Learn to Conduct Divergent (Discriminant) Validity in SPSS With Data From Northern Ireland Life and Times Survey (2014)

Learn to Conduct Divergent (Discriminant) Validity in SPSS With Data From Northern Ireland Life and Times Survey (2014)

How-to Guide for IBM® SPSS® Statistics Software

Introduction

In this guide, you will learn how to produce a test for divergent validity in IBM® SPSS® Statistics software (SPSS) using a practical example to illustrate the process. You will find links to the example dataset, and you are encouraged to replicate this example. An additional practice example is suggested at the end of this guide. This example assumes that you have already opened the data file in SPSS.

Contents

1 Divergent Validity

Divergent validity allows researchers to confirm that two separate measures can in fact be regarded as separate measures. In social science research, we often use surveys to measure psychological constructs such as attitudes, beliefs, and perceptions. However, we have to ensure that in doing so, the tools we use to measure such abstract concepts are reliable. This is often done through Factor Analysis, which allows researchers to determine the reliability and validity of their measures. Whilst we determine reliability of measures, we must also ensure they are valid. One of the ways we can validate separate measures as being separate is by establishing that they are in fact measuring two distinct constructs from one another. This is known as divergent validity. Similarly, we determine the validity of a construct itself when distinguishing it from another construct using factor loadings, also in Factor Analysis (please see the SAGE Methods Datasets guide on Factor Loadings for more information).

---

2 An Example in SPSS: Establishing Divergent Validity Between Two Separate Constructs of Attitudes Towards People Suffering With Dementia

This example will demonstrate how to establish divergent validity between two separate constructs of attitudes towards people suffering with dementia. The data are from the 2014 Northern Ireland Life and Times Survey. This extract includes 1,211 respondents. The five variables we examine are:

- C7a “And what about someone who has had dementia for a long time. Do you think in most cases they should continue to live alone?” (dmlalone)
- C7b “And what about someone who has had dementia for a long time. Do you think in most cases they should … continue to manage their own medication?” (dmlmedic)
- C7c “And what about someone who has had dementia for a long time. Do you think in most cases they should … continue to drive?” (dmldrive)
- C9a “How much do you agree or disagree with each of these statements…

caring for someone with dementia is often very lonely.” (carelone)
- C9c “How much do you agree or disagree with each of these statements… caring for someone with dementia often means that your own health suffers.” (carerewa)

Each of the five variables consists of a statement to which respondents may answer via a 5-point Likert scale.

2.1 The SPSS Procedure

Divergent validity is appropriate when certain assumptions have been met. To look for normal distribution, we must first carry out the appropriate analysis for each of the variables. As the variables being used here are ordinal with a Likert scale, it would be appropriate to analyse both frequency tables and use measures of central tendency as a guide to assess the similarities in answers between each variable.

Univariate analysis can be carried out by selecting the following on SPSS:

Analyze → Descriptive Statistics → Frequencies

To carry out univariate analysis, select the variables you wish to analyse and place them in the box. As we are also going to observe how the scales have been answered, we also need to then select “statistics” to ensure we have the information highlighted in Figure 1. A histogram is also useful as it allows us to visualise the distribution. This is done by selecting “charts” and “histogram” along with “show normal curve on histogram” as shown in Figure 2.

Figure 1: Measures of Central Tendency.

