

Reliability Analysis

Measures of Reliability

Reliability: the fact that a scale should consistently reflect the construct it is measuring.

One way to think of reliability is that other things being equal, a person should get the same score on a questionnaire if they complete it at two different points in time (test-retest reliability). Another way to look at reliability is to say that two people who are the same in terms of the construct being measured, should get the same score. In statistical terms, the usual way to look at reliability is based on the idea that individual items (or sets of items) should produce results consistent with the overall questionnaire.

The simplest way to do this in practice is to use split half reliability. This method randomly splits the data set into two. A score for each participant is then calculated based on each half of the scale. If a scale is very reliable a person's score on one half of the scale should be the same (or similar) to their score on the other half: therefore, across several participants scores from the two halves of the questionnaire should correlate perfectly (well, very highly). The correlation between the two halves is the statistic computed in the split half method, with large correlations being a sign of reliability. The problem with this method is that there are several ways in which a set of data can be split into two and so the results could be a product of the way in which the data were split. To overcome this problem, Cronbach (1951) came up with a measure that is loosely equivalent to splitting data in two in every possible way and computing the correlation coefficient for each split. The average of these values is equivalent to **Cronbach's alpha**, α , which is the most common measure of scale reliability (This is a convenient way to think of Cronbach's alpha but see Field, 2005, for a more technically correct explanation).

There are two versions of alpha: the normal and the standardized versions. The normal alpha is appropriate when items on a scale are summed to produce a single score for that scale (the standardized α is not appropriate in these cases). The standardized alpha is useful though when items on a scale are standardized before being summed.

Interpreting Cronbach's α (some cautionary tales ...)

You'll often see in books, journal articles, or be told by people that a value of 0.7-0.8 is an acceptable value for Cronbach's alpha; values substantially lower indicate an unreliable scale. Kline (1999) notes that although the generally accepted value of 0.8 is appropriate for cognitive tests such as intelligence tests, for ability tests a cut-off point of 0.7 is more suitable. He goes on to say that when dealing with psychological constructs values below even 0.7 can, realistically, be expected because of the diversity of the constructs being measured.

However, Cortina (1993) notes that such general guidelines need to be used with caution because the value of alpha depends on the number of items on the scale (see Field, 2005 for details).

Alpha is also affected by reverse scored items. For example, in our **SAQ** from last week we had one item (question 3) that was phrased the opposite way around to all other items. The item was 'standard deviations excite me'. Compare this to any other item and you'll see it requires the opposite response. For example, item 1 is 'statistics make me cry'. Now, if you don't like statistics then you'll strongly agree with this statement and so will get a score of 5 on our scale. For item 3, if you hate statistics then standard deviations are unlikely to excite you so you'll strongly disagree and get a score of 1 on the scale. These reverse phrased items are important for reducing response bias) participants will actually have to read the items in case they are phrased the other way around. In reliability analysis these reverse scored items make a difference: in the extreme they can lead to a negative Cronbach's alpha! (see Field, 2005 for more detail).

Therefore, if you have reverse phrased items then you have to also reverse the way in which they're scored before you conduct reliability analysis. This is quite easy. To take our SAQ data, we have one item which is currently scored as 1 = strongly disagree, 2 = disagree, 3 = neither, 4 = agree, and 5 = strongly agree. This is fine for items phrased in such a way that agreement indicates statistics anxiety, but for item 3 (standard deviations excite me), disagreement indicates statistics anxiety. To reflect this numerically, we need to reverse the scale such that 1 = strongly agree, 2 = agree, 3 = neither, 4 = disagree, and 5 = strongly disagree. This way, an anxious person still gets 5 on this item (because they'd strongly disagree with it).

To reverse the scoring find the maximum value of your response scale (in this case 5) and add one to it (so you get 6 in this case). Then for each person, you take this value and subtract from it the score they actually got. Therefore, someone who scored 5 originally now scores $6-5 = 1$, and someone who scored 1 originally now gets $6-1 = 5$. Someone in the middle of the scale with a score of 3, will still get $6-3 = 3$! Obviously it would take a long time to do this for each person, but we can get SPSS to do it for us by using **Transform**⇒**Compute...** (see your handout on Exploring data).




- ✓ Load the **SAQ.sav** data from last week.
- ✓ Using what you know about the Compute command (use your handout on exploring data for help if you need to), reverse score question 3.

