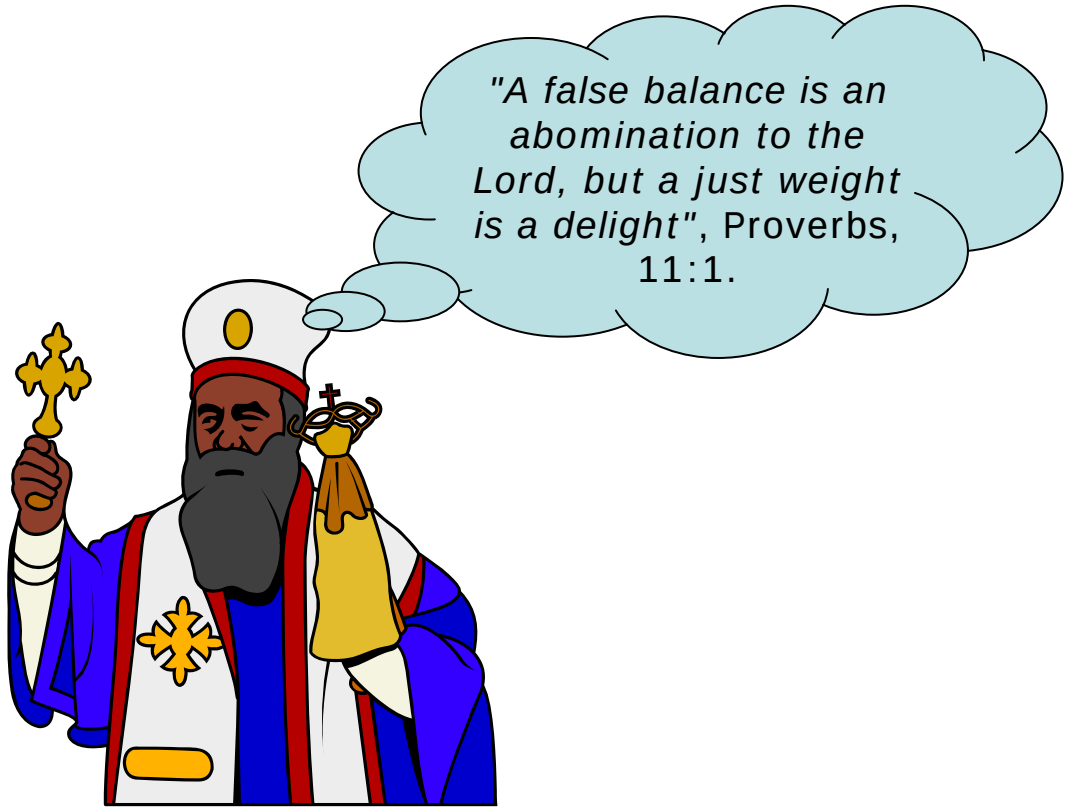
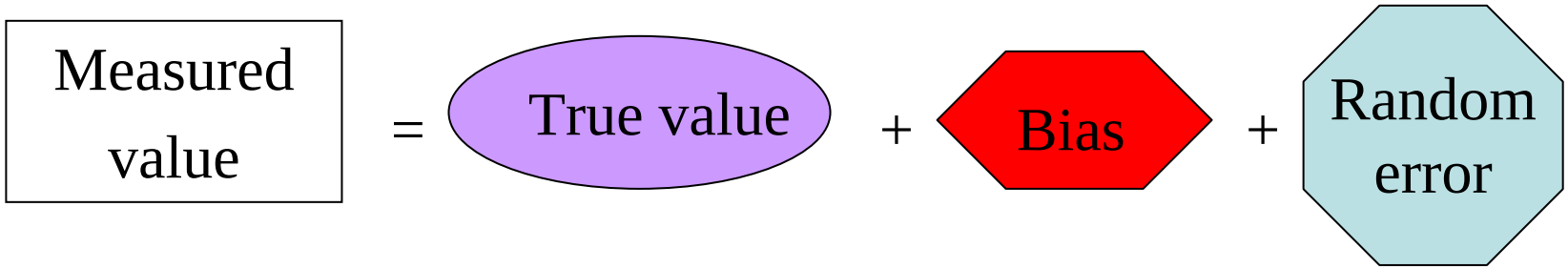
Scales, Scaling and Indices



