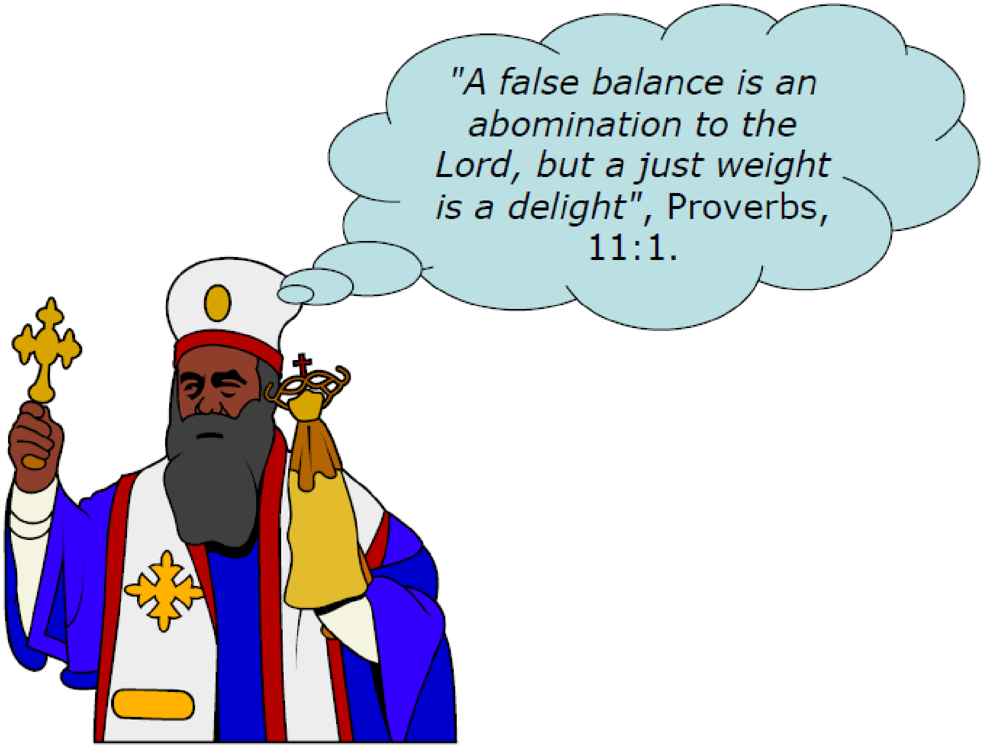
Measurement 2

http://users.ox.ac.uk/~sfos0015


"A false balance is an abomination to the Lord, but a just weight is a delight", Proverbs, 11:1.

1. Classical Test Theory

2. Measurement Error

3. Scales

4. Reliability again

Classical Test Theory

$x_i = t_i + e_i$

Observed score on X for individual i is equal to the true score T of individual i plus random measurement error.

$E(e) = 0$

The expected value of the errors is zero.

$\sigma_{te} = 0$

The covariance of the true score and the error is zero.

$E(X) = E(T)$

The expected value of the observed is equal to the expected value of the true score.

Classical Test Theory

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

The variance of the observed score is equal to the variance of the true score plus the variance of the error plus twice the covariance of the true score and the error.

$$\sigma_{te} = 0$$

By assumption.

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

$$\frac{\sigma_t^2}{\sigma_x^2}$$

The correlation of the true score with the observed score. The ratio of true score variance to observed score variance. The proportion of observed score variance that is true score variance. **The Reliability**.

Classical Test Theory

How to estimate reliability

1. Needed:
    1. At least two measures.
    2. Some assumptions.

$$\sigma_e^2 = \sigma_{e'}^2$$

The error variance of the first measure (e) is equal to the error variance of the second measure (e'). Called **tau equivalence**.

$$\rho_{xx'} = \frac{\sigma_{xx'}}{\sigma_x \sigma_{x'}} \equiv \frac{Cov(xx')}{SD(x)SD(x')}$$

The correlation between two measures, x and x', is the ratio of their covariance to the product of their standard deviations.