Reliability Analysis on SPSS

Let's test the reliability of the SAQ using the data in **SAQ.sav**. Now, you should have reverse scored item 3 (see above). Remember also that I said we should conduct reliability analysis on any subscales individually. If we use the results from our orthogonal rotation (look back at your handout on Factor Analysis), then we have 4 subscales:

1. Subscale 1 (Fear of computers): items 6, 7, 10, 13, 14, 15, 18
2. Subscale 2 (Fear of statistics): items 1, 3, 4, 5, 12, 14, 10, 21
3. Subscale 3 (Fear of mathematics): items 8, 11, 17
4. Subscale 4 (Peer evaluation): items 2, 9, 19, 22, 23

To conduct each reliability analysis on these data you need to follow the **Analyze**⇒**Scale**⇒**Reliability Analysis ...** menu path to display the dialog box in Figure 1. Select any items from the list that you want to analyze (to begin with let's do the items from the fear of computers subscale) on the left hand side of the dialog box and transfer them to the box labelled Items by clicking on .

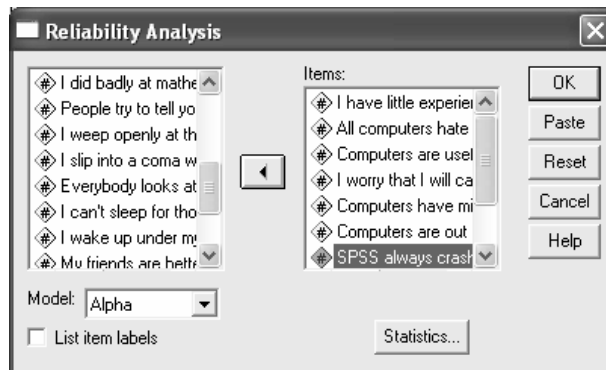


Figure 1

Selecting the *List item labels* checkbox will list all of the variable labels for each variable (which can be useful for checking to which items your variables relate). There are several reliability analyses you can run, but the default option is Cronbach's alpha, which is the one we want.

If you click on **Statistics...** you can access the dialog box in Figure 2. In the statistics dialog box you can select several things, but the one most important for questionnaire reliability is: *Scale if item deleted*. This option provides a value of Cronbach's alpha for each item on your scale. It tells us what the value of alpha would be if that item were deleted. If our questionnaire is reliable then we would not expect any one item to greatly affect the overall reliability. In other words, no item should cause a substantial decrease in alpha. If it does then we have serious cause for concern and you should consider dropping that item from the questionnaire. As 0.8 is seen as a good value for alpha, we would hope that all values of *alpha if item deleted* should be around 0.8 or higher.

Use the simple set of options in Figure 2, to run a basic reliability analysis. Click on **Continue** to return to the main dialog box and then click **OK** to run the analysis.

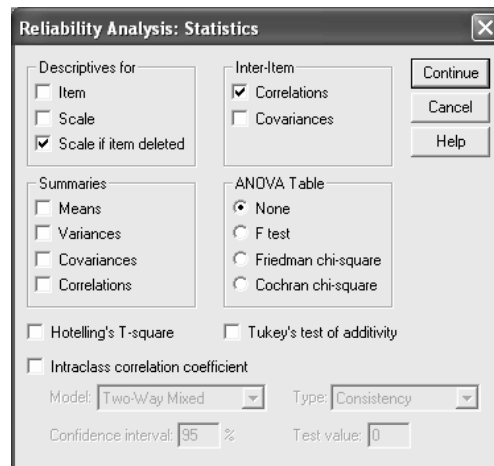


Figure 2

Interpreting the Output

SPSS Output 1 shows the results of this basic reliability analysis for the fear of computing subscale. The values in the column labelled *Corrected Item-Total Correlation* are the correlations between each item and the total score from the questionnaire. In a reliable scale all items should correlate with the total. So, we're looking for items that don't correlate with the overall score from the scale: if any of these values are less than about .3 (depends slightly on your sample size—with bigger samples smaller correlation coefficients are acceptable) then we've got problems because it means that a particular item does not correlate very well with the scale overall. Items with low correlations may have to be dropped. For these data, all data have item-total correlations above .3, which is encouraging.