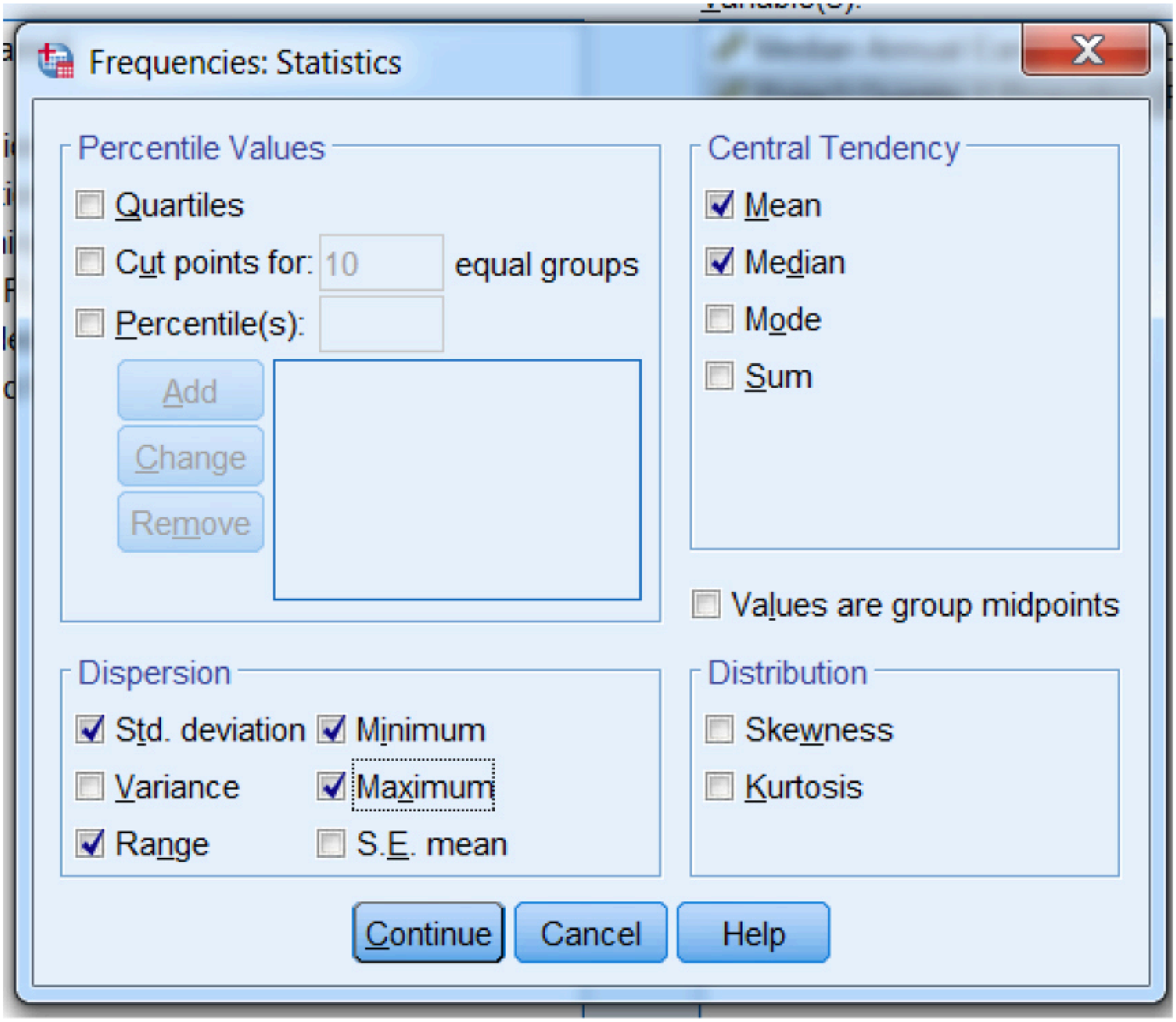
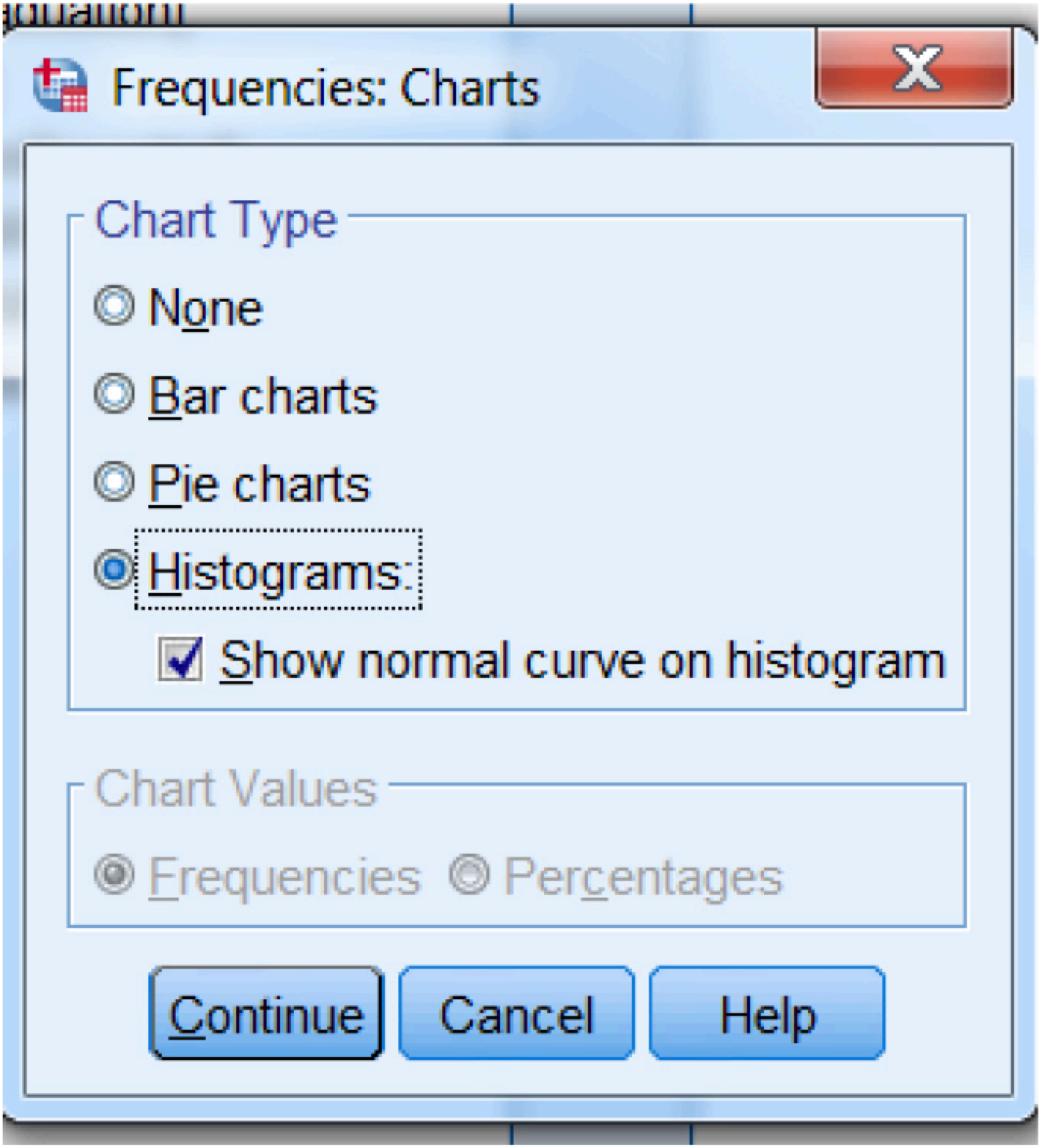


Figure 2: Selecting a Histogram.



Tables 1 and 2 provide the frequency distribution for each of the five variables. Figures 3 and 4 indicate that the Likert scales are slightly skewed. However, C7 variables are negatively skewed and C9 variables are positively skewed. Looking at the questions and wording of answers, it seems that the skews both reflect positive answers such as “Strongly agree” or “Definitely.”

Table 1: C7 Frequency Distribution.

Answer	Frequency (valid percent)		
	C7a	**C7b**	**C7c**
Definitely	11 (.9)	9 (.8)	9 (.8)
Probably	26 (2.1)	23 (2)	11 (.9)
It depends	210 (17.3)	168 (14.4)	133 (11.4)
Probably not	309 (25.5)	273 (23.4)	232 (19.9)
Definitely not	610 (50.4)	693 (59.4)	778 (66.9)
Total (*N*)	1,166 (100)	1,166 (100)	1,163 (100)

Figure 3: Histogram of C7a.

