Paper 194-26

A Correction for Unbalanced Kappa Tables SAS 6.12®

Philip E. Crewson, Ph.D., American College of Radiology, Reston, VA
January 2001

THIS IS A TITLE PAGE ONLY
Paper continues on the next page.

Abstract

The kappa statistic is a commonly used method for estimating agreement between two raters using the same set of rating categories. Kappa requires a balanced contingency table where the rating categories used by one rater (represented in the columns) are directly comparable to the rating categories used by a second rater (represented in the rows). Situations where there are no responses in one or more of the rating categories by either or both raters can cause SAS 6.12 to 1) not report kappa because the table is unbalanced, or 2) report an erroneous kappa that is based on a table that is balanced but the categories used by the raters are not consistent. To solve this problem, a simple method was developed to compensate for raters who did not use all categories of a rating scale. The solution entails adding dummy observations to the data set to establish at least one observation in each rating category for each rater and then setting these dummy observations to missing so as not to bias the kappa coefficient.

The Problem

Kappa is a commonly used method for estimating paired interrater agreement for nominal and/or ordinal data (Fleiss 1981). To estimate kappa, both raters must use the same number of rating criteria so that the number of rows representing the rating categories used by Rater1 equals the number of columns representing the rating categories used by Rater2. An unbalanced contingency table will result if the rows (categories used by one rater) do match the columns (categories used by a second rater). However, as has been previously reported (Liu and Hays 1999), SAS Proc Freq does not correct for unbalanced contingency tables. As an example, Rater1 may have used all of five possible rating criteria while Rater2 only used four out of five. As shown in Table 1, SAS Proc Freq will produce a contingency table but appropriately does not report kappa statistics.

TABLE 1: UNCORRECTED KAPPA TABLE OF RATER1 BY RATER2

RATER1(USED ALL CRITERIA) RATER2(NEVER USED PROB-BEN)

Frequency	BENIGN	INCONCLU	PROB-CAN	CANCER	Total
BENIGN	5	1	0	0	6
PROB-BEN	3	2	0	0	5
INCONCLU	0	7	5	0	12
PROB-CAN	0	1	1	0	2
CANCER	0	0	1	4	5
Total	8	11	7	4	30

If the paired raters failed to use all rating criteria but used the same number (e.g., 4 out of 5 categories), SAS will output kappa statistics. As long as the raters are using the same set of rating criteria, the kappa statistics will be correct. However, as shown in Table 2, if the raters used the same number of criteria but different categories, an erroneous kappa will be produced. In this example, although both Rater2 and Rater3 used the same number of criteria (four), Rater2 never used the probably benign (PROB-BEN) criterion and Rater3 never used the probably cancer (PROB-CAN) criterion. As a result, PROB-BEN ratings by Rater3 are inappropriately matched with Inconclusive ratings by Rater2, and Inconclusive ratings by Rater3 are inappropriately matched with the PROB-CAN ratings of Rater2.

TABLE 2: UNCORRECTED KAPPA TABLE OF RATER2 BY RATER3

RATER2(NEVER USED PROB-BEN) RATER3(NEVER USED PROB-CAN)

Frequency	BENIGN	PROB-BEN	INCONCLU	CANCER	Total
BENIGN	4	3	1	0	8
INCONCLU	1	2	7	1	11
PROB-CAN	0	0	5	2	7
CANCER	0	0	0	4	4
Total	5	5	13	7	30

STATISTICS FOR TABLE OF RATER2 BY RATER3

Test of Symmetry

Statistic = 12.000 DF = 6 Prob = 0.062

Kappa Coefficients

Statistic	Value	ASE	95% Confidence Bounds	
Simple Kappa	0.344	0.116	0.117	0.571
Weighted Kappa	0.555	0.104	0.352	0.758

Sample Size = 30

A Solution

A simple solution to this problem is to add dummy observations to the data set for each rater to compensate for those who did not use all categories of a rating scale. For Rater2, one observation would have to be added that used the PROB-BEN criterion. For Rater3 one observation would have to be added that used the PROB-CAN criterion. Although this approach balances the table, it also adds fictitious data to the sample. As a result, once dummy observations have been used to dimension a balanced table, they must be excluded from the cell frequencies.