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

Expand the covariance and eliminate all terms assumed to be zero. If we can assume at least tau equivalence then reliability is estimated by the correlation between two measures.

Classical Test Theory

$$\sigma_t^2 = \sigma_x^2 \rho_{xx'}$$

If you accept the assumptions then it follows that the true score variance is the product of the observed score variance and the inter-item correlation.

Random measurement errors will attenuate the correlation between two variables.

From which it follows that if you know the reliability of a measure you can correct for random measurement error.

Consequences of measurement error

Consider two variables Y and X drawn from a bivariate normal distribution with $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $\rho = 0.4$.

This implies that $y_i = 0.4X_i + \varepsilon_i$. This is the data generating model.

Now add random error to Y so the Y* = Y + E. E is drawn from N(0,1)

This implies that $\mathrm{y}_\mathrm{i}^* = 0.4X_i + \varepsilon_i$.

Adding random measurement error to Y does not affect β (but it does affect the residual variance)

Now add random error to X so the X* = X + E. E is drawn from N(0,1)

This implies that $\mathrm{y}_\mathrm{i} = 0.2X_i^* + \varepsilon_i$.

Adding random measurement error to X does affect β. It biases it towards 0.

Consequences of measurement error

Consider three variables Y, X and Z drawn from a trivariate normal distribution with:

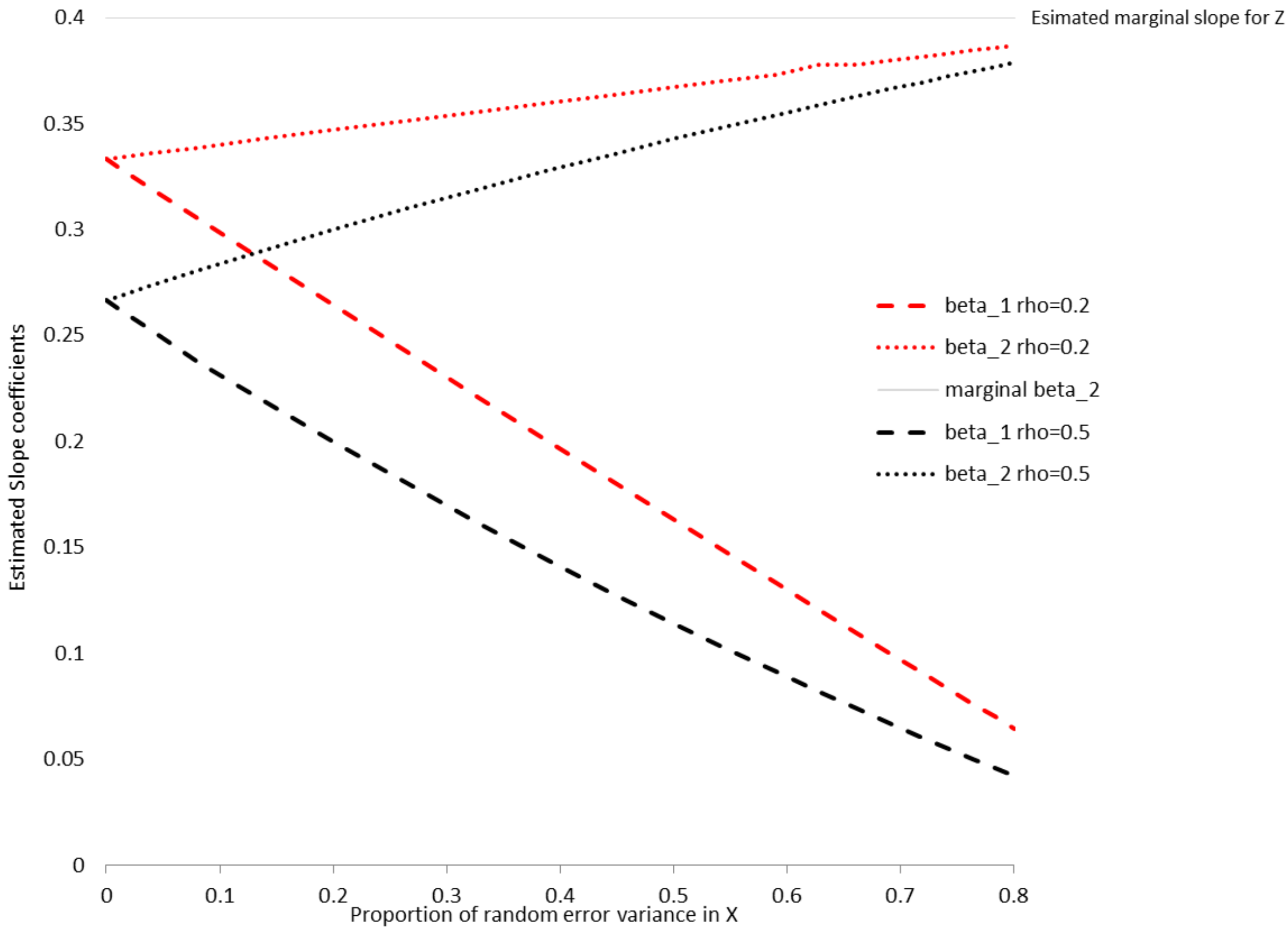
$\mu_1 = \mu_2 = \mu_3 = 0$; $\sigma_1 = \sigma_2 = \sigma_3 = 1$ and $\rho_{YX} = \rho_{YZ} = 0.4$ and $\rho_{XZ} = 0.2$

