

Interrater Reliability Using SPSS Dropdown Menus

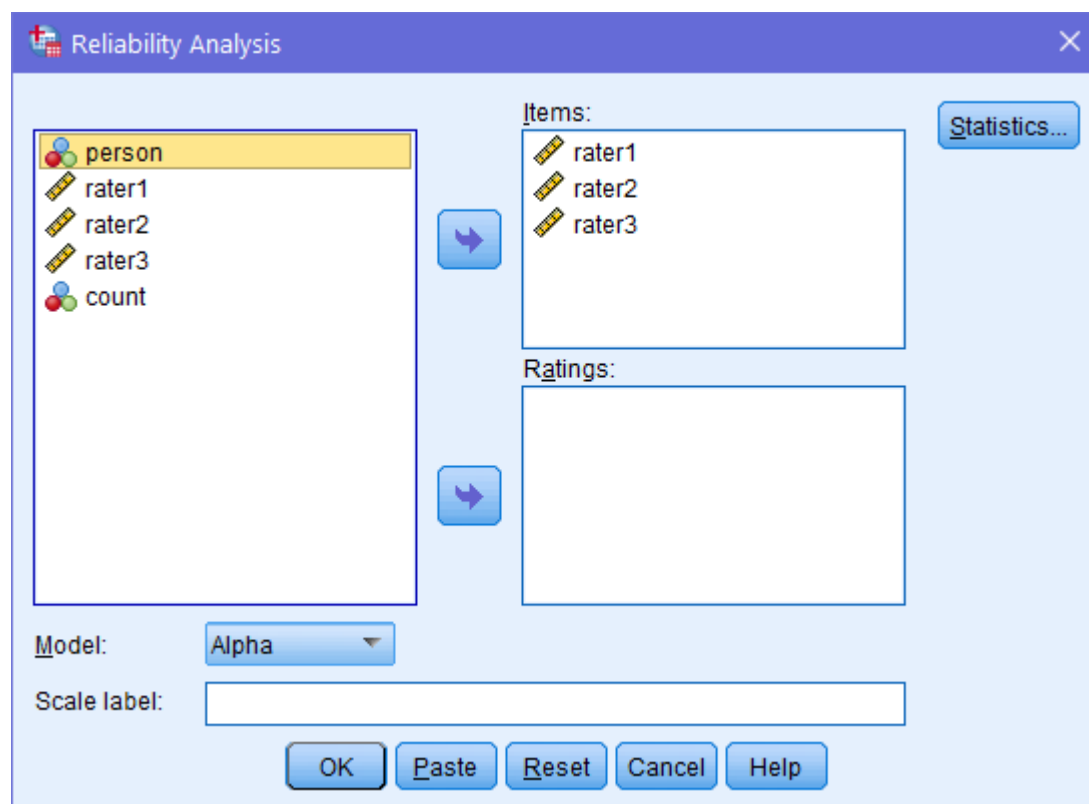
In this document I explain how to use SPSS to obtain Interrater reliability indices using SPSS dropdown menus. For SPSS syntax, see the document “Interrater Reliability Using SPSS Syntax.”

The data for these examples are taken from Table 9.1 in the book (p. 211) and are in the SPSS dataset “**rater data.sav**.”

In the sections below I demonstrate how to obtain values of coefficient alpha and of the intraclass correlation.

Coefficient Alpha for Interrater Reliability

To obtain values of coefficient alpha, choose **scale**, then **reliability analysis** from the **analyze** menu. Then choose the variables to be analyzed and click the arrow to put them into the box labeled **Items**. (in this example, these are the three raters).



Click *OK* and you will obtain the output below.

Reliability Statistics

Cronbach's Alpha	N of Items
.957	3

Note that this is the same (once rounded) as the value of .96 reported on page 215 of the text. The “**N of Items**” is actually the number of raters.

Intraclass Correlation for Interrater Reliability

The intraclass correlation is also obtained from the **reliability analysis** menu. Choose **scale** then **reliability analysis** as before. Choose the variables to be analyzed and then click on **statistics**. You will get the screen below:

Reliability Analysis: Statistics

Descriptives for

- ☒ Item
- ☐ Scale
- ☐ Scale if item deleted

Inter-Item

- ☒ Correlations
- ☒ Covariances

Summaries

- ☒ Means
- ☒ Variances
- ☒ Covariances
- ☒ Correlations

ANOVA Table

- ☒ None
- ☐ F test
- ☐ Friedman chi-square
- ☐ Cochran chi-square

Interrater Agreement: Fleiss' Kappa

- ☒ Display agreement on individual categories
- ☒ Ignore string cases
- ☒ String category labels are displayed in uppercase

Asymptotic significance level (%): 95

Missing

- ☒ Exclude both user-missing and system missing values
- ☐ User-missing values are treated as valid

☐ Hotelling's T-square

☐ Tukey's test of additivity

☒ Intraclass correlation coefficient

Model: Two-Way Random Type: Consistency

Confidence interval: 95 % Test value: 0

Continue Cancel Help

Click on **Intraclass correlation coefficient**. Then choose the appropriate options in the boxes next to **Model:** and **Type**. See pp. 217-221 in the text for more explanation of these options. Choosing **Two-Way Random** and **Consistency** will yield the same value as coefficient alpha, as shown below:

Intraclass Correlation Coefficient

	Intraclass Correlation ^b	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.881 ^a	.699	.966	23.316	9	18	.000
Average Measures	.957	.874	.988	23.316	9	18	.000

Two-way random effects model where both people effects and measures effects are random.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

The value in the row labeled “Average Measures” is the value of the intraclass correlation for the two-way random consistency coefficient: .957, which is identical to the value of coefficient alpha obtained previously for the three raters. The 95% confidence interval for this estimate is [.874, .988].

Choosing “Absolute Agreement” instead of “Consistency” will result in the table below:

Intraclass Correlation Coefficient

	Intraclass Correlation ^b	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.797 ^a	.408	.944	23.316	9	18	.000
Average Measures	.922	.674	.981	23.316	9	18	.000

Two-way random effects model where both people effects and measures effects are random.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

The value in the row labeled “Average Measures” is the value of the two-way random absolute coefficient: .922, with a confidence interval of .674 to .981.

Because absolute coefficients measure exact agreement among raters, whereas consistency coefficients measure the degree to which raters’ scores are in the same rank order, absolute coefficients will be lower than consistency coefficients, as seen here.