Measurement error



Measured value — what you see or record

True value — true underlying state of affairs

Bias — (systematic error) things measured unintentionally

- response sets
- yea saying
- memory

Random error — response errors that are as likely to be big/small + as big/small -

Classical Test Theory (CTT)

$$X_i = T_i + e_i$$

Observed score X for individual i is equal to i's true score (T) plus random measurement error (e)

Assumptions: $E(e) = \rho_{te} = \rho_{e_1 t_2} = \rho_{e_1 e_2} = 0$

1) Average value of the random error is zero;

2) True score and error values are uncorrelated;

3) Error score of one measure uncorrelated with true score of another;

4) Error score of one measure uncorrelated with true score of another.

3) and 4) important in context of summated scales

CTT

From which it follows that: $E(X) = E(T)$

Three statisticians
went duck hunting. A duck was
approaching and the first
statistician shot, and missed
the duck by being a foot too
high. The second shot and was a
foot too low. The third
cried, “We hit it!”

CTT

But what about the variance of X?

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2 + 2\sigma_{te}$$

but by assumption $\sigma_{te} = 0$

So:

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2$$

And

$$\text{Reliability} = \frac{\sigma_t^2}{\sigma_x^2}$$

Reliability is the ratio of true score variance to observed score variance;

The numerator is unobserved!

CTT

Q. How to estimate reliability? A: With at least 2 measures and some assumptions.

Minimal assumption: $\sigma_e^2 = \sigma_{e'}^2$ Bizarrely called “tau equivalence”

(If in addition we assume E(T) = E(T’) then the measures are said to be “parallel”)

Define the correlation between the measures as: $\rho_{xx'} = \frac{\sigma_{xx'}}{\sigma_x \sigma_{x'}} \equiv \frac{Cov(xx')}{SD(x)SD(x')}$

$$\equiv \frac{\sigma_t^2 + \sigma_{te} + \sigma_{te'} + \sigma_{ee'}}{\sigma_x \sigma_{x'}} \equiv \frac{\sigma_t^2}{\sigma_x^2}$$

So in the CTT set up with parallel measures or tau equivalence reliability is estimated by the correlation between the measures (which probably isn’t too surprising!)

What does it matter?

If you believe all the assumptions then: $\sigma_t^2 = \sigma_x^2 \rho_{xx'}$

Which could be good to know if you wanted to “correct” for (random) measurement error.

It is also worthwhile noting that the correlation between two variables – say X and Y can never be greater than their reliabilities.

Poor measurement means the estimated relationship between variables is attenuated.

What about bias?

$$X_i = T_i + B_i + R_i$$

Where B is bias and R is random error

$$\mathrm{E(X)} = \mathrm{E(T)} + \mathrm{E(B)}$$

Take expectations

$$\sigma_x^2 = \sigma_t^2 + \sigma_b^2 + \sigma_r^2 + 2\sigma_{tb} + 2\sigma_{tr} + 2\sigma_{br}$$

$$= \sigma_t^2 + \sigma_b^2 + \sigma_r^2 + 2\sigma_{tb}$$

Think about variances

$$\text{Reliability} = \frac{\sigma_t^2 + \sigma_b^2 + 2\sigma_{tb}}{\sigma_x^2}$$

$$\text{Validity} = \frac{\sigma_t^2}{\sigma_x^2}$$

(In the absence of bias, reliability and validity are the same thing)

