

Cohen's Kappa

EDP 612 Week 8

Dr. Abhik Roy



Reliability and Validity

Reliability

being consistent

Validity

on target

Recall Scenarios

Not **Reliable** and *not* **Valid**



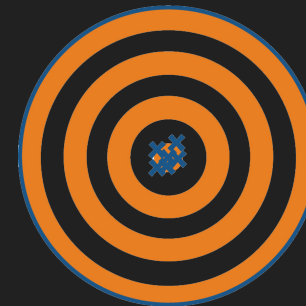
Reliable but *not* **Valid**



Not **Reliable** but **Valid**



Reliable and **Valid**



Basic Tenant



A test can be **Reliable** without being **Valid**

A test cannot be **Valid** unless it is **Reliable**

Reliability

Intra-rater Reliability

the degree of agreement between different measurements done by the same person

Inter-rater Reliability

degree of agreement between between different measurements done by multiple people

Cohen's Kappa κ

- *Officially.* Measure of the agreement between two raters who each classify N items into C mutually exclusive categories
- *Basic idea.*
 - Quantitative measure of reliability for two raters that are rating the same thing
 - With a correction for how often that the raters may agree by chance
- *Lay terms.* Measure of how well do different people agree

Evaluating



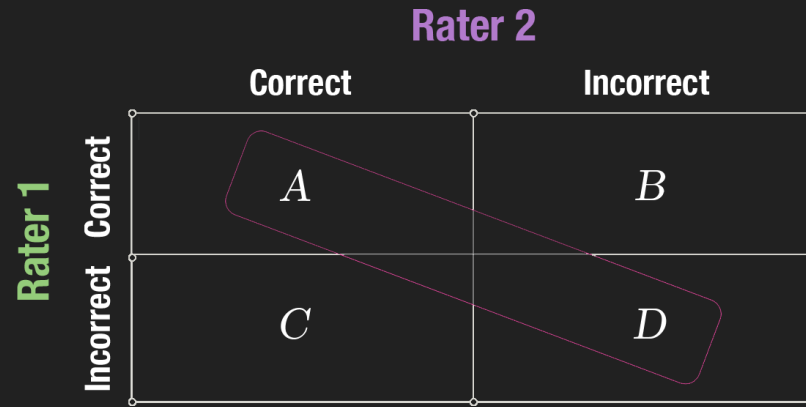
$$\kappa < 0$$

random agreement among raters

$$\kappa = 1$$

complete agreement among raters

Decision Matrix



A

Agreement

The total number of instances that both Raters said were correct

B

Disagreement

The total number of instances that Rater 2 said was incorrect, but Rater 1 said were correct

C

Disagreement

The total number of instances that Rater 1 said was incorrect, but Rater 2 said were correct

D

Agreement

The total number of instances that both Raters said were incorrect

Calculations



Probability of *Agreement* P_0



Number in Full *Agreement* / Total

$$\frac{A + D}{A + B + C + D}$$

Probability of Correct Random *Agreement* $P_{correct}$

Number Correct in Full or Partial *Agreement* / Total

$$\frac{A + B}{A + B + C + D} \cdot \frac{A + C}{A + B + C + D}$$

Probability of Incorrect Random *Agreement* $P_{incorrect}$



Number Incorrect in Full or Partial *Agreement* / Total

$$\frac{C + D}{A + B + C + D} \cdot \frac{B + D}{A + B + C + D}$$

Probability of Random *Agreement*

$$P_e = P_{correct} + P_{incorrect}$$

Kappa κ

$$\kappa = \frac{P_0 - P_e}{1 - P_e}$$

Interpretation



Value of κ	Strength of Agreement
≤ 0.20	Poor
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Good
≥ 0.81	Very Good

That's it!

Any questions?

