The following resources are associated: The dataset *’Birthweight.csv’*  and ‘*Getting started in Jamovi’*



Creating and recoding variables in

There are several situations where adjustments to variables in the dataset need to be made. For example: creating a mean of a scale, combining categories or creating categories from continuous data. This sheet explains how to:

* calculate new variables,
* recode existing continuous variables into categories
* reduce the number of categories (groups)
* select subsets of data

**Data:** The data set ‘*birthweight.csv’* contains details of 42 babies and their parents at birth such as Birth weight (lbs), gestational age of the baby at birth (weeks), mothers height (inches) and weight (lbs) prior to pregnancy and whether or not the mother smokes (smoker). Downloaded the csv file from the ‘[Practice Datasets](https://maths.shu.ac.uk/mathshelp/SSupport_Practice.html)’ page of the Sheffield Hallam Maths support website: *Resources 🡺 Stats Helpsheets 🡺 Practice Datasets page.*

## Data editing options in Jamovi

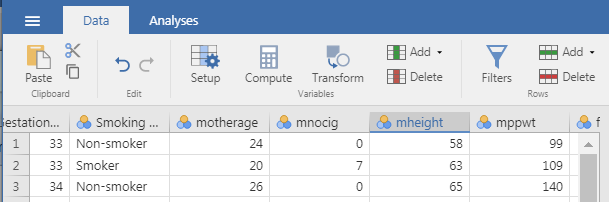
The basic Jamovi package has only two main tabs; the ‘**Data’** tab is where variables are defined and created and the ‘**Analysis’** tab which clearly shows the range of statistical techniques offered. The options within the ‘Data’ tab are covered in this sheet.

Click on ‘**Setup’** to label variables and values

Select subsets of the data using ‘**Filters’**

Create new variables using ‘**Compute’**

Add or delete variables here



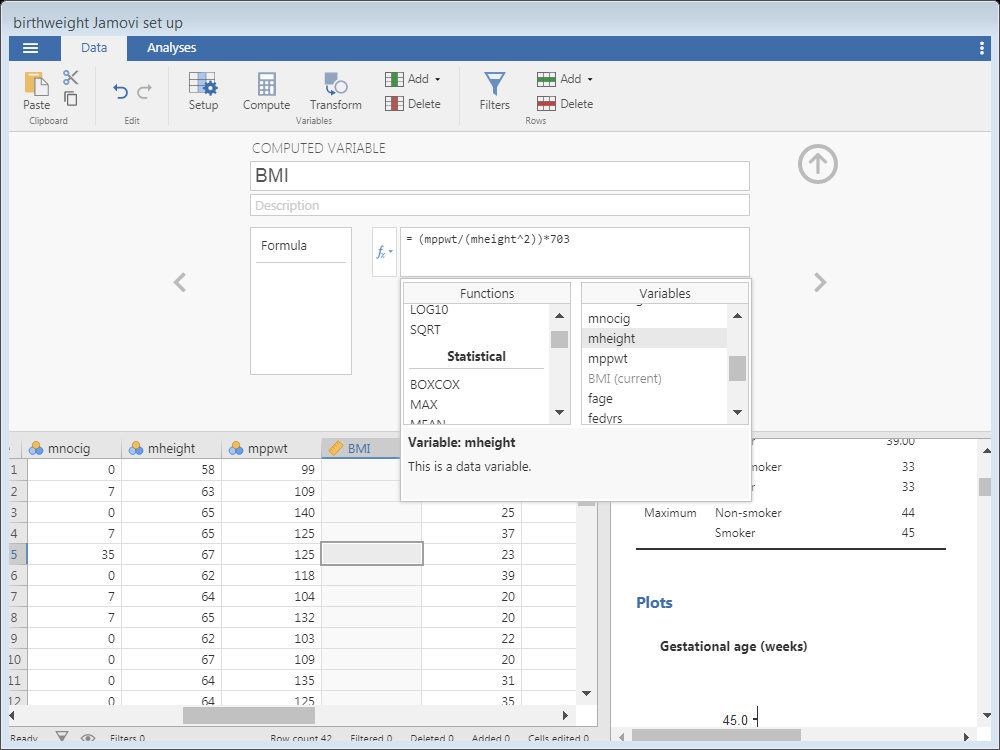
mppwt: Pre-pregnancy weight (lbs)

Recode variables using ‘**Transform’**

## Calculations using existing variables

First, the calculation of BMI using the height (inches) and weight (pounds) of the mother prior to pregnancy is used to demonstrate how to carry out a calculation using the **Compute** menu. Calculate BMI using pounds and inches using:

BMI =

To create a new variable, select the ‘**Compute’** option from the data tab

Change the **variable name** to ‘BMI’

Click on the **Function** button to view the **Functions** and **Variables** drop down menus.