Internal Consistency

Q. How do we examine the reliability of multiple indicators – say a set of items that will form a summated scale?

A. Various “split-half” techniques and variants and derivatives thereof one of which is a measure first proposed by Guttman but most commonly called Cronbach’s alpha

$$\alpha = \frac{N}{N-1}\left[1 - \frac{\Sigma\sigma^2(Y_i)}{\sigma_x^2}\right]$$

Where:

N = number of items

Numerator = sum of item variances

Denominator = variance of the composite scale

Cronbach's Alpha

If the items have roughly equal variances then an approximation is valid

$$\alpha = \frac{N\bar{r}}{1+(N-1)\bar{r}}$$

where:

N = number of items in the scale

$\bar{r}$ = average inter-item correlation

Gives a number between 0 and 1 – higher is "better".

Cronbach's alpha

Alpha approaches 1 (at a diminishing rate) as more items are added as long as the addition of an item does not reduce the average inter-item correlation

Alpha is widely and ritualistically used in "snoopy blanket" fashion.

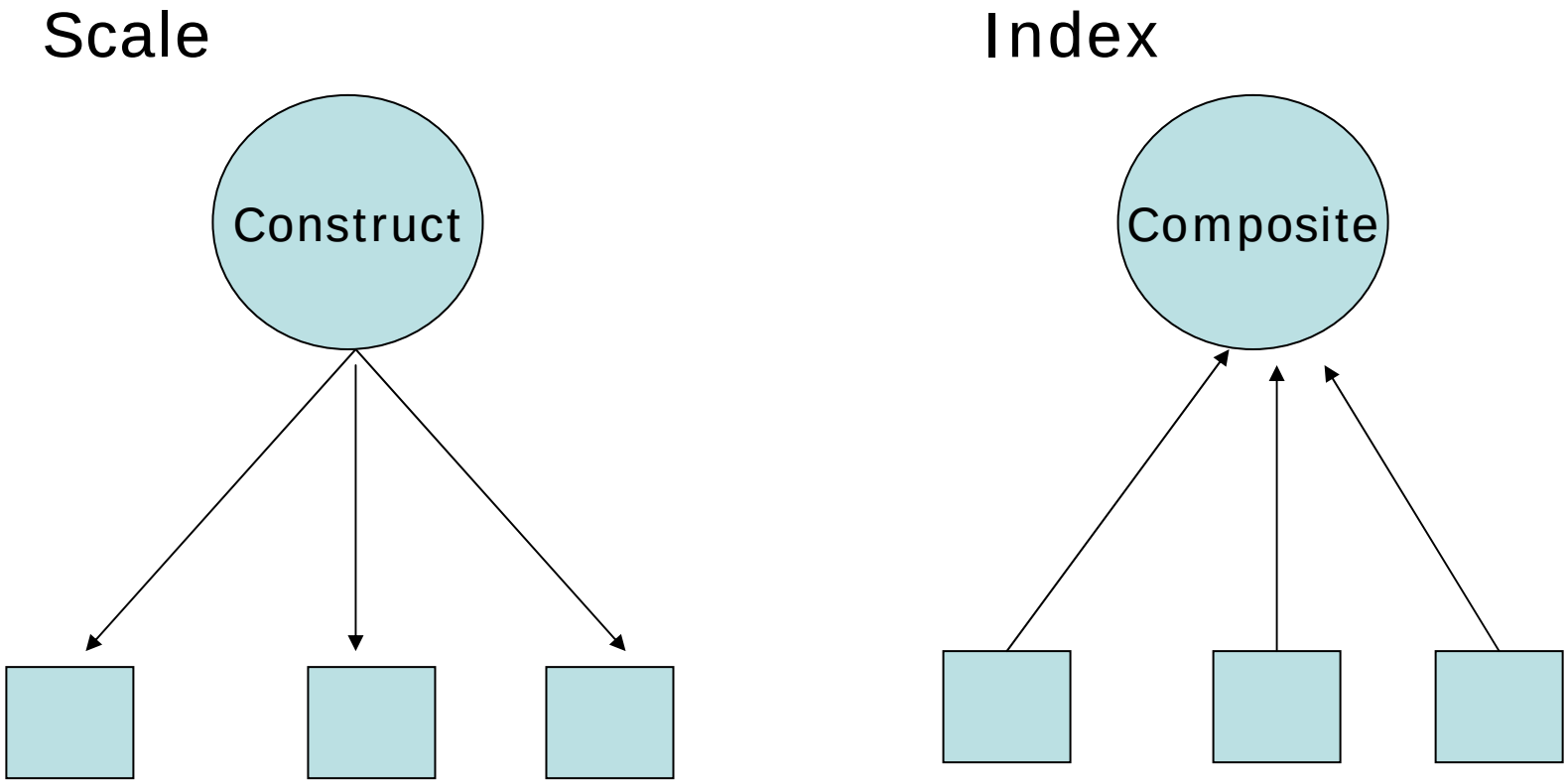
Is best regarded as setting a lower bound on reliability – which was clearly Guttman's original intention.

