

## Validity of Classification

### Part 1: Answer - Reliability (stability) of Classifications

Create this data file (1 = master; 0 = non master)

Test	Retest
1	1
0	1
0	0
1	0
1	1
0	1
0	0
1	1
1	1
0	0

Determine the proportion of agreement and Interpret

Proportion of Agreement =  $.30 + .40 = .70$

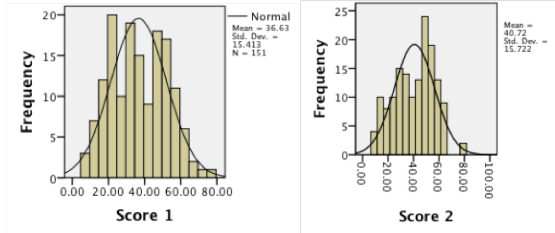
		retest classification		Total
		nonmaster	master	
test classification	nonmaster	30.0%	20.0%	50.0%
	master	10.0%	40.0%	50.0%
Total		40.0%	60.0%	100.0%

Interpretation: 70% of the classifications made from test to retest were in agreement.

### Part 2: Answer - Reliability (stability) of Classifications

Use the Hoola Data File

Set a good cut score that can be used to classify individuals as masters & non masters



Decided to try a cut score of 50. Worked fairly well. Proportion of agreement = .82. However, note that performance improved in general from hoola 1 to hoola 2 so, some (22) individuals classified as non masters based on their hoola 1 score are then classified as masters based on their hoola 2 scores.

		hoola 2 classification			
		nonmaster	master	Total	
hoola 1 classification	nonmaster	Count	91	22	113
		% of Total	60.3%	14.6%	74.8%
	master	Count	6	32	38
		% of Total	4.0%	21.2%	25.2%
Total	Count	97	54	151	
	% of Total	64.2%	35.8%	100.0%	

Recode scores to classifications based on cut score selected.

Determine both the proportion of agreement and Kappa;

$$Pag = .603 + .212 = .815.$$

		Value	Asymp. Std. Error <sup>a</sup>	Approx. T <sup>b</sup>	Approx. Sig.
Measure of Agreement	Kappa	0.568	0.071	7.203	0.000
N of Valid Cases	151				

Interpretation: Stability of classifications not good. While Pag high, stability with chance taken into account = .568. This suggests that more needs to be done to get a more accurate measure of ability at the outset. Eg. provide practice to minimize learning effect, improve counting perhaps by having an observer count.

Moving cut score up to 60, improves Pag but Kappa remains low.

### Part 3: Answer: Validity of Classifications

Create this data file

Test	Criterion (expert's classifications)
M	M
NM	M
M	M
M	NM
M	M
NM	NM
NM	NM

Determine concurrent validity; Interpret

		criterion classification		Total
		nonmaster	master	
test classification	nonmaster	2	1	3
	master	1	3	4
Total		3	4	7

		Value	Approx. Sig.
Nominal by Nominal	Phi	0.417	0.270
	Cramer's V	0.417	0.270
N of Valid Cases		7	

Interpretation: Based on Phi (.42), concurrent validity weak. However, note that most people are in the main diagonal which is where there is a match between the expert's evaluation and the test's evaluation of performance. Difficult to draw firm conclusions since group size so small.