

Paper 201-2012

Calculating Multi-Rater Observation Agreement in Health Care Research Using the SAS® Kappa Statistic

Abbas S. Tavakoli, DrPH, MPH, ME, Bo Cai, PhD, Rita Snyder, PhD, RN, & Nathan Huynh, PhD

Abstract

This paper describes procedures for calculating multi-rater observation agreement using the SAS® Kappa statistic in a health care research study of the medication administration process (MAP). In the study, Registered Nurses (RNs) modeled the oral MAP 27 times in a simulated laboratory environment. All 27 MAP modeling sessions were video recorded. Four individuals observed each of the 27 modeling sessions and recorded the MAP functions and tasks that they observed for each session. Three nurse subject matter experts (observers 1, 2 and 3) watched video recordings of the 27 MAP sessions and recorded their observations on a paper Video Observation Scoring Form. The fourth observer (observer 4) watched all 27 MAP sessions in real-time in the simulation laboratory and recorded his observations on a hand-held device. The content on the scoring form and hand-held device was identical. Inter-rater reliability (IRR) between the four observers was examined using the Kappa statistic, calculated using the SAS® PROC FREQ, MEANS, and PRINT procedures. The Kappa statistic and 95% CI for observers were calculated and the overall IRR was calculated by averaging pairwise Kappa agreements. Pairwise Kappa agreements ranged from 0.28 (observers 3 and 4) to 0.58 (observers 1 and 2). Agreements were good for all pairwise comparisons except for a fair agreement for observer 4. The overall IRR between observers 1, 2 and 3 was 0.46 (average of 0.58, 0.41, and 0.40). This paper provides an example of how to calculate percentage agreement with the Kappa statistic with a 95% CI using SAS® PROC FREQ, MEANS, and PRINT for multiple raters with multiple observation categories. This technique expands the current functionality of the SAS® PROC FREQ procedure to support application of the Kappa statistic for more than two raters and several categories.

Keywords:

Multi-rater Observation Agreement, Kappa Statistic, Health Care.

Introduction

In observational studies it is often necessary to assess multi-rater agreement for multiple observation categories. The Kappa statistic is one of the most common agreement measures for categorical data (Altman, 1991). The Kappa statistic estimates the percentage of agreement among raters after removing the percentage of agreement which would occur by chance. The SAS® PROC FREQ procedure supports application of the Kappa statistic for two raters and several categories. However, calculation is not straightforward for more than two raters as data must first be manipulated in a square form table in order to use the SAS® PROC FREQ procedure.

Purpose

The purpose of this paper is to describe an approach for using the SAS® PROC FREQ, MEANS, and PRINT procedures to calculate multi-rater observation agreement with the Kappa statistic.

Background

An observational study of the medication administration process (MAP) was conducted in a clinical simulation laboratory (CSL). Three Registered Nurses (RNs) modeled the oral MAP 27 times. All 27 MAP modeling sessions were video recorded. Four individuals observed each of the 27 modeling sessions and recorded the MAP functions and tasks that they observed for each session. Three nurse subject matter experts (observers 1, 2 and 3) watched video recordings of the 27 MAP sessions and recorded their observations on a paper Video Observation Scoring Form. The fourth observer (observer 4) watched all 27 MAP sessions in real-time in the CSL and recorded his observation on a hand-held device. The content on the scoring form and hand-held device was identical.

Data Analysis

Analysis of observational data was done using SAS/STAT® statistical software, version 9.2 (SAS, 2008). The SAS® PROC FREQ procedure was used for Kappa statistical calculations. Table 1 shows a subset of the tasks recorded. The maximum number of tasks that could be selected by each observer was 70. Some variability existed in the number and order of tasks recorded by each observer. Table 2 is part of the original data for the first three videos and shows the sequence of tasks that were selected by each observer. Steps were used in the data analysis process.

