More Examples of KR21 and KR20

Sec 01 Data - A great example of very poor reliability

KISS Information Standard Skills Exam (KISSE) Questions.									
Subject	Q1	Q2	Q3	Q4	Q5	Q6		X	X^2
s1	1	0	1	0	0	1		3	9
s2	1	0	1	1	1	0		4	16
s3	0	1	0	0	0	1		2	4
s4	0	1	1	1	0	0		3	9
s5	1	1	1	0	1	0		4	16
p	0.6	0.6	0.8	0.4	0.4	0.4	Σ	16	54
q	0.4	0.4	0.2	0.6	0.6	0.6	Mean	3.2	
pq	0.24	0.24	0.16	0.24	0.24	0.24			

KR21 - Steps

Step 1 - Find the Mean of respondents' totaled scores (X).

Note: each subject's score for X represents the number of questions they got Correct.

$$\overline{X} = \frac{\sum X}{n} = \frac{16}{5} = 3.2$$

Step 2 - Find the Variance (σ^2) for respondents' totaled scores.

- Computational Example:

$$\sigma^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{n}}{n} = \frac{54 - \frac{16^2}{5}}{5} = \frac{54 - \frac{256}{5}}{5} = \frac{54 - 51.2}{5} = \frac{2.8}{5} = .56$$

- Mean Deviation Example of Finding Variance:

Mean Deviation Example of Finding Variance			
	X	(x-mean)	$(x\text{-mean})^2$
	3	-0.2	0.04
	4	0.8	0.64
	2	-1.2	1.44
	3	-0.2	0.04
	4	0.8	0.64
Σ	16		2.8
Mean	3.2		
σ^2 =	(2.8)/5 =	0.56	

$$\sigma^2 = \frac{\sum \left(X - \overline{X}\right)^2}{n} = \frac{2.8}{5} = .56$$

Step 3 - Plug and Chug:

$$KR21 = \left(\frac{K}{K-1}\right)\left(1 - \frac{\overline{X}(K - \overline{X})}{K(\sigma^2)}\right) = \left(\frac{6}{6-1}\right)\left(1 - \frac{3.2(6 - 3.2)}{6(.56)}\right) = \left(\frac{6}{5}\right)\left(1 - \frac{3.2(6 - 3.2)}{6(.56)}\right) = (1.2)\left(1 - \frac{3.2(6 - 3.2)}{6(.56)}\right)$$

$$(1.2)\left(1 - \frac{3.2(2.8)}{6(.56)}\right) = (1.2)\left(1 - \frac{8.96}{6(.56)}\right) = (1.2)\left(1 - \frac{8.96}{3.36}\right) = (1.2)(1 - 2.6667) = (1.2)(-1.6667) = -2.0000$$

- The reliability for this measure (with this data) is appallingly poor. In fact statistic is uniterpretable.

Sec 01 Data Continued - KR20

KISS Information Standard Skills Exam (KISSE) Questions.									
Subject	Q1	Q2	Q3	Q4	Q5	Q6		X	X^2
s1	1	0	1	0	0	1		3	9
s2	1	0	1	1	1	0		4	16
s3	0	1	0	0	0	1		2	4
s4	0	1	1	1	0	0		3	9
s5	1	1	1	0	1	0		4	16
p	0.6	0.6	0.8	0.4	0.4	0.4	Σ	16	54
q	0.4	0.4	0.2	0.6	0.6	0.6	Mean	3.2	
pq	0.24	0.24	0.16	0.24	0.24	0.24			

- KR20 is basically the same as 21. However, KR20 makes use of data from each question.
Specifically - KR20 looks at the number of people who got each question correct and the number who got it incorrect.
p = Number of Correct Responses for an item (question) / Number of People taking the test.
q = Number of Incorrect Responses for an item (question) / Number of People taking the test.
Note: Since p is a ratio (# correct/n), q can be defined as (1-p).
- In the data above:
p for item 1 is .60: 3 people got it correct, 5 people took the text, so 3/5 = .60
q for item 1 is .40: 2 people got it correct, 5 people took the text, so 2/5 = .40
Or q = (1-p) = 1-.6 = .4.

- Looking at the KR20 Formula:

$$KR20 = \left(\frac{K}{K-1}\right)\left(1 - \frac{\sum pq}{\sigma^2}\right)$$

- The formula instructs us to multiply the *p* and *q* values for each item (we obtain *pq*) and then sum the *pq* values across all of the times of the test (we obtain Σ*pq*).
pq for item 1 is .24: *p*q* = .6 * .4 = .24
pq for item 2 is .24: *p*q* = .6 * .4 = .24
pq for item 3 is .16: *p*q* = .8 * .2 = .16
pq for item 4 is .24: *p*q* = .4 * .6 = .24
pq for item 5 is .24: *p*q* = .4 * .6 = .24
pq for item 6 is .24: *p*q* = .4 * .6 = .24

Σ*pq* = .24 + .24 + .16 + . 24 + . 24 + .24 = 1.36

- Plug & Chug:

$$KR20 = \left(\frac{K}{K-1}\right)\left(1 - \frac{\sum pq}{\sigma^2}\right) = \left(\frac{6}{6-1}\right)\left(1 - \frac{1.36}{.56}\right) = \left(\frac{6}{5}\right)\left(1 - \frac{1.36}{.56}\right) = (1.2)\left(1 - \frac{1.36}{.56}\right) =$$

$$(1.2)(1 - 2.4286) = (1.2)(-1.4286) = -1.7143$$

- Again we get a terrible estimate of reliability for this data. The reason for this is that the responses to the questions are essentially random. This generates tremendous Between item variability (error) with very little between subject variability (everyone pretty much got the same score). In extreme cases of unreliability KR20 and KR21 can give us uninterpretable results.

