MN 0011-01

APPROVED

By the Methodological council on 5 February 2021

Minutes No. 21/1.1-20/1

Work in the PSPP programme

Methodological material

Iveta Ludviga, Dr.sc.adminstr.

2021

Contents

1. Introduction to PSPP

The PSPP programme for statistical data analysis is similar to SPSS. Over many iterations and years, SPSS (officially IBM SPSS Statistics) has served as a widely used standard for quantitative data analysis. PSPP is a simplified version of SPSS, which is freely available at no cost.

If PSPP is not yet installed on your computer, now is the time to do so. The easiest way to download PSPP is to click on the link

Windows: https://sourceforge.net/projects/pspp4windows/files/2018-11-09/pspp-20181109-daily-64bits-setup.exe/download

and looking for the 'Downloads' section. Then download the latest version in either 32-bit or 64-bit format. If you're not sure which version to download, go to the Control Panel, click on 'System' and check your system type. Then follow the instructions to download and install the Programme.

Mac: https://en.freedownloadmanager.org/Mac-OS/PSPP-FREE.html

1.1. Opening an Excel file in PSPP

Most often, you will have a data file that has been saved as an Excel file.

It is important to remember that **PSPP can only open Excel data files that have been saved as a comma-separated file (.csv)**. Here's how you can save an Excel file in the required format:

• Open the Excel file and click on File; Save as

• Select CSV (comma-separated) (*.csv)

• Click Save (save it to a location you know).

Sometimes the computer does not recognise the CSV file in the PSPP programme if it is saved in a folder, so it is recommended to save the CSV file on the desktop.

We recommend giving the CSV file a simple name, without commas or long marks typical of the Latvian language, as these can cause problems with file recognition in the PSPP programme.

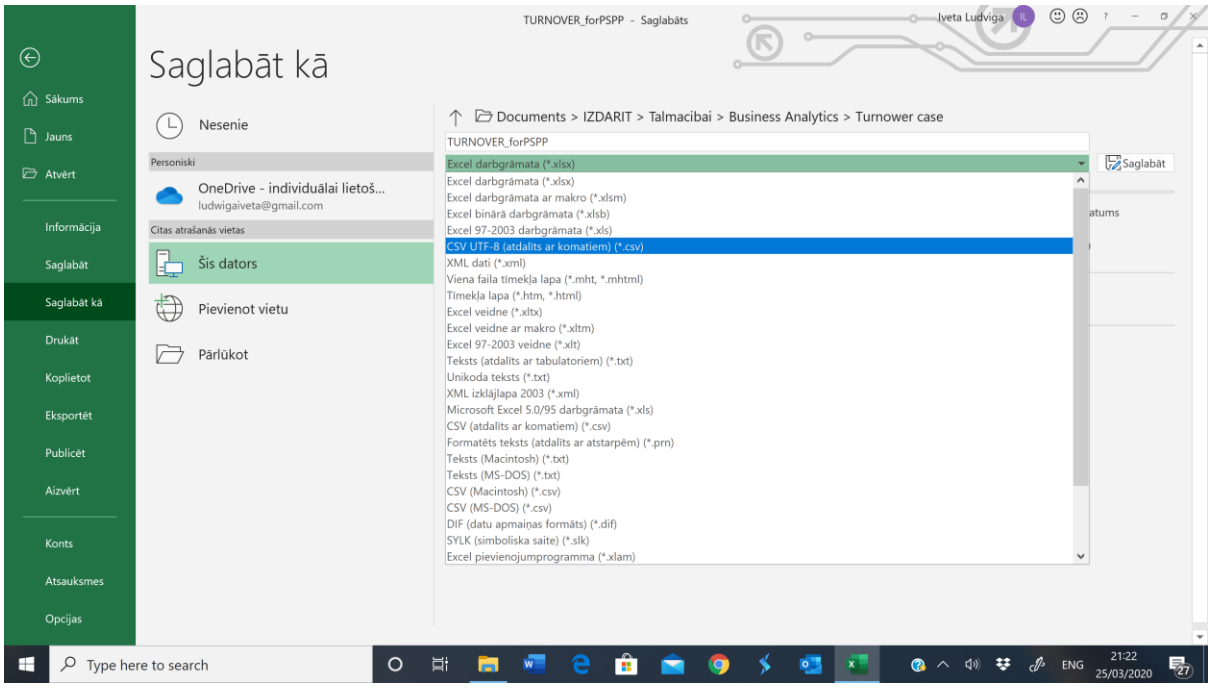


Figure 1.1. Save as...

Once the file is ready, you can open it in the PSPP programme. To do this, proceed as follows:

• Open PSPP by clicking on the PSPP icon.

• In the menu bar at the top of the screen, click on FILE.

• In the drop-down menu, click on IMPORT DATA

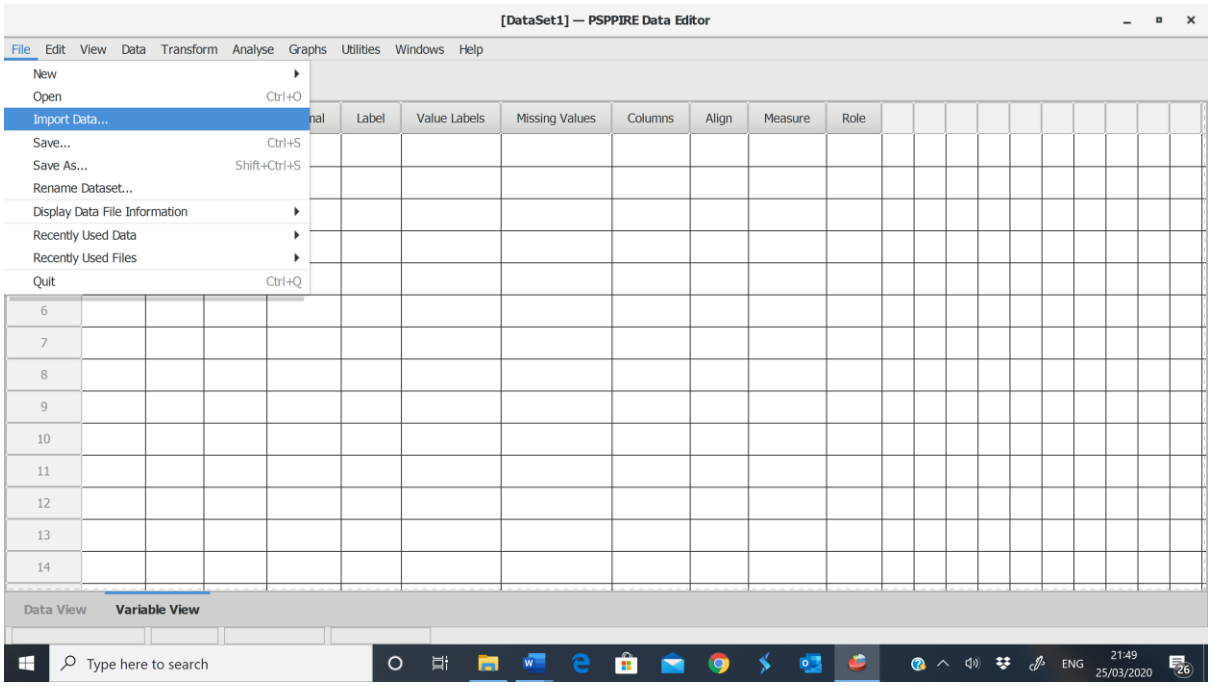


Figure 1.2 Data import.

• Locate the previously prepared .csv file on your computer and select the file name. Now click on NEXT in the bottom right-hand corner of the screen.

• Select the data rows you wish to open. This will usually be ‘all cases’ (all data). Then click NEXT.

• Select the first row of the file containing data, then click NEXT. Usually, this will be the second row of your data file, as the first row will contain the variable names (see the red arrow).

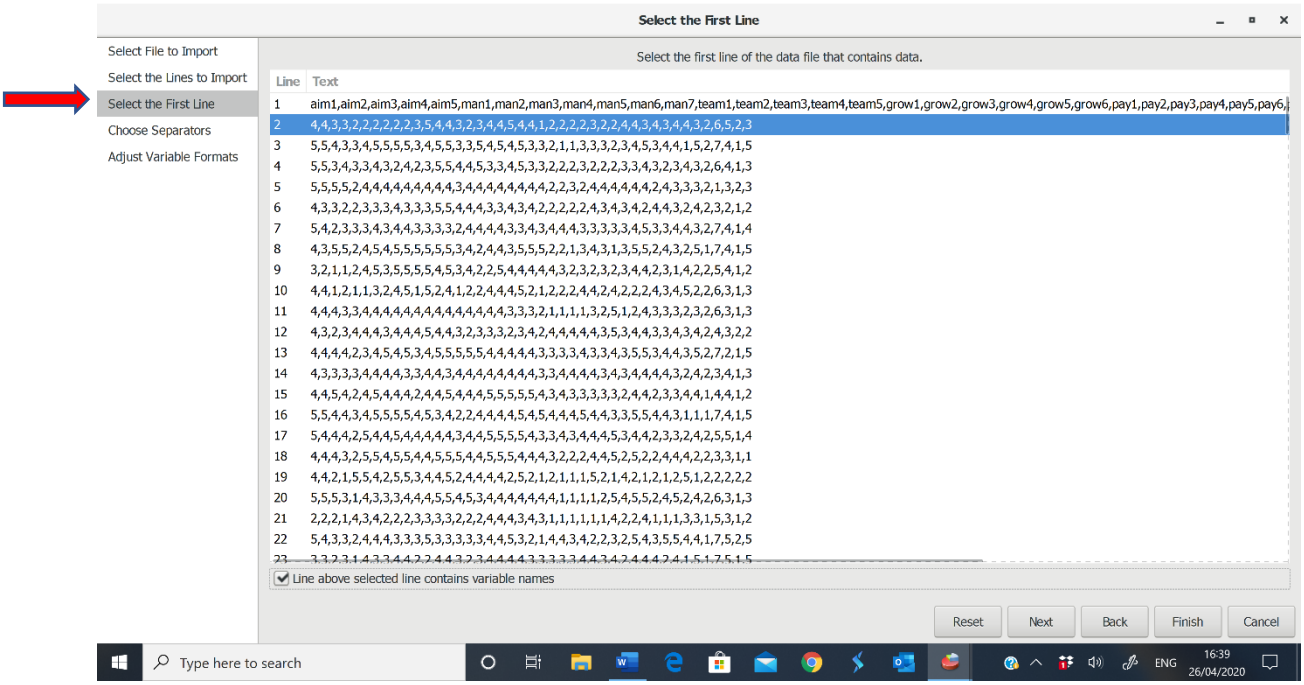


Figure 1.3. Data import (continued).

• At the bottom of the page, click on “The row above the selected row contains variable names” (see red arrow)

• The next screen asks for *‘characters that separate your variables’*. This should already be filled in. For example, in a comma-separated file (.csv), this will be a comma. Click NEXT.

• The next screen allows you to customise variable formats (which you do not need to do). Click APPLY.

• Your file should now be open in PSPP. You can save it so that you do not have to check it the next time you open the file.

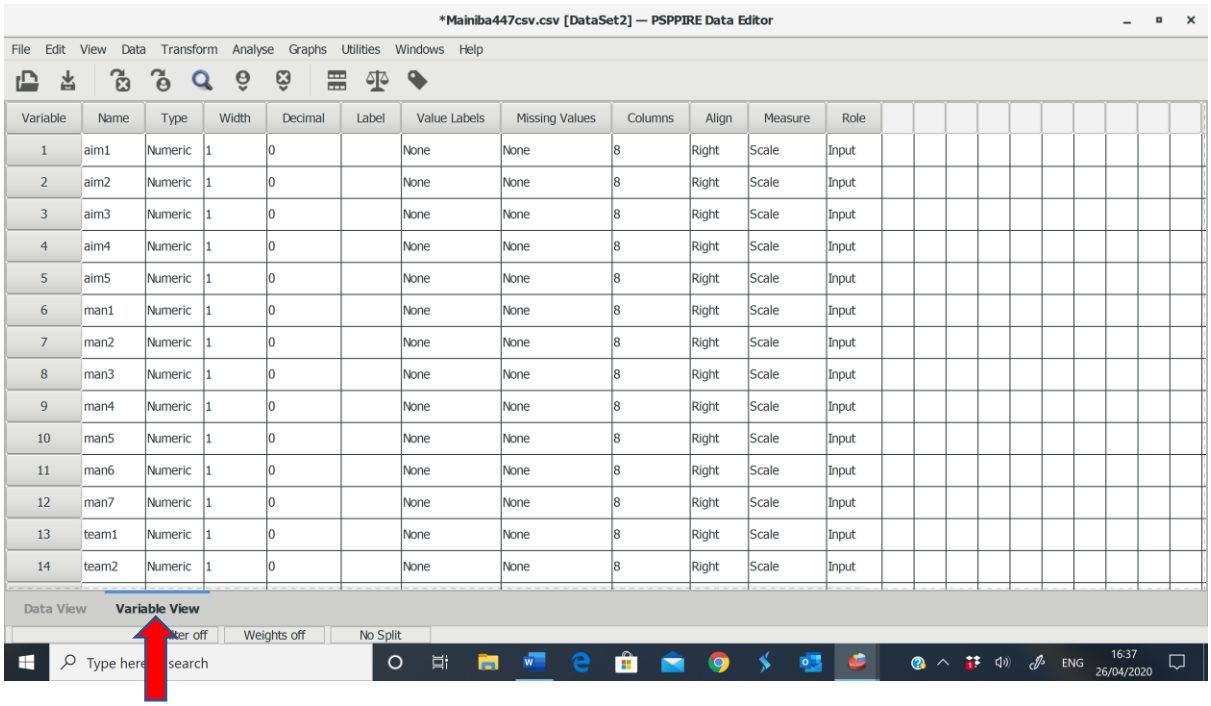


Figure 1.4. PSPP features a Variable View and a Data View (just like SPSS)

• In the Variable View, all variables are displayed in rows with their names in the ‘Name’ column. The data type is ‘Numeric’ – the data consists of numbers.

• The Data View is similar to Excel, with all survey items arranged as columns

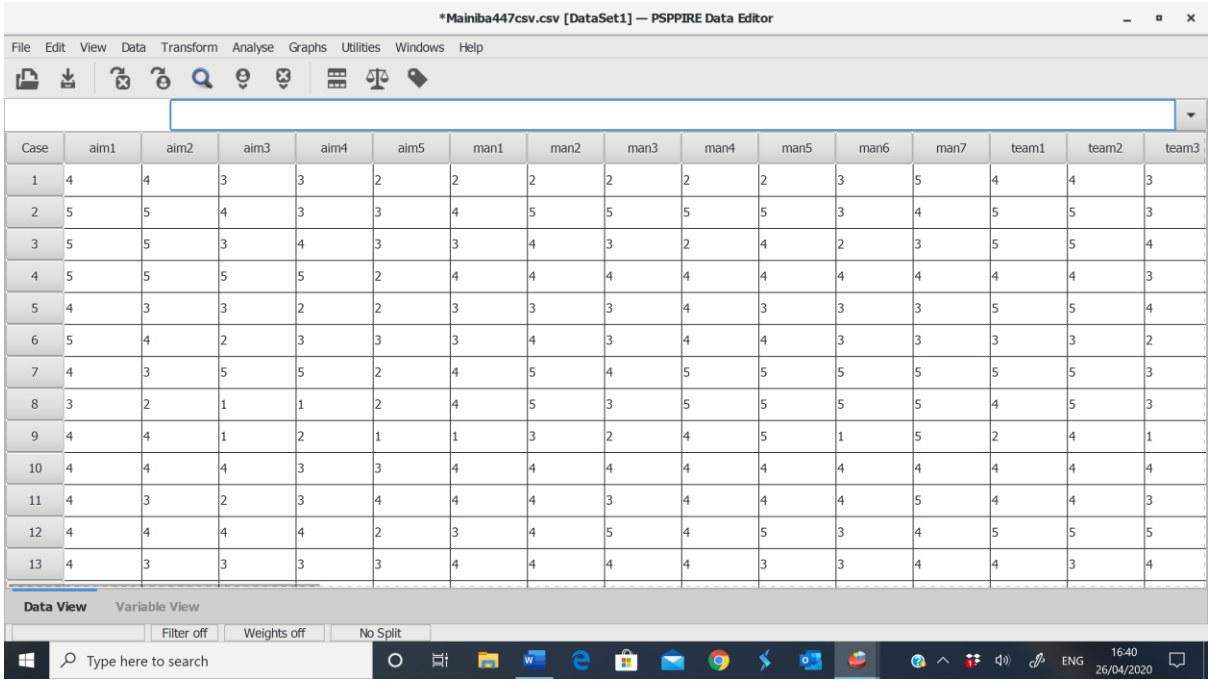


Figure 1.5 Data View.

1.2. Formatting the PSPP file before starting the analysis

The first thing we need to do is format the PSPP data file so that we understand what we are analysing – this means that names should be added to the demographic variables.

In the TURNOVER data file, the demographic variables are ‘position’, ‘tenure’, ‘income’, ‘gender’ and ‘age’ – you can find these by scrolling down the page in the Variable View.

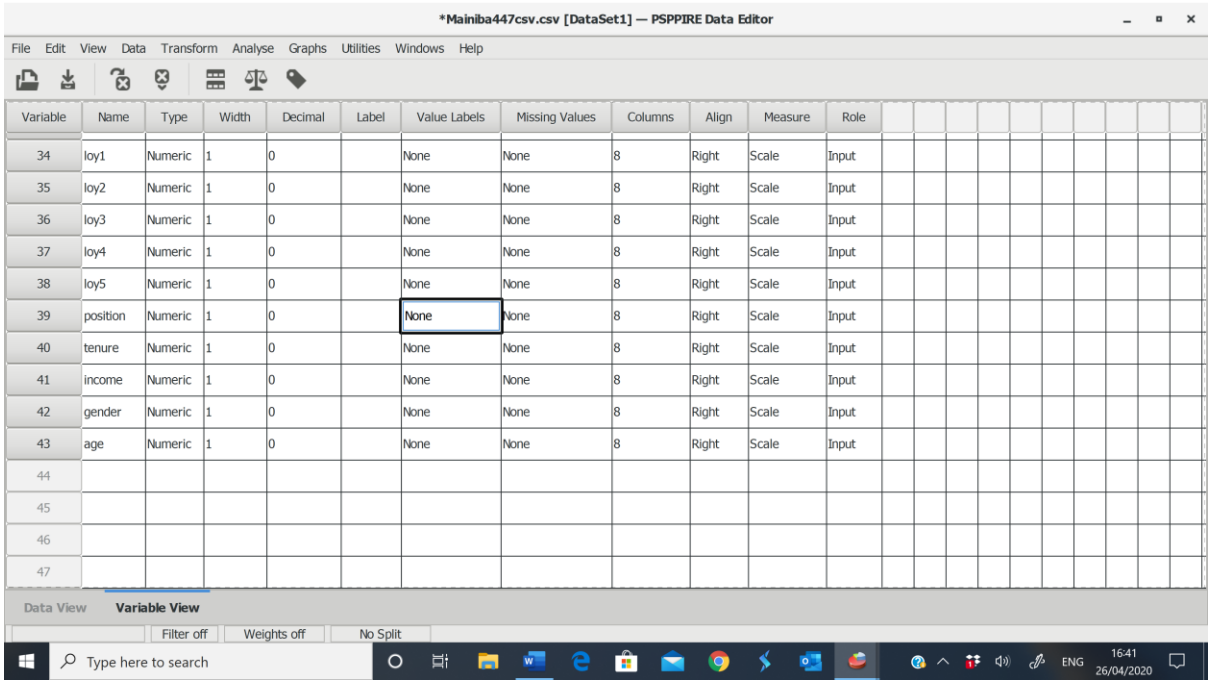


Figure 1.6. In the Variable View, under the column heading Value Labels, you can enter names for the different options of demographic data responses.

For example, job title. In the case of TURNOVER, the job title options are as follows:

1. Manager

2. Employee/Specialist

To do this, proceed as follows:

• Double-click on a cell in the Value Labels column and the “position” row

• In the window that opens, enter 1 as the Value and “Manager” as the Value Label, then click ADD

• Do the same for another value: enter 2 as the value and “Employee” as the value label, then click ADD

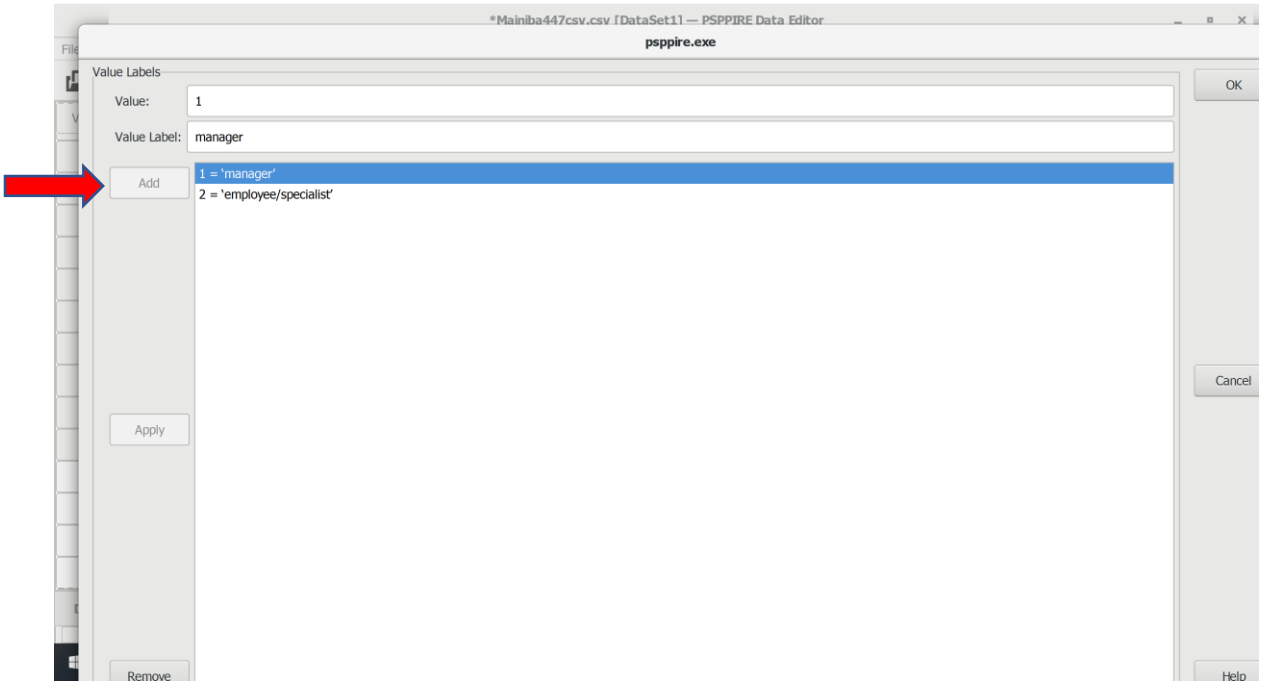


Figure 1.7. Data coding

Do the same for the other demographic variables. Below is an example for length of service groups.