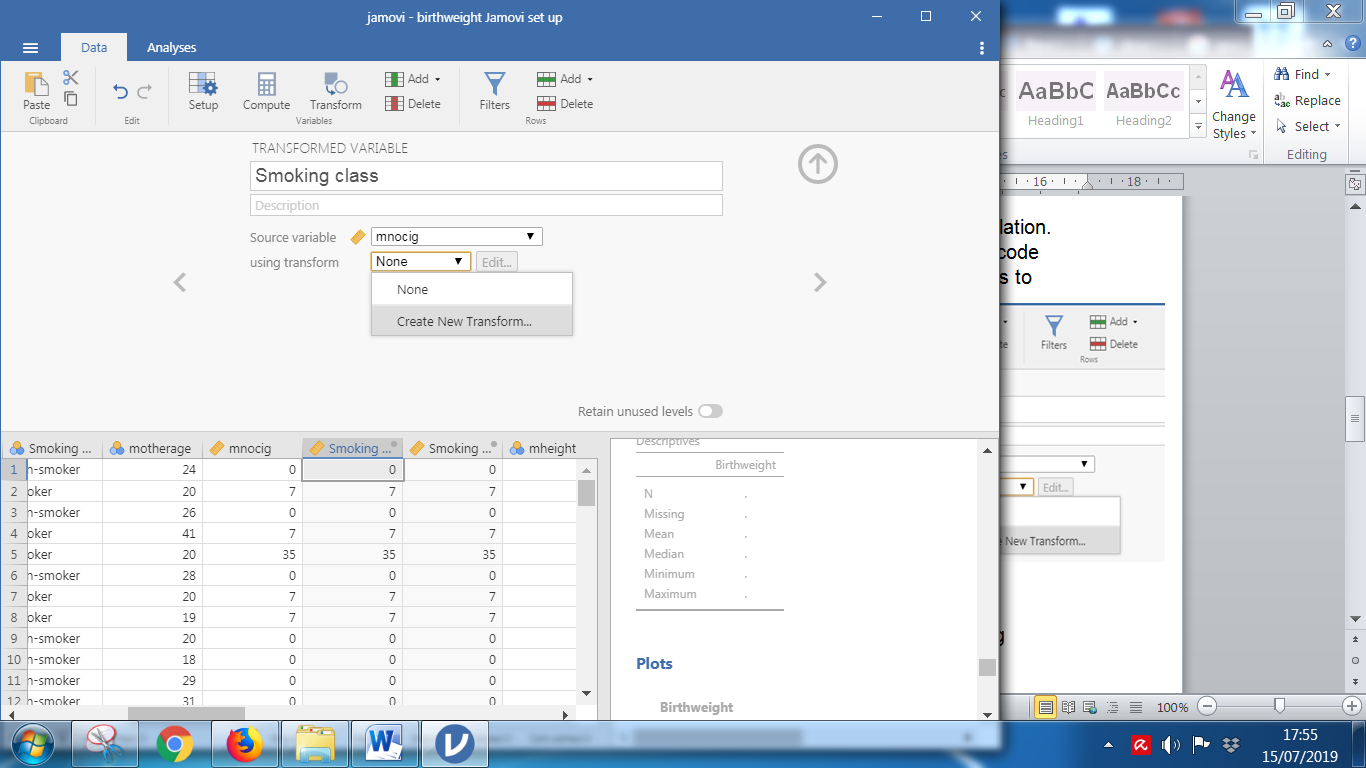
Select from the pre-defined functions such as the mean or type formulae in directly. Move variables to be used from the ‘**Variables’** pull down menu into the calculation.

Double click to move variables of interest into the formula

Standard operators such as / for divide, ^ for a power and \* for multiply are used in the formula here.