Is possible to get identical alphas from correlation matrices with very different structures.

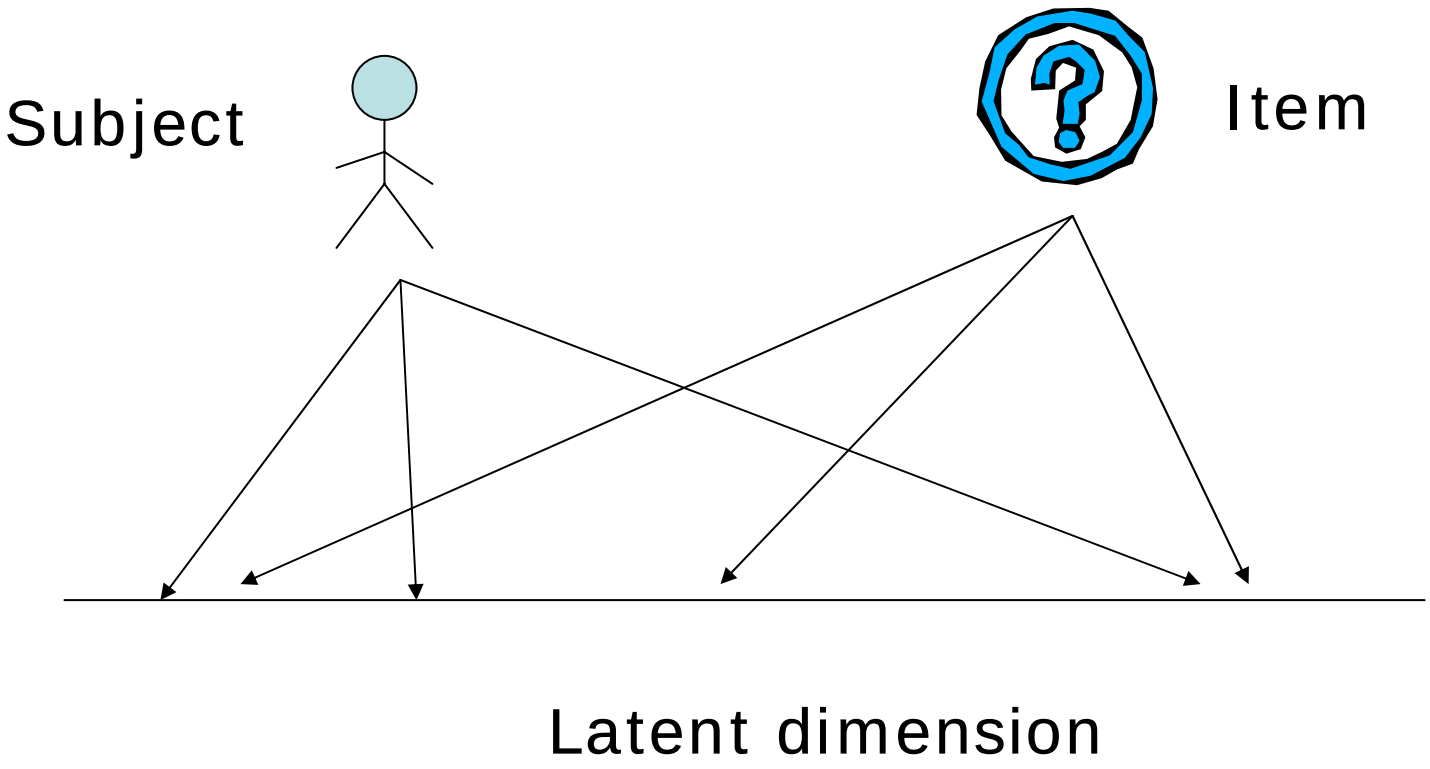
There are (arguably) better (model based) alternatives.