To exclude the dummy observations, a third variable (Control) is created and added to the Proc Freq Tables statement to produce a 3-way table. This third variable is used to classify the original observations as 1 and the dummy observations as missing. Without the control, SAS Proc Freq will produce a table of sufficient dimension to accommodate the merged data file. This table will contain both the original observations and the dummy observations. With the incorporation of the control variable, the table retains the dimensions created for the merged data file along with the original observations and sets the dummy observations to missing. The SAS Proc Freq program syntax for placement of the Control variable is given below:

```
PROC FREQ;TABLES CONTROL*RATER2*RATER3
/NOCOL NOROW NOPERCENT KAPPA;
```

As shown in Table 3, correcting the unbalanced table will result in a square contingency table with empty rows and columns inserted for those categories not used by the raters

TABLE 3: RATER2 BY RATER3 CONTROLLING FOR CONTROL=1

RATER2(NEVER USED PROB-BEN) RATER3(NEVER USED PROB-CAN)

Frequency	BENIGN	PROB-BEN	INCONCLU	PROB-CAN	CANCER	Total
BENIGN	4	3	1	0	0	8
PROB-BEN	0	0	0	0	0	0
INCONCLU	1	2	7	0	1	11
PROB-CAN	0	0	5	0	2	7
CANCER	0	0	0	0	4	4
Total	5	5	13	0	7	30

Test of Symmetry

Statistic = 13.000 DF = 10 Prob = 0.224

Kappa Coefficients

Statistic	Value	ASE	95% Confidence Bounds	
Simple Kappa	0.347	0.105	0.141	0.552
Weighted Kappa	0.600	0.091	0.421	0.779

Sample Size = 30

SAS Program Steps

The following SAS program provides the basic programming details. The additional programming needed to create the solution is in enlarged bold text. The Appendix provides a more detailed program and represents the common application where an external SAS data file is the primary source of rater data. In both programs, a sufficient number of dummy observations are created to cover all possible rating criteria (i.e., a five category rating scale requires five observations with data for each rater). In practice, this removes the necessity of ascertaining the missing criteria for every rater.

```
DATA OLD;
INPUT RATER1 1 RATER2 2 RATER3 3 RATER4 4;

CONTROL=1;

CARDS;
1111
2122
3334
4454
5555
1112
2132
3323
4334
5455
1311
1122
1111
2322
2122
2333
5554
5555
5555
3334
3333
3433
3333
3434
3432
3353
3333
3432
3433
1111
RUN;

LABEL
  RATER1='USED ALL CRITERIA'
  RATER2='NEVER USED PROB-BEN'
  RATER3='NEVER USED PROB-CAN'
  RATER4='USED ALL CRITERIA';

PROC FORMAT;
  VALUE RATING 1='BENIGN ' 2='PROB-BEN ' 3='INCONCLU'
               4='PROB-CAN' 5='CANCER';

DATA DUMMY;
INPUT RATER1 1 RATER2 2 RATER3 3 RATER4 4;
CARDS;
1111
2222
3333
4444
5555
RUN;

DATA NEW;
SET OLD DUMMY;

PROC FREQ;
  TABLES CONTROL*RATER1*(RATER2 RATER3 RATER4)
         CONTROL*RATER2*(RATER3 RATER4)
         CONTROL*RATER3*RATER4
         /NOCOL NOROW NOPERCENT KAPPA;
         FORMAT RATER1--RATER4 RATING.;
RUN;
```

Conclusion

This solution is useful in situations where bi-rater kappas must be estimated from data that include incomplete use of all rating criteria. Although not as elegant as a macro, it requires little additional programming and can be easily adjusted to balance any size contingency table.

References

Fleiss, J.L. (1981). Statistical Methods for Rates and Proportions. John Wiley, New York.

Liu, H. and Hays, R.D. (1999). Measurement of Interrater Agreement: A SAS/IML Macro Kappa Procedure for Handling Incomplete Data. SUGI24, Paper 280.

SAS/STAT User's Guide, SAS Institute Inc., SAS Campus Drive, Cary, NC.