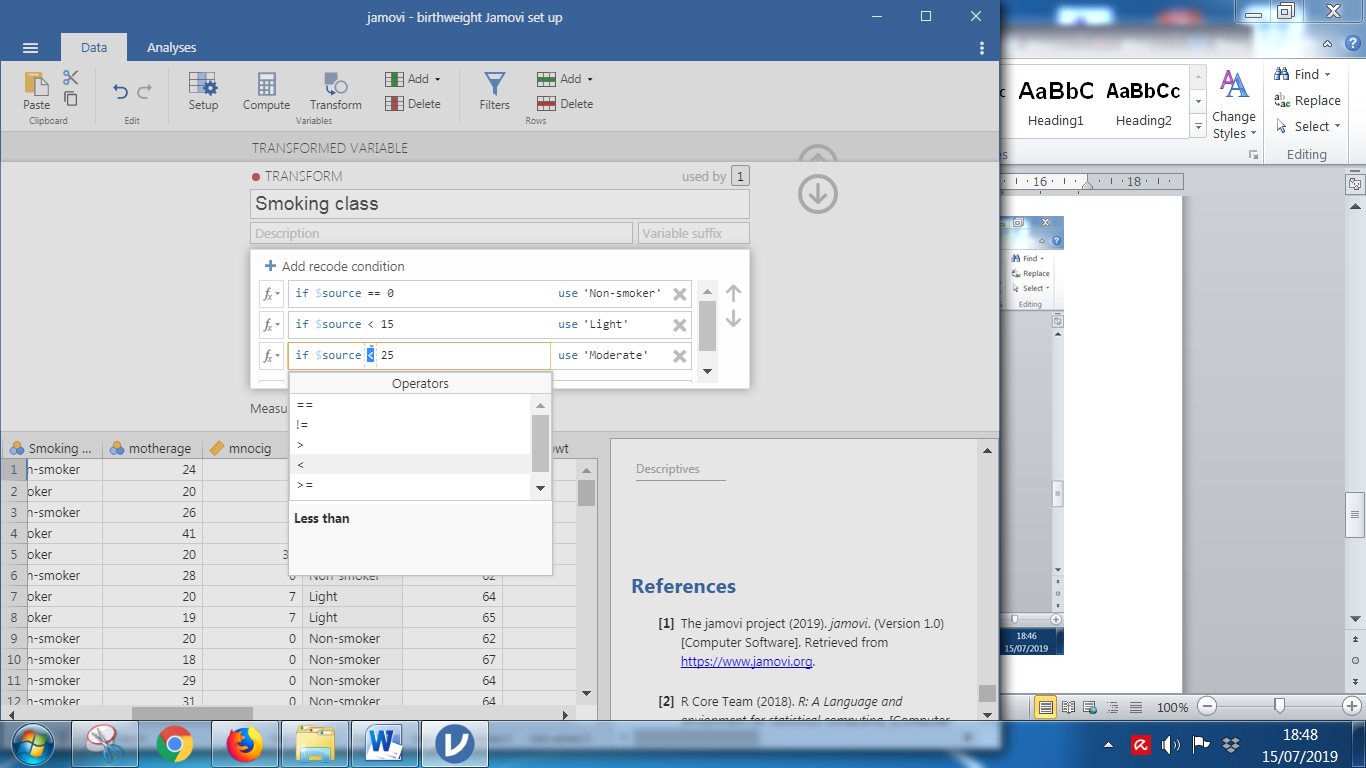
## Recoding one variable to create a new one

The Transform option in the Data tab can be used for many types of data manipulation. For example, to recode categorical variables to create fewer categories, reverse code items from scales, create categories from continuous variables or use if conditions to create new combinations. Here the number of cigarettes smoked per day will be converted to the following categories:

Highlight the variable to be transformed (mnocig) and select the **Transform** option. Give the new variable a name e.g. *smoking class* which can then be applied to any variable of interest.

Make sure the ‘**Source’** variable is the one which needs recoding (mnocig here).

If a transformation has been previously defined, it can be selected from the ‘Using transform’ menu. If not, select **‘Create New Transform’**.

Click on the **‘Add recode condition’** to add another line to be edited.

$source refers to the variable selected as the source variable in the previous step (mnocig). The first category ‘Non-smoker’ is for those smoking no cigarettes. Choose the operator == and type in ‘Non-smoker’ after the ‘use’. Then add the second category; if $source < 15 use ‘Light’. For the last category, use if $source >= 25 use ‘Heavy’. The ELSE IF option should only be used if there are no missing values.

