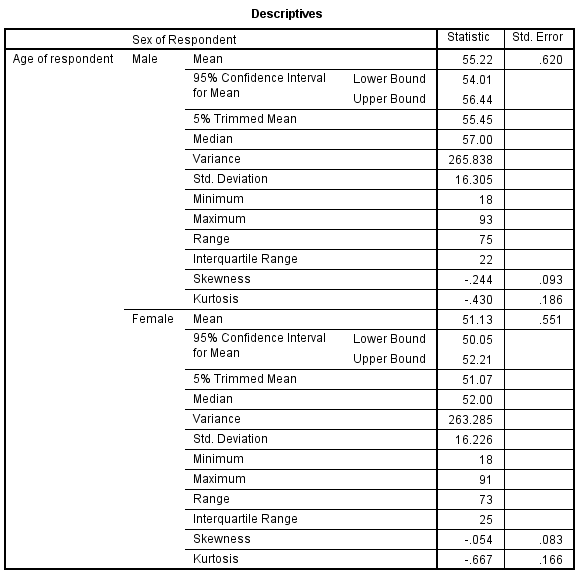
When we want to examine the relationship between two variables, it would be useful to compare descriptive statistics of one variable by groups defined by another variable. For example, comparison of hours worked per week between men and women could suggest some relationships between gender and working hours.

In this section, we will compare the distribution of age between men and women. Go **Go to Analyze > Descriptive Statistics > Explore**. In the popped-up window, 1) **select variables** for which you want to compute descriptive statistics (in our case, it is *age*) and **move it to the pane of *Dependent List***. 2) **select a group variable** (in our case, it is *sex*) by which descriptive statistics are compared and **move it to the pane of *Factor List***. Then, **click *Plots***. As you did in <Figure 29>, **tick only *Histogram***. Then, **click *Continue***. You will be back to the previous window. **Click *OK*** at the bottom.

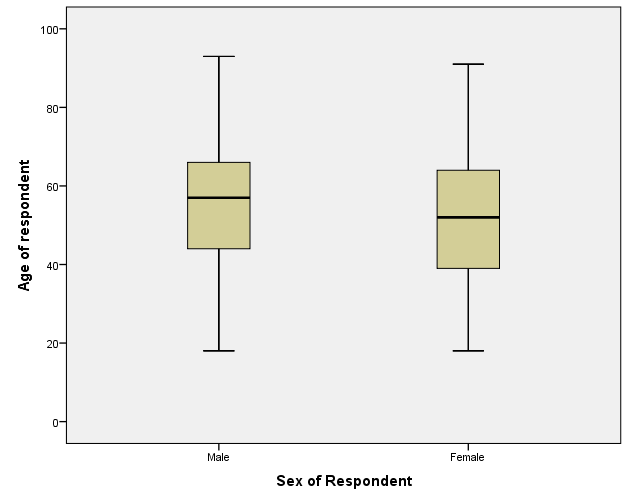


<Figure 31>

In the output window, you will see the descriptive statistics of both males and females.(see <Figure 32>). Also, you will see the box plot of age by gender which is very helpful for comparing the distribution between groups (see <Figure 33>).



<Figure 32>



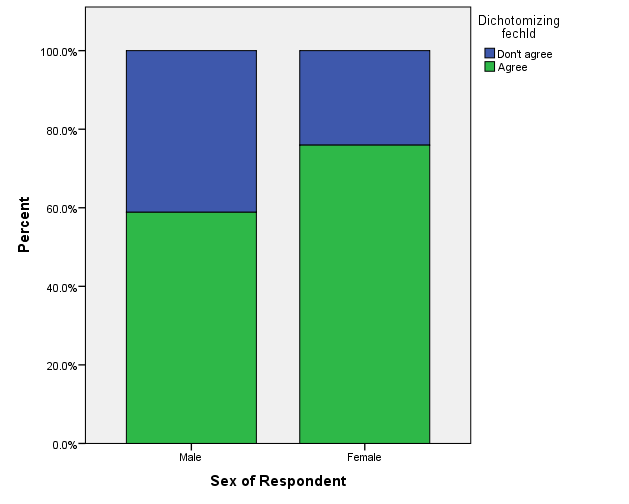
<Figure 33>

**Workshop Activities**

The website of the 2012 AuSSA presents two selected findings. The activity question Q1 and Q2 ask you to replicate these findings using the 2012 AuSSA dataset. You can see the original findings here: <https://www.acspri.org.au/aussa/2012>

Q1. Read the following report excerpted from the website of the 2012 AuSSA.

*For the question "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work", 59% of male respondents agreed or strongly agreed, but 76% of female respondents agreed or strongly agreed.*



**Replicate this finding using *fechld***. You need to dichotomise *fechld* in a way those who strongly agree or agree will be coded as 1, otherwise as 0 (read the report carefully). Then, create a stacked bar plot using the newly created variable and *sex* (see pp. 16—20 in workshop 2).

Q2. Read the following report excerpted from the website of the 2012 AuSSA.

*When asked "Who do you think should PRIMARILY cover the costs of childcare for children under school age?" - 73% of respondents chose "The family", 20% chose "The government" and small remainder chose "the employer".*

**Replicate this finding using *careprov*.** You need to recode *careprov* in a way that the option of Non-profit organisation and Private childcare providers are treated as missing values (read the report carefully). Then make a frequency table of this newly created variable.

Q3. The 2012 AuSSA asked five questions to study people’s attitudes toward working moms. Respondents are asked to indicate whether they agree or disagree with each statement:

* ***fechld***: A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.
* ***fepresch***: A preschool child is likely to suffer if his or her mother works.
* ***famsuffr***: All in all, family life suffers when the woman has a full-time job.
* ***homekid***: A job is all right, but what most women really want is a home and children.
* ***housewrk***: Being a house wife is just as fulfilling as working for pay.

Using descriptive statistics of these five variables, **find a variable in which respondents show the most progressive (or non-traditionalist) view of working moms.** Note that agree to the statement does not always translate into more progressive view of working moms. Also, **find a variable in which respondents show the most varying view of working moms**.