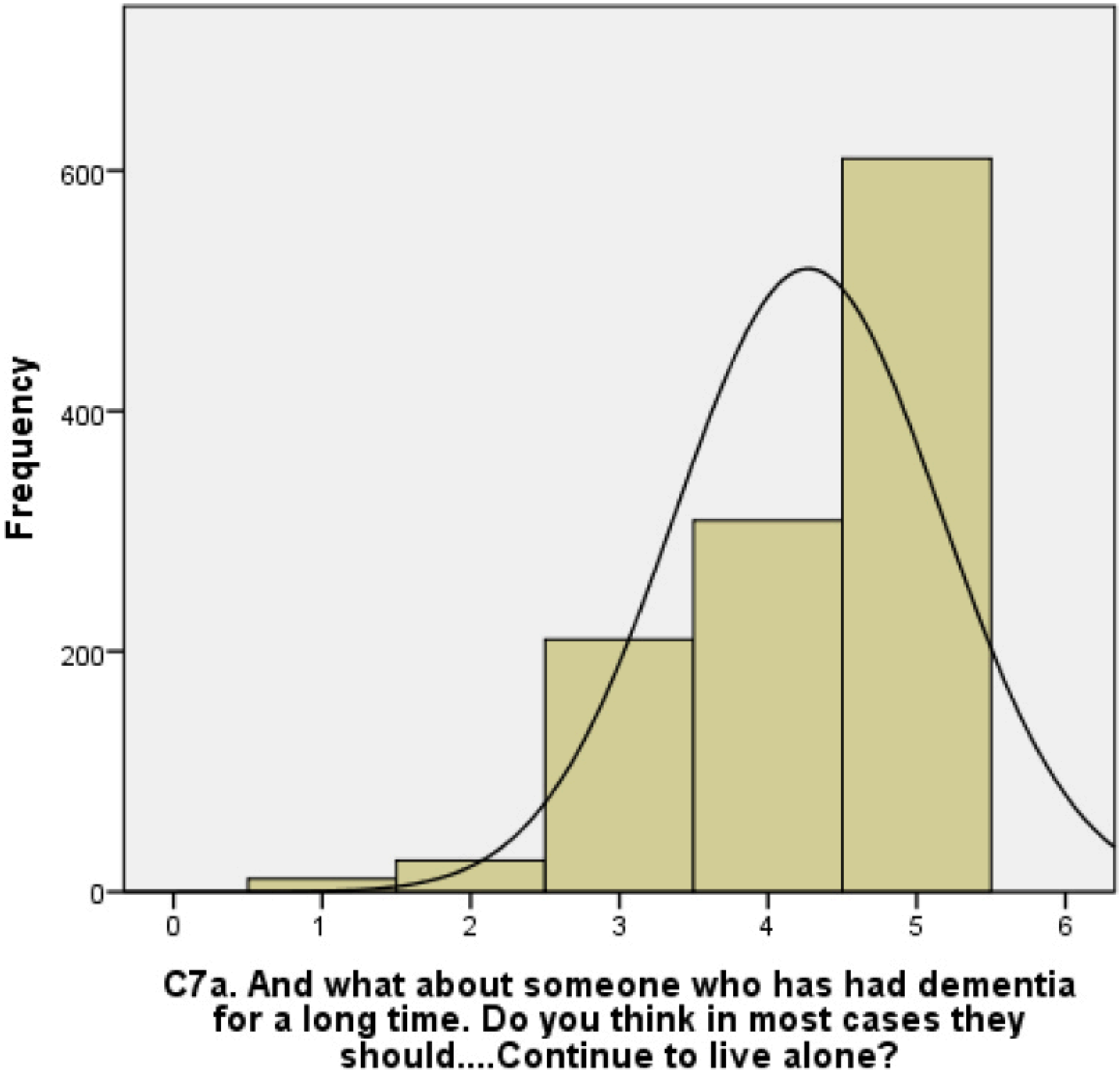
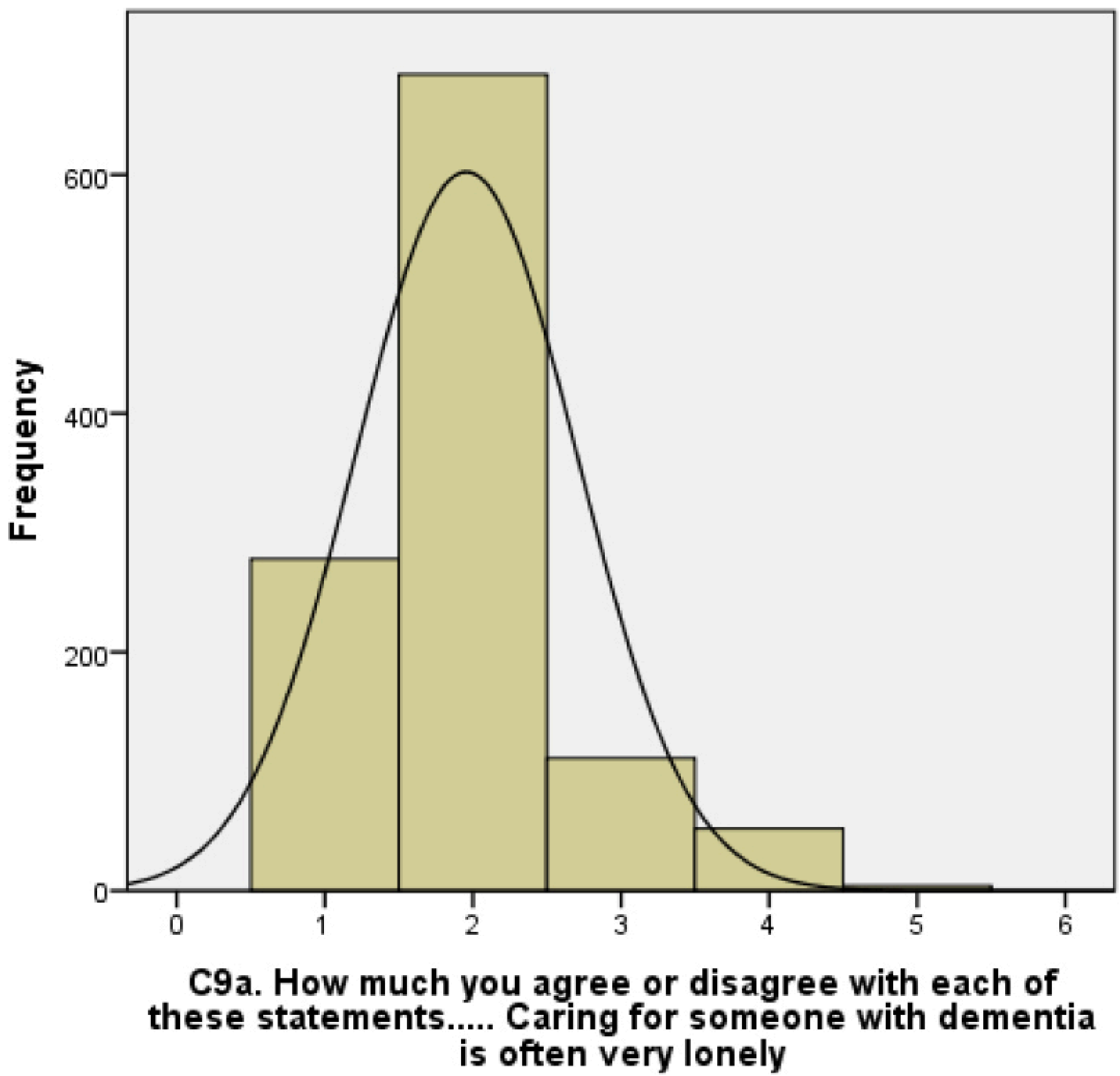


Table 2: C9 Frequency Distribution.

Answer	Frequency (valid percent)	
	C9a	C9c
Strongly agree	278 (24.6)	251 (22.3)
Agree	684 (60.6)	605 (53.7)

Neither	111 (9.8)	167 (14.8)
Disagree	52 (4.6)	96 (8.5)
Strongly disagree	4 (.4)	8 (.7)
Total (N)	1,129 (100)	1,127 (100)

Figure 4: Histogram of C9a.