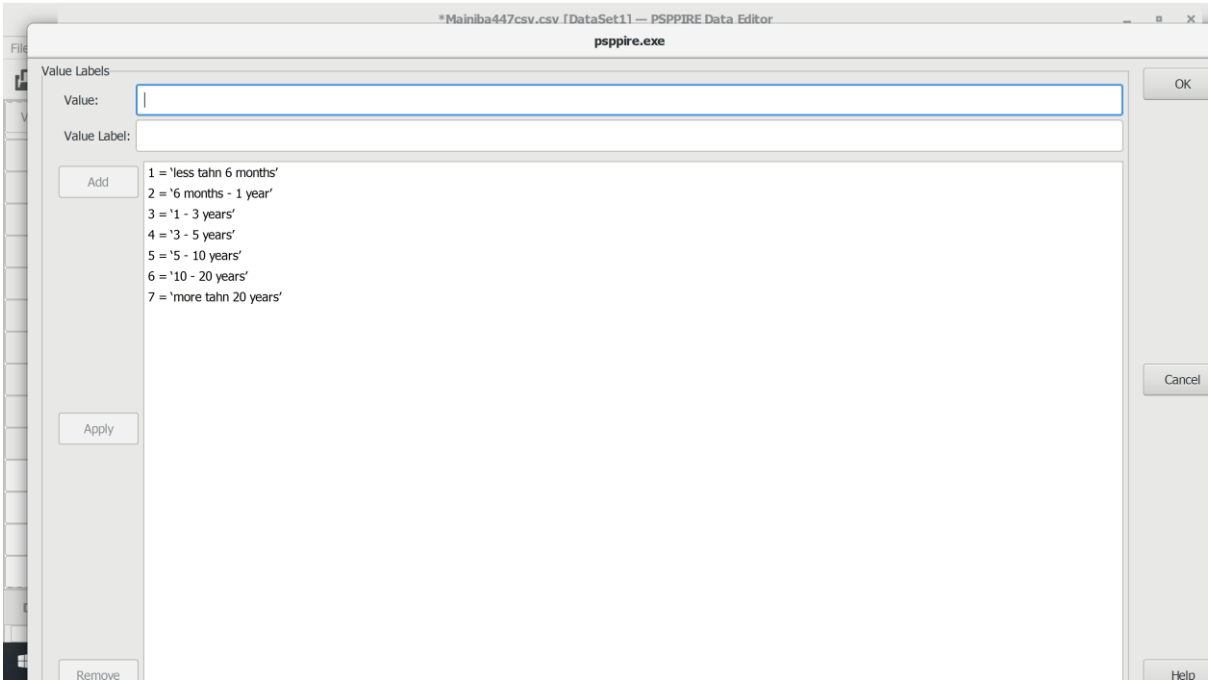
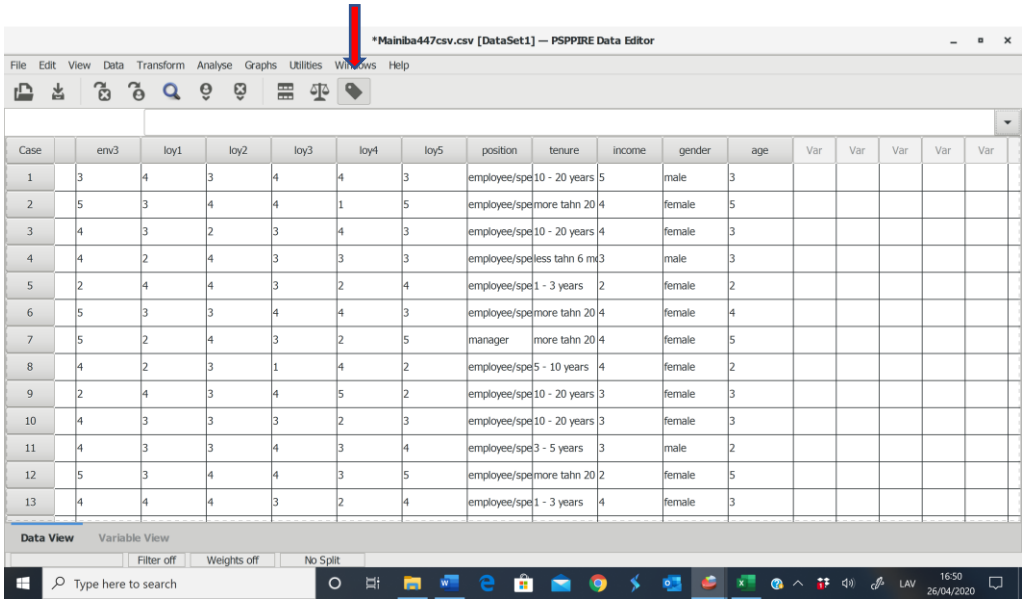


Figure 1.8. Data coding (continued).

In Data View, you can see the nominal values corresponding to the numbers by clicking on the ‘show/hide value labels’ icon:



Case	env3	loy1	loy2	loy3	loy4	loy5	position	tenure	income	gender	age
1	3	4	3	4	4	3	employee/spe	10 - 20 years	5	male	3
2	5	3	4	4	1	5	employee/spe	more tahn 20	4	female	5
3	4	3	2	3	4	3	employee/spe	10 - 20 years	4	female	3
4	4	2	4	3	3	3	employee/spe	less tahn 6 m	3	male	3
5	2	4	4	3	2	4	employee/spe	1 - 3 years	2	female	2
6	5	3	3	4	4	3	employee/spe	more tahn 20	4	female	4
7	5	2	4	3	2	5	manager	more tahn 20	4	female	5
8	4	2	3	1	4	2	employee/spe	5 - 10 years	4	female	2
9	2	4	3	4	5	2	employee/spe	10 - 20 years	3	female	3
10	4	3	3	3	2	3	employee/spe	10 - 20 years	3	female	3
11	4	3	3	4	3	4	employee/spe	3 - 5 years	3	male	2
12	5	3	4	4	3	5	employee/spe	more tahn 20	2	female	5
13	4	4	4	3	2	4	employee/spe	1 - 3 years	4	female	3

Figure 1.8. ‘Show/hide value labels’ icon.

2. Charts and graphs

When beginning the analysis of survey data, the first step is usually to create a respondent profile. In a research report, the most convenient way to present data on respondents is in the form of charts.

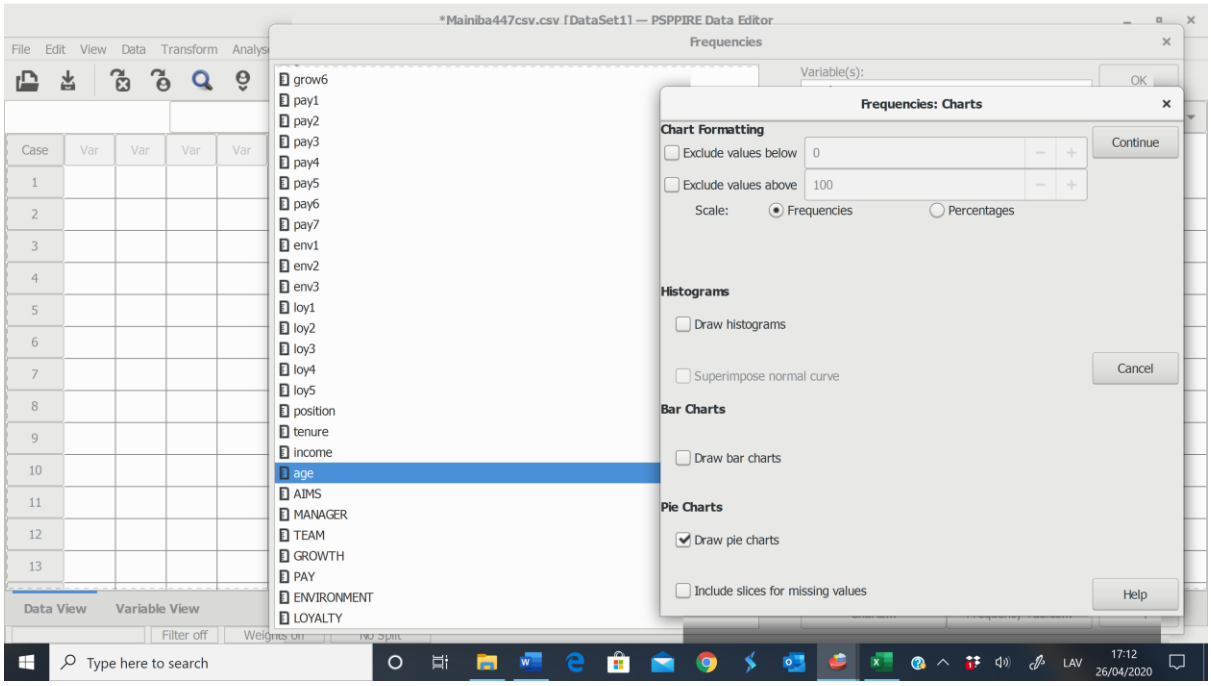
As part of the FREQUENCIES procedure, PSPP will construct pie charts, bar charts and histograms. Bar charts and histograms can also be created by selecting GRAPHS from the menu bar.

A pie chart is a diagram that shows the frequency or percentages of a variable with a small number of categories. It is depicted as a circle divided into a series of sectors. The area of each slice is proportional to the number of cases or the percentage of cases in each category. It is typically used with nominal or ordinal variables, but can be used with interval or ratio variables that have a small number of categories.

To create a chart in PSPP, click the CHARTS button and tick the checkbox for the desired chart type. You can create a sector chart and a frequency distribution.

If you do not want a pie chart and frequency distribution, but a pie chart instead, proceed as follows:

- Open the Analyse, Descriptives, FREQUENCIES section
- In the analysis window, drag the variable for which you want to create a graph – in this example, ‘gender’
- Click on Charts
- Click on Draw next to the chart



2.1.Figure. Creating a pie chart.

The result in the OUTPUT log will look like this:

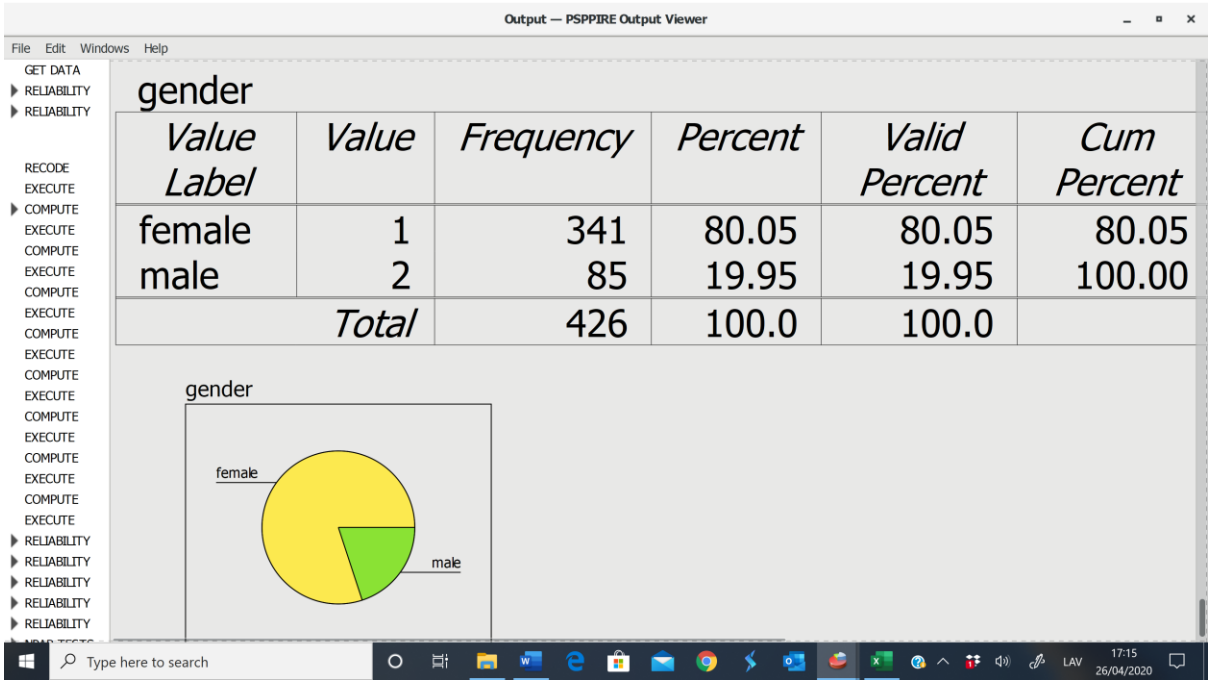


Figure 2.2. Creating a pie chart (continued).

As can be seen, there are 341 respondents, of whom 80% are women and 85, or 20%, are men. The proportion is heavily in favour of women.

To determine whether the gender proportion of the sample corresponds to the organisation's gender proportion, the Chi-square test can be used. We will do this later – see the section on the Chi-square test.

3. Calculation of Cronbach's alpha

Cronbach's alpha is used to assess the internal consistency of variables measured using Likert scales.

To do this, proceed as follows:

• Under the Analyse menu, click Reliability ...

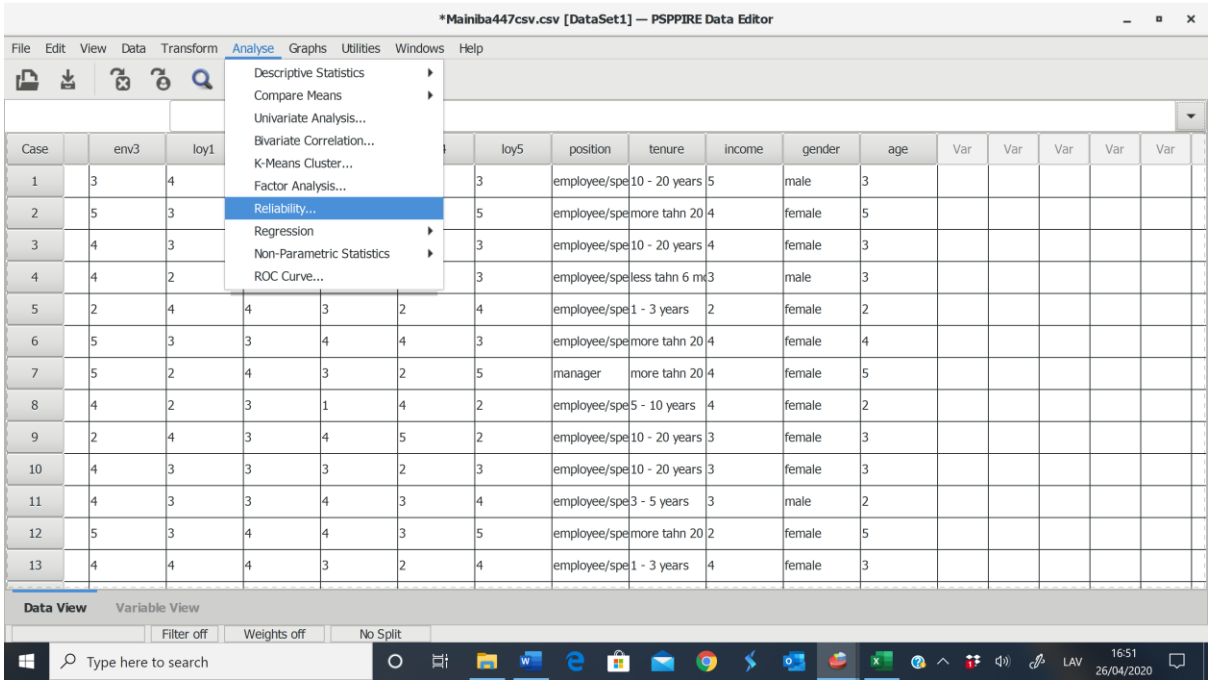


Figure 3.1. Calculation of Cronbach's alpha

• In the window that opens, select all items of a single variable ('aim' in this example) and click the arrow between the windows to move the items to another window.

• At the bottom of the page, click on 'Show descriptive for all scales if item deleted'.

• Click OK

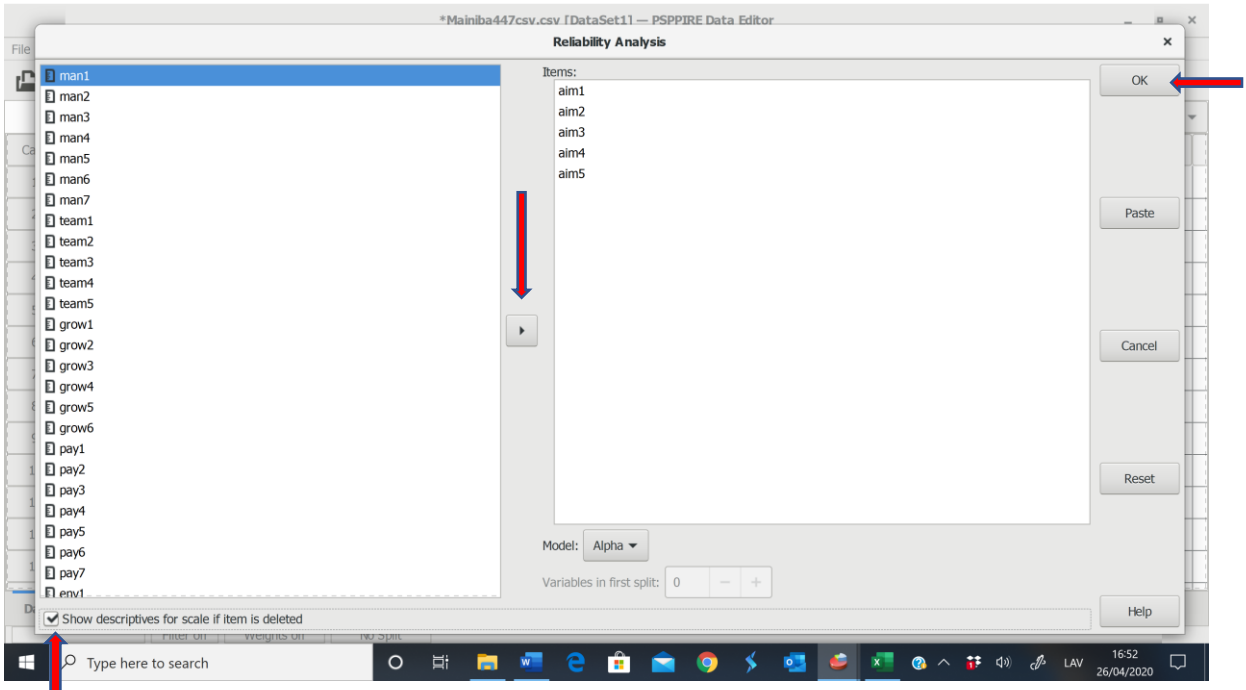


Figure 3.2. Show descriptives for scale if item deleted.

The OUTPUT log appears – this can be accessed at the bottom of the page:

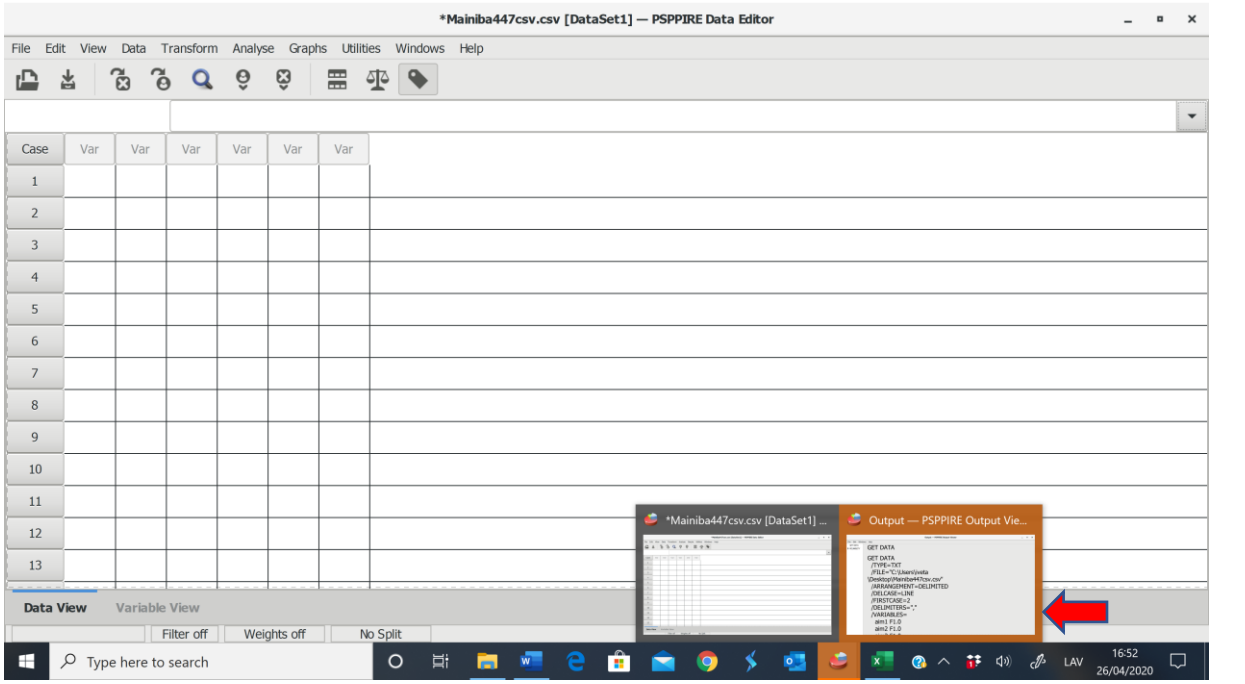


Figure 3.3. Opening and viewing the output log.

The OUTPUT log displays all analysis results.

The OUTPUT log is cumulative. When you request various analyses from PSPP, it appends them to the bottom of the Output file. Therefore, you will need to scroll down to see what you have done recently.

