Interrater Agreement Using SPSS Syntax

In this document I explain how to use SPSS to obtain interrater agreement indices 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**"

For interrater agreement, SPSS will calculate Cohen's kappa and will produce tables such as that shown in Table 9.2 in the book. SPSS versions 26 and higher can also compute Fleiss's (1971) multirater kappa statistic.

SPSS does not calculate nominal agreement directly, but I discuss two ways to obtain these values using SPSS.

The first way is to obtain a crosstabulation table for the ratings from the two raters of interest. Then simply count the numbers on the diagonal of the table and divide by the total to obtain the nominal agreement index.

The second way is to use the **COMPUTE** menu to calculate agreement between pairs of raters and then use the **DESCRIPTIVES** command to obtain the mean of these agreements. This mean is the nominal agreement.

Nominal Agreement

The first way of obtaining values of nominal agreement is to use the **CROSSTABS** procedure to obtain the necessary table using the basic syntax below. Only two variables at a time can be analyzed. Here, I have chosen to assess the agreement between raters 1 and 2.

```
CROSSTABS
  /TABLES=rater1 BY rater2.
```

This will produce the following table.

rater1 * rater2 Crosstabulation

Count

		rater2					
		1	2	3	4	5	Total
rater1	1	1	0	0	0	0	1
	2	1	1	0	0	0	2
	3	0	1	2	0	0	3
	4	0	0	1	1	0	2
	5	0	0	0	0	2	2
Total		2	2	3	1	2	10

This table is the same as Table 9.2 in the book. I have highlighted the values that indicate exact agreement; these appear on the diagonal. Nominal agreement is simply the sum of these agreements divided by the total number of people being rated (10).

This is equal to:

$$P_O = \frac{1}{N}\sum_{i}^{C} n_{ii} = \frac{1}{10}(1+1+2+1+2) = \frac{1}{10}(7) = .7$$

P_0 is the nominal agreement, *N* is the number of people being rated, *C* is the number of rating categories, and n_{ii} represents the number on the diagonal of the matrix.

Calculating agreement using COMPUTE

To calculate agreement between two raters, use the syntax below.

```
COMPUTE count =0.
  IF (rater1 = rater2 ) count = 1.
EXECUTE.

DESCRIPTIVES VARIABLES=count/STATISTICS=mean.
```

The **COMPUTE** command creates the variable “count” with a value of 0 for all cases. The subsequent **IF** subcommand changes the value of “count” to 1 if the scores of rater1 and rater2 are equal (i.e., if the two raters assign the same score).

The mean of “count” is the average of these agreements, or the value of nominal agreement.

Descriptive Statistics

	N	Mean
count	10	.7000
Valid N (listwise)	10	

The value of .7000 under “mean” is the value for nominal agreement. Note that this is the same value obtained by counting the entries on the diagonal and dividing by the total number of ratings.

Cohen's Kappa

Values of Cohen's kappa can be obtained by adding the **STATISTICS** subcommand to the previous syntax, as shown below:

```
CROSSTABS
  /TABLES=rater1 BY rater2
  /STATISTICS KAPPA.
```

These commands will yield the table below:

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.620	.182	3.922	.000
N of Valid Cases		10			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

The value of kappa is .620, as calculated in the book on page 214.

Fleiss's Multirater Kappa

In many cases, agreement between more than two raters is of interest. This is true for the data in Table 9.2, for which there are three raters. Fleiss's multirater kappa computes the level of agreement among two or more raters.

For two raters, Fleiss's kappa will produce the value of kappa obtained previously. The basic syntax is shown below:

```
FLEISS MULTIRATER KAPPA rater1 rater2.
```

Note that the value of kappa in the table below is the same as that provided by the **CROSSTABS** command, although the standard error is somewhat smaller, yielding a larger value of the test statistic.

Overall Agreement[a]

	Kappa	Asymptotic Standard Error	z	Sig.	Asymptotic 95% Confidence Interval Lower Bound	Upper Bound
Overall Agreement	.618	.162	3.814	.000	.608	.628

a. Sample data contains 10 effective subjects and 2 raters.

To obtain the multirater kappa value for all three raters, simply add rater3 to the list of variables in the command:

FLEISS MULTIRATER KAPPA rater1 rater2 rater3.

This will result in the table below. Although kappa was .62 for raters 1 and 2, it drops to .36 across all three raters. This is not surprising, because a) it is more difficult to obtain agreement across three raters than across two raters, and b) rater 3 provided somewhat higher ratings than the other two raters for most people, thus decreasing the overall agreement level.

Overall Agreement[a]

	Kappa	Asymptotic Standard Error	z	Sig.	Asymptotic 95% Confidence Interval Lower Bound	Upper Bound
Overall Agreement	.363	.094	3.850	.000	.357	.369

a. Sample data contains 10 effective subjects and 3 raters.