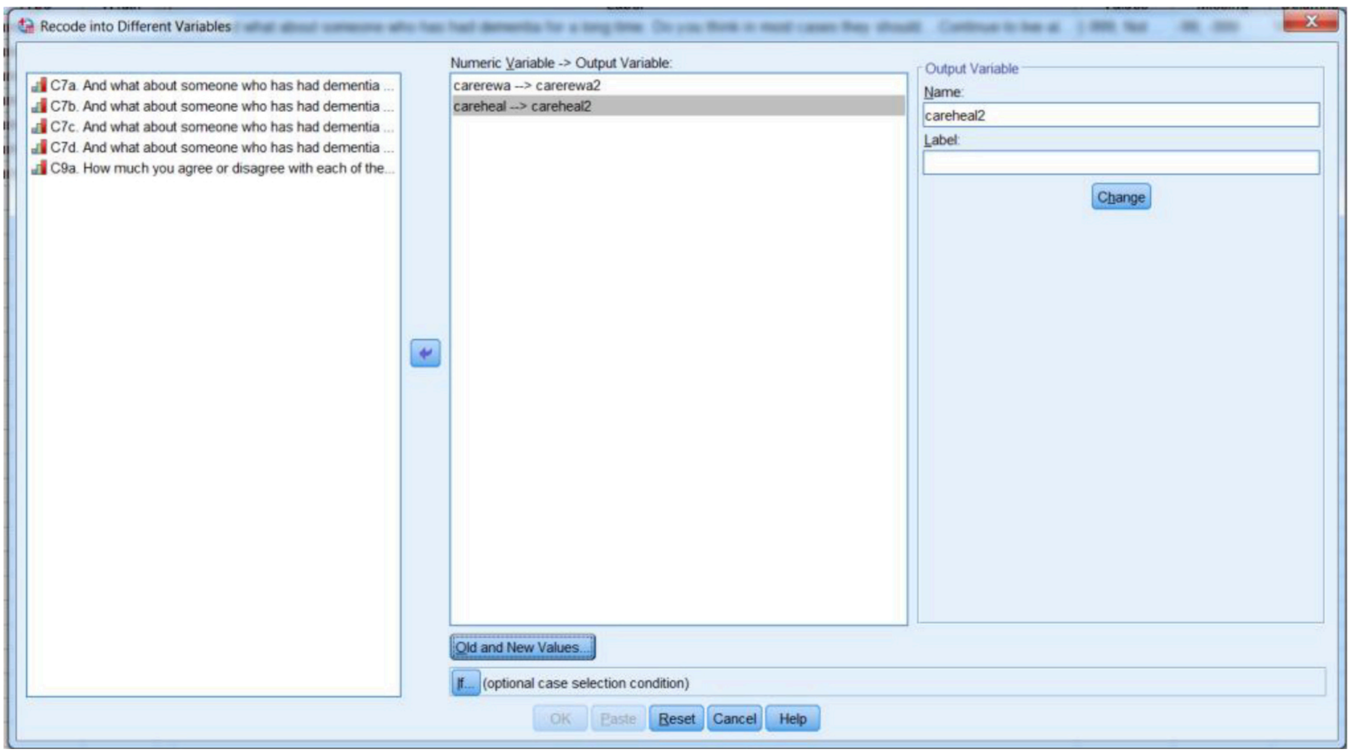
Before carrying out Factor Analysis, it would be more appropriate to recode the C9 variables, so that the numbering of the Likert scale reflects similar answers to the C7 variables, where a higher score reflects a more negative answer.

To recode an existing variable into a new variable, select the following on SPSS:

Transform → Recode into Different Variables

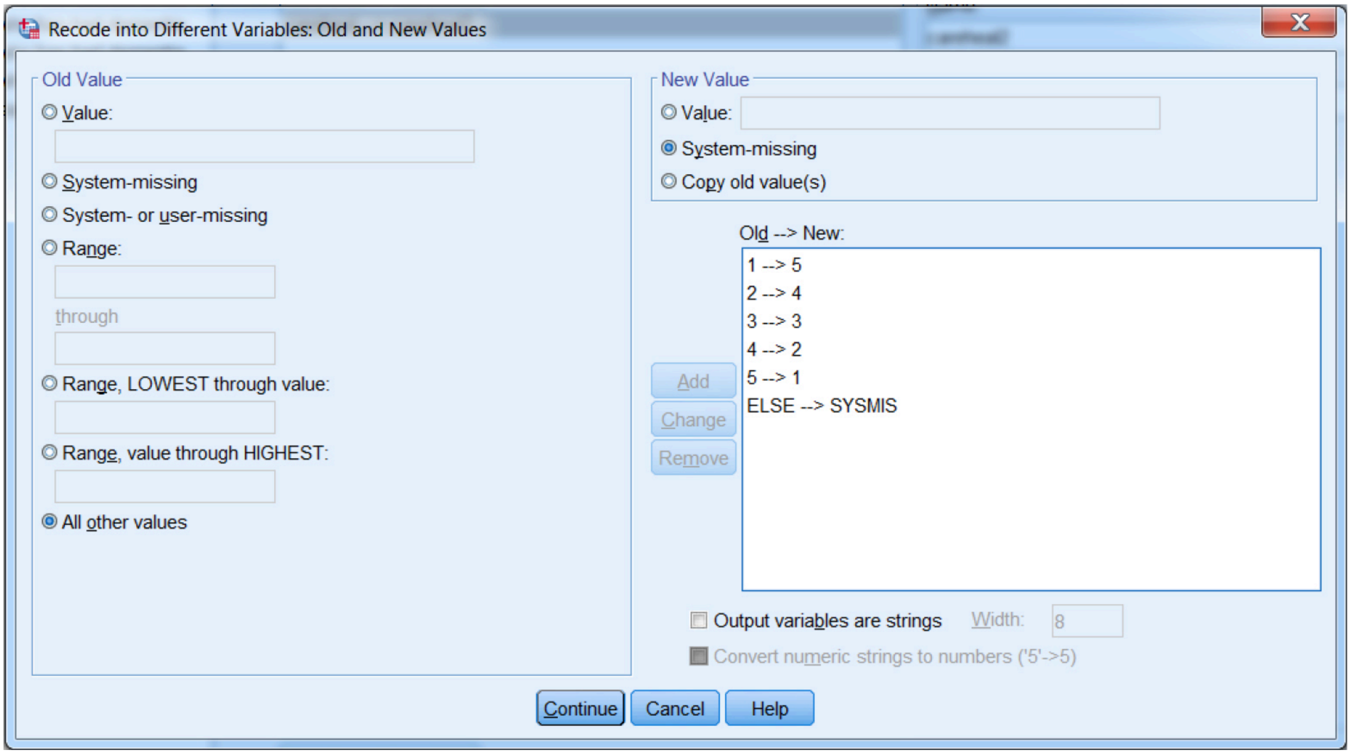
Place the two variables that need to be recoded in the “Numeric Variable > Output Variable” box. Simply rename the variables by providing them with the same name and adding “2.” Figure 5 shows what this looks like in SPSS.

Figure 5: Selecting Variables for Recoding.



Click the “Old and New Values” box and then enter 1 into the old value box and 5 into the new value box. Figure 6 shows what the reversed coding should look like in SPSS.

Figure 6: Reverse Coding.



Your new variable is now ready to be computed. Simply click **Continue**, and in the "Recode into different variable" dialogue box click **OK**. You should find your new variable at the bottom of the list in the Variable View window. Ensure that you input the new values to reflect the correct answer. Figure 7 shows what this looks like in SPSS.

Figure 7: Selecting Labels for New Values.