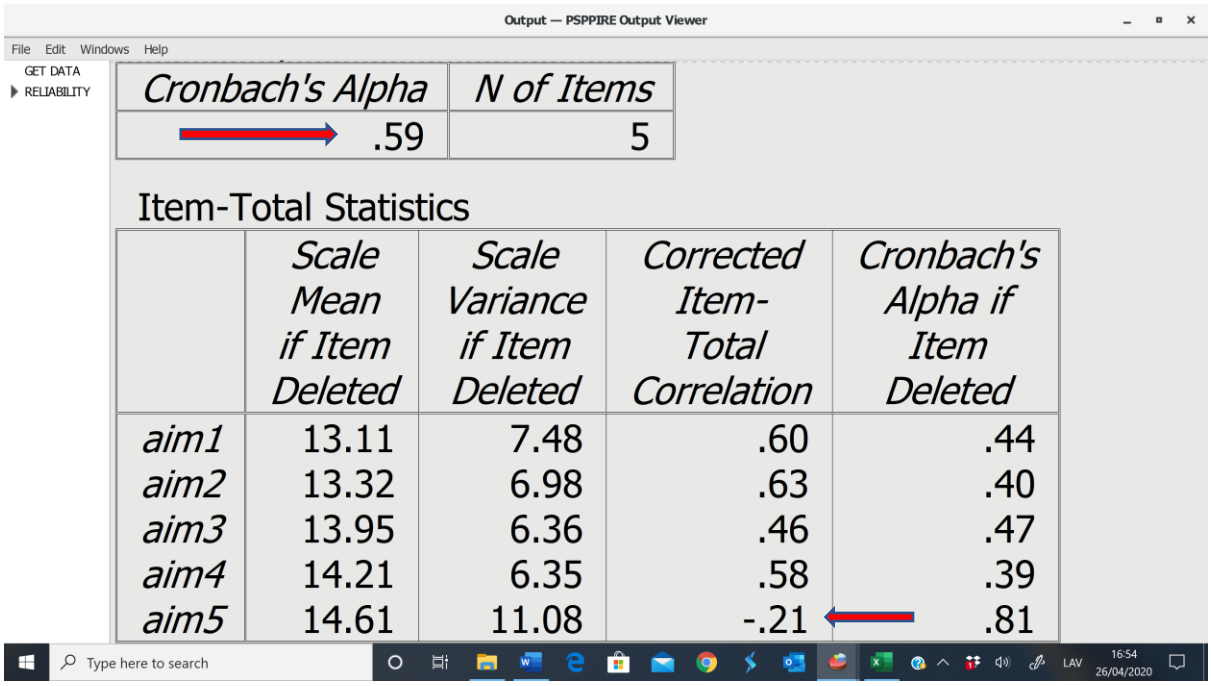


Figure 3.4. Output log for Cronbach's alpha

Table 3.1

Interpretation of Cronbach's α

Cronbach's α coefficient	Interpretation
α > 0.9	Excellent
α > 0.8	good
α > 0.7	acceptable
α > 0.6	on the question
α > 0.5	weak
α < 0.5	unacceptable

When measuring respondents' personality traits (e.g. stress resilience), Cronbach's α coefficient may be lower; in this case, it must be α > 0.42. The study must then provide appropriate justification that the trait being measured is considered a personality trait.

We see that Cronbach's alpha is 0.59, and the scale has 5 items. This coefficient is too low; it can be concluded that the scale is NOT very reliable.

We also see that item aim5 has a negative (-0.21) Corrected Item-Total Correlation and a Cronbach's Alpha of 0.781 if the item is deleted This means that item 5 is questionable – the Researcher must decide, after reviewing the text, whether this item should be excluded from the analysis.

In this example, we will choose to remove it and not use it in the subsequent analysis.

Cronbach's Alpha must be calculated for all variables. In the TURNOVER example, we have 7 variables: goals (AIM); manager (MANAGER); team (TEAM); growth opportunities (GROWTH); pay (PAY); work environment (ENVIRONMENT) and loyalty (LOYALTY).

For the OUTPUT variable Loyalty, it will display:

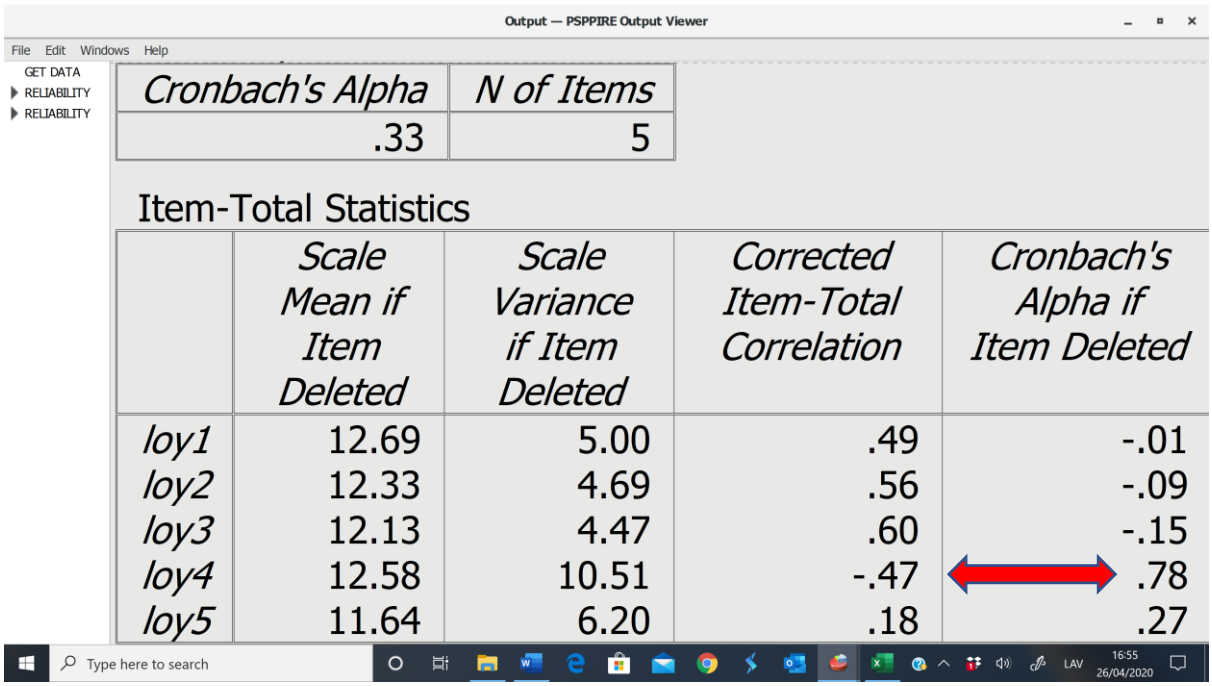


Figure 3.5. Increasing the Cronbach's alpha coefficient.

Item loy4 shows a negative Corrected Item-Total Correlation (-0.47), indicating that this item measures something else.

Upon reviewing the questionnaire item, it is evident that this is a reverse-scored item and must be recoded.

4. Recoding survey items

Sometimes questionnaires include items that are "reversed". In the example, in the TURNOVER variable "Loyalty", the fourth statement (loy4) measures the intention to leave, not the intention to stay. This is a "reversed" item and must be recoded before the Cronbach's Alpha coefficient is calculated. This means that a value of 1 must be changed to 5; 2 -> 4; 3 -> 3; 4 -> 2; and 5 -> 1.

To do this, proceed as follows:

• In the Data View, under the Transform command, click on Recode into Same Variable

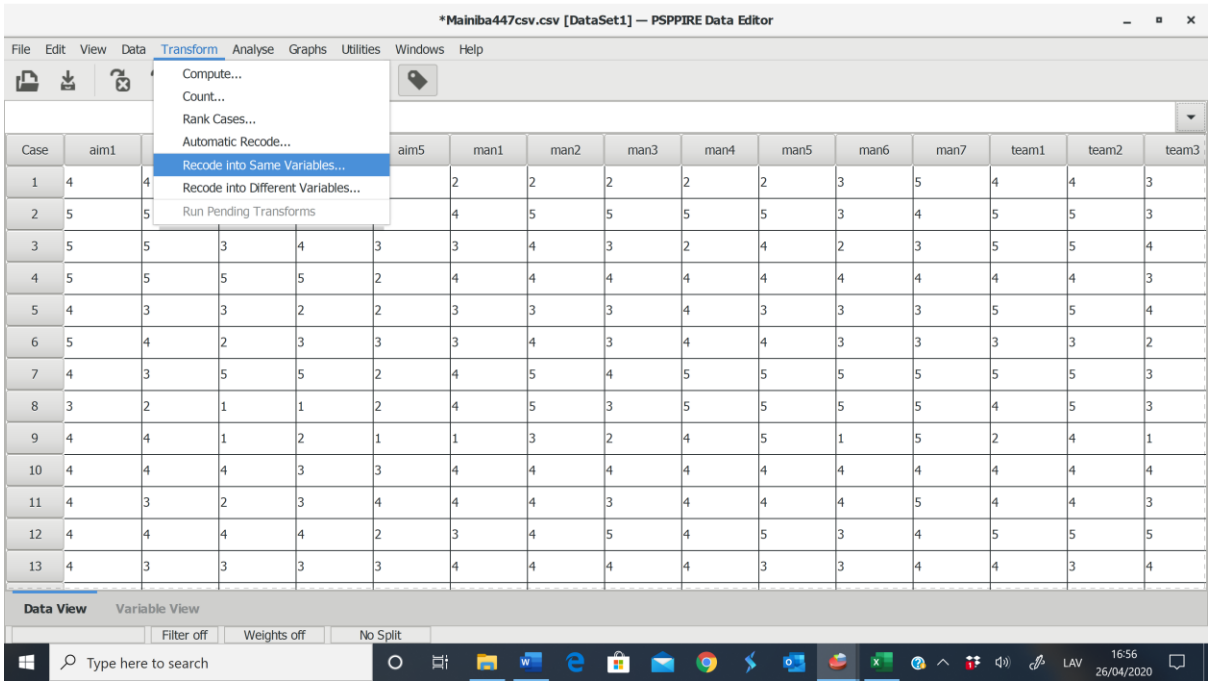


Figure 4.1. Location of the survey item recoding function in PSPP.

• In the window that appears, select the variable to be recoded (in the case of TURNOVER, this is loy4) and, using the arrow, move it to the right-hand pane.

• Click on ‘Old and New Values’

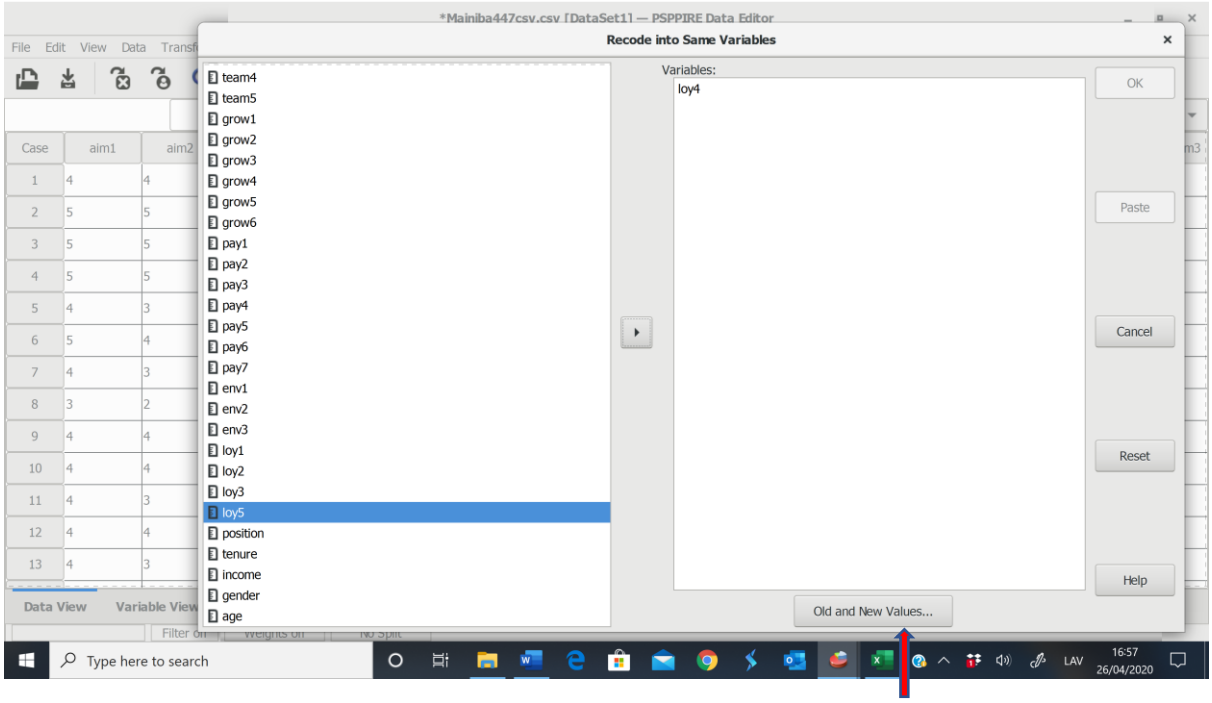


Figure 4.2. Old and New Values.

• In the new window that appears, enter the value pairs – Value 1 on the left and New value 5 on the right

• Click on ADD

• Do this for all 5 pairs

• Click Continue

• Click OK

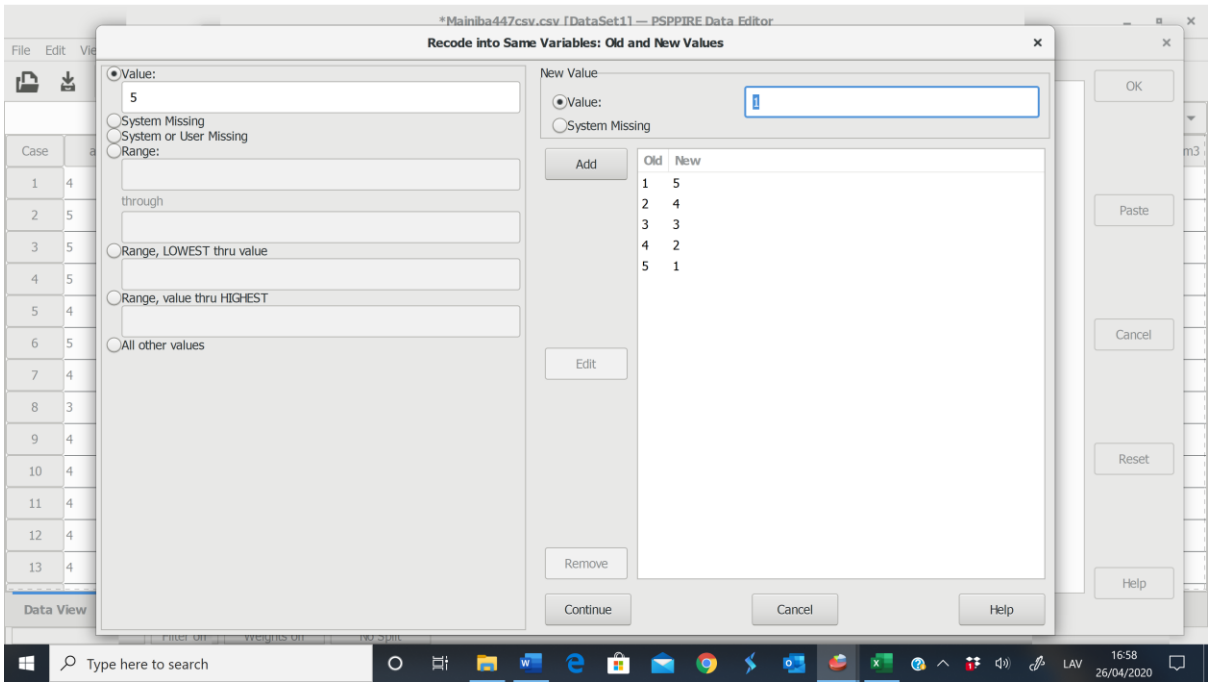


Figure 4.3. Entering the recoding values.

In the Output window, you will see that the entry has been executed:

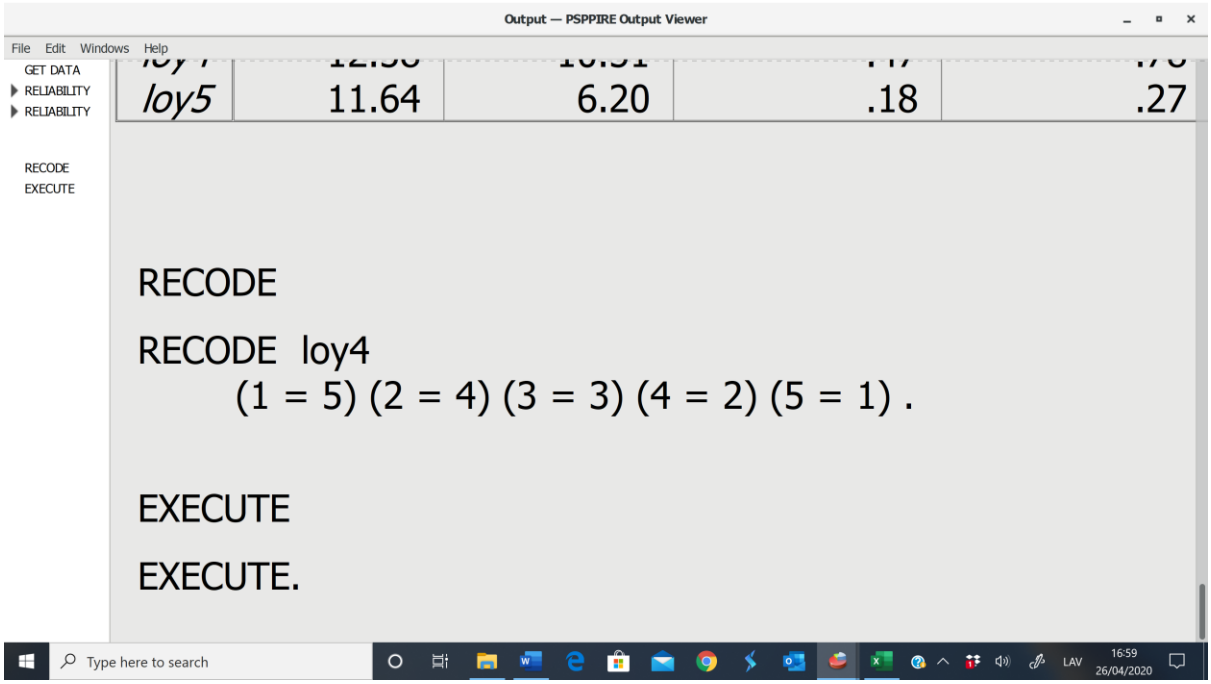


Figure 4.4. Execution of the entry.

You can now also calculate the Cronbach's alpha coefficient for the loyalty variable.

The TRANSFORM command can also be used for other purposes. For example, if you wish to group variables, you can use ‘Recode into different variables’. The new variable must also be given a unique name. In a PSPP file, just like in SPSS, names can consist of only one word (spaces are not allowed), and an underscore can be used.

5. Creating variables (Summary table)

For further analysis, a single variable must be created from several items that measure the same variable. A summary table is created for each variable by calculating the mean (MEAN value).

In the case of TURNOVER, this will be a table of 7 variables: goals (AIM); manager (MANAGER); team (TEAM); growth opportunities (GROWTH); pay (PAY); working environment (ENVIRONMENT) and loyalty (LOYALTY).

To do this, proceed as follows:

• Under the Transform tab, click on Compute…

• In the Functions section, find MEAN and double-click it

• The MEAN function appears in the Numeric Expressions window

• Enter the variable terms one by one in the brackets, separating them with commas

• Click OK

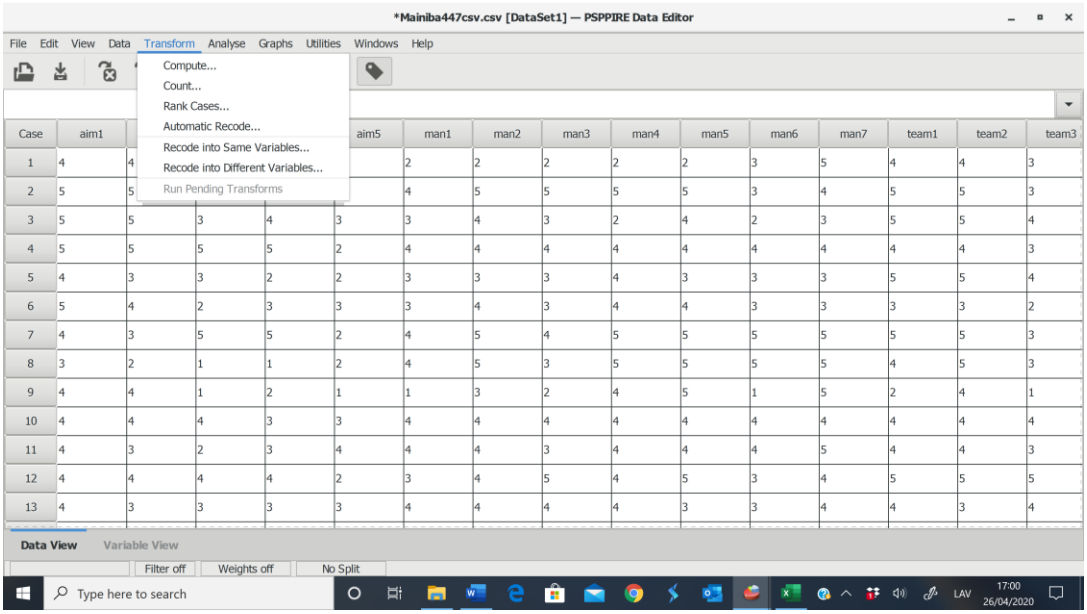


Figure 5.1. Location of the summary data function in the PSPP programme.

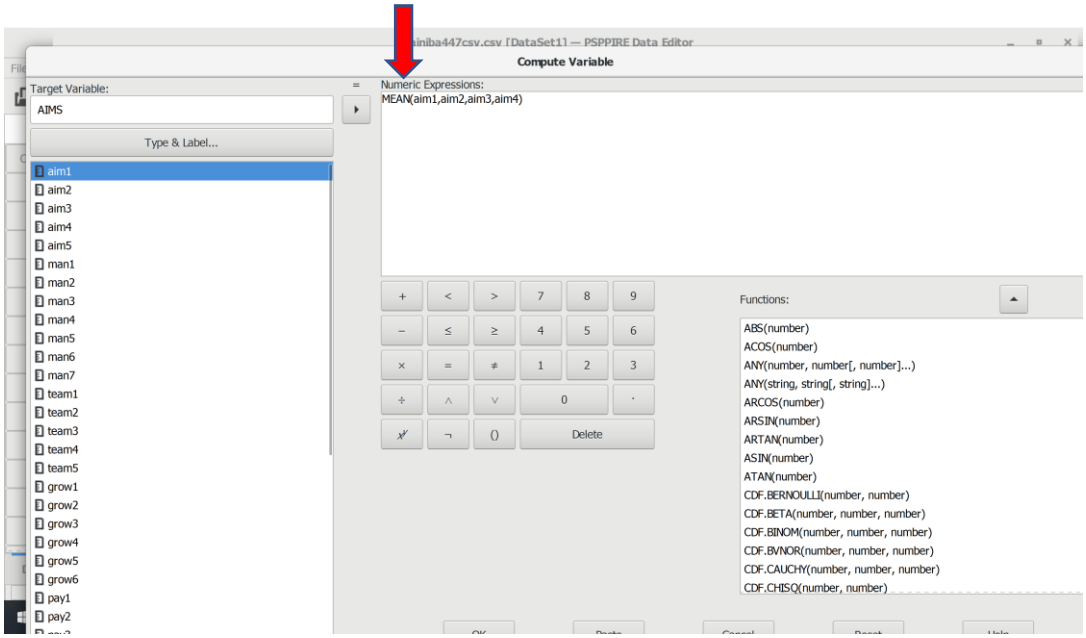


Figure 5.2. Creation of aggregated data.