1. A square form table for observers using pseudo-observations was created using original data (Table 2).
 - An Array statement was used to verify if tasks were selected from the task sequence chosen by each observer (Attachment: Part 1 SAS Syntax: Data Two).
 - Selected tasks were assigned a code of 1. If no task was selected, a missing value code was assigned (Attachment: Part 1 SAS Syntax: Data Two). The method adopted by Uebersax (2002) to replace the missing flag values with the value 'M' was used (Attachment: Part 1 SAS Syntax: Data Replace).
 - A weight of 1 for real observations and a weight of .0000000001 for pseudo-observations was assigned. Pseudo-observations ensure responses for every task assigned by any other observers and the small weight does not have any effect on the Kappa statistic (Attachment: Part 1 SAS Syntax: Data Wt).
2. Data Replace (real data) and Data Wt (pseudo-data) were concatenated (Attachment: Part 1 SAS Syntax: Data WtALL).
3. Original data were transformed by replacing the flag values of 1 with any given task and assigning a value of 1 to the weight variable. Observations were deleted when no task was selected by either observer. PROC FREQ was used to calculate the Kappa statistic for each pairwise observer. The overall IRR was calculated by averaging the pairwise Kappa agreements.
4. The number of agreements, the actual percentage of agreement, the Kappa statistic with a 95% CI for each video, and the average of Kappa statistics overall videos ignoring the individual video were calculated together (Attachment: Part 2 SAS Syntax).
 - The OUT option in the table statement (Tables btsk1*btsk2/ out=Store norow nocol nopercnt agree) and the OUTPUT statement were used to obtain the Kappa statistic and the number of agreements (Attachment: Part 2 SAS Syntax: output out =Kappafile agree).
 - The Store and Kappa file datasets were created to keep only the most important columns from the Kappa output and the diagonal columns from the Store dataset (Attachment: Part 2 SAS Syntax: Data Kappafile and Store).
 - The ODS statement was used to construct tables which contain the number of agreements for each pair of observers (e.g., Attachment: Part2 SAS Syntax: Data Tableconst12).
 - The percentage of agreement was calculated as Percentagree= NumAgree/TotCap.
5. Data for each pairwise observation were combined to produce the final output (Attachment: Part2 SAS Syntax: Data Combinevids).
6. PROC MEANS and OUTPUT options were used to produce a mean value for the combined data (Attachment: Part 2 SAS Syntax: Data Combinevids and Means).
7. PROC PRINT and ODS statements were used to construct the final outputs.

Results

Table 3 shows part of the output created for each video for the number of observer agreements, percentage of agreement, and Kappa with a 95% CI. For example, the Kappa agreement for video 1 between observers 1 and 2 was 0.52 (95% CI 0.31-0.74). It is shown that the agreement between two raters varies across different videos. Table 4 shows the complete output for the number of agreements, percentage of agreement, and Kappa with 95% CI ignoring individual videos. As indicated in Table 4, the Kappa between two observers ranged from 0.28 (observers 3 and 4) to 0.58 (observers 1 and 2) with an overall IRR between observer 1, 2, and 3 of 0.46 (average of 0.58, 0.41, and 0.40) (Attachment: SAS syntax part 2). According to Landis and Koch (1977), all the pairs had moderate agreements except that observers 3 and 4 had a fair agreement. Table 5 shows the complete output for the number of agreements, percentage of agreement, and Kappa with 95% CI by averaging individual videos. (Attachment: SAS syntax Part 2). Table 5 reveals all pairwise comparisons had good agreements.

Conclusion

This paper provides an example of how to calculate percentage agreement with the Kappa statistic with 95% CI using SAS ® PROC FREQ, MEANS, and PRINT for multiple raters with multiple observation categories. This technique expands the current functionality of the SAS ® PROC FREQ procedure to support application of the Kappa statistic for more than two raters and several categories.

References

1. Altman, D. (1991). Practical Statistics for Medical Research. Chapman and Hall.
2. Landis, J. & Koch, G. (1977). The Measurement of Observer Agreement for Categorical Data". *Biometrics*, 33:159-174.
3. Uebersax, J. (2002). Calculating Kappa with SAS: <http://www.john-uebersax.com/stat/saskappa.htm>

Acknowledgements

The authors wish to express their sincere appreciation of Ms. Gene Meding's editorial support in the development of this paper.