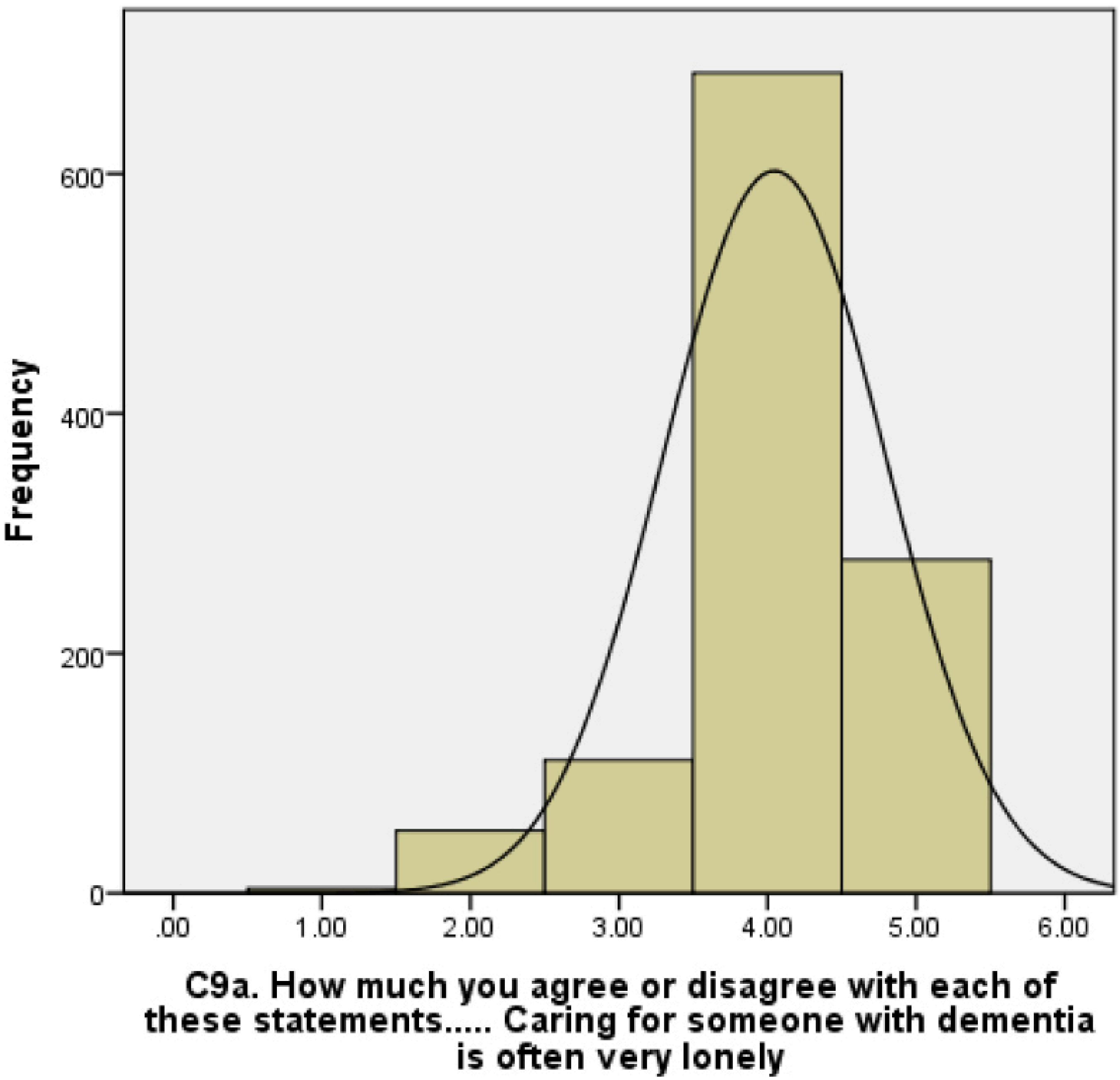
Table 3 provides the new frequency distribution for the two recoded variables. Figure 8 also shows how the skew matches the skew of the C7 variables.

Table 3: Frequency Distribution for Recoded Variables.

Answer	Frequency (valid percent)	
	C9a recoded	C9c recoded
Strongly disagree	4 (.4)	8 (.7)
Disagree	52 (4.6)	96 (8.5)
Neither	111 (9.8)	167 (14.8)
Agree	684 (60.6)	605 (53.7)
Strongly agree	278 (24.6)	251 (22.3)

Total (*N*)	1,119 (100)	1,127 (100)

Figure 8: Histogram of C9a.



We can now look at the measures of central tendency to see how the range of responses for all five variables compare with one another. Table 4 provides the results.

Table 4: Measures of Central Tendency for all Five Variables.

Variable	C7a	C7b	C7c	C9a	C9c
Mean	4.27	4.39	4.51	4.05	3.88
Standard deviation	.897	.859	.793	.748	.873
Range	4	4	4	4	4
Minimum	1	1	1	1	1
Maximum	1	1	1	1	1

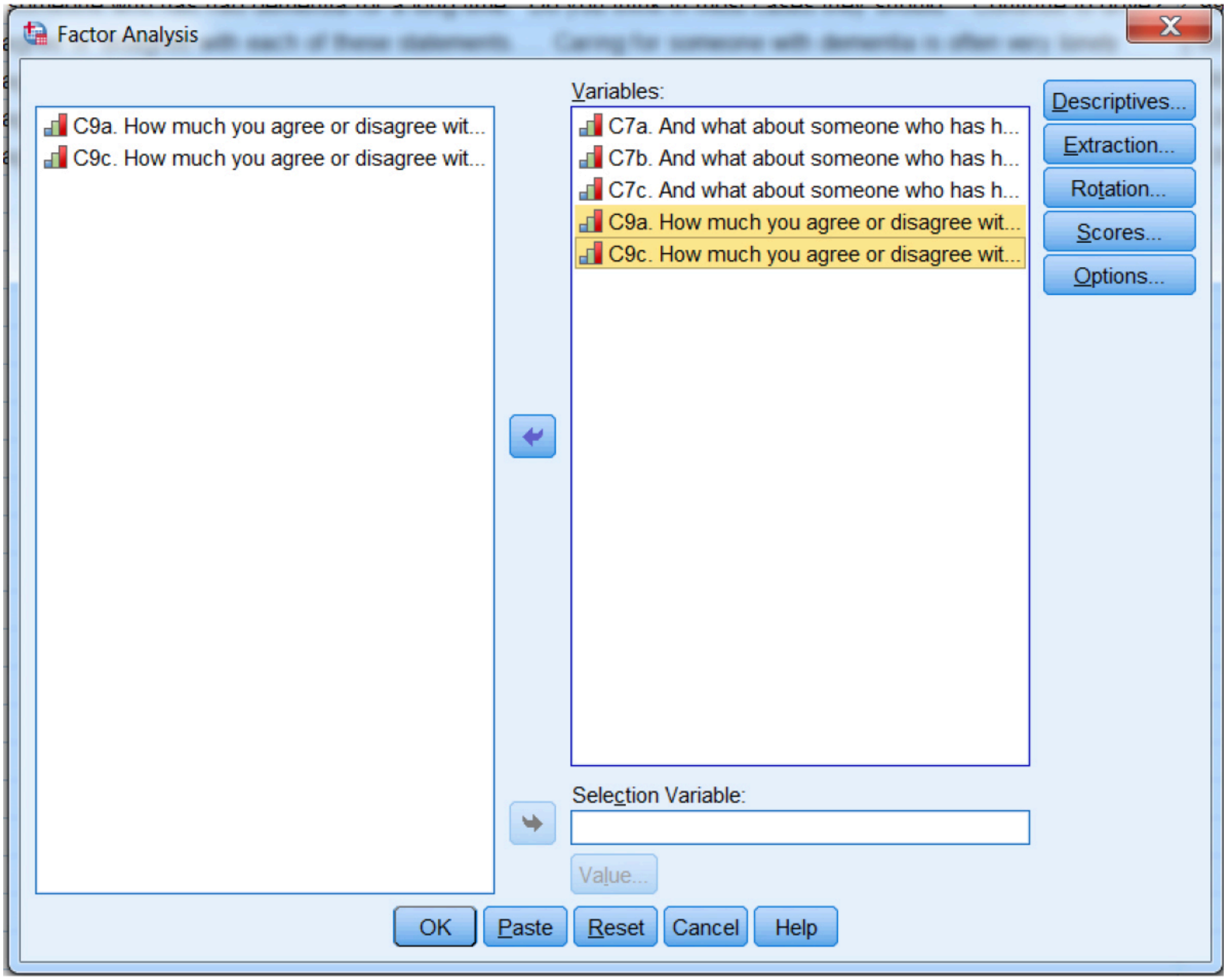
Table 4 shows that there is a similar mean of responses for each of the variables all sharing a 1–5 scale. All five variables indicate a negative skew with the majority of responses being toward the higher end of the scale indicating a more negative response rate when answering questions regarding people suffering with dementia.