Rules of thumb about "acceptable" values of alpha are largely fatuous (though editors and referees like them)

Scale and Index



Measuring subjective phenomena objectively



Thurstone scale items

- Underlying trait to be measured:
  - Favourability towards the EU
- Typical items
  - I believe the EU is a powerful agency for promoting the economic welfare of all Europeans.
  - I believe each European state should be free to defend its interests without excessive interference from Brussels.
  - I think the EU is an affront to democracy.
  - etc
- Response scale
  - Respondent to endorse **only** those items she agrees with.

Steps to make Thurstone scale

- Create an item pool - items spanning the complete favourable-unfavourable range - N circa 100.
- Circa 50 judges sort items into say 11 sets. Sets are to be 'equal interval‘ along the target dimension.
  - Sets are scored say 1 to 11.
  - Items receive the median score over the 50 judgments.
- Items are with large variance are discarded.
  - Select about 20 which over the entire range
- Items put into questionnaire (in random order).
- Respondent asked to endorse only those they agree with .
- Respondent's scale score set equal to to average of the values of the endorsed items.

Guttman scaling

- Item responses should be reproducible from total scores
- Respondent replies in affirmative to all items below a threshold
- Above the threshold respondent replies to all items in the negative

Management should turn a blind eye to…

- Taking home pens and stationery?
- Making occasional private phone calls?
- Using a company computer to make a private purchase from the net?
- Sending personal greetings cards in the company mail?

Guttman Scale

A Perfect Guttman Scale

Items

Subjects	1	2	3	4	5	6	Scale Score
A	1	1	1	1	1	1	6
B	1	1	1	1	1	0	5
C	1	1	1	1	0	0	4
D	1	1	1	0	0	0	3
E	1	1	0	0	0	0	2
F	1	0	0	0	0	0	1
G	0	0	0	0	0	0	0

Four-item Guttman scale with two types of error assignment

	Response Pattern				Assignment of Error: Guttman	Assignment of Error: Edwards-GI
*	+	+	+	+	0	0
*	+	+	+	-	0	0
	+	+	-	+	1	2
*	+	+	-	-	0	0
	+	-	+	+	1	2
	+	-	+	-	1	2
	+	-	-	+	1	2
*	+	-	-	-	0	0
	-	+	+	+	1	2
	-	+	+	-	1	2
	-	+	-	+	2	2
	-	+	-	-	1	2
	-	-	+	+	2	4
	-	-	+	-	1	2
	-	-	-	+	1	2
*	-	-	-	-	0	0

* Perfect scale type

Summated ratings (Likert) scale

- Likert items
  - The EU is an affront to democracy.

    1. Agree strongly
    2. Agree
    3. Neither agree nor disagree
    4. Disagree
    5. Disagree strongly
  - The EU works in the best interests of all Europeans

    1. Agree strongly
    2. Agree
    3. Neither agree nor disagree
    4. Disagree
    5. Disagree strongly

Summated ratings

- For each individual add up the scores for set of items
  - Make sure (higher/lower scores) mean the same thing!
- Total score for subject is her scale value
- Looks as though it shouldn’t work
  - But often more reliable than Thurstone or Guttman scales