Remember that the fifth term of the variable MEŖĶI (AIM) did not fit on the scale, so we will not use it.

A summary data table is created – in the case of TURNOVER, it consists of 7 columns.

These variable columns, together with the demographic variables, will be used for further analysis.

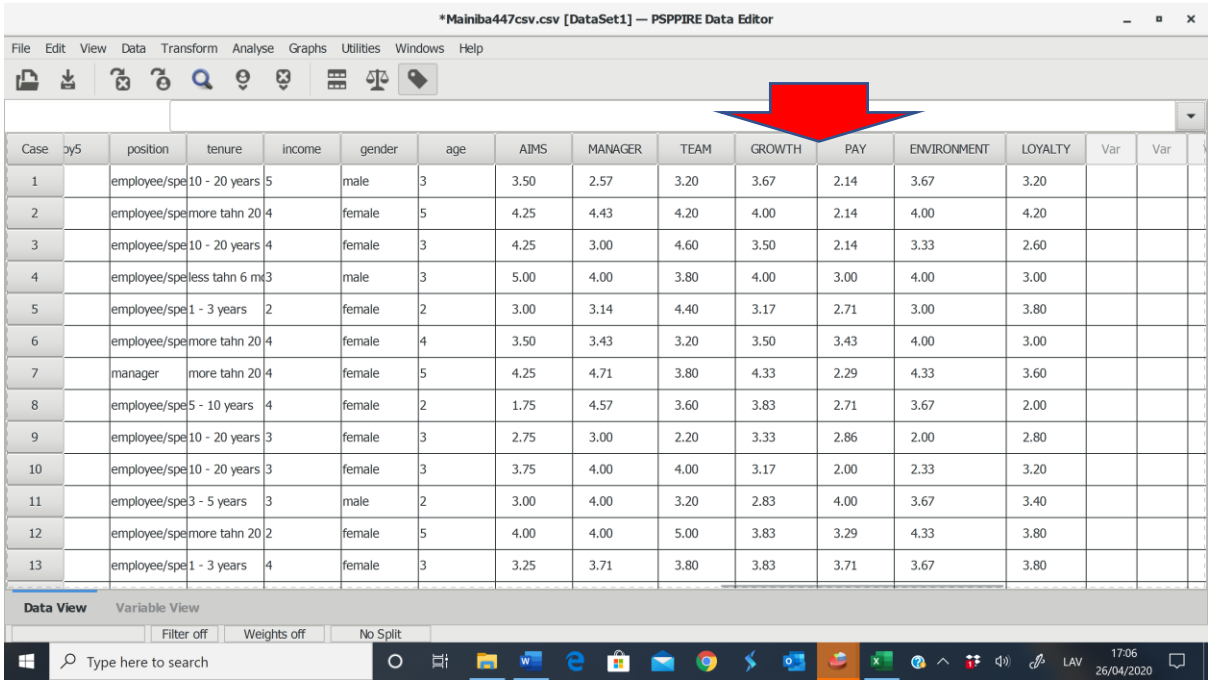


Figure 5.3. Variable columns to be used for further analysis.

6. Test for normality of data distribution (K-S test)

Before we begin an in-depth statistical analysis, we must check whether the data conforms to a normal distribution. Before we start analysing relationships or differences, we need to know which methods to use – parametric or non-parametric.

This test is performed on variables, not on individual elements.

To do this, proceed as follows:

• Under the Analyse menu, click on Non-Parametric Statistics and 1 Sample K-S

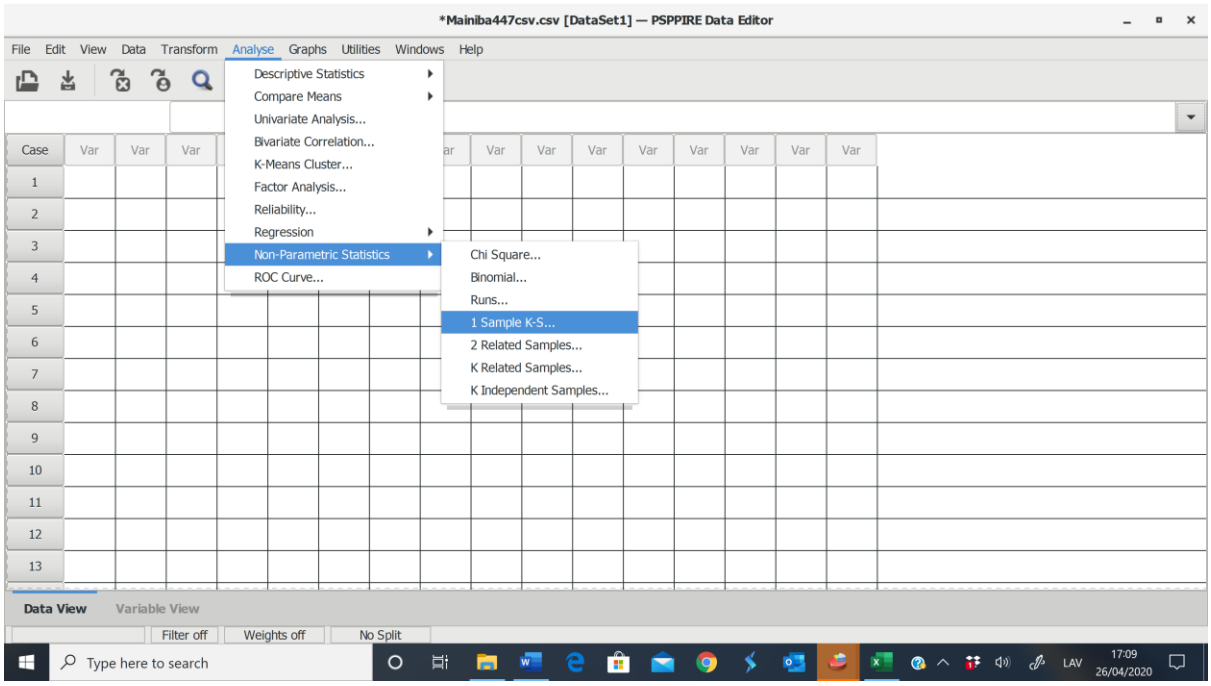


Figure 6.1. The location of the Kolmogorov–Smirnov (K–S test) function in PSPP.

- In the Analysis window, enter all variables
- Below this window, click Normal
- Click OK

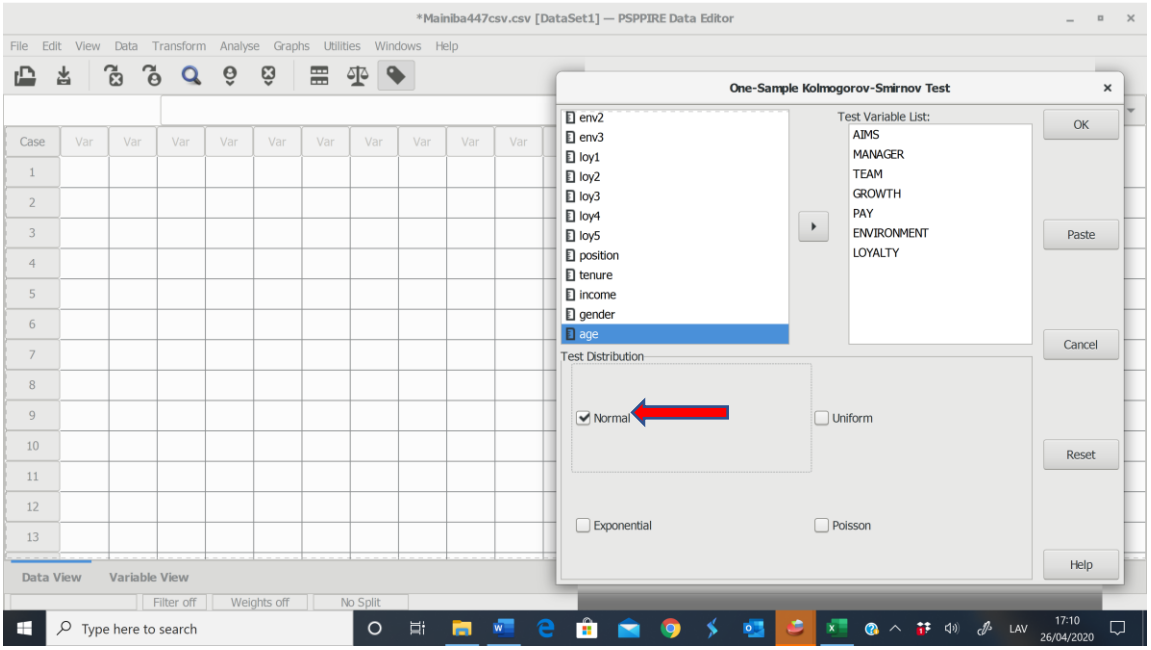


Figure 6.2. Location of the 'Normal' button in the context of the K-S test.

The result shown below appears in the Output

For each variable, you will see the arithmetic mean (Mean); standard deviation (Std. Deviation); Kolmogorov-Smirnov Z and Sig values. Interpretation is based on the Sig value.

Significance (sig) of the K-S test values:

- $p \geq 0.05$ – the data follow a normal distribution, and the results can be considered representative; parametric tests should be applied for further analysis
- $p < 0.05$ – the data do not follow a normal distribution; non-parametric tests should be used for further analysis

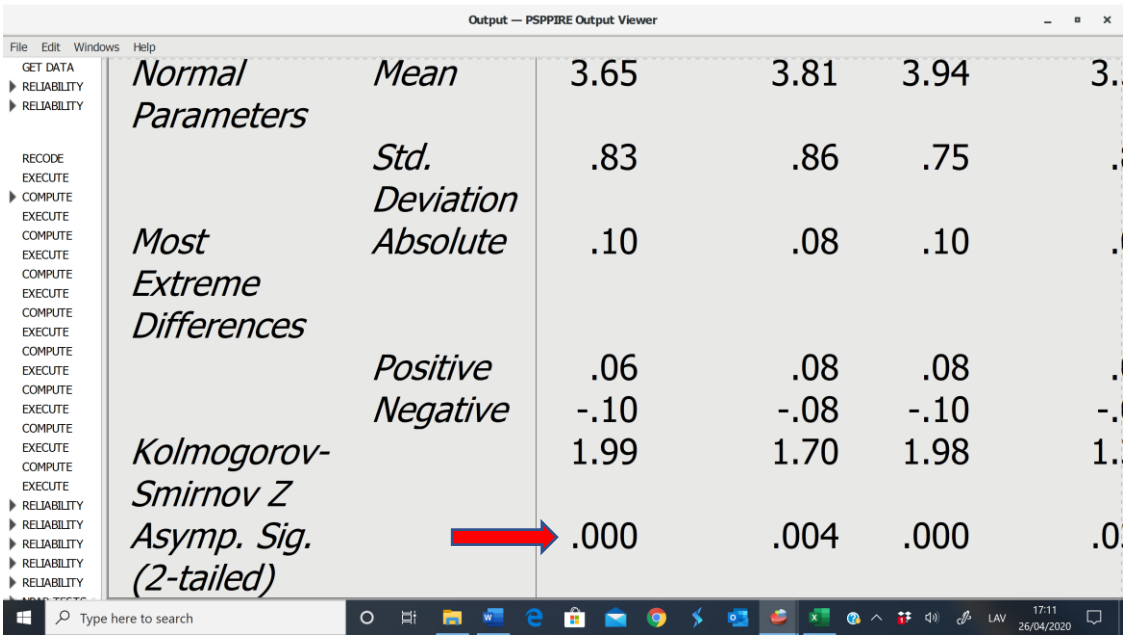


Figure 6.3. Output logs for the K-S test and the Sig value.

Depending on the results of the K-S test, the data analysis methods differ:

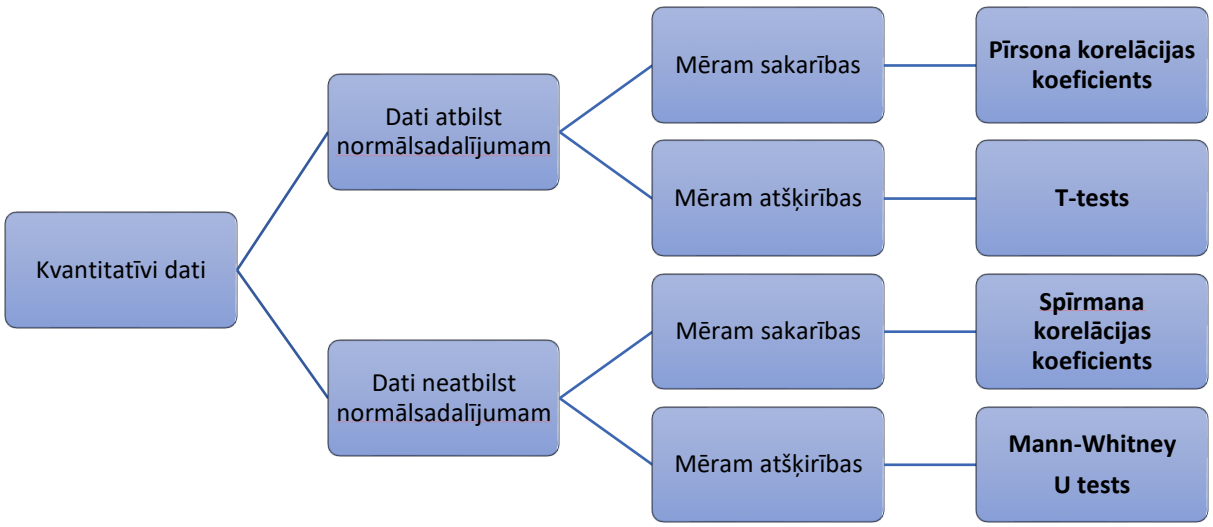


Figure 6.4. Data analysis functions for determining variable relationships.

Parametric or non-parametric test?

Parametric tests are used:

1) if the data follow a normal distribution *(according to the K-S test)*
2) the data is measured on a continuous scale

These include: t-test; one-way ANOVA; Pearson correlation coefficient

Non-parametric tests are used:

1) if the data do not follow a normal distribution *(K-S test)*

These are: Mann-Whitney U-test; Kruskal-Wallis test; Spearman rank correlation coefficient;

2) a nominal scale is used

Chi-square test

However, the PSPP programme only allows for parametric tests, so regardless of the result, we will use parametric tests for further analysis: Pearson's correlation and the t-test.

Histogram

The conformity of the data to a normal distribution can also be visualised. To do this:

- Under the Graphs menu, select Histogram
- In the window that opens, select the variable and drag it to the Variable
- Tick Show nominal curve on histogram
- Click OK

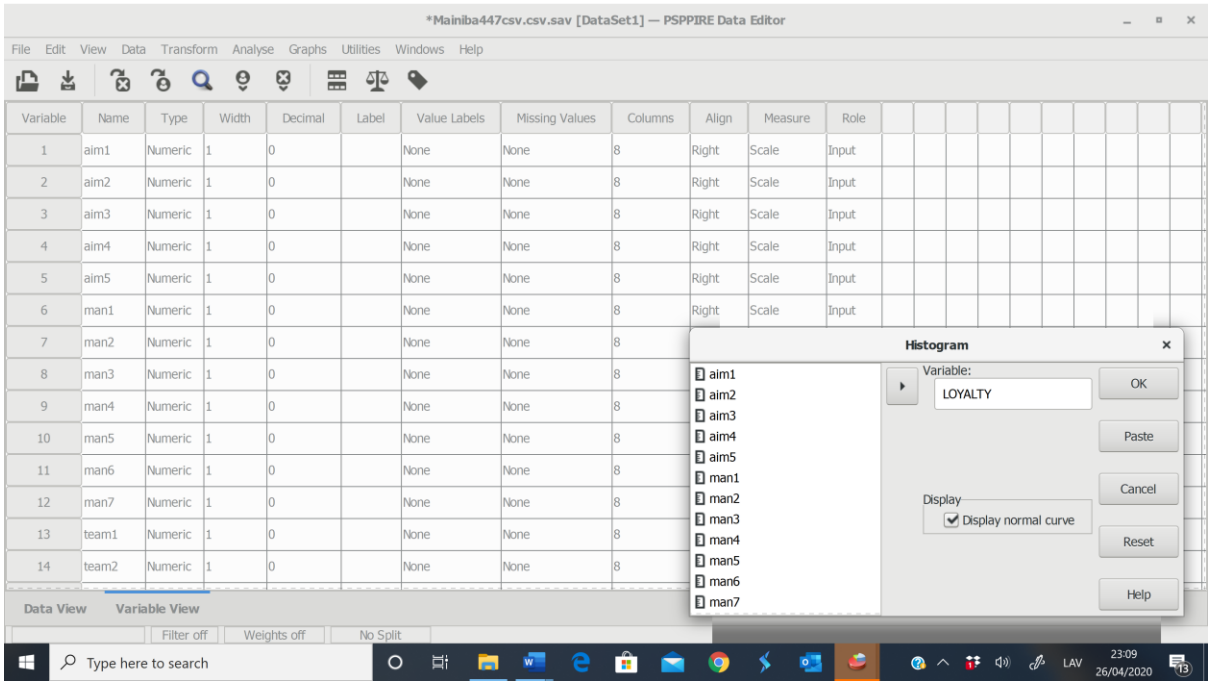


Figure 6.5. The histogram creation function in PSPP.

The result will look like this:

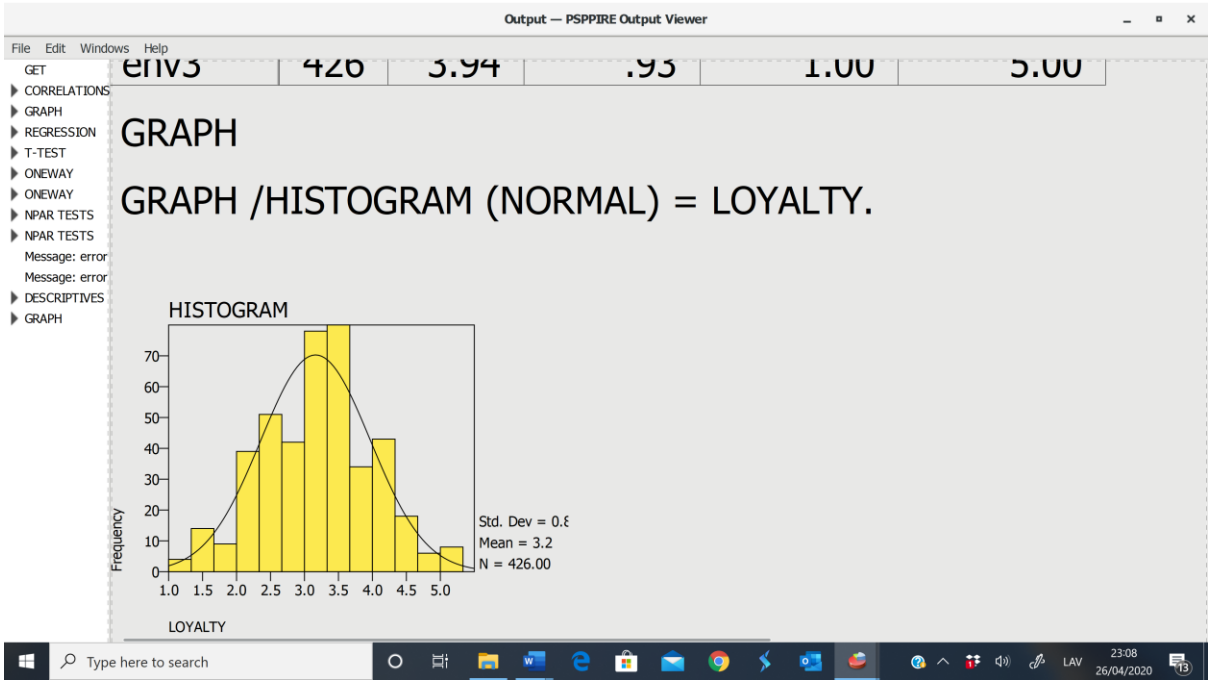


Figure 6.6. Histogram in PSPP programme.

The histogram shows that it is quite close to a normal distribution curve. This image can serve as justification for choosing parametric analysis methods.

The histogram shows the distribution of responses on the scale – it can be seen that the average Loyalty grade is 3.2, with most responses falling between 3 and 3.6.

7.

To measure the degree to which two variables are interrelated or vary together, we can calculate the correlation coefficient. Unless otherwise stated, the correlation coefficient refers to the 'Pearson product-moment correlation coefficient', denoted by the symbol 'r'. This coefficient requires data on an interval or ratio scale (Likert scales are considered valid).

Pearson's r is useful if you want to find out to what extent, as one variable changes, the value of the other variable increases or decreases linearly. The correlation coefficient ranges from 0.00 (no relationship) to ±1.00 (perfect positive or negative relationship). A positive value indicates that as one variable changes, the other also increases. A negative value indicates that as one variable increases, the other decreases. The relationship can also be illustrated graphically, as shown below.

To calculate the correlation coefficient:

• Click on ANALYSIS in the menu bar. Then click on BIVARIATE CORRELATION.