To check the assumptions before testing for divergent validity, we do so by running the Factor Analysis. This is done by selecting the following on SPSS:

Analyze → Dimension Reduction → Factor Analysis

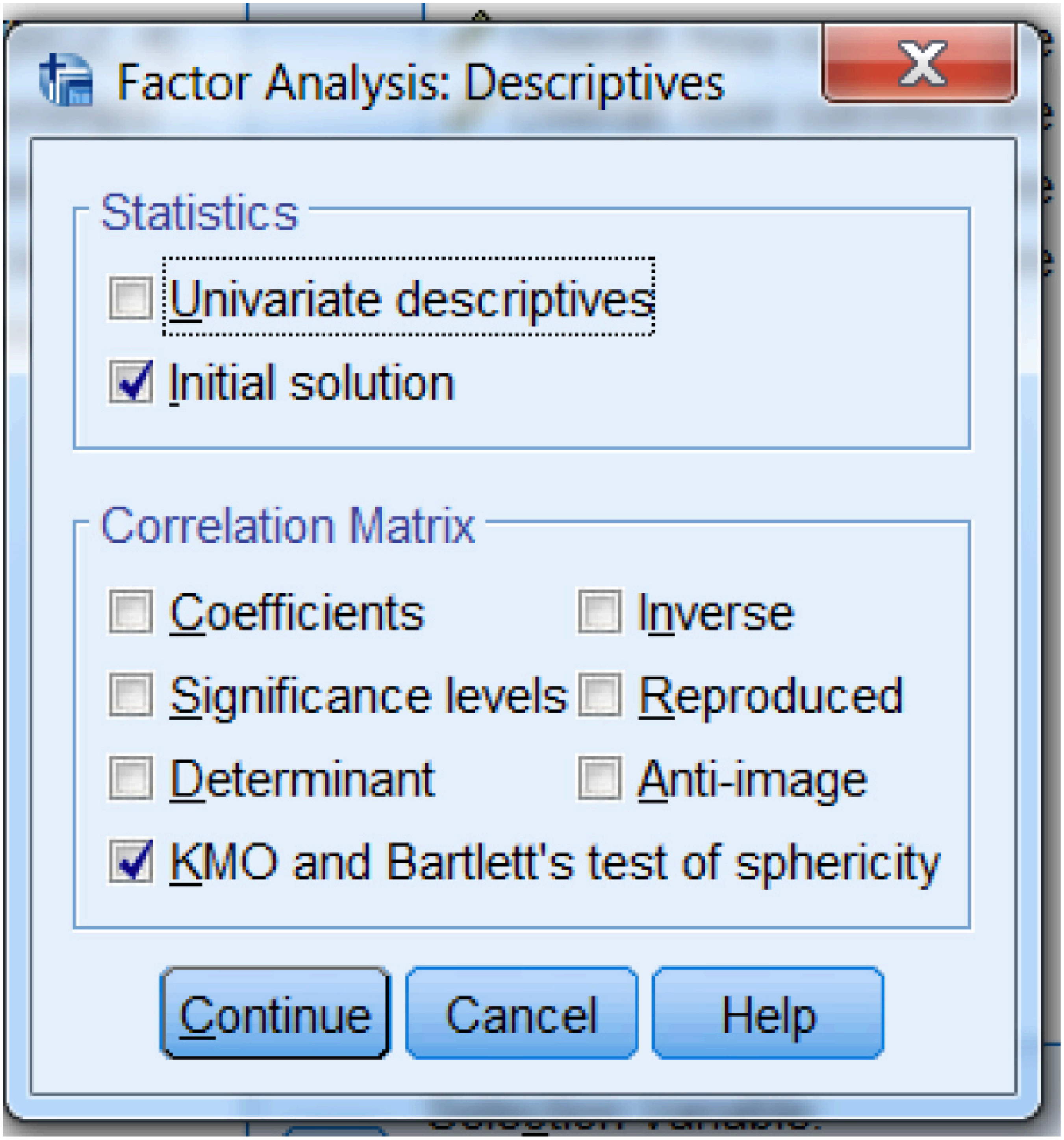
You can now add your variables to the variables box. Make sure you place the recoded variables in the box and not the original variables. Figure 9 shows what this looks like in SPSS.

Figure 9: Factor Analysis Dialog Box.



We must now select the "Descriptives" dialog box to ensure we test the assumptions required before examining the factor loadings. We must therefore select "Coefficients," and "KMO and Bartlett's Test of Sphericity." Figure 10 shows what this looks like in SPSS.

Figure 10: Selecting the Required Assumptions.



We must then select our extraction method by opening the “extraction” dialog box. As we are conducting a Factor Analysis, we must change our extraction method to principal analysis factoring. We will also extract eigenvalues greater than 0.7. Figure 11 shows what this looks like in SPSS.

We must then finally select our rotation method. In order to identify which method of rotation is most appropriate, we must first run a Direct Oblimin Rotation, which assumes the factors are correlated. Figure 12 shows what this looks like in SPSS.

Figure 11: Selecting Extraction Method.