Contact Information

Abbas S. Tavakoli, DrPH, MPH, ME
 College of Nursing
 University of South Carolina
 1601 Greene Street
 Columbia, SC 29208-4001
 Fax: (803) 777-5561
 E-mail: abbas.tavakoli@sc.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

Table 1. Sample list of MAP study tasks.

1='Greet patient'
2='Pre-wash hands'
3='Chat with patient'
4='Scan personal tag'
5='Log-in to MAR'
6='Review MAR'
.
.
.
.
.
68= 'At floor'
69= 'Introduce self'
70='At computer'

Table 2. Partial data set for first three video observations by four observers.

OBS	ID	TASK1	SEQ1	SEQ2	SEQ3	SEQ4
1	1	Greet patient	2	1	1	.
2	1	Pre-wash hands	3	3	3	.
3	1	Chat with patient	4	.	4	.
4	1	Scan personal tag	5	4	5	.
5	1	Log-in to MAR	6	5	6	.
6	1	Review MAR	7	10	10	2
.

OBS	ID	TASK1	SEQ1	SEQ2	SEQ3	SEQ4
71	2	Greet patient	2	1	1	.
72	2	Pre-wash hands	4	3	3	.
77	2	Obtain medications	18	13	13	10
78	2	Verify correct medications	.	.	.	11
.
142	3	Pre-wash hands	4	3	3	.
143	3	Chat with patient	5	.	4	.
144	3	Scan personal tag	7	5	5	4

Key – Table 2

OBS = Observation

SEQ = Sequence

Table 3. Partial output for MAP observer task selection for individual videos.

Video	Rater*	Number of agreement	Total tasks	Percentage of agreement	Kappa agreement	Lower 95% CI** Kappa	Upper 95% CI** Kappa
1	Rater 1 vs. Rater 2	15	70	0.21429	0.52212	0.30633	0.73792
2	Rater 1 vs. Rater 2	18	70	0.25714	0.51518	0.32137	0.70899
3	Rater 1 vs. Rater 2	19	70	0.27143	0.55448	0.37063	0.73834
.
1	Rater 1 vs. Rater 3	17	70	0.24286	0.63380	0.43368	0.83393
2	Rater 1 vs. Rater 3	17	70	0.24286	0.48083	0.28392	0.67773
3	Rater 1 vs. Rater 3	19	70	0.27143	0.55448	0.37063	0.73834
.
1	Rater 1 vs. Rater 4	9	70	0.12857	0.29577	0.06296	0.52859
2	Rater 1 vs. Rater 4	10	70	0.14286	0.27425	0.08117	0.46732
3	Rater 1 vs. Rater 4	10	70	0.14286	0.25515	0.07116	0.43914
.

*Raters 1, 2, and 3 are RN videotape observers. Rater 4 is hand-held device observer.

**CI = Confidence interval

Table 4. Output for MAP observers ignoring individual videos.

Obs	Rater*	Number of agreement	Total tasks	Percentage of agreement	Kappa agreement	Lower 95% CI** Kappa	Upper 95% CI** Kappa
1	Rater 1 vs. Rater 2	25	70	0.35714	0.58006	0.53744	0.62268
2	Rater 1 vs. Rater 3	25	70	0.35714	0.40750	0.36212	0.45289
3	Rater 1 vs. Rater 4	19	70	0.27143	0.41707	0.37078	0.46337
4	Rater 2 vs. Rater 3	21	70	0.30000	0.39376	0.34573	0.44180
5	Rater 2 vs. Rater 4	21	70	0.30000	0.56391	0.51850	0.60932
6	Rater 3 vs. Rater 4	17	70	0.24286	0.27826	0.22289	0.33363
7	Overall	21.33333	70	0.30476	0.44010	0.39291	0.48728

*Raters 1, 2, and 3 are RN videotape observers. Rater 4 is hand-held device observer.

**CI = Confidence interval

Table 5. Output for averaged MAP Observations of individual videos.