The values in the column labelled *Alpha if Item is Deleted* are the values of the overall alpha if that item isn't included in the calculation. As such, they reflect the change in Cronbach's alpha that would be seen if a particular item were deleted. The overall alpha is .823, and so all values in this column should be around that same value. We're looking for values of alpha greater than the overall alpha because if the deletion of an item increases Cronbach's alpha then this means that the deletion of that item improves reliability. None of the items here would substantially affect reliability if they were deleted. The worst offender is question 10: deleting this question would increase the alpha from .823 to .824. Nevertheless this increase is not dramatic and both values reflect a reasonable degree of reliability.

Finally, and perhaps most important, the value of *Alpha* at the very bottom is The Cronbach's alpha: the overall reliability of the scale. To re-iterate we're looking for values in the magnitude of .7 to .8 (or there about) bearing in mind what we've already noted about effects from the number of items. In this case alpha is slightly above .8, and is certainly in the region indicated by Kline, so this probably indicates good reliability. As a final point, it's worth noting that if items do need to be removed at this stage then you should re-run your factor analysis as well to make sure that the deletion of the item has not affected the factor structure!

R E L I A B I L I T Y A N A L Y S I S - S C A L E (A L P H A)								
Correlation Matrix								
	Q06	Q07	Q10	Q13	Q14	Q15	Q18	
Q06	1.0000							
Q07	.5136	1.0000						
Q10	.3222	.2837	1.0000					
Q13	.4664	.4421	.3020	1.0000				
Q14	.4022	.4407	.2547	.4498	1.0000			
Q15	.3599	.3914	.2952	.3422	.3801	1.0000		
Q18	.5133	.5009	.2925	.5329	.4983	.3429	1.0000	
N of Cases =		2571.0						
Item-total Statistics								
	Scale Mean if Deleted	Scale Variance if Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Alpha if Deleted			
Q06	15.8650	17.6141	.6187	.3981	.7906			
Q07	15.1684	17.7370	.6190	.3949	.7905			
Q10	15.8114	20.7360	.3999	.1665	.8239			
Q13	15.6429	18.8086	.6067	.3844	.7937			
Q14	15.2159	18.7188	.5768	.3504	.7980			
Q15	15.3259	19.3217	.4913	.2497	.8119			
Q18	15.5235	17.8324	.6474	.4475	.7855			
Reliability Coefficients		7 items						
Alpha =		.8234		Standardized item alpha =		.8214		

SPSS Output 1



Cronbach's alpha is very easy to report. We could report this example as:

- ✓ The fear of computing subscale of the SAQ appeared to have good internal consistency, $\alpha = .82$.

If you wanted to add some detail you could add:

- ✓ All items appeared to be worthy of retention: the greatest increase in alpha would come from deleting item 10, but removal of this item would increase alpha only by .005. All items correlated with the total scale to a good degree (lower $r = .40$).

Guided Example



- ✓ Using the same data set, compute Cronbach's alpha for the *Fear of statistics* subscale of the SAQ.

How well do the items correlate with each other?

Your Answer:

Is the subscale reliable? (quote relevant statistics in APA format and comment on them).

Your Answer:

Would the reliability of the subscale benefit from deleting any items?

Your Answer:

The answers can be found in Field (2005) Chapter 15.

Unguided Example 1



- ✓ Using the same data set, compute Cronbach's alpha for the *Fear of Mathematics* subscale of the SAQ.
- ✓ Is this subscale reliable? (Explain your answer).

The answers can be found in Field (2005) Chapter 15.

Unguided Example 2



- ✓ Using the same data set, yes you've guessed it, compute Cronbach's alpha for the *Peer Evaluation* subscale of the SAQ.
- ✓ Is this subscale reliable? (Explain your answer).

The answers can be found in Field (2005) Chapter 15.

Unguided Example 3



- ✓ Using the **TOSSE-R.sav** data from last week (see your factor analysis handout) conduct a reliability analysis on each of the subscales of this questionnaire.
- ✓ Are these subscales reliable? (Explain your answer).

Multiple Choice Questions



Go to <http://www.sagepub.co.uk/field/multiplechoice.html> and test yourself on the multiple choice questions for Chapter 15. If you get any wrong, re-read this handout (or Field, 2005, Chapter 15) and do them again until you get them all correct.

This handout is an abridged version of Chapter 15 of Field (2005) and so is copyright protected.

Field, A. P. (2005). *Discovering statistics using SPSS (2nd edition)*. London: Sage.