Section 02 Data - A good example of Adequate Reliability and the Manner in Which KR21 under estimates reliability compared to KR20.

KISS Information Standard Skills Exam (KISSE) Questions.									
Subject	Q1	Q2	Q3	Q4	Q5	Q6		X	X^2
s1	1	0	1	1	0	1		4	16
s2	0	1	1	1	1	1		5	25
s3	0	1	0	0	0	1		2	4
s4	1	1	1	1	1	1		6	36
s5	0	0	0	0	1	0		1	1
p	0.4	0.6	0.6	0.6	0.6	0.8	Σ	18	82
q	0.6	0.4	0.4	0.4	0.4	0.2	Mean	3.6	
pq	0.24	0.24	0.24	0.24	0.24	0.16			

KR21 -

Step 1 - Find the Mean of the summed scores.

$$\overline{X} = \frac{\sum X}{n} = \frac{18}{5} = 3.6$$

Step 2 - Find the Variance of the summed scores. (See page 1 for an example with the mean deviation formula

$$\sigma^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{n}}{n} = \frac{82 - \frac{(18)^2}{5}}{5} = \frac{82 - \frac{324}{5}}{5} = \frac{82 - 64.8}{5} = \frac{17.2}{5} = 3.44$$

Step 3 - Plug and Chug

$$KR21 = \left(\frac{K}{K-1}\right)\left(1 - \frac{\overline{X}\left(K - \overline{X}\right)}{K\sigma^2}\right) = \left(\frac{6}{6-1}\right)\left(1 - \frac{3.6(6-3.6)}{6(3.44)}\right) = (1.2)\left(1 - \frac{3.6(6-3.6)}{6(3.44)}\right) =$$

$$(1.2)\left(1 - \frac{3.6(2.4)}{6(3.44)}\right) = (1.2)\left(1 - \frac{8.64}{20.64}\right) = (1.2)(1-.4186) = (1.2)(.5814) = .6977$$

In this example - the reliability is far more impressive. You should notice in the raw data above that there is far less variability in scores between items within subjects. Respondent 1, 2, & 4 got most of the questions correct, and Respondent 3 and 5 got most of the questions wrong. This means the amount of error attributable to unreliability present in the responses is far less than the previous example (that gave us negative reliability). Also, notice that there is far more variability between the respondents' totaled scores. Thus, less error + more total variability = greater overall reliability.

Sec -2 Data Continued.

KISS Information Standard Skills Exam (KISSE) Questions.									
Subject	Q1	Q2	Q3	Q4	Q5	Q6		X	X^2
s1	1	0	1	1	0	1		4	16
s2	0	1	1	1	1	1		5	25
s3	0	1	0	0	0	1		2	4
s4	1	1	1	1	1	1		6	36
s5	0	0	0	0	1	0		1	1
p	0.4	0.6	0.6	0.6	0.6	0.8	Σ	18	82
q	0.6	0.4	0.4	0.4	0.4	0.2	Mean	3.6	
pq	0.24	0.24	0.24	0.24	0.24	0.16			

KR20 -

Step 1 - Find p and q for each question

pq for item 1 is .24: $p*q = .4 * .6 = .24$
pq for item 2 is .24: $p*q = .6 * .4 = .24$
pq for item 3 is .16: $p*q = .6 * .4 = .24$
pq for item 4 is .24: $p*q = .6 * .4 = .24$
pq for item 5 is .24: $p*q = .6 * .4 = .24$
pq for item 6 is .24: $p*q = .8 * .2 = .16$

$\Sigma pq = .24 + .24 + .24 + . 24 + . 24 + .16 = 1.36$

Step 2 - Find the Variance for the respondents totaled scores.

$$\sigma^2 = \frac{\sum X^2 - \frac{\left(\sum X\right)^2}{n}}{n} = \frac{82 - \frac{(18)^2}{5}}{5} = \frac{82 - \frac{324}{5}}{5} = \frac{82 - 64.8}{5} = \frac{17.2}{5} = 3.44$$

Step 3 - Plug and Chug :

$$KR20 = \left(\frac{K}{K-1}\right)\left(1 - \frac{\sum pq}{\sigma^2}\right) = \left(\frac{6}{6-1}\right)\left(1 - \frac{1.36}{3.44}\right) = (1.2)(1-.3953) = (1.2)(.6047) = .7256$$

Again, the reliability for this data set is far superior to what was found for the first data set. Also notice that the KR20 is a little higher than the KR21 for the exact same data. This illustrates the observation that KR21 typically underestimates the reliability of a test. In this example the underestimation is not terribly large.