EDUCATIONAL MEASUREMENT (EPSY 6303) - Vinh Nguyen (vinh.nguyen@ttu.edu)

ASSIGNMENT 5

1. **What is the level of exact agreement between each pair of raters? (Hint: Use Cohen's Kappa or weighted Kappa).**

ANSWER:

Code

```
================================================================

install.packages("irr")
install.packages("foreign")
install.packages("haven")
library("irr")
library("foreign")
library("haven")
assigment512 <-assigment5[,c(1,2)]
assigment513 <-assigment5[,c(1,3)]
assigment514 <-assigment5[,c(1,4)]
assigment523 <-assigment5[,c(2,3)]
assigment524 <-assigment5[,c(2,4)]
assigment534 <-assigment5[,c(3,4)]
kappa2(assigment512, weight="equal")
kappa2(assigment513, weight="equal")
kappa2(assigment514, weight="equal")
kappa2(assigment523, weight="equal")
kappa2(assigment524, weight="equal")
kappa2(assigment534, weight="equal")
```

==========

References:

Cohen (1960) suggested the following guidelines

- k<= 0 => no agreement,
- BTW 0.001 and 0.20 => no to slight agreement
- 0.21 to 0.40 => fair agreement
- 0.41 to 0.60 => moderate agreement
- 0.61 to 0.80 => substantial agreement
- 0.81 to 1.00 => nearly perfect agreement

Our result

Raters	1	2	3
1			
2	0.487		
3	0.414	0.544	
4	0.32	0.404	0.484

Most of the level of exact agreement falls between 0.4 to 0.6 which is moderate agreement suggested by Cohen (1960), except one pair (rater 1 vs rater 4) which has fair level of agreement (0.32).

2. **What is the level of exact agreement across the 4 raters? (Hint: Use Fleiss Kappa)? .**

```
> kappam.fleiss(assigment5,detail = TRUE)
 Fleiss' Kappa for m Raters

 Subjects = 3376
   Raters = 4
    Kappa = 0.304

        z = 73.4
  p-value = 0

   Kappa      z p.value
0  0.307 43.731   0.000
1  0.420 59.776   0.000
2  0.158 22.492   0.000
3  0.202 28.757   0.000
4  0.452 64.353   0.000
```

Result from the experiment showed that the level of exact agreement across the 4 raters is 0.304 which is **the fair agreement**

3. **Calculate ICCs.**

    a. The 4 raters will be the only individuals scoring and we are primarily interested in average score and agreement among the raters, as opposed to consistency

```
> icc(assigment5, model = c("oneway"), type=("agreement"), unit=c("average"))
 Average Score Intraclass Correlation

   Model: oneway
   Type : agreement

   Subjects = 3376
     Raters = 4
     ICC(4) = 0.833

 F-Test, H0: r0 = 0 ; H1: r0 > 0
F(3375,10128) = 6 , p = 0

 95%-Confidence Interval for ICC Population Values:
  0.824 < ICC < 0.842
```

The ICC score is 0.833 which is an excellent score (falls in range 0.75 and 1.00)

b. The 4 raters were NOT the only people who could provide the scores and we are primarily interested in average score and agreement among the raters, rather than the consistency

```
> icc(assigment5, model = c("twoway"), type=("agreement"), unit=c("average"))
 Average Score Intraclass Correlation

   Model: twoway
   Type : agreement

   Subjects = 3376
     Raters = 4
   ICC(A,4) = 0.835

 F-Test, H0: r0 = 0 ; H1: r0 > 0
F(3375,1152) = 6.32 , p = 1.12e-226

 95%-Confidence Interval for ICC Population Values:
  0.818 < ICC < 0.849
```

The ICC score is 0.835 which also indicates an excellent score (falls in range 0.75 and 1.00)

c. The 4 raters were NOT the only people who could provide the scores and we are primarily interested in average score and consistency among the raters, rather than the agreement.

```
> icc(assigment5, model = c("twoway"), type=("consistency"), unit=c("average"))
 Average Score Intraclass Correlation

   Model: twoway
   Type : consistency

   Subjects = 3376
     Raters = 4
   ICC(C,4) = 0.842

 F-Test, H0: r0 = 0 ; H1: r0 > 0
F(3375,10125) = 6.32 , p = 0

 95%-Confidence Interval for ICC Population Values:
  0.833 < ICC < 0.85
```

The ICC score is 0.842 which also indicates an excellent score (falls in range 0.75 and 1.00)

4. Write up a brief summary of your findings

Overall, all ICCs are really high (between 0.75 and 1.0) which indicate an excellent inter-rater agreement. Of these scores, the ICC is somewhat larger when we are interested in average score and consistency among the raters, rather than the agreement.