Obs	Rater*	Number of agreement	Total tasks	Percentage of agreement	Kappa agreement	Lower 95% CI** Kappa	Upper 95% CI** Kappa
1	Rater 1 vs. Rater 2	14.2963	70	0.20423	0.58675	0.37005	0.80346
2	Rater 1 vs. Rater 3	14.11111	70	0.20159	0.61579	0.40160	0.82998
3	Rater 1 vs. Rater 4	10.77778	70	0.15397	0.42173	0.19082	0.65263
4	Rater 2 vs. Rater 3	13.33333	70	0.19048	0.59931	0.37611	0.82238
5	Rater 2 vs. Rater 4	12.11111	70	0.17302	0.57557	0.35430	0.79528
6	Rater 3 vs. Rater 4	10.11111	70	0.14444	0.47159	0.23018	0.71301
7	Overall	12.45679	70	0.17795	0.54512	0.32051	0.76946

*Raters 1, 2, and 3 are RN videotape observers. Rater 4 is hand-held device observer.

**CI = Confidence interval

Attachment A

SAS Syntax

Part 1

```
DATA ONE;
Set vtr.vtr10;
If 0<task1<71;
```

Label

```
Id = 'Obs number'
Task1 = ' Task for rater'
seq1 = ' Activity Sequence for rater 1'
seq2 = ' Activity Sequence for rater 2'
seq3 = ' Activity Sequence for rater 3'
```

```

seq4 = ' Activity Sequence for rater 4/ device';
format func1-func4 funcf. task1-task4 taskf. loc1-loc4 locf. ;

data two;    set one;
array items seq1-seq4;
array itemb btask1-btask4 ;
do over items; if items ne . then itemb=1;    else itemb=.; end;
proc sort;  by id;

ods rtf; ods listing close;
*** calculate Kappa    FOR Selecting UP THE TASKS BY RATER ***;
data REPLACE (keep=id task1  btsk1 btsk2 btsk3 btsk4 wgt) ;
set two ; by id ;
length btsk1 btsk2 btsk3 btsk4 $ 3;
if btask1 =1 then btsk1=task1 ;
else btsk1=' M ' ;
if btask2 =1 then btsk2=task1 ;
else btsk2=' M ' ;
if btask3 =1 then btsk3=task1 ;
else btsk3=' M ' ;
if btask4 =1 then btsk4=task1 ;
else btsk4=' M ' ;
* ASSIGN WEIGHT OF '1' TO REAL RECORDS ;
wgt = 1 ; run ;
Proc sort;  by id;

* CREATE WT DATA RECORDS TO ENSURE SQUARE TABLE FOR OBTAINING
KAPPAS. ASSIGN A TINY WEIGHT SO DUMMY OBSERVATIONS DO NOT EFFECT
KAPPA VALUES. INCLUDE ROW OF MISSING ('M') VALUES FOR EACH VIDEO ;

Data WT (keep=id task1 btsk1 btsk2 btsk3 btsk4 wgt);
set two ; by id ;
length btsk1 btsk2 btsk3 btsk4 $ 3;
btsk1=task1 ; btsk2=task1 ; btsk3=task1 ; btsk4=task1; wgt = .0000000001 ; output;
if last.id then do ; btsk1=' M ' ; btsk2=' M ' ; btsk3=' M ' ; btsk4=' M ' ;
output ; end ; run ;

* CONCATENATE REPLACE & WT DATA & SORT BY VIDEO;

data wtALL;  set REPLACE WT ; run ;
proc sort data=wtALL;  by id; run;

* CREATE DATA SET FOR COMPARING OBSERVERS 1 & 2 ;

Data RATEAB; set wtALL;
* DELETE REAL RECORDS WHERE task CODE WAS NOT USED BY EITHER OBSERVER ;
if btsk1=' M ' and btsk2=' M ' and wgt=1 then delete ; run ;

* PRODUCE FREQ TABLE COMPARING OBSERVERS 1 & 2. USE WEIGHT STATEMENT TO
OBTAIN SQUARE TABLES, BUT USE NON-WEIGHTED KAPPA RESULTS ;

Proc freq data=RATEAB; weight wgt;
tables btsk1*btsk2 / norow nocol nopercnt agree ;
title ' Kappa calculation / RATER 1 AND 2'; title2 ' Medication error study'; run ;

* USE ABOVE STEPS TO CALCULATE KAPP STATISTICS FOR RATER 1&3 , RATER 1 & 4, RATER 2 &
3 , RATER 2& 4, AND RATER 3 &4 **

ODS rtf close; ods listing; quit; run;

