Lecture 4 - Introduction to Reliability

Goal of psychological measurement – detect psychological differences

↳ to what degree are differences in observed test scores consistent with differences in the true levels of a psychological trait?

↳ one way to answer this is to assume an underlying mathematical model of our measurements

↳ Classical test theory

* Lord & Novick (1968)
* Allen & Yen (1979)

In CTT, we assume all observed scores X can be written as the sum of a true score T and an error score E

$$X = T + E$$

Further we assume:

* error scores are random
* the mean of the error scores is 0
* the correlation between T and E is 0

In general, true scores are latent — that is, they cannot be observed. But for now, let us ignore this and consider the following example:

Participant	Observed score X	$=$ True score T	$+$ Error score E
A	115	125	−10
B	135	115	20
C	95	105	−10
D	85	95	−10
E	105	85	20
F	65	75	−10
Mean	100	100	0
Variance	590	350	240
SD	24.29	18.708	15.492

Some things to note: (check these in JASP)

(1) mean error score = 0

(2) the correlation between T and E is 0.

(3) $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$

↳ why is this true?

– think about the variance of a composite!

So what about reliability?

Two definitions:

(1) reliability index:

ρ_{XT} = correlation between observed scores X and true scores T

(2) reliability coefficient:

$$\rho_{xx'} = \frac{\text{true score variability}}{\text{observed score variablity}}$$

$$= \frac{\sigma_T^2}{\sigma_X^2} = \frac{\text{signal}}{\text{signal + noise}}$$

Let's compute these for our example in JASP:

(1) reliability index: $\rho_{XT} = 0.77$

↳ but notice: $\frac{\sigma_T}{\sigma_X} = \frac{18.708}{24.290} = 0.77$ also!

↳ Does this always happen?

$\boxed{\text{Yes – you can prove that under the assumptions of CTT, } \rho_{XT} = \frac{\sigma_T}{\sigma_X}}$

(2) reliability coefficient

$$\rho_{xx'} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{350}{590} = 0.59$$

Interpretation: 59% of the variability in observed scores is due to variability in the underlying true scores.

→ implies that 41% of the observed variability comes from measurement error!

→ clearly, we want reliability coefficients to be closer to $\underline{1}$ = 100%

Why two measures of reliability?

↳ they are intimately related:

$$\boxed{\rho_{xx'} = \frac{\sigma_T^2}{\sigma_X^2} = \left(\frac{\sigma_T}{\sigma_X}\right)^2 = \rho_{XT}^2}$$

"Reliability = squared correlation between T and X"

Other equivalent ways to think about reliability:

(3) $$\rho_{XX'} = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_X^2 - \sigma_E^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$$

Interpretation: reliability = what is left over after removing error variability.

Example: $$\rho_{XX'} = 1 - \frac{\sigma_E^2}{\sigma_X^2} = 1 - \frac{240}{590}$$
$$= 0.59$$

(4) $$\rho_{XX'} = 1 - \frac{\sigma_E^2}{\sigma_X^2} = 1 - \left(\frac{\sigma_E}{\sigma_X}\right)^2 = 1 - \rho_{XE}^2$$

Interpretation: reliability = lack of correlation between observed scores and error scores.

Example: $\rho_{XX'} = 1 - \rho_{XE}^2 = 1 - (0.638)^2 = 0.59$

<u>Summary:</u> reliability =

(1) proportion of variance in observed scores that is due to variability among true scores.

$$\rho_{XX'} = \frac{\sigma_T^2}{\sigma_X^2}$$

(2) squared correlation between observed and true scores

$$\rho_{XX'} = \rho_{XT}^2$$

(3) what is left over after removing error variability

$$\rho_{XX'} = 1 - \frac{\sigma_E^2}{\sigma_X^2}$$

(4) lack of correlation between observed and error scores

$$\rho_{XX'} = 1 - \rho_{XE}^2$$

What's it good for?

* beyond being a diagnostic for your test, it turns out that reliability can directly index the expected measurement error of your test!

Recall that reliability = lack of error variance

$$\rho_{xx'} = 1 - \frac{\sigma_E^2}{\sigma_x^2}$$

Solve for σ_E:

$$\frac{\sigma_E^2}{\sigma_x^2} = 1 - \rho_{xx'}$$

$$\longrightarrow \quad \sigma_E^2 = \sigma_x^2 \left(1 - \rho_{xx'}\right)$$

$$\longrightarrow \quad \sigma_E = \sigma_x \sqrt{1 - \rho_{xx'}}$$

For our example: $\rho_{xx'} = 0.59$. So

$$\sigma_E = 24.29 \sqrt{1 - 0.59} = 15.55 \text{ points}$$

Thus 95% margin of error $= 1.96\,\sigma_E = 1.96(15.55)$

$= 30.48$ points