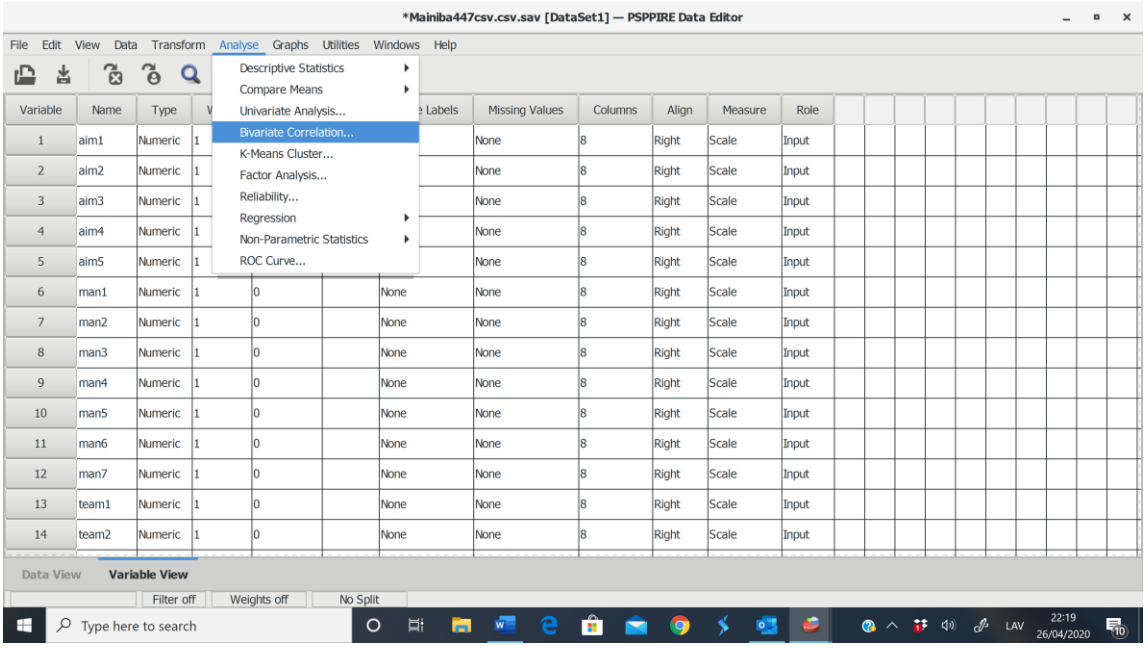


Figure 7.1. Location of the correlation function in PSPP.

From the menu that appears, select the variables in the left-hand pane and drag them to the right-hand pane.

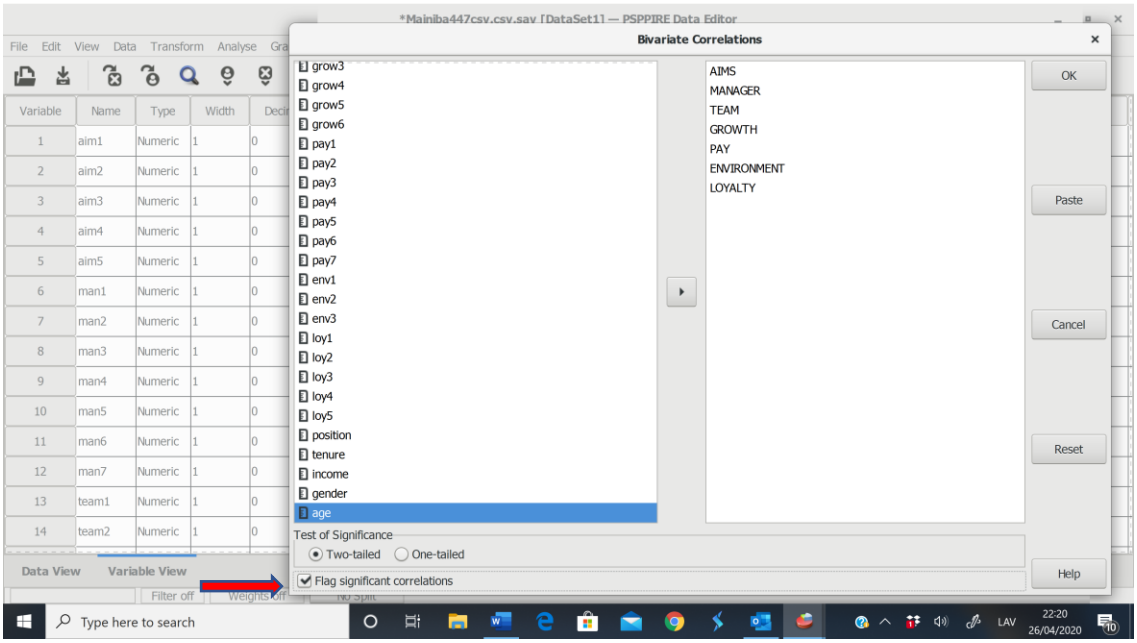


Figure 7.2. Flagging significant correlations.

At the bottom of the window, click on Flag significant correlations

• Click OK

You will receive the following correlation table (see Table 7.1).

Table 7.1

Correlation table in PSPP

		AIMS	MANAGER	TEAM	GROWTH	PAY	ENVIRONMENT	LOYALTY
AIMS	**Pearson Correlation**	1.00	.43	.38	.57	.39	.52	.56
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000
	N	426	426	426	426	426	426	426
MANAGER	**Pearson Correlation**	.43	1.00	.45	.49	.37	.38	.48
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000
	N	426	426	426	426	426	426	426
TEAM	**Pearson Correlation**	.38	.45	1.00	.41	.27	.36	.40
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000
	N	426	426	426	426	426	426	426
GROWTH	**Pearson Correlation**	.57	.49	.41	1.00	.42	.55	.51
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000
	N	426	426	426	426	426	426	426
PAY	**Pearson Correlation**	.39	.37	.27	.42	1.00	.49	.48
	Sig. (2-	.000	.000	.000	.000		.000	.000

		AIMS	MANAGER	TEAM	GROWTH	PAY	ENVIRONMENT	LOYALTY
	tailed)							
	N	426	426	426	426	426	426	426
ENVIRONMENT	Pearson Correlation	.52	.38	.36	.55	.49	1.00	.57
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000
	N	426	426	426	426	426	426	426
LOYALTY	Pearson Correlation	.56	.48	.40	.51	.48	.57	1.00
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	
	N	426	426	426	426	426	426	426

The correlation between Loyalty and Work Environment (r=0.57; sig=0.000) indicates a moderately strong positive statistically significant correlation.

Table 7.2

Interpretation of the numerical values of the correlation coefficient

Numerical value of the correlation coefficient	Interpretation
0.8 to 1	Very strong correlation
0.6 to 0.8	Strong correlation
0.4 to 0.6	Moderate correlation
0.2 to 0.4	Weak correlation
0 to 0.2	Very weak or no correlation

Scatterplots – a visual representation of correlation

We can also create a graph showing the relationship between any two variables measured on at least an interval scale. For example, we might wish to investigate (visualise) the extent to which our respondents' loyalty is related to their pay.

- In the ribbon, click on GRAPHS.
- In the left-hand dialogue box that appears, select 'Loyalty' and drag it into the right-hand pane under "Y-axis".
- Move the other variable (in this example, 'Goals') under "X axis".
- Click OK

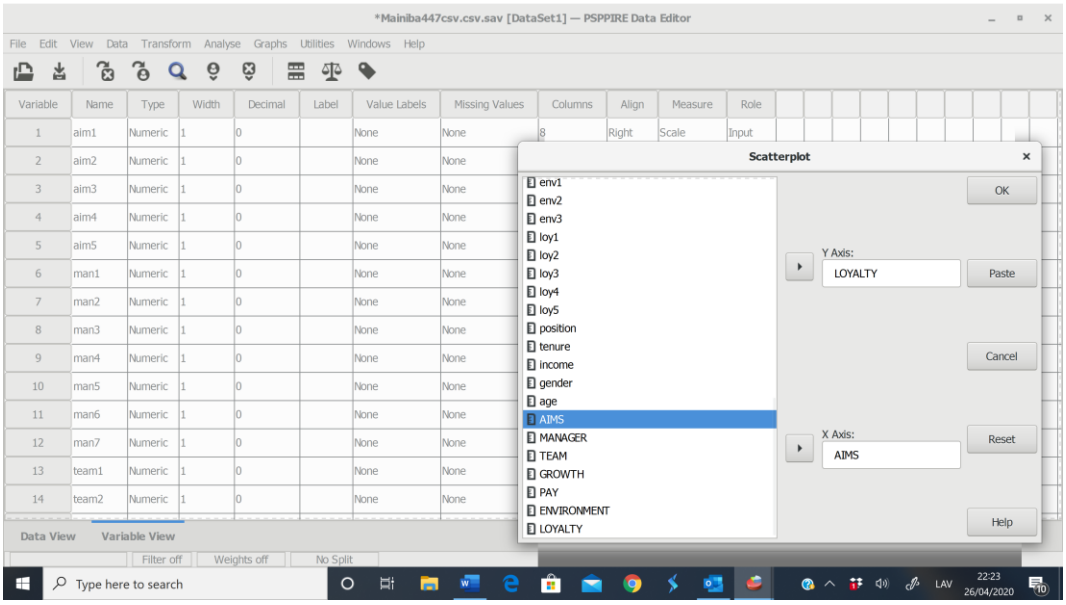


Figure 7.3. Location of the Scatterplot function in PSPP.

In the graph below, we can see that, generally speaking, the higher the AIMS value (the higher the X value), the higher the loyalty (the higher the Y value).

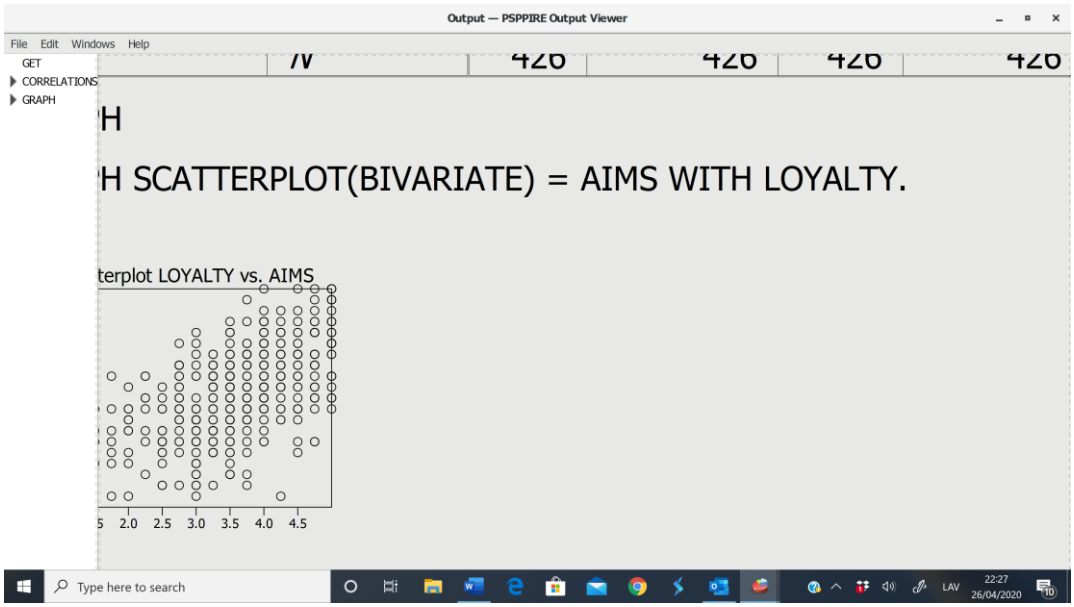


Figure 7.4. Visual representation of the scatterplot in PSPP

8. Regression

Let us assume that we wish to determine the extent to which all independent variables influence the dependent variable (in the case of TURNOVER, loyalty). Our independent variables will be environment, growth, leadership and remuneration.

To create a regression model:

- Click on ANALYSIS in the menu bar.
- Then click on REGRESSION and LINEAR.

- In the menu that appears, select Loyalty in the left-hand pane and move it to the top-right pane under DEPENDENT.
- Move Environment, Growth, Management and Remuneration to the right-hand pane under INDEPENDENT. Your screen should look like this

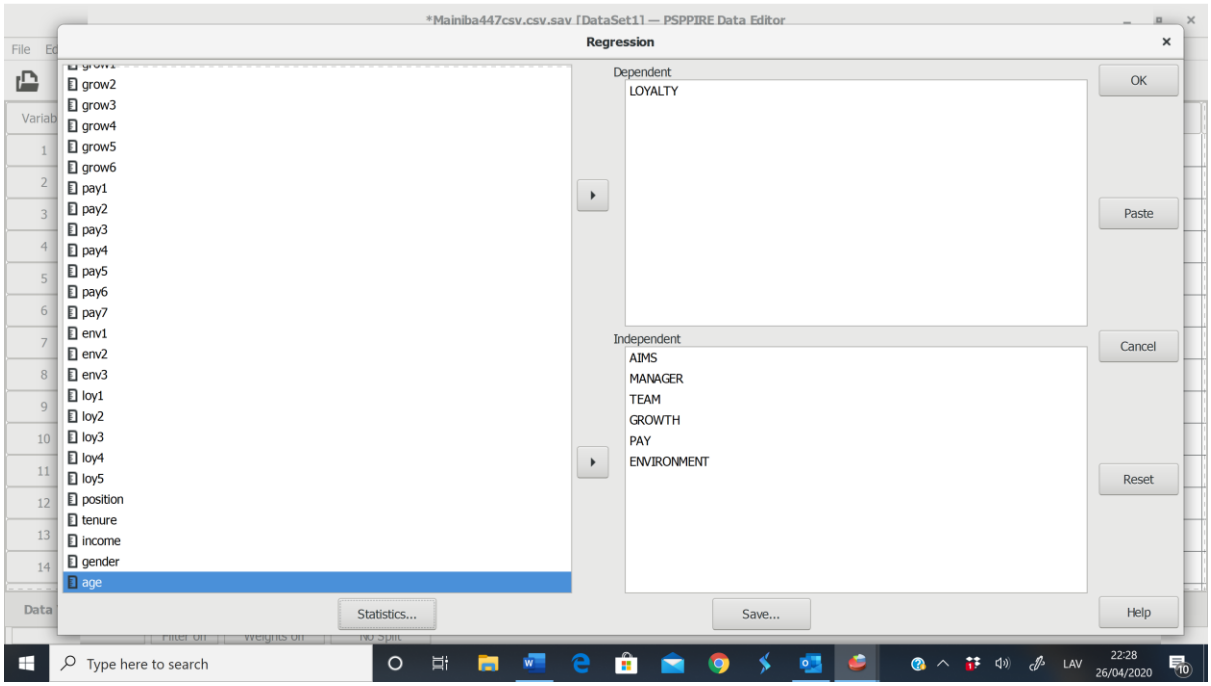


Figure 8.1 Location of the Regression function in PSPP

- Click OK.
- The results should look like this

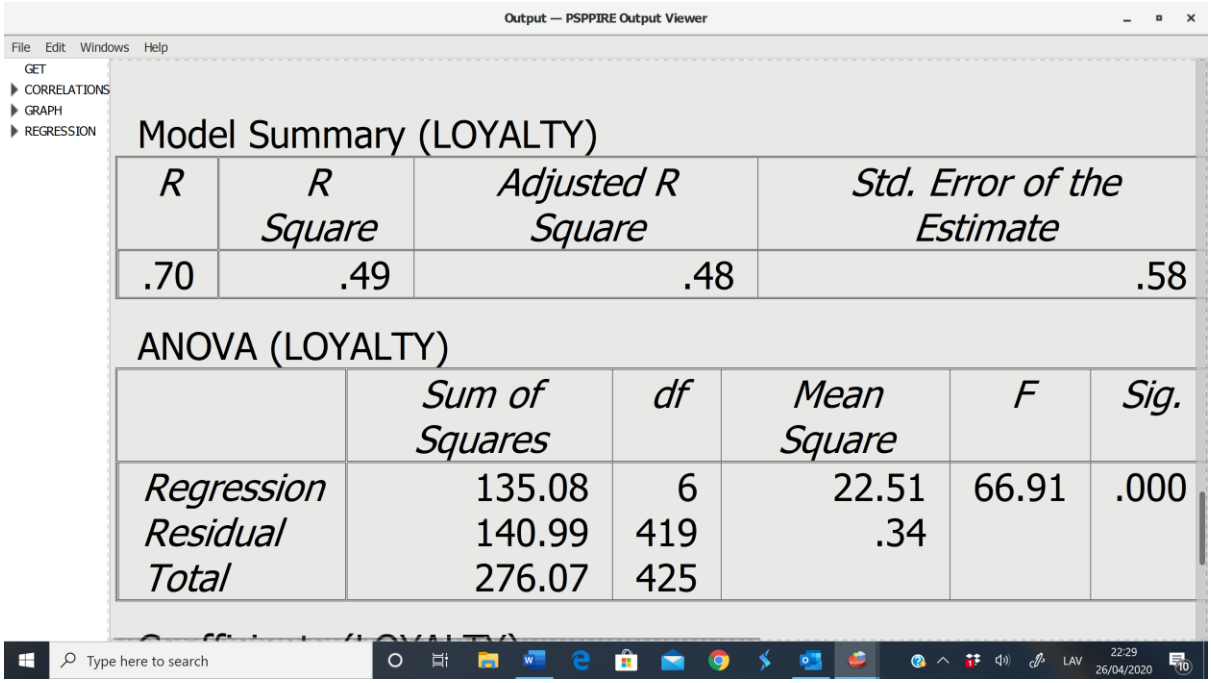


Figure 8.2. Regression results output.

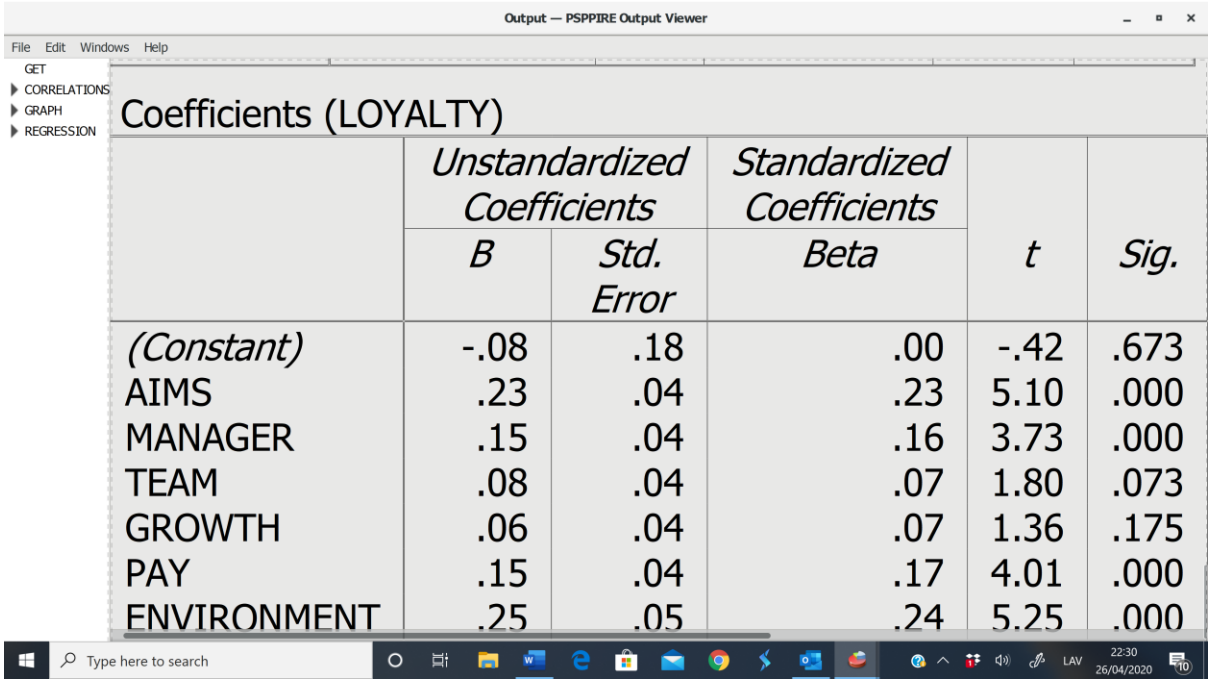


Figure 8.3. Regression results output (continued).

Let's look at just the main results.

From the first table (model summary), we obtain the R, R^2 and adjusted R^2 values (which take into account the number of variables in the equation)

Adjusted R^2 is 0.48. This means that all four independent variables together explain approximately half (48%) of the total variation in loyalty. Note that in multiple regressions (i.e. with more than one independent variable), the correlation coefficients are displayed in capital letters.

From the second (ANOVA) table, we can see that the model as a whole is statistically significant (sig < 0.05).

In the third table, we find the following.

- The B coefficients (also known as effect coefficients) show the change in the dependent variable in response to a one-unit change in any independent variable, all of which are expressed as ratios in our model. Since these coefficients are expressed in the units used to code the respective variables, they are referred to as 'non-standard' regression coefficients.
- t is a significance test that determines whether the coefficients are significantly greater than zero.
- The significance level for each t-value is sig. This tells us whether the relationship between the dependent variable and each independent variable is statistically significant, whilst controlling for all other independent variables.

The larger the B coefficient, provided it is statistically significant, the greater the influence of the respective factor on the dependent variable.

9. Descriptive statistics

The regression results show that the greatest influence on Loyalty is the working environment (ENVIRONMENT), so it would be interesting to find out what should be done to improve the environment.

To calculate the grade for each statement for the ENVIRONMENT variable:

- In the Analyse menu, select Descriptive Statistics and Descriptives

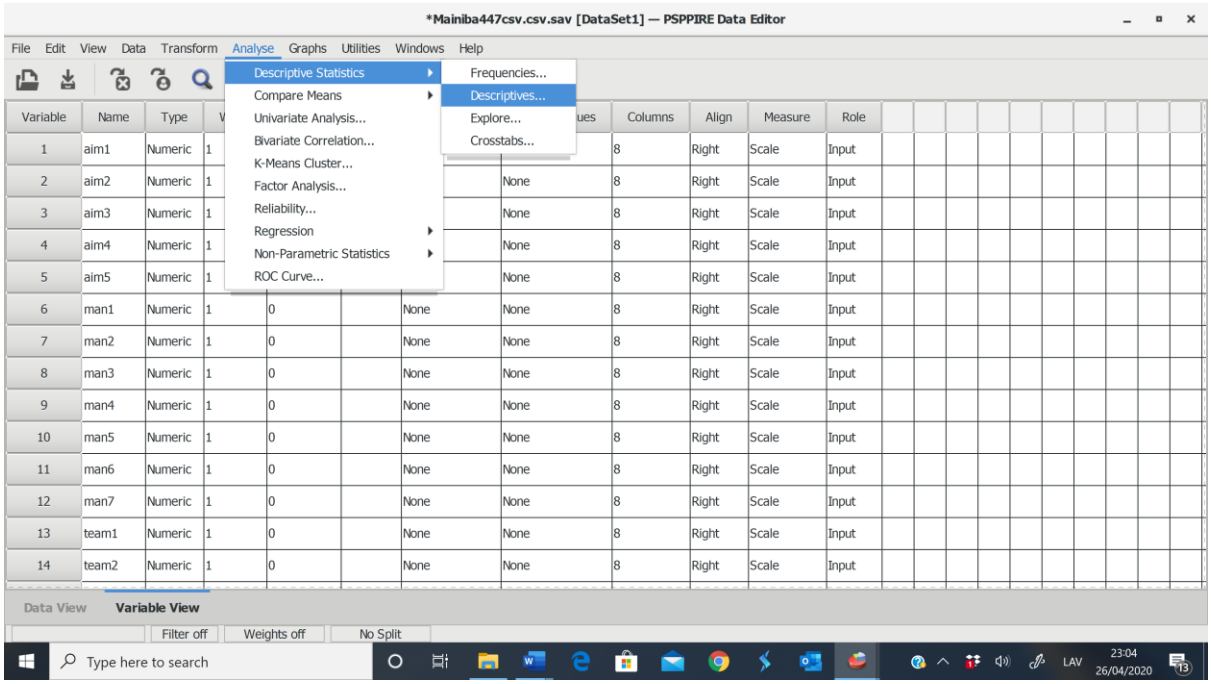


Figure 9.1. Location of the Descriptive Statistics function in PSPP.