```

Part 2

```
*Programs that put number of percentage of agreement with Kappa statistics and 95 % CI
of Kappa: CREATE DATA SET FOR COMPARING OBSERVERS 1 & 2 FOR EACH VIDEO ;
```

```
Data testAB; set wtId;
```

```
* DELETE REAL RECORDS WHERE task CODE WAS NOT USED BY EITHER OBSERVER ;
```

```
If btsk1=' x ' and btsk2=' x ' and wgt=1 then delete; run; quit;
```

```
* PRODUCE FREQ TABLE COMPARING OBSERVERS 1 & 2. USE WEIGHT STATEMENT TO OBTAIN SQUARE
TABLES, BUT USE NON-WEIGHTED KAPPA RESULTS. Also Obtain kappa statistics and number of
agreements and output into their own dataset;
```

```
Proc sort data=testAB; by id;
```

```
proc freq data=testAB noprint;
```

```
weight wgt;
```

```
tables btsk1*btsk2 / out=Store norow nocol nopercnt agree; by id;
```

```
output out=kappafile agree;
```

```
Title 'Kappa calculation / RATER 1 AND 2'; title2 'Medication error study';
```

```
Run; quit;
```

```
/* Keep only the most important columns from the kappaoutput */
```

```
data kappafile; set kappafile; keep ID _Kappa_ U_Kappa L_Kappa; run; quit;
```

```
/* Keep only the diagonal columns and adjust the count so the
numbers are more pure. */
```

```
Data store; set store;
```

```
if (btsk1 = ' x ') then delete; if (btsk2 = ' x ') then delete;
```

```
if (COUNT < .9) then COUNT = 0; if (COUNT >= 1) then COUNT = 1; run; quit;
```

```
/* Sum up the and obtain the total number of agreements. This is obtained by summing
up the diagonals of the matrix. The more 1's means the more agreements. So the
higher the number, the more the two raters agreed at that video. This code also puts
the frequencies into a table. */
```

```
ODS Listing Close; ODS Output OneWayFreqs = tableconst12;
```

```
proc freq data = store ; by id;
```

```
tables COUNT / norow nocol nopercnt; run; quit; ODS Listing;
```

```
/* Get all the agreement frequencies as well as constructing a rater variable to
tell us which raters we are comparing and the total number of sequencies within the
dataset (also known as the total number of comparisons. */
```

```
Proc sort data=tableconst12; by ID;
```

```
data tableconst12; set tableconst12; by ID ;
```

```
if (last.ID); if (count = 0) then Frequency = 0; Rater1 = 12; totCap = 70;
```

```
keep id Rater1 Frequency totcap; run; quit;
```

```
/* Merge the frequencies and the kappa statistics into 1 file. this
combines the two datasets so all the information for each video is in one place. */
data tableconst12; merge tableconst12 kappafile; by ID; run; quit;
```

```
.
```

```
.
```

```
* CREATE DATA SET FOR COMPARING OBSERVERS 1 & 2 FOR OVERALL VIDEOS ;
```

```
data testAB; set WtALL;
```

```
* DELETE REAL RECORDS WHERE task CODE WAS NOT USED BY EITHER OBSERVER;
```

```
If btsk1=' M ' and btsk2=' M ' and wgt=1 then delete; run; quit;
```

```
* PRODUCE FREQ TABLE COMPARING OBSERVERS 1 & 2. USE WEIGHT STATEMENT TO
```