Choose operator from menu

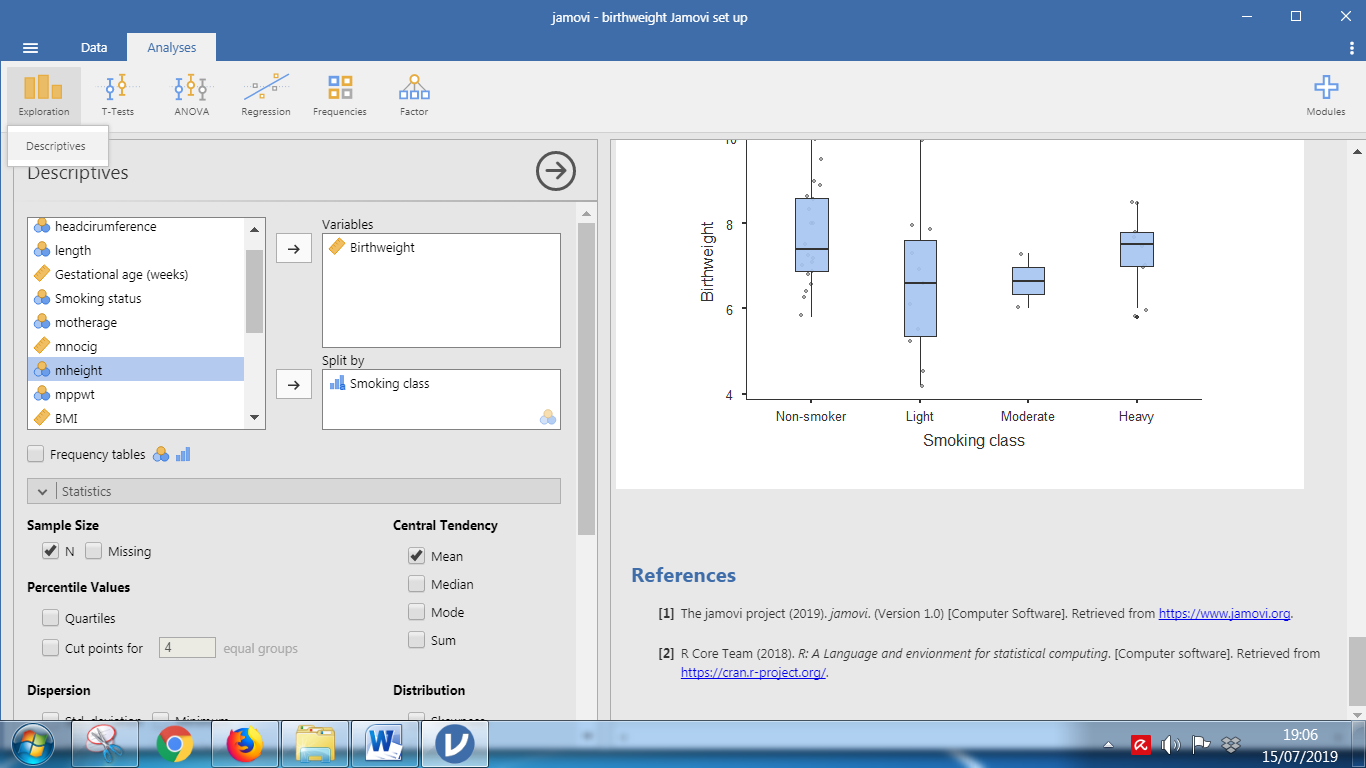
Put the new value here. Use labels within ‘label’. Jamovi does not currently label numeric values

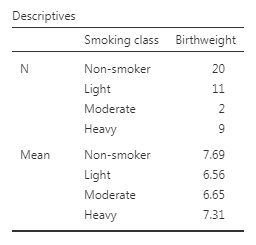
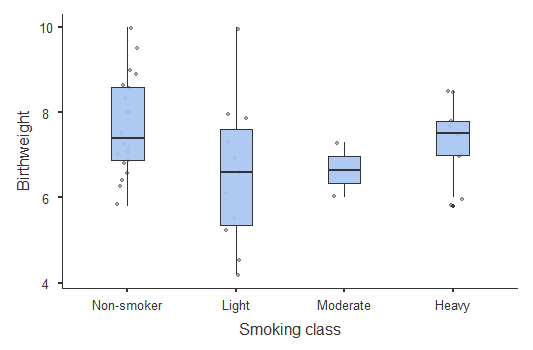
State condition here

Don’t forget to classify the variable as ordinal .

Usually, it is better to give categories numbers and then label them particularly for ordinal variables. However, currently Jamovi doesn’t allow the labelling of transformed variables which is very frustrating! Therefore, labels rather than numeric values have to be entered.