- Move all 3 Environment items into the analysis window
- Click OK

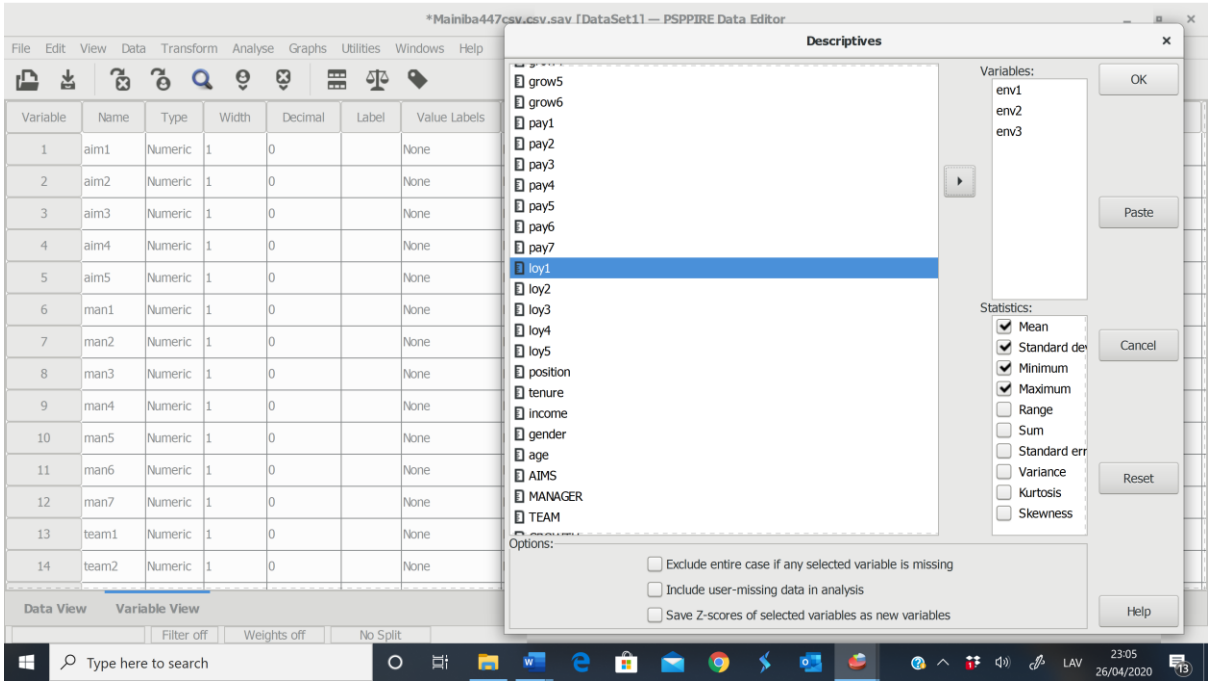


Figure 9.2. Marking central tendency indicators.

The result shows the arithmetic mean (MEAN) for each item

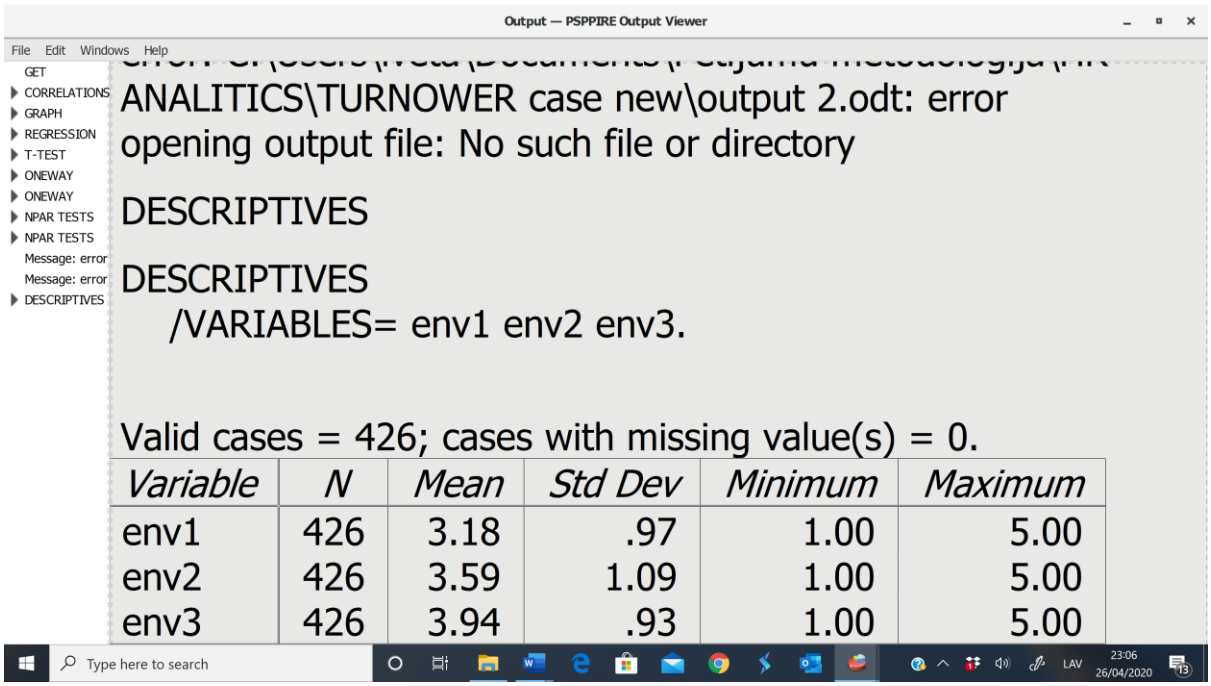


Figure 9.3. Output of descriptive statistics.

It can be seen that env1 is the lowest-rated item, so a data-driven decision would be to improve the situation regarding it.

10.Saving PSPP analysis results

When working in PSPP, results are accumulated in the Output log, but when the programme is closed, they disappear and it is not possible to save the Output as a separate file. To avoid losing the results before closing the PSPP programme, they must be exported.

To save the accumulated PSPP Output results, proceed as follows:

- Open File, click on Export
- In the window that appears, click on ‘Infer the type from extensions’
- Select OpenDocument
- Name the document and choose where you want to save it
- Click Save

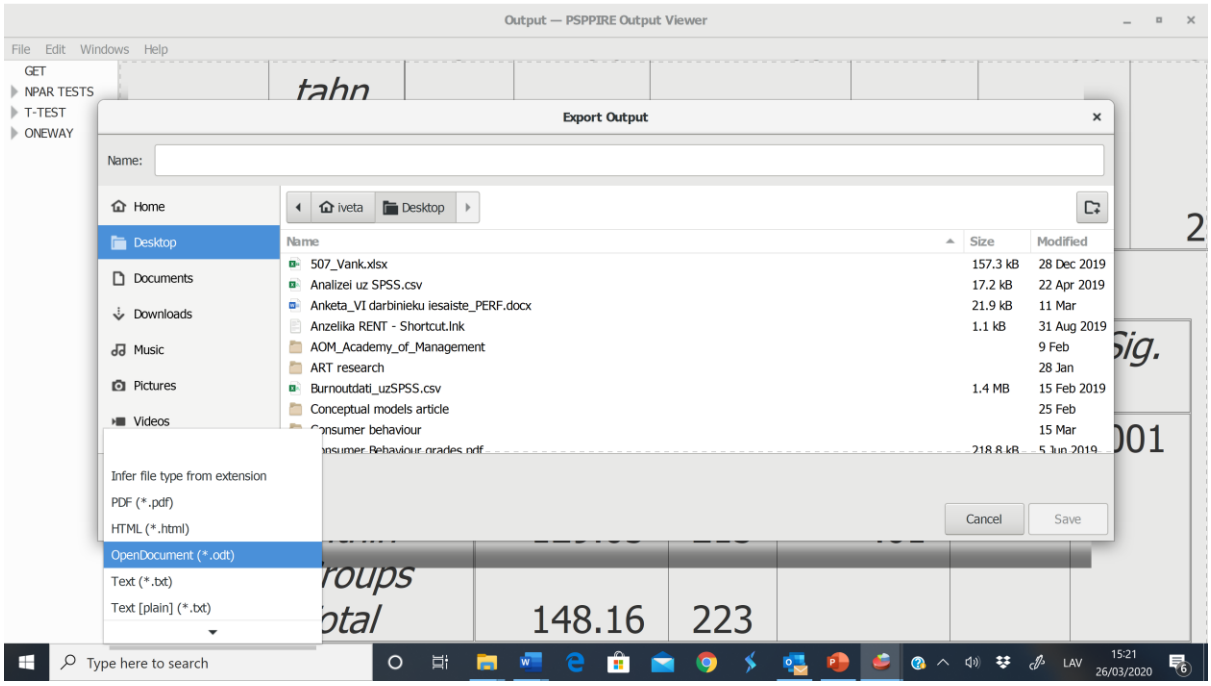


Figure 10.1. Saving PSPP data.

This will save all the PSPP analysis results in the editable Word document. These results will be required for inclusion in the research report (including the Master's thesis).

The results can also be exported in PDF format.

11.Comparing group grade scores (Tests of difference)

PSPP will calculate the arithmetic means (Mean) for the variables you have specified, and will also calculate whether the differences between the mean grades of different groups are significant.

Click on ANALYSIS and then on COMPARE MEANS. The window that opens will offer the following tests:

Means...
One Sample T Test...
Independent Samples T Test...
Paired Samples T Test...
One Way ANOVA...

11.1. Tests for differences between independent groups

In the case of TURNOVER, we can compare whether employees' views differ from those of managers. These are two independent groups.

What do we mean by independent groups? Independent groups are groups where the composition of one group does not influence the composition of the other group.

The tests for differences will provide answers to the research questions:

RQ1: Do employees and managers have different attitudes towards the environment, growth, leadership and remuneration?

RQ2: Are there significant differences between employee and manager loyalty?

The t-test is used when you wish to compare two groups. The grouping variable defines these two groups. The variable 'position' is a nominal variable. It has only two categories – employee (value 1) and manager (value 2)

Do the following:

- Under Analyse, go to Compare means
- Independent Samples t-Test

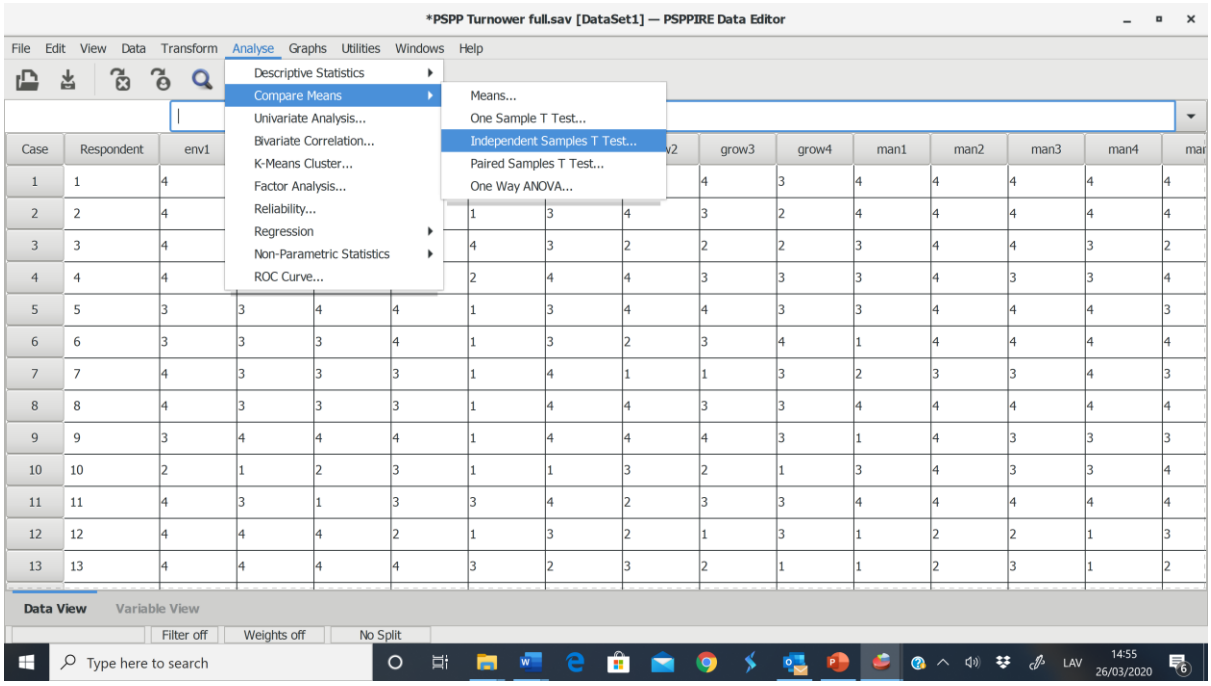


Figure 11.1. Location of the t-test function in PSPP.

• Move the variable into the test window

• Transfer the grouping variable (in this case, position) to the window below (Grouping variable)

• Click on Define groups

• Enter the codes in both boxes. Codes for the TURNOVER case: 1 = employee; 2 = manager

• Click Continue

• Click OK

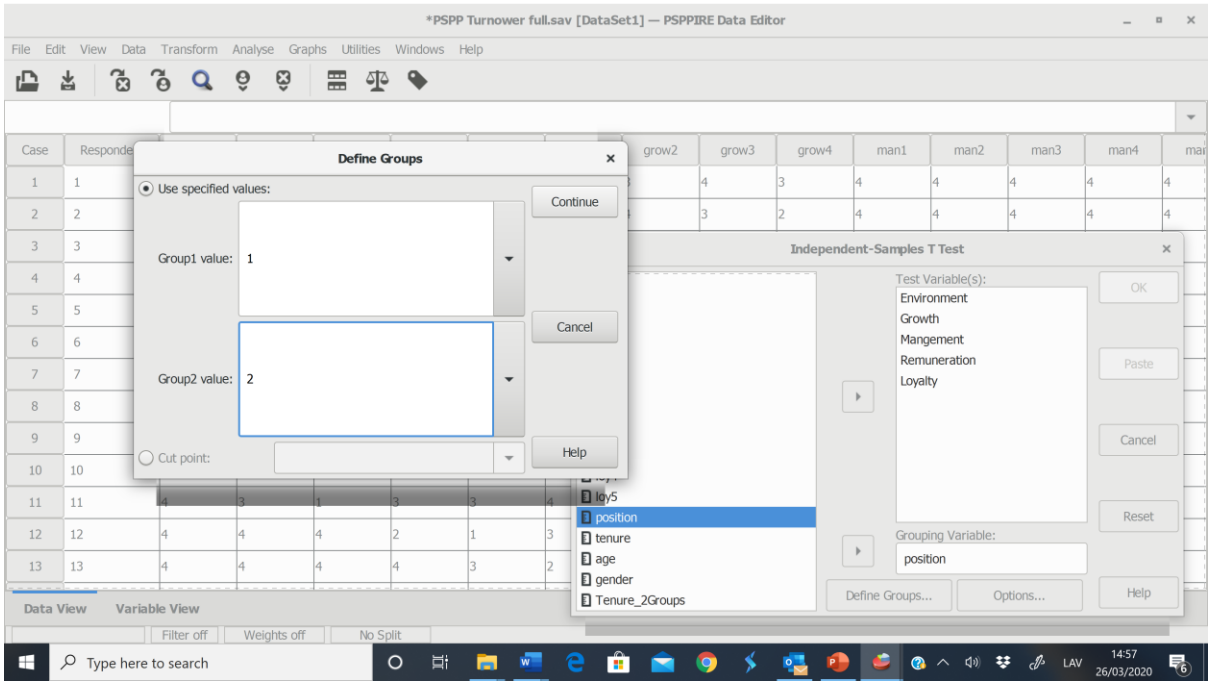


Figure 11.2. Defining groups.

Interpret the T-test sig values for equal means: significance of the test (two-tailed):

- sig > 0.05: Differences are not statistically significant
- sig < 0.05: Differences are statistically significant

11.2. One-way analysis of variance (ANOVA)

In this section, we will examine whether the grades for employees with different lengths of service differ and whether these differences are statistically significant. This is the task for a one-way analysis of variance (ANOVA).

The research question to be answered is as follows:

PJ: Is the length of service of respondents within the organisation associated with differences in opinion? Which employees belong to 'risk groups'?

Proceed as follows:

• In the menu bar, click on ANALISE, then on COMPARE MEANS, and finally on ONE WAY ANOVA.

• Select the variable Loyalty and move it to the DEPENDENT variable box

• Then select the variable Tenure and move it to the FACTOR box.

• In the Statistics box, click on Descriptives.

Your screen should look like this:

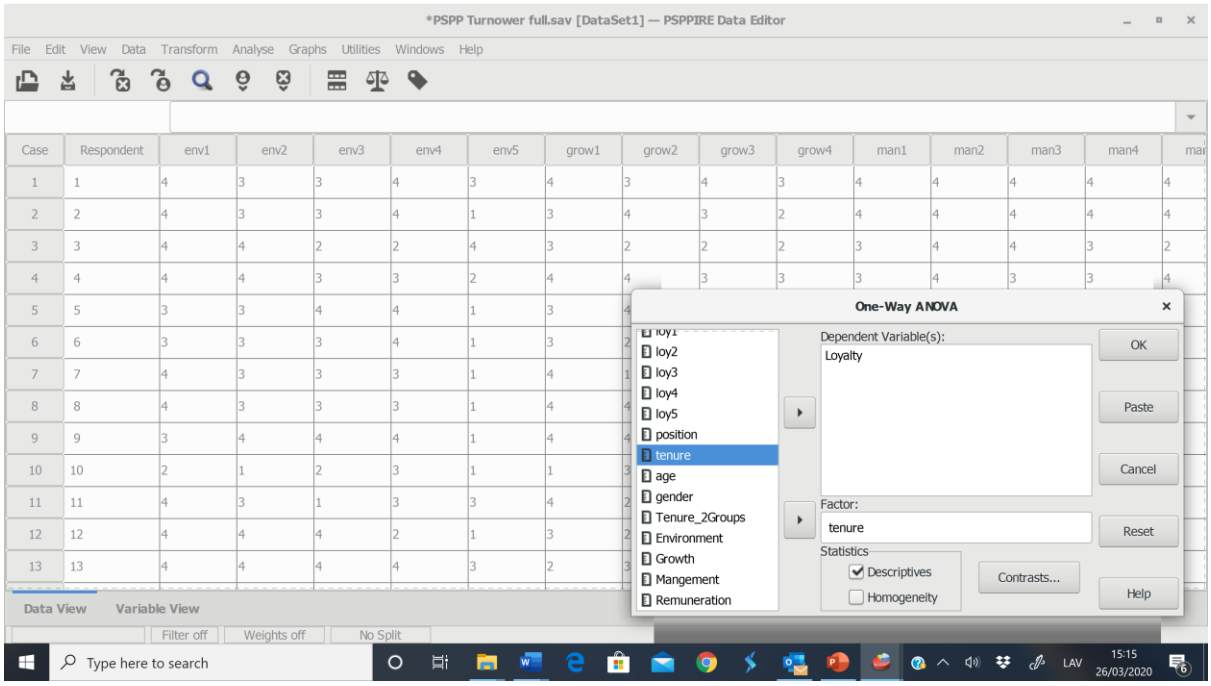


Figure 11.3. Location of the one-way analysis of variance (ANOVA) function in PSPP.

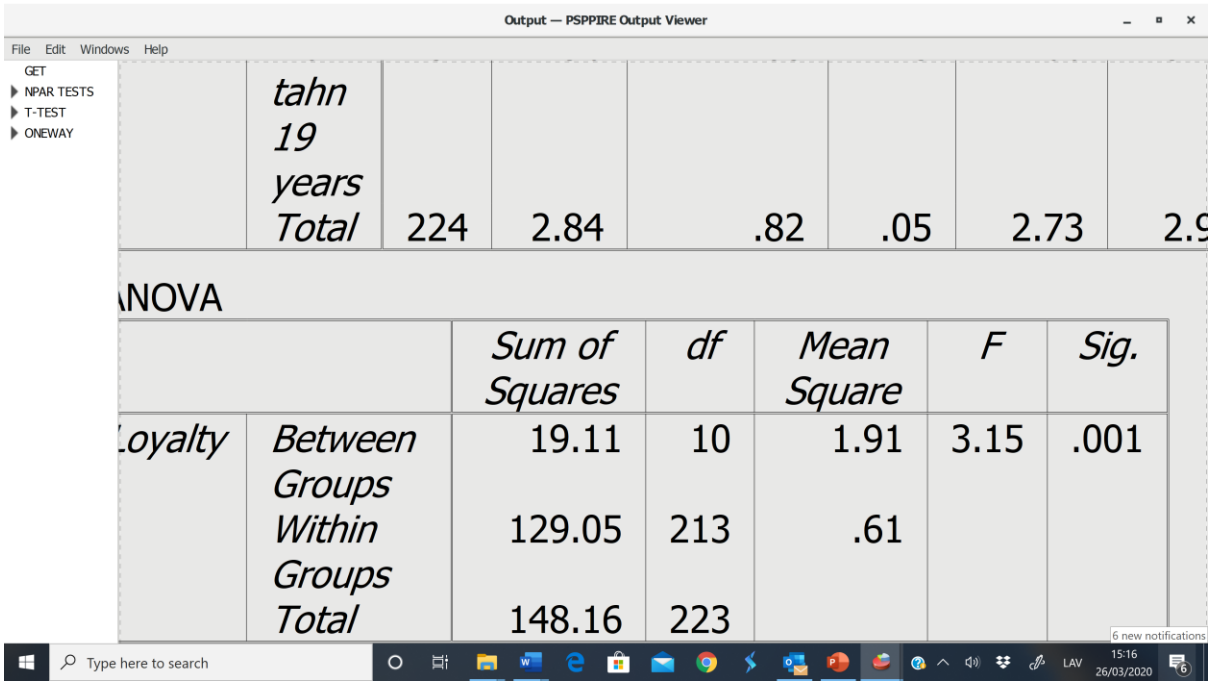


Figure 11.4. ANOVA results output.