OBTAIN SQUARE TABLES, BUT USE NON-WEIGHTED KAPPA RESULTS. Also Obtain kappa statistics and number of agreements and output into their own dataset;

```
Proc sort data=testAB; by id;
proc freq data=testAB noprint;
weight wgt;
tables btsk1*btsk2 / out=Store norow nocol nopercnt agree; output out=kappafile
agree;
Title 'Kappa calculation / RATER 1 AND 2'; title2 'Medication error study';
Run; quit;
```

/* Keep only the most important columns from the kappaoutput */

```
Data kappafile; set kappafile; keep _Kappa_ U_Kappa L_Kappa; run; quit;
```

/* Keep only the diagonal columns and adjust the count so the numbers are more pure. */

```
Data store; set store;
if (btsk1 = ' M ') then delete; if (btsk2 = ' M ') then delete;
if (COUNT < .9) then COUNT = 0; if (COUNT >= 1) then COUNT = 1; run; quit;
```

/* Sum up the and obtain the total number of agreements. This is obtained by summing up the diagonals of the matrix. The more 1's means the more agreements. So the higher the number, the more the two raters agreed at that video. This code also puts the frequencies into a table. */

```
ODS Listing Close; ODS Output OneWayFreqs = tableconst12;
Proc freq data = store; tables COUNT / norow nocol nopercnt; run; quit;
ODS Listing;
```

/* Get all the agreement frequencies as well as constructing a rater variable to tell us which raters we are comparing and the total number of sequences within the dataset (also known as the total number of comparisons. Assign code to rater for example rater 1 vs 2 code as 12 . Total tasks here is 70.*/

```
Data tableconst12; set tableconst12; if (count = 1); Rater1 = 12; totCap = 70;
keep Rater1 Frequency totcap; run; quit;
/* Merge the frequencies and the kappa statistics into 1 file. This combines the two
datasets so all the information is in one place. */
data tableconst12; merge tableconst12 kappafile; run; quit;
.
```

/* CREATE DATA SET FOR COMPARING OBSERVERS 1 & 3; OBSERVERS 1 & 4; OBSERVERS 2 & 3
OBSERVERS 2 & 4; AND OBSERVERS 3 & 4 THE SAME AS ABOVE PROGRAM
/* Combine all the raters into one table. This is all based on videotape. */

```
Data combineVids; set tableconst12 tableconst13 tableconst14 tableconst23 tableconst24
tableconst34;
REName ID = Video Frequency = NumAgree Rater1 = Rater; run; quit;
```

/* Delete datasets */

```
Proc datasets nodetails; delete Clean Dummy Kappafile Mappings Store Tableconst12
Tableconst13 tableconst14 Tableconst23 Tableconst24 Tableconst34 Testab Testac Testad
Testbc Testbd Testcd Two; run; quit;
```