## Reducing the number of categories

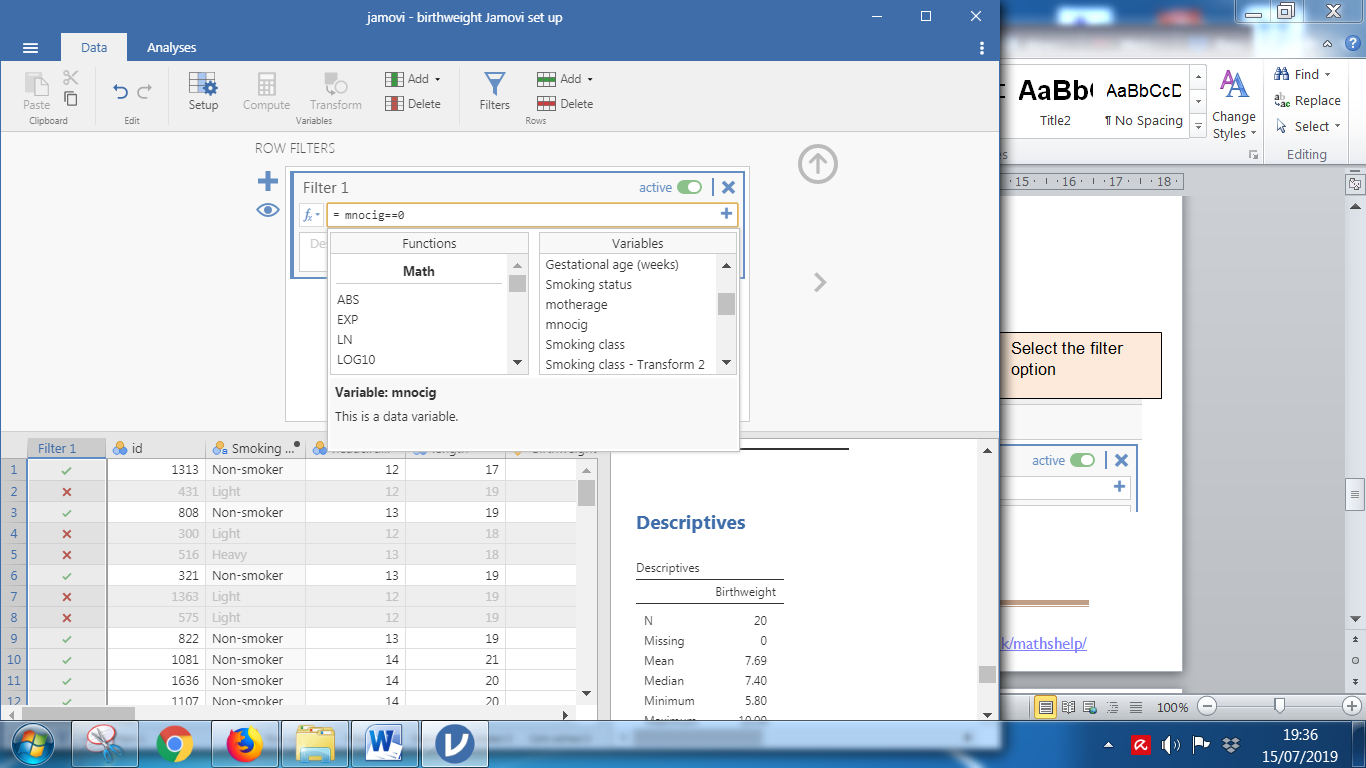
When testing for group differences, it is important that there are no groups with a small number of subjects in. Summarise variables of interest by the groups to assess whether there any small group sizes and similarities between groups. For example, if birthweight by smoking group os being investigated, use Analyses 🡪 Exploration 🡪 Descriptives and select N, Mean and a box-plot.



There are only two people in the ‘Moderate’ group so the group should be merged with either the Light or Heavy smoking categories. The means and box-plot suggest that the moderate group is more similar to the ‘Light’ group. In this situation, the transform in the previous section can be adjusted but for demonstration purposes, the recode will be carried out in a new transform here. Follow the process in the previous section but make the adjustment to the Moderate category of the original variable only and allow everything else to stay the same through the ‘else if’ box.



## Using a subset of the data

The ‘Filter’ option enables analysis to be carried out for particular groups only. For example, perhaps only non-smokers are of interest. This is carried out in a similar way to recoding but restricting to only the groups or range of values of interest using the ‘**Filter’** option.

As with the previous examples, the ‘Functions’ and ‘Variables’ drop down menus appear when selecting the function sign.