The output provides the results of the one-way analysis of variance. The interpretation is based on significance (sig):

- sig > 0.05: the differences are not statistically significant
- sig < 0.05: there are statistically significant differences somewhere in the sample

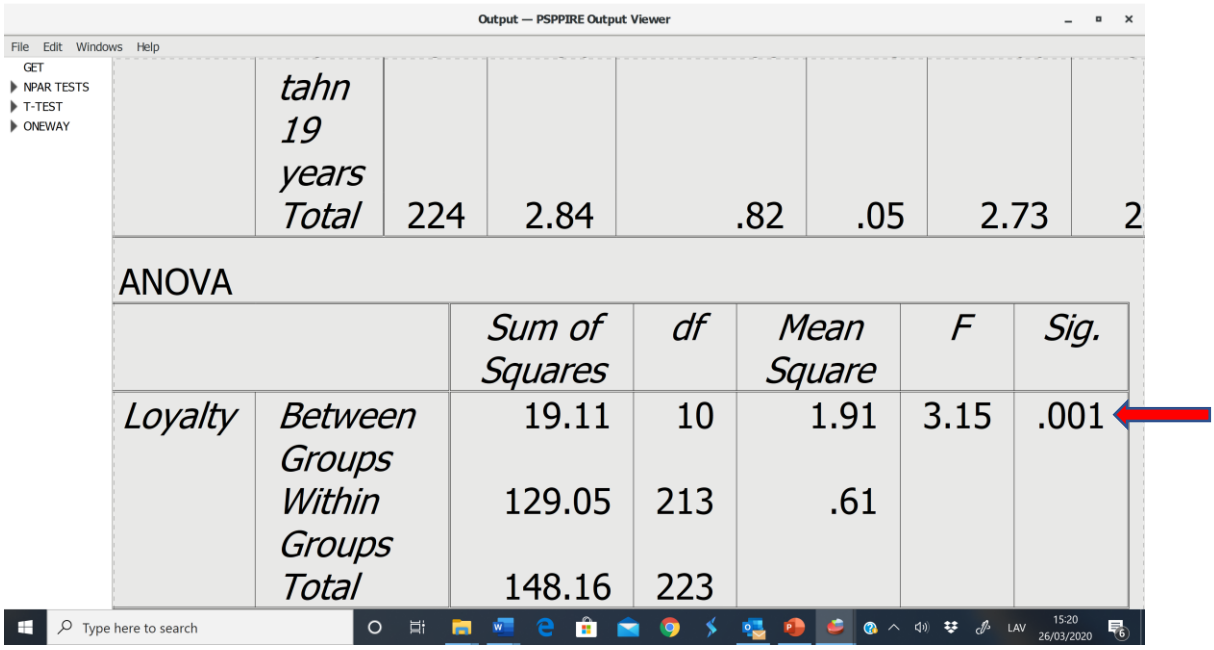


Figure 11.5. Interpreting ANOVA results.

If the differences are statistically significant, the arithmetic means (mean) of the groups must also be interpreted. For example, by displaying the arithmetic mean of each group in the form of a bar chart.

12.Multi-group analysis

If an analysis (e.g. ANOVA or t-tests) has identified several groups where the differences between groups are statistically significant, it is advisable to examine each of these groups separately.

For this reason, you can split the data into groups and, with a single click, perform an analysis by group.

The split is performed as follows

- Click on the “Split the active dataset” icon

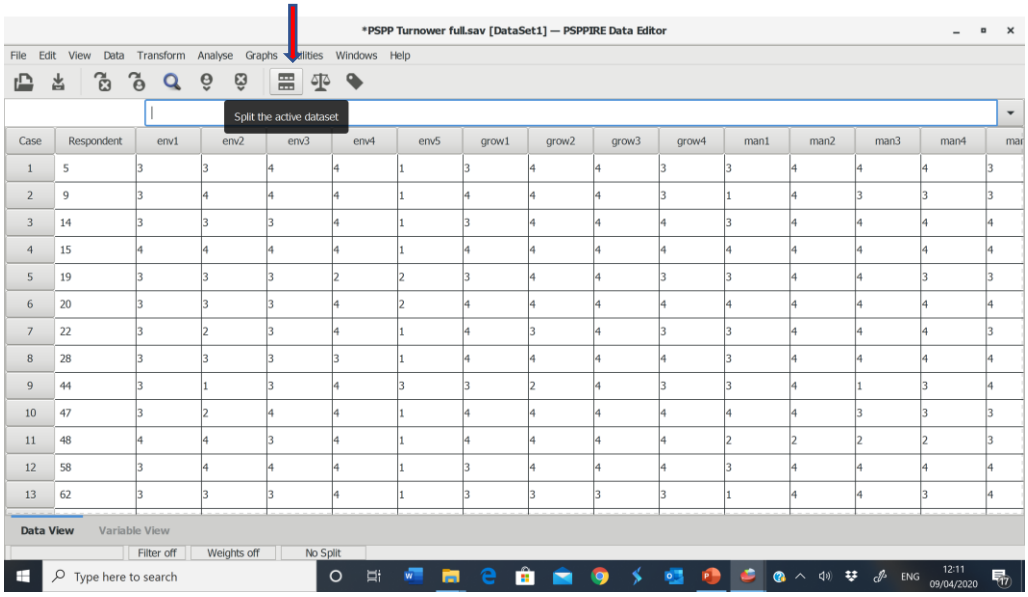


Figure 12.1. Location of the group split function in PSPP.

- Select the variable for grouping (usually on a nominal scale) – for example, ‘Seniority’ in the example
- Move it to the right-hand pane using the arrow
- Click “Organise output by groups”

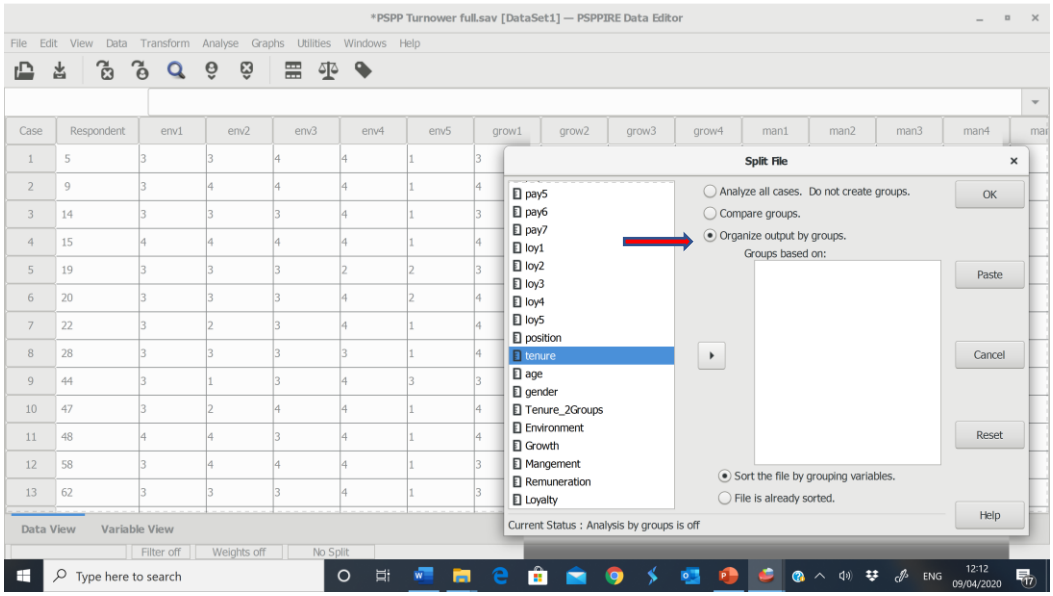


Figure 12.2. Adding a grouping variable.

- Click OK

In the output window, you will see that the grouping command has been executed. You can now run any test of your choice, and the results for each group will be displayed separately.

In this example, we have 11 groups based on length of service, which means you will obtain 11 different sets of results.

For example, to understand how the independent variables (environment, growth, leadership and remuneration) are related to the dependent variable (loyalty), you can use correlation analysis.

To calculate the correlations for each group, proceed in the same way as before (as shown in Chapter 7)

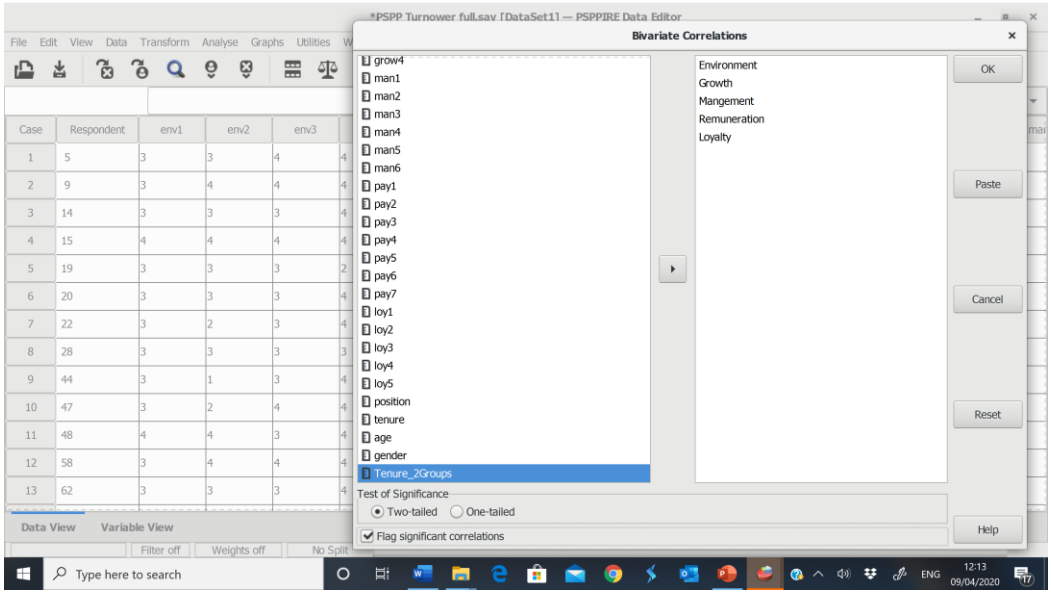


Figure 12.3. Correlation coefficient for each group.

The result will be 11 separate correlation tables for each group. The tables will appear in the order in which the groups are listed:

Table 12.1

First correlation: For the length of service group: less than 1 year (n = 30)

		Environment	Growth	Management	Remuneration	Loyalty
Environment	**Pearson Correlation**	1.00	.29	.11	.40	.52
	Sig. (2-tailed)		.124	.577	.031	.003
	N	30	30	30	30	30
Growth	**Pearson Correlation**	.29	1.00	.11	.54	.44
	Sig. (2-tailed)	.124		.560	.002	.014
	N	30	30	30	30	30
Management	**Pearson Correlation**	.11	.11	1.00	.18	.15
	Sig. (2-tailed)	.577	.560		.351	.425
	N	30	30	30	30	30
Remuneration	**Pearson Correlation**	.40	.54	.18	1.00	.65
	Sig. (2-tailed)	.031	.002	.351		.000
	N	30	30	30	30	30
Loyalty	**Pearson Correlation**	.52	.44	.15	.65	1.00
	Sig. (2-tailed)	.003	.014	.425	.000	
	N	30	30	30	30	30

You can see that for new employees, three factors are significantly associated with loyalty – remuneration has the strongest influence (r = 0.65 ***), followed by the working environment (r = 0.53 ***) and career development (r = 0.44 *)……..

Correlation for Group 4: 3–5 years (n=8)

Table 12.2

Correlation for Group 4: Length of service: 3–5 years (n = 8)

		Environment	Growth	Management	Remuneration	Loyalty
Environment	**Pearson Correlation**	1.00	.85	-.03	.44	.73
	Sig. (2-tailed)		.008	.940	.270	.040
	N	8	8	8	8	8
Growth	**Pearson Correlation**	.85	1.00	.41	.59	.62
	Sig. (2-tailed)	.008		.310	.120	.099
	N	8	8	8	8	8
Management	**Pearson Correlation**	-.03	.41	1.00	.48	.20
	Sig. (2-tailed)	.940	.310		.225	.633

		Environment	Growth	Management	Remuneration	Loyalty
	N	8	8	8	8	8
Remuneration	**Pearson Correlation**	.44	.59	.48	1.00	.85
	Sig. (2-tailed)	.270	.120	.225		.008
	N	8	8	8	8	8
Loyalty	**Pearson Correlation**	.73	.62	.20	.85	1.00
	Sig. (2-tailed)	.040	.099	.633	.008	
	N	8	8	8	8	8

Table 12.3

Correlation for Group No. 6: Length of service: 9–11 years (n=32)

		Environment	Growth	Management	Remuneration	Loyalty
Environment	**Pearson Correlation**	1.00	.33	.19	.36	.49
	Sig. (2-tailed)		.067	.286	.041	.004
	N	32	32	32	32	32
Growth	**Pearson Correlation**	.33	1.00	.49	.58	.69
	Sig. (2-tailed)	.067		.004	.001	.000
	N	32	32	32	32	32
Management	**Pearson Correlation**	.19	.49	1.00	.48	.28
	Sig. (2-tailed)	.286	.004		.005	.117
	N	32	32	32	32	32
Remuneration	**Pearson Correlation**	.36	.58	.48	1.00	.60
	Sig. (2-tailed)	.041	.001	.005		.000
	N	32	32	32	32	32
Loyalty	**Pearson Correlation**	.49	.69	.28	.60	1.00
	Sig. (2-tailed)	.004	.000	.117	.000	
	N	32	32	32	32	32

You can see that for those who have been working for 9–11 years, three factors are significantly associated with loyalty – the strongest influence is growth (r = 0.69 ***), followed by remuneration (r = 0.60 ***) and the working environment (r = 0.6 ***). Therefore, it might be advisable to establish career development programmes to improve employees' opportunities for growth and only then to address other factors.

The output file contains 11 such correlation tables, and you can draw conclusions for each employee group.

The same can be done using regression analysis.

13.Chi-square test

Research often makes use of nominal data to which the statistical analysis methods discussed above cannot be applied. However, it is necessary to determine whether there is a relationship between variables measured on a nominal or categorical scale.

Used when the variables are nominal (words; categories).

The chi-square test is used to check whether:

1) Two statistical characteristics are independent of one another – a test of independence;

2) The empirical distribution corresponds to the theoretical distribution – a one-variable test (goodness-of-fit test).

13.1. Test of independence

Can be used to check whether two statistical variables are independent of one another – a two-sample test (**test of independence**)

In the example, it might be interesting to answer the question of whether gender is related to an employee's position – to determine whether gender discrimination is taking place and whether men have a greater chance of reaching a managerial position.

To perform a Chi-square test:

Under the ANALYSE menu, go to Descriptive Statistics and Crosstabs…

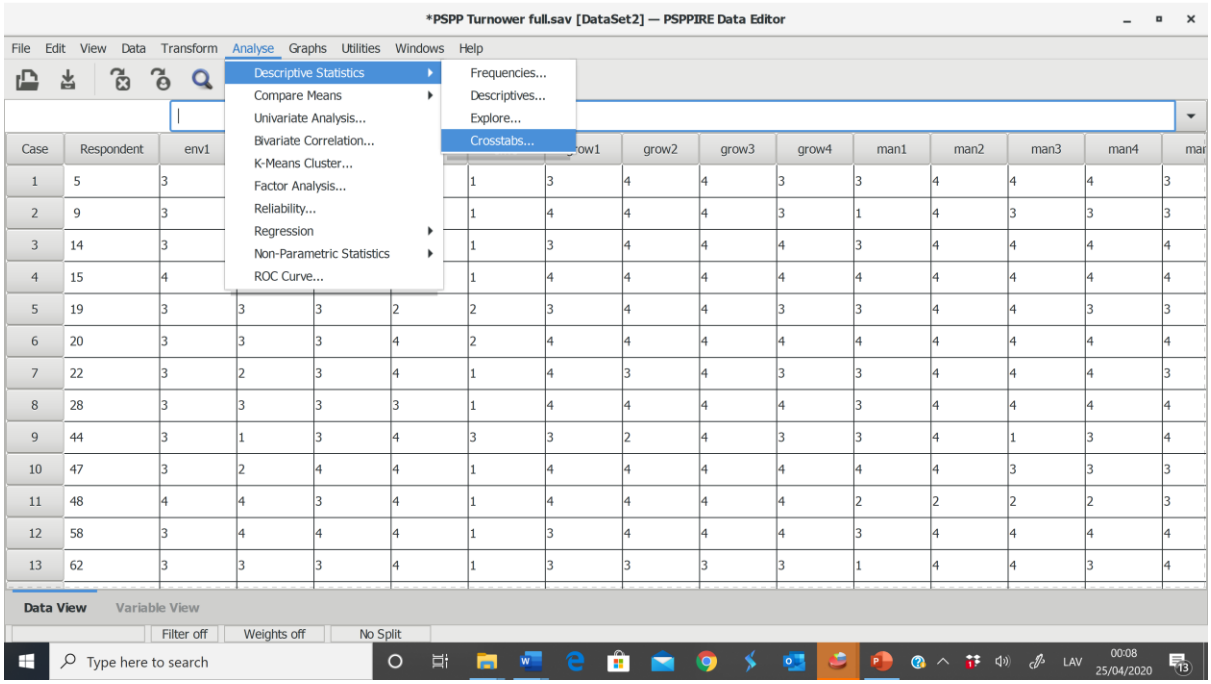


Figure 13.1. Location of the independent samples (Chi-square test) in PSPP.

- In the window that opens, select the variables and drag them into the Rows and Columns boxes using the mouse. It does not matter which variable is in which box.

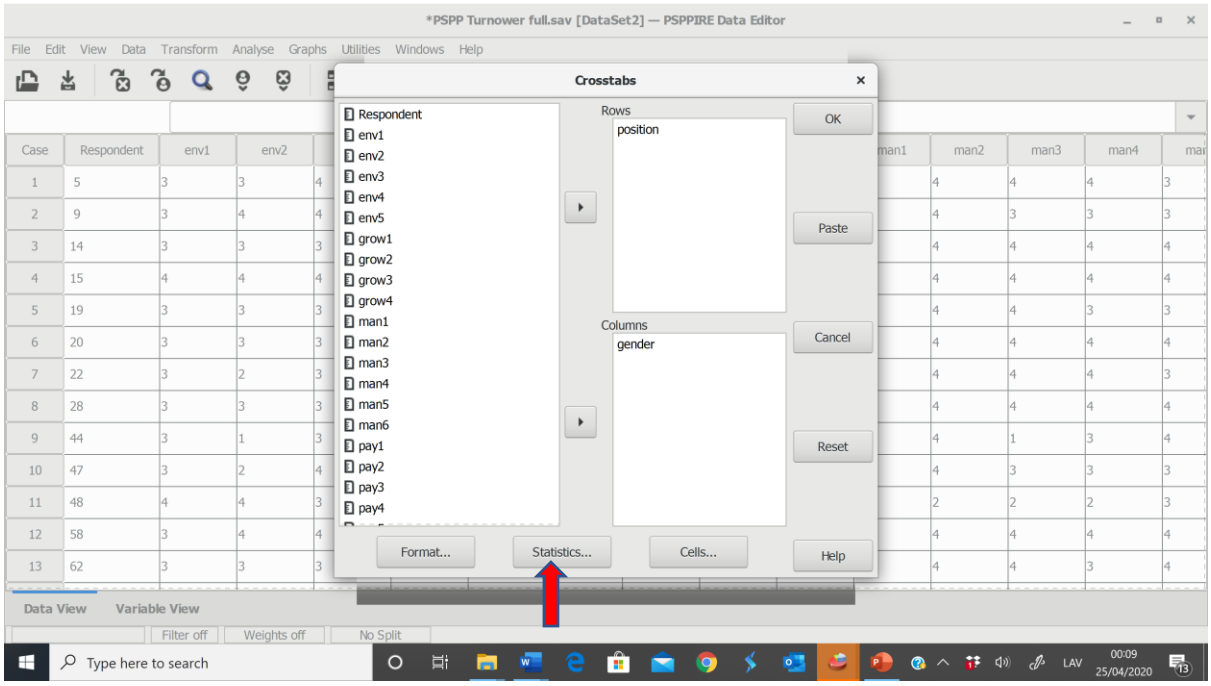


Figure 13.2. Introduction to groups.

- Clicking on Statistics will open the window shown below
- Select Phi

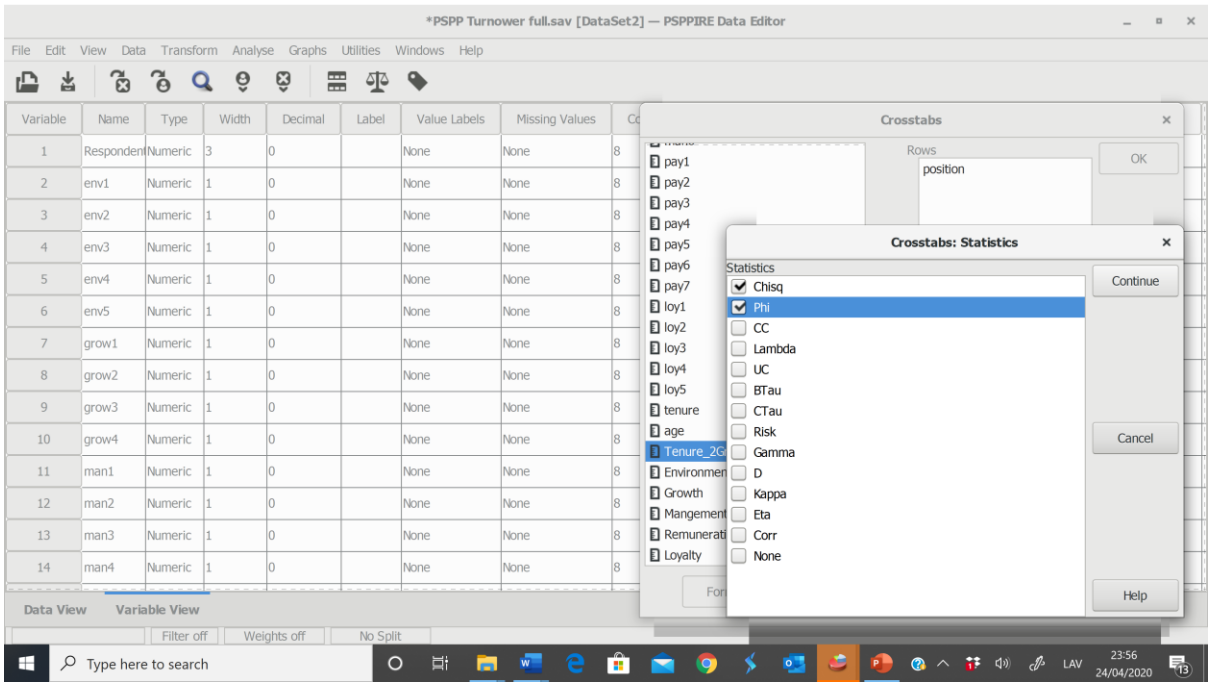


Figure 13.3. Selecting Phi.