/* Delete any videos with no agreements. */

```
Data combineVids; set combineVids; if NumAgree = 0 then delete;
PercentAgree = NumAgree/totCap; run; quit;
```

```

/* Sum up the number of agreements, the total number of comparisons, the Kappa
statistic and its two CI intervals. This is combining all the information
corresponding to the videos. Then output all the means */

Proc means data=combineVids noprint; var NumAgree totCap _Kappa_ L_Kappa U_Kappa;
class Rater; output out = means; run; quit;

/* Delete all extra information that comes from PROC Means. Also calculate the
percent agreement. */

Data means; set means; if (_STAT_ = 'MEAN'); if (Rater =.) then delete;
PercentAgree = NumAgree/totCap;
Keep Rater NumAgree _Kappa_ L_Kappa U_Kappa totCap PercentAgree; run; quit;

* Sum up all the rater values. This will sum up the raters into one overall score;

Proc means data=means noprint; var NumAgree totCap _Kappa_ L_Kappa U_Kappa;
output out=OverallFinal; run; quit;

/* Adjust the output and delete all extra information that comes from Proc Means.
Also calculate the Percent Agreement. */

Data OverallFinal; set OverallFinal; if (_STAT_ = 'MEAN'); Rater = 100;
PercentAgree = NumAgree/totCap;
Keep Rater NumAgree _Kappa_ L_Kappa U_Kappa totCap PercentAgree; run; quit;

/* Combine the dataset that holds all the raters separately with the overall
calculation. */

Data OverallSummery; set means OverallFinal; run; quit;

/* Combine all the raters into one table. This is all based on ALL videotape. */
data combineNoV; set tableconst12 tableconst13 tableconst14 tableconst23
tableconst24 tableconst34;
REName Frequency = NumAgree Rater1 = Rater; run; quit;

/* Delete datasets */

proc datasets nodetails; delete Clean Dummy Kappafile Mappings Store Tableconst12
Tableconst13 Tableconst14 Tableconst23 Tableconst24 Tableconst34 Testab Testac Testad
Testbc Testbd Testcd Two; run; quit;

/* Delete any videos with no agreements. */

Data combineNoV; set combineNoV; if NumAgree = 0 then delete;
PercentAgree = NumAgree/totCap; run; quit;

/* Sum up the number of agreements, the total number of comparisons, the Kappa
statistic and its two CI intervals. This is combining all the information
corresponding to the videos. Then output all the means */

Proc means data=combineNoV noprint; var NumAgree totCap _Kappa_ L_Kappa U_Kappa;
output out = means; run; quit;

/* Delete all extra information that comes from PROC Means. Also calculate the
percent agreement. */
Data means; set means; if (_STAT_ = 'MEAN'); PercentAgree = NumAgree/totCap;
Rater = 100;
Keep Rater NumAgree _Kappa_ L_Kappa U_Kappa totCap PercentAgree; run; quit;

```

```

/* Combine the dataset that holds all the raters separately with the overall
calculation. */

Data OverallNoVSummery;      set means combineNoV;   run; quit;

/* Delete datasets */

Proc datasets nodetails;    delete Means Overallfinal Combinenov;   run; quit;

/* Create a format so it is eaiser to read the output. */

PROC FORMAT;
VALUE FRater
12='Rater 1 vs. Rater 2'
13='Rater 1 vs. Rater 3'
14='Rater 1 vs. Rater 4'
23='Rater 2 vs. Rater 3'
24='Rater 2 vs. Rater 4'
34='Rater 3 vs. Rater 4'
100='Overall'      ;   run; quit;
ods rtf; ods listing close;

/* Print out our final sets of output.  These two outputs correspond to taking into
account the videos. */

proc sort data=combineVids; by Rater;
proc print data=combineVids split = '*';
var video rater numagree totcap percentagree _kappa_ l_kappa u_kappa;

label          numagree = ' number of *agreememt'
totcap        = ' total tasks'
percentagree='percentage of * agreement'
_kappa_      = 'Kappa agreement'
l_kappa      = ' lower 95%*CI kappa'
u_kappa      = 'Upper 95%*CI Kappa' ;
title 'Raters along individual videos';      format Rater FRater.;   run; quit;

proc sort data=OverallSummery; by Rater;
proc print data=OverallSummery split='*';

var  rater numagree totcap percentagree _kappa_ l_kappa u_kappa;
label          numagree = ' number of *agreememt'
totcap        = ' total tasks'
percentagree='percentage of * agreement'
_kappa_      = 'Kappa agreement'
l_kappa      = ' lower 95%*CI kappa'
u_kappa      = 'Upper 95%*CI Kappa';
title 'Raters along individual videos - Summed';      format Rater FRater.;   run; quit;

/* Print out our final sets of output  This output corresponds to ignoring the
separate videos. */

proc sort data=OverallNoVSummery; by Rater;
proc print data=OverallNoVSummery split='*';
var  rater numagree totcap percentagree _kappa_ l_kappa u_kappa;
label          numagree = ' number of *agreememt'
totcap        = ' total tasks'
percentagree='percentage of * agreement'
_kappa_      = 'Kappa agreement'
l_kappa      = ' lower 95%*CI kappa'
u_kappa      = 'Upper 95%*CI Kappa';
title 'Raters ignoring individual videos - Summed';      format Rater FRater.;

```

```
run; quit;  
ods rtf close; ods listing; quit; run;
```