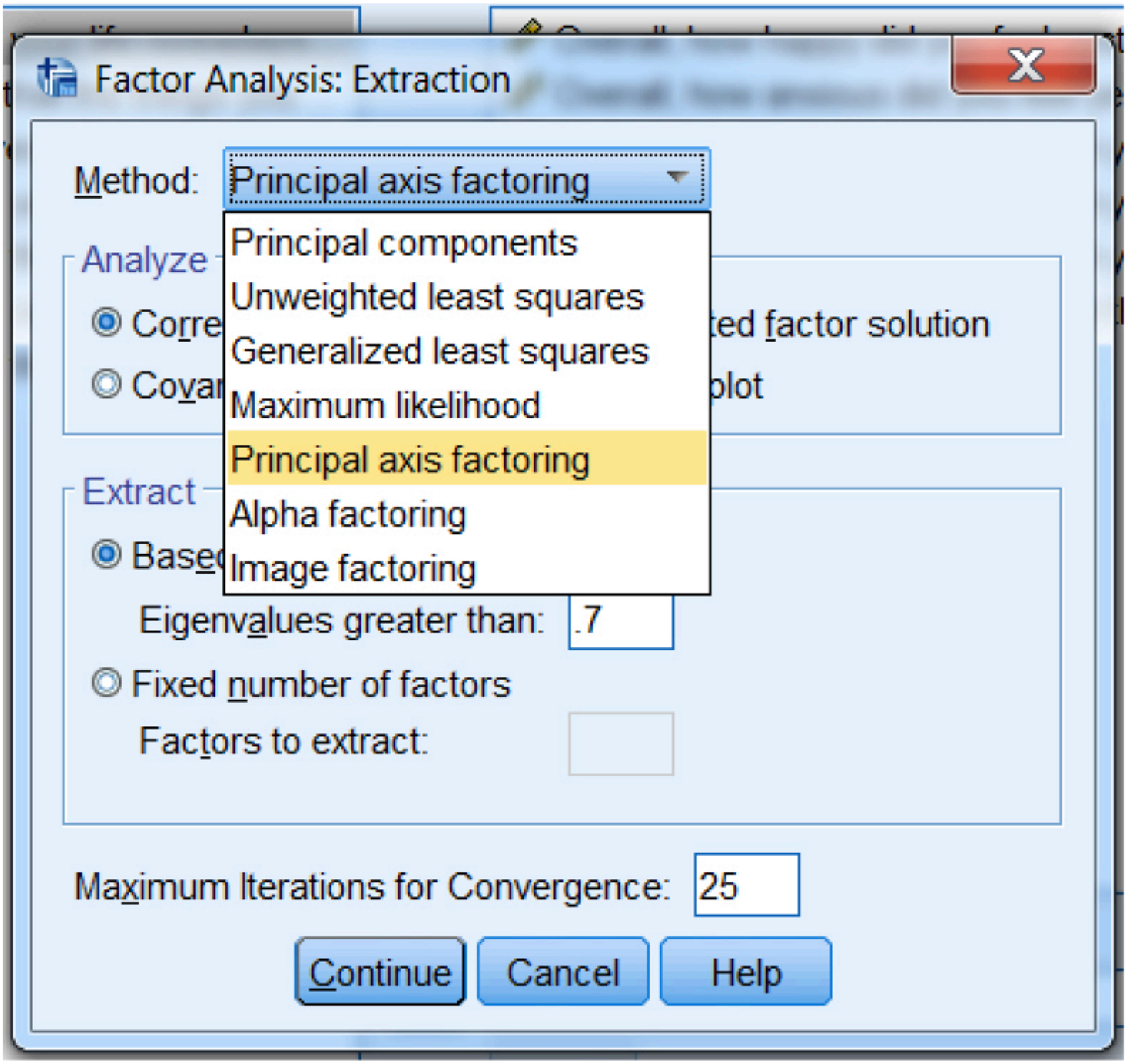
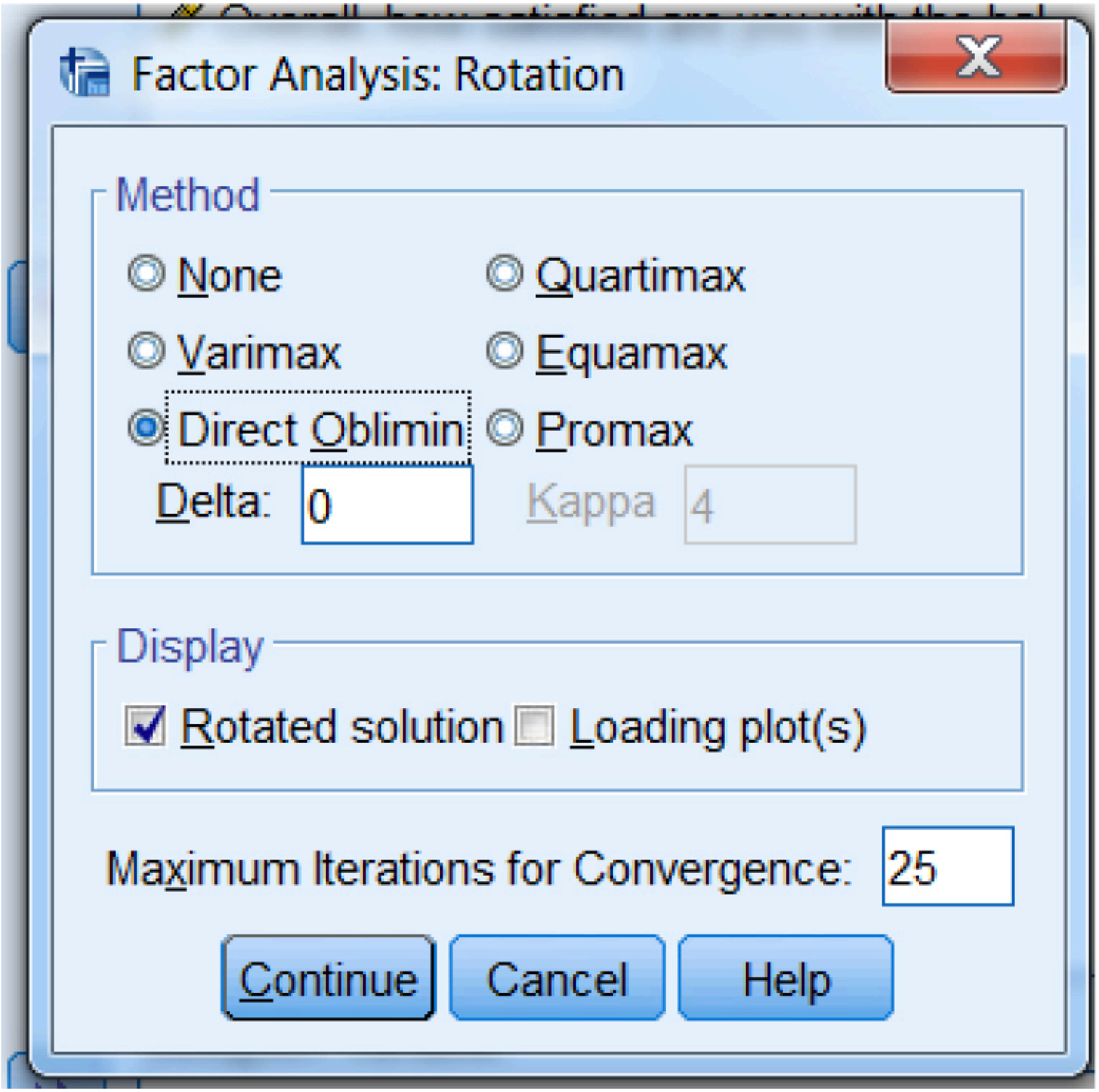
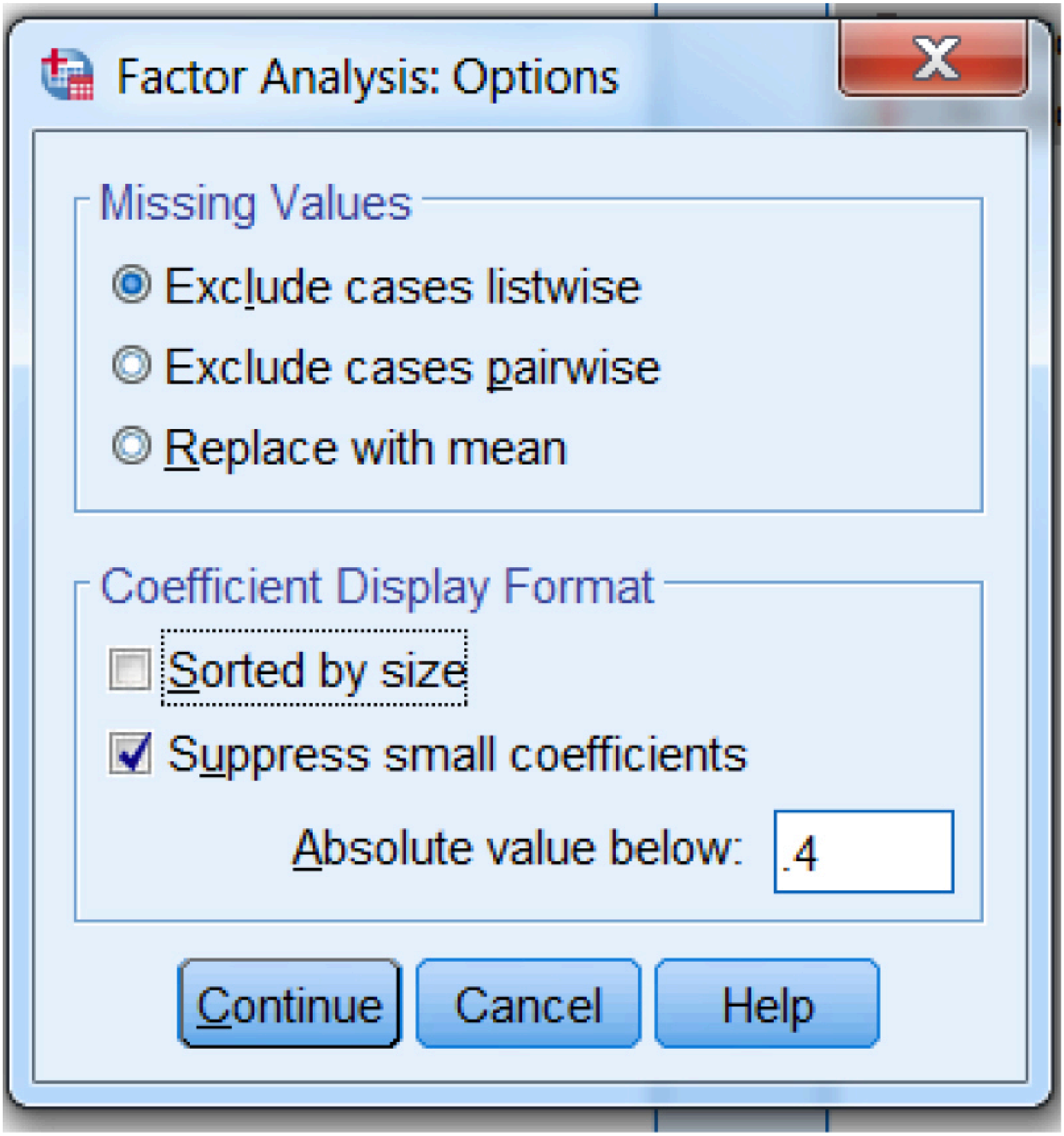