Appendix

```
TITLE1 'SASOUT     FILE: KAPPA.SAS';
TITLE2 'CORRECTION FOR INCOMPLETE KAPPA TABLES';

/*
**************************************************
*                                                *
*   CORRECTING INCOMPLETE/UNBALANCED KAPPA TABLES   *
*         (EXPLANATION OF SAS PROGRAM STEPS)       *
*                                                *
*Step 1: LOAD ORIGINAL DATA FILE                   *
*                                                *
*Step 2: CREATE CONTROL VARIABLE. This is used     *
*        to exclude dummy observations from the    *
*        final kappa analysis.                     *
*                                                *
*Step 3: CREATE DUMMY DATA FILE TO FILL            *
*        INCOMPLETE DATA.  A matching variable     *
*        (using same variable name) must be        *
*        created for each rater in the kappa       *
*        analysis. The number of observations      *
*        must match the number of categories       *
*        used in the kappa tables to insure        *
*        each rater in the dummy data file has one *
*        response in each rating category.         *
*                                                *
*                                                *
*Step 4: MERGE ORIGINAL DATA FILE WITH DUMMY DATA. *
*                                                *
*Step 5: RUN PROC FREQ FOR KAPPA USING THE CONTROL.*
*        Since the control variable is missing for *
*        the dummy observations, only the original *
*        data are used for kappa computations.     *
*        However, the tables are adjusted to       *
*        accommodate the complete range of ratings *
*        thus including rows and columns that      *
*        otherwise would have been excluded due    *
*        to zero observations.                     *
*                                                *
*Note: Kappa and standard weighted Kappa           *
*      statistics will be unaffected by this       *
*      adjustment. However, symmetric statistics   *
*      may be incorrect.                         *
*                                                *
**************************************************/

****************************************;
*  STEP 1   LOAD ORIGINAL DATA FILE    *;
****************************************;

LIBNAME IN 'C:\WORK\KAPPA';
DATA OLD;
SET IN.KAPPA;

****************************************;
*  STEP 2   CREATE CONTROL VARIABLE    *;
****************************************;

CONTROL=1;

*  PRODUCES EXAMPLE OF UNALTERED DATA  ;

LABEL
  RATER1='USED ALL CRITERIA'
  RATER2='NEVER USED PROB-BEN'
  RATER3='NEVER USED PROB-CAN'
  RATER4='USED ALL CRITERIA';

PROC FORMAT;
  VALUE RATING 1='BENIGN ' 2='PROB-BEN ' 3='INCONCLU'
               4='PROB-CAN' 5='CANCER';

TITLE3 'UNCORRECTED KAPPA TABLES';
PROC FREQ;TABLES RATER1*(RATER2 RATER3 RATER4)
                 RATER2*(RATER3 RATER4)
                 RATER3*RATER4
          /NOCOL NOROW NOPERCENT KAPPA;
          FORMAT RATER1--RATER4 RATING.;
RUN;

****************************************;
*  STEP 3   CREATE DUMMY DATA FILE     *;
****************************************;

DATA DUMMY;
INPUT RATER1 1 RATER2 2 RATER3 3 RATER4 4;
CARDS;
1111
2222
3333
4444
5555
RUN;

****************************************;
*  STEP 4   MERGE DATA FILES           *;
****************************************;

DATA NEW;
SET OLD DUMMY;


****************************************;
*  STEP 5   RUN PROC FREQ WITH CONTROL  *;
****************************************;

TITLE3 'CORRECTED KAPPA TABLES';
PROC FREQ;TABLES CONTROL*RATER1*(RATER2 RATER3 RATER4)
                 CONTROL*RATER2*(RATER3 RATER4)
                 CONTROL*RATER3*RATER4
          /NOCOL NOROW NOPERCENT KAPPA;
          FORMAT RATER1--RATER4 RATING.;
RUN;
```

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Philip E. Crewson, Ph.D.
American College of Radiology
1891 Preston White Drive
Reston, VA 20191
Phone: 703-648-8983
Fax: 702-264-2443
E-mail: philipc@acr.org
Website: www.acr.org