- Clicking on Cells will open the window shown below
- Select Expected

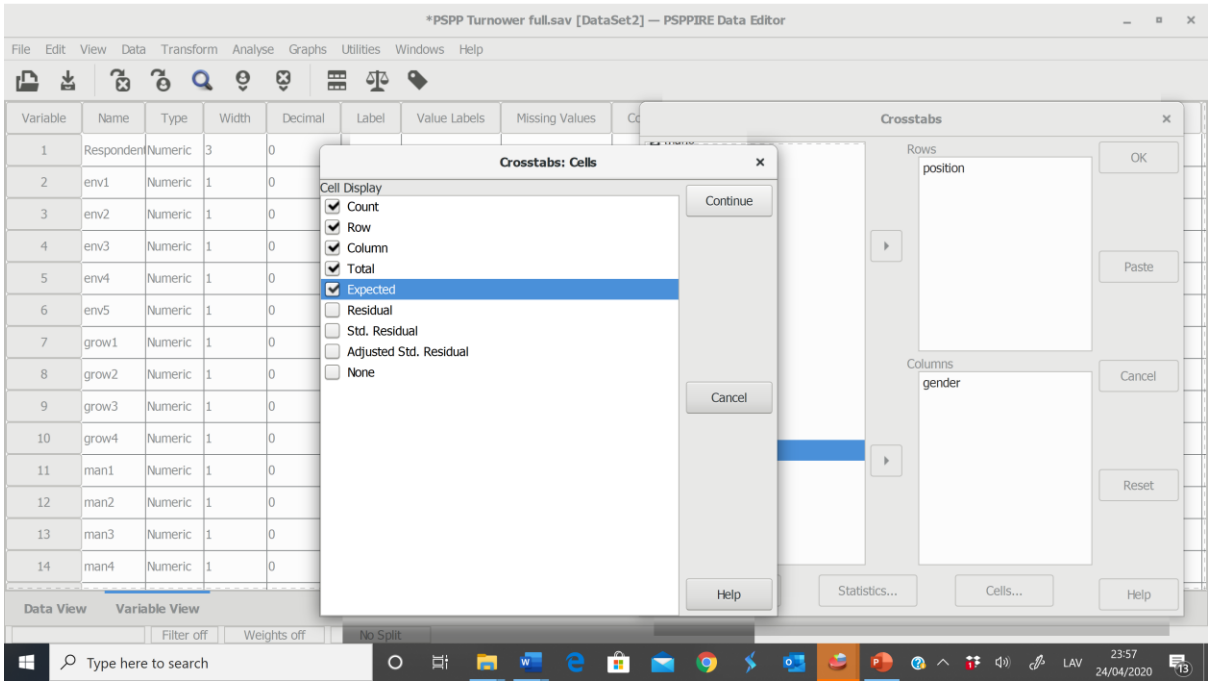


Figure 13.4. Selecting Expected.

- Click OK

OUTPUT will generate the following table:

CROSSTABS

CROSSTABS /TABLES= position BY gender /FORMAT=AVALUE TABLES PIVOT /STATISTICS=CHISQ PHI /CELLS=COUNT ROW COLUMN TOTAL EXPECTED.

Summary.

	N	Percent	N	Percent	N	Percent
position * gender	224	100.0%	0	0.0%	224	100.0%

position * gender [count, row %, column %, total %, expected].

position	**female**	**male**	**Total**
employee	169.00	31.00	200.00
	166.96	33.04	.00
	84.50%	15.50%	100.00%
	90.37%	83.78%	89.29%
	75.45%	13.84%	89.29%
manager	18.00	6.00	24.00
	20.04	3.96	.00
	75.00%	25.00%	100.00%
	9.63%	16.22%	10.71%

This row indicates that the actual number of employees is 169 women and 31 men

This row indicates the expected number that would result if, taking into account the difference in the samples, gender had no effect – there would then be 167 female and 33 male employees.

The chi-square test measures whether the differences between the actual number and the expected number are statistically significant.

The sig=0.247>0.05 shown below indicates

position	**female**	**male**	**Total**
	8.04%	2.68%	10.71%
Total	187.00	37.00	224.00
	83.48%	16.52%	100.00%
	100.00%	100.00%	100.00%
	83.48%	16.52%	100.00%

Chi-square tests.

Statistic	**Value**	**df**	**Asymp. Sig. (2-tailed)**	**Exact Sig. (2-tailed)**	**Exact Sig. (one-tailed)**
Pearson Chi-Square	1.40	1	.236		
Likelihood Ratio	1.27	1	.260		
Fisher's Exact Test				.247	.183
Continuity Correction	.80	1	.372		
Linear-by-Linear Association	1.40	1	.237		
N of Valid Cases	224				

Symmetric measures.

Category	**Statistic**	**Value**	**Asymptotic standard error**	**Approx. T**	**Approx. Sig.**
Nominal by Nominal	**Phi**	.08			
	Cramer's V	.08			
N of Valid Cases		224			

13.2. Goodness-of-fit test

Another option in the Chi-square test is the Goodness-of-fit test: to determine whether the sample corresponds to the population

Go to Analyse -> Nonparametric Statistics -> Chi-square

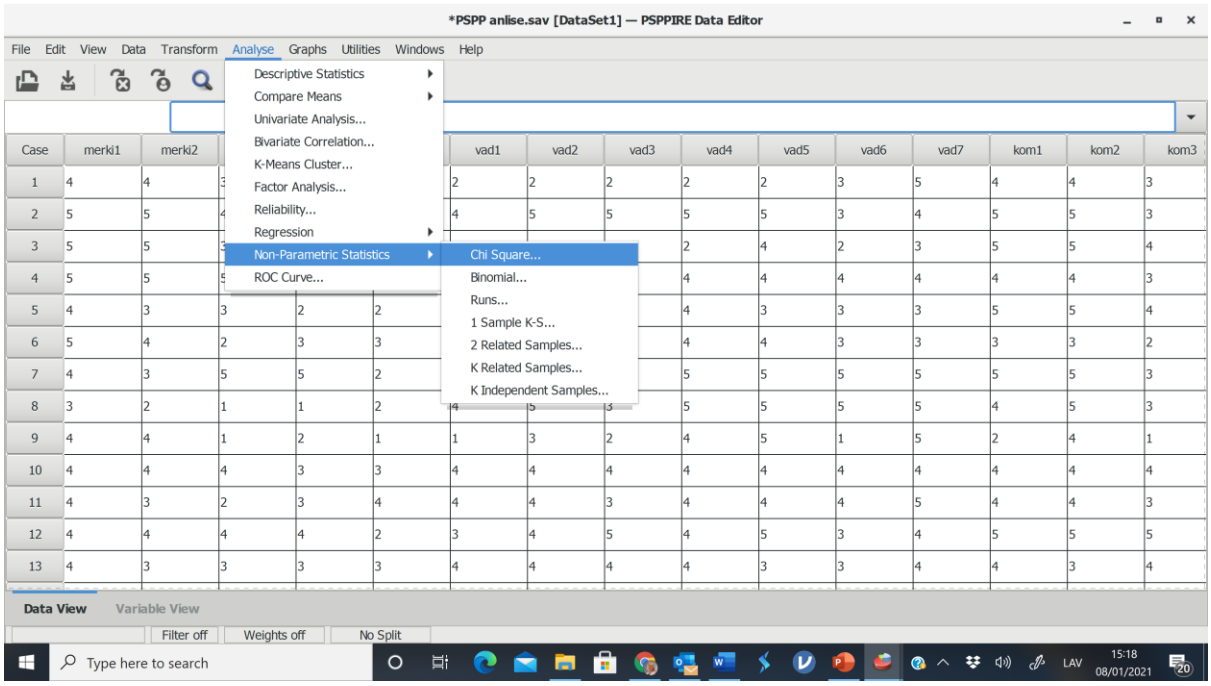


Figure 13.5. Location of the Goodness-of-fit function in PSPP.

In the analysis window, enter the variable “Gender”

By leaving the “All categories equal” option active in the Expected values window, the test will detect deviations from a 50%/50% proportion.

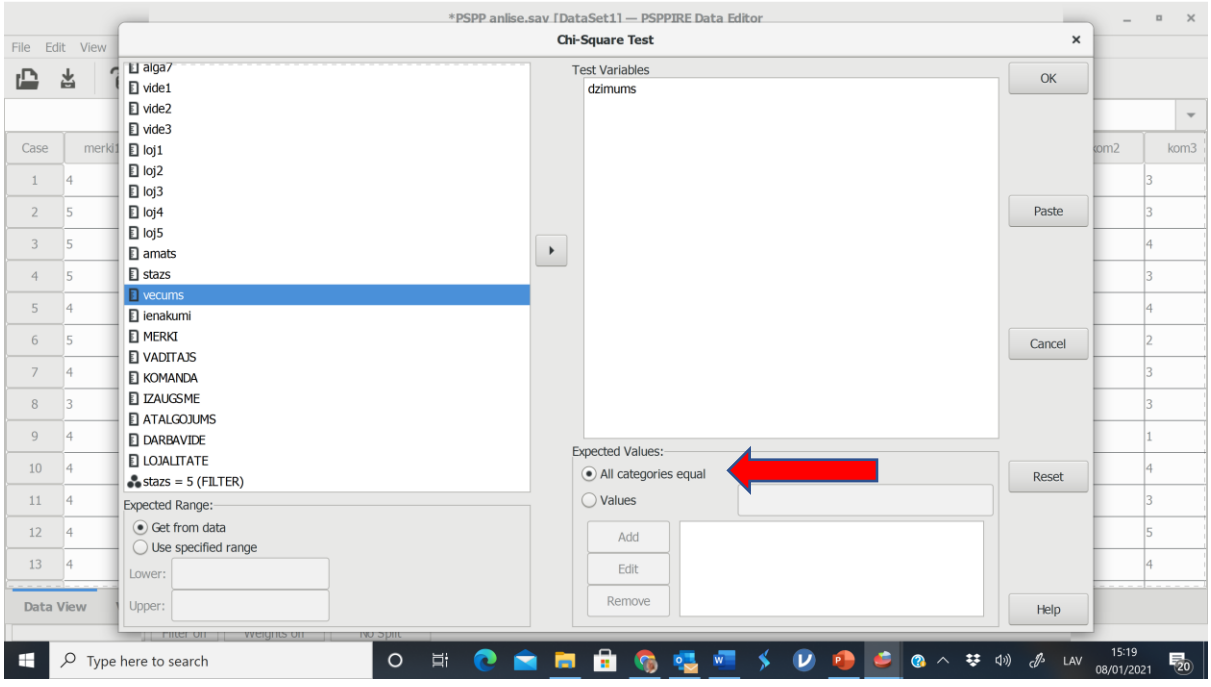


Figure 13.6. Selecting the ‘All categories are equal’ function.

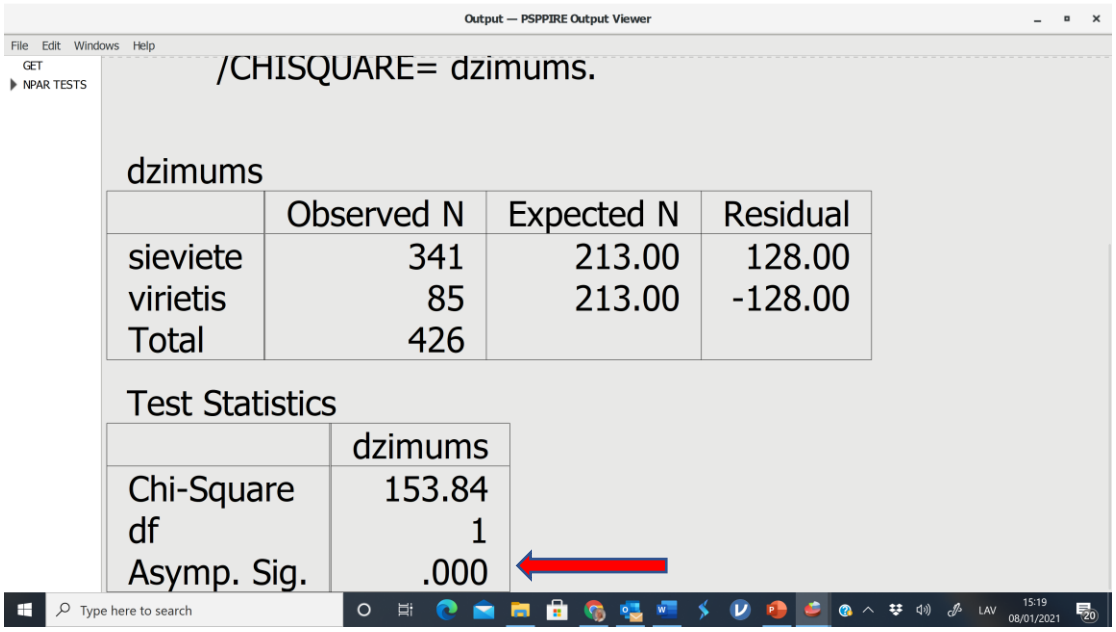


Figure 13.7. Results output.

Second option:

If the gender ratio in the organisation is 80% women and 20% men

In the Expected values log, select “Values” and enter the percentage corresponding to code 1 (women), click Add, then enter the percentage corresponding to code 2 (men), and click Add.

The test will determine the deviation from the 80%/20% ratio.

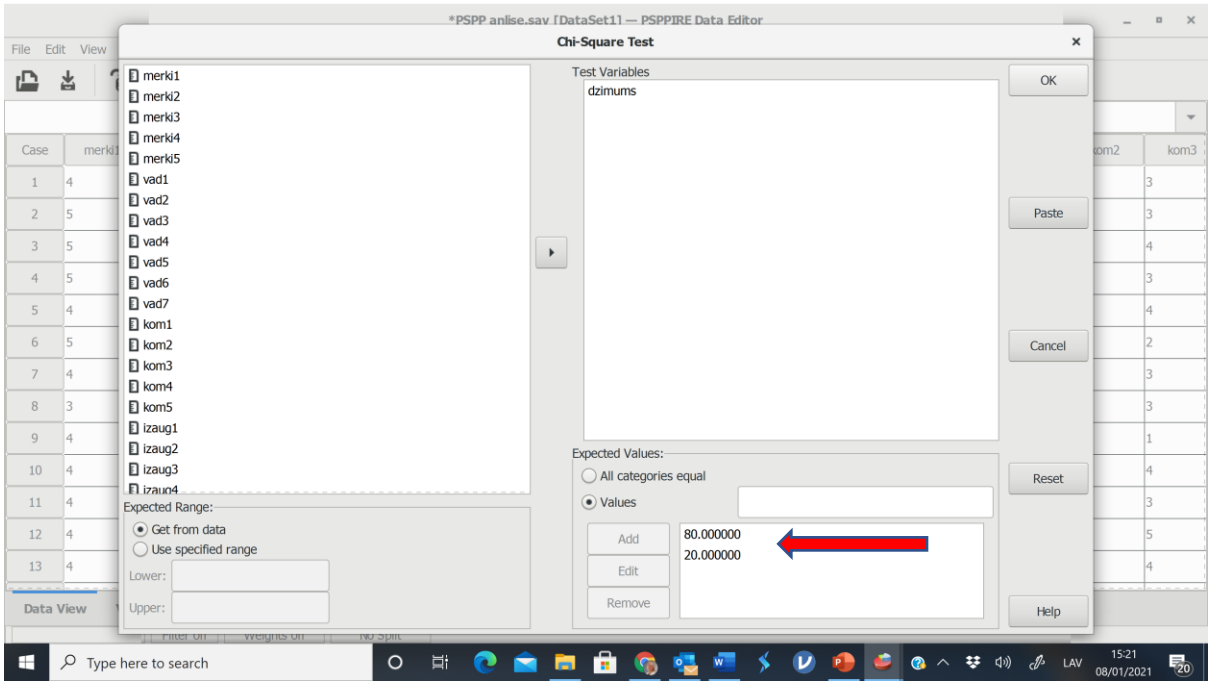


Figure 13.8. Adding proportions.

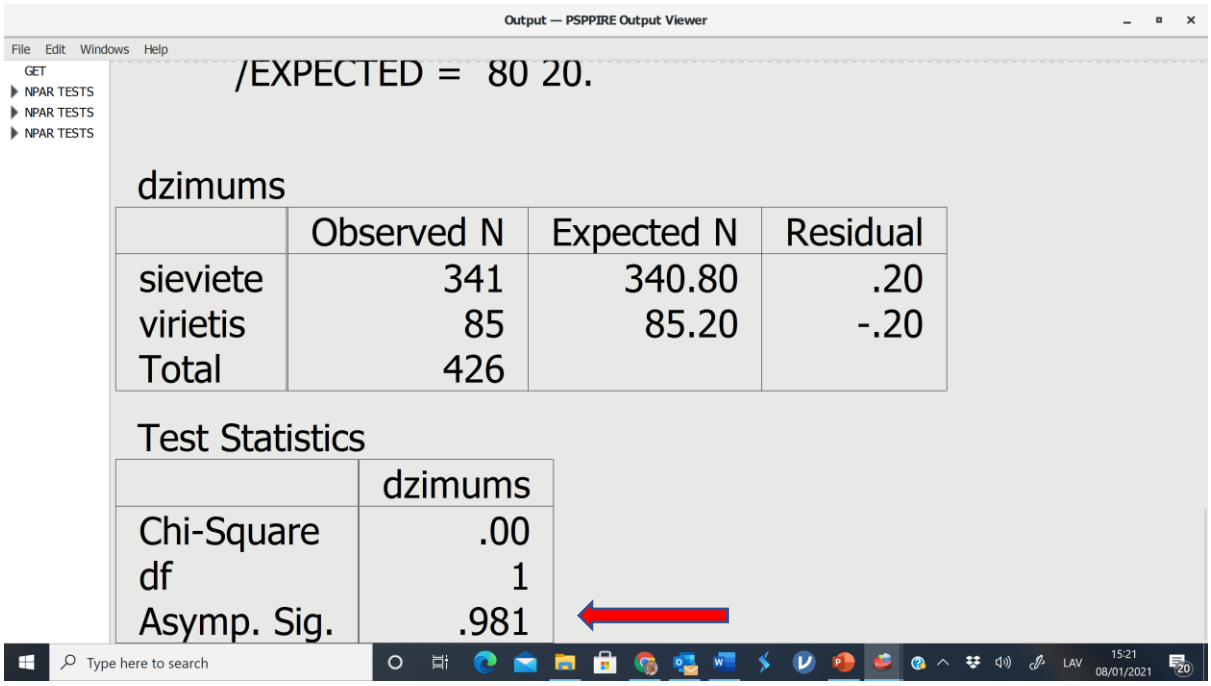


Figure 13.9. Output results.

14.Saving the results of the PSPP analysis

When you have finished, don't forget to save your results. Unlike SPSS, PSPP does not allow you to save the Output as a separate file. You must therefore save the results by exporting them as an OpenDocument or PDF file.

To save the accumulated results of PSPP Output, proceed as follows:

• Open the File menu, click on Export

• In the window that appears, click on 'Infer the type from extensions'

• Select OpenDocument

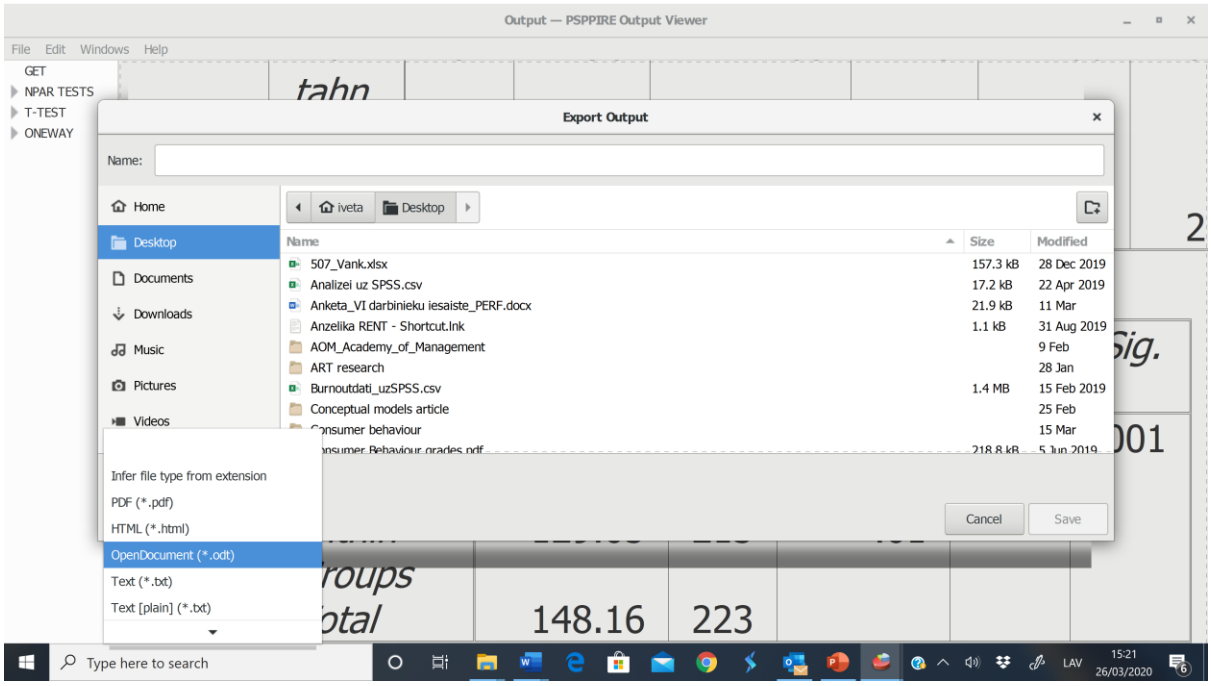


Figure 14.1. Saving results.

This will save all the results of the PSPP analysis in the editable Word document. These results will be required for inclusion in your research report (including your Bachelor's or Master's thesis).

Please note that the saved document will contain absolutely everything you have done, including the commands. Only the necessary information should be included in the research report or Master's thesis, and the tables should be formatted so that they are visually easy to understand.

For example, a correlation table in a research report should look like this:

Table 14.1

Pearson's correlation coefficients between variables

	Environment	*Growth*	*Manager*	*Remuneration*
Environment	1.00			
Growth	*.28***	1.00		
Manager	*.28***	*.46***	1.00	
Remuneration	*.38***	*.52***	*.45***	1.00
Loyalty	*.28***	*.59***	*.48***	*.58***

*** sig<0.000

The results of other tests should be treated in a similar manner.

When including the results of statistical analyses in a Study Paper, the terms must be translated into Latvian. The original PSPP tables and untranslated text may only be used in the appendices.