Figure 12: Selecting Rotation Method.



We can also suppress coefficients that indicate a weak relationship between the item and the factor. This is done by selecting "options" followed by "suppress small coefficients" and entering .4. Figure 13 shows what this looks like in SPSS.

Figure 13: Suppressing Small Coefficients.



Once you have clicked **Continue**, you are ready to carry out your Factor Analysis. Click **OK**.

2.2 Exploring the SPSS Output

Step 1: Examine the tests for assumptions. This includes: Bartlett's test for sphericity and

Kaiser–Meyer–Olkin statistics

The KMO value of .7 indicates that the data are adequate for Factor Analysis. The Bartlett's test of Sphericity indicates significance ($p < .001$) and provides evidence to suggest the data are normally distributed.

Step 2: Examine the factor loadings and component correlation coefficient

Table 5 provides the component correlation coefficient for the two factors. This can be found at the bottom of the input in SPSS known as the "Factor Correlation Matrix."

Table 5: Factor Correlation Matrix.

Component correlation coefficient	.17

Table 5 revealed the component correlation coefficient between factors to be very low at .17. as this is very low and the rotation method assumes factors are correlated, we must change the rotation method to Varimax. This can be done by selecting varimax from the 'Rotation' tab in the Factor Analysis Dialog box (Figure 9).

Table 6: Factor Loadings From Varimax Rotation.

Item	Component 1	Component 2
C7a	.89	
C7b	.988	
C7c	.871	
C9a		.727
C9c		.724
Component correlation coefficient	−.201	

Step 3: Calculate the average loading per factor and identify convergent validity

Table 7 provides calculations of the average loading per factor, indicating convergent validity.

Table 7: Calculations for Establishing Divergent Validity.

	Average loading	Variance extracted
C7	.916	.840
C9	.726	.526
Between C7 and C9	NA	.683
Correlation		−.201
Correlation squared		.04

Step 4: Calculate the variance extracted per factor and between factors

Table 3 also provides the results of the calculations of the variance extracted for factors and between factors.

- Average loading is computed by adding the loadings and dividing by the number of items.
- The variance extracted for each factor is the average loading squared.
- The variance extracted between factors is the sum total variance extracted divided by the number of factors.

Step 5: Calculate the component correlation coefficient squared

Table 7 also provides the calculation of the component correlation coefficient squared.

Step 6: Identify whether the variance extracted between components is higher than the component correlation coefficient squared

The variance extracted between both components is .683, higher than the calculated correlation squared of .04; divergent validity is therefore established. This is because we have established that the square root of every average variance extracted (in this case 2) is higher than the correlation amongst the pair of latent constructs (the two separate extracted factors). Therefore, the variance within each construct is explained by the items rather than the other construct. Because of running Factor Analysis and calculating the information in Table 7, divergent validity has been established, and there is evidence to suggest that the five variables measure two distinct, reliable measures of attitudes to people suffering with dementia.

---

3 Your Turn

You can download this sample dataset along with a guide showing how to produce a Factor Analysis and determine divergent validity using statistical software. See whether you can reproduce the results presented here whilst also adding the two additional variables (dmlelect and carerewa) in the dataset.