To turn the filter off, make it ‘Inactive’

Type the condition in here remembering to use == for equals

Specify just the group of interest using e.g. mnocig==0

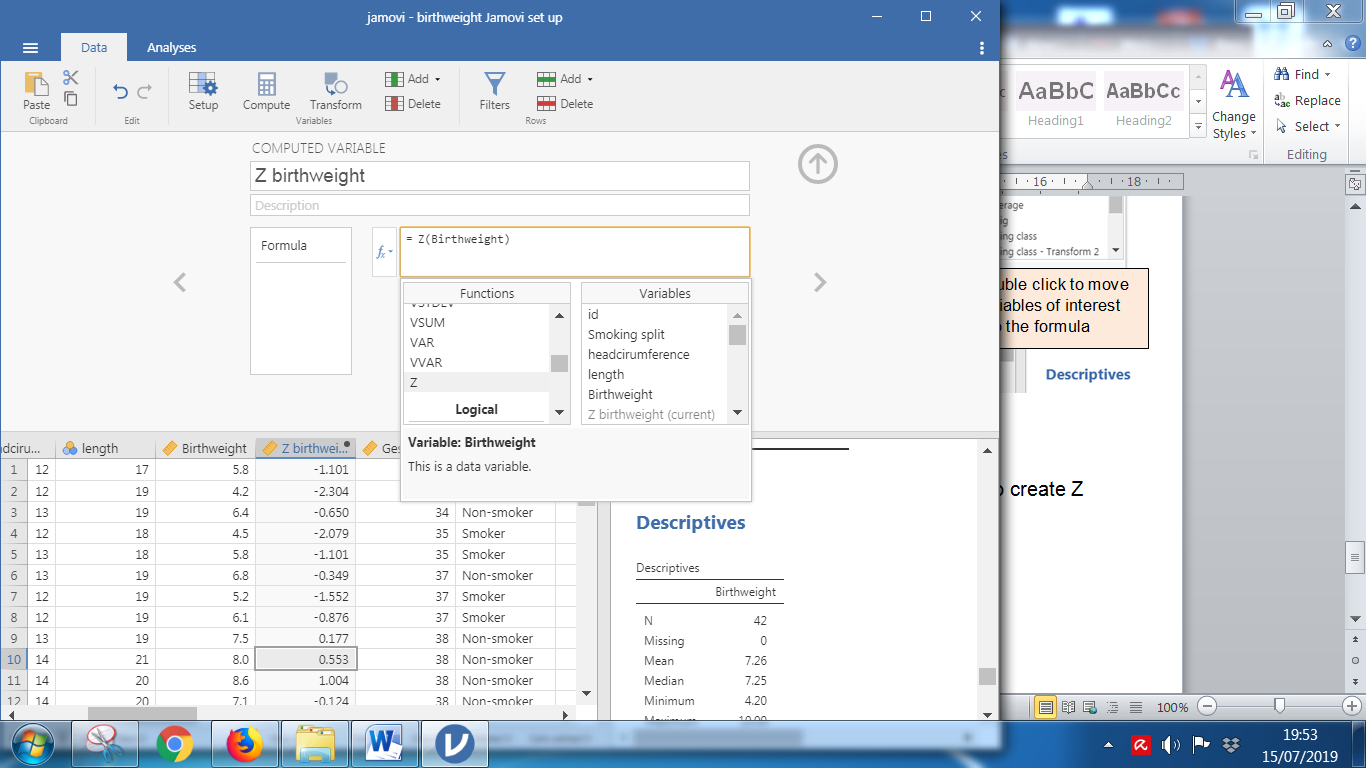
A filter column appears with only those in the group of interest ticked.

Double click to move variables of interest into the formula

Any further analysis will be for this group only until the filter is turned off.

## Calculating Z scores

Some disciplines use standardized values (Z scores) of original variables (X) to help identify outliers. Outliers can have a substantial impact on some statistical techniques particularly for smaller sample sizes and in some cases could be data entry errors. For example, outliers can increase the standard deviation considerably which reduces the likelihood of finding a significant difference between groupsZ scores follow a normal distribution and 95% of Z scores should be between -1.96 and +1.96. Z scores less than -3 or more than +3 are considered to be extreme outliers and should be investigated in case they have a substantial effect on results of tests.



A standardised value is calculated as

To create Z scores, use the **Compute** menu and select ‘Z’ from the **Functions** menu. Then place the variable of interest inside the brackets.

A new variable will be added to the dataset. Look first at the minimum and maximum values to see if any are outside the range -3 < Z < 3. Investigate any outside this range.