This implies that $y_i = 0.33X_i + 0.33Z_i + \varepsilon_i$. This is the data generating model.

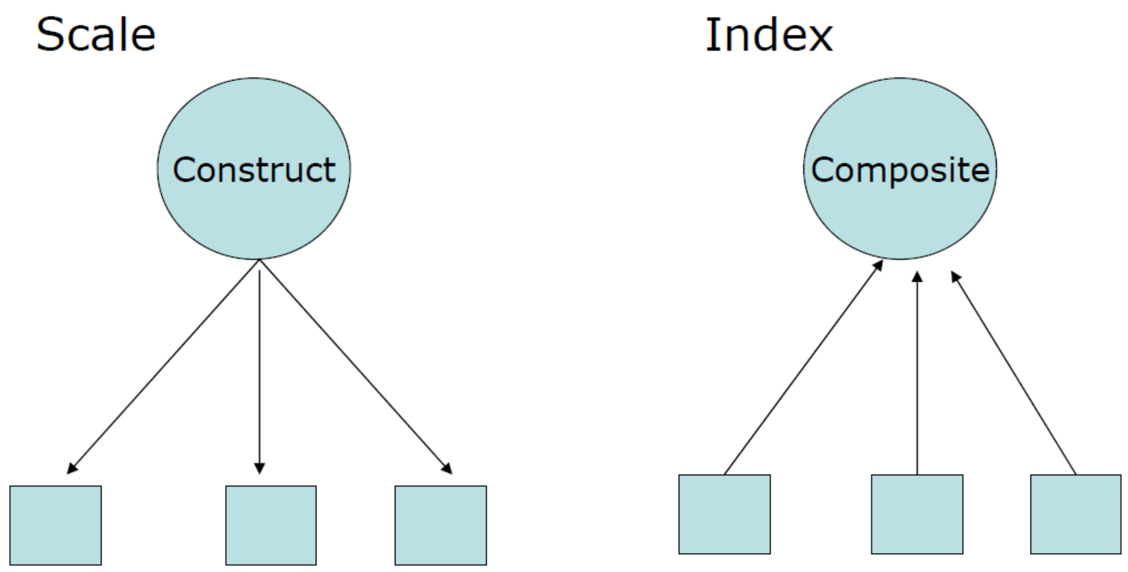
Now add random error to X so the X* = X + E. E is drawn from N(0,1)

This implies that $\mathrm{y_i} = 00.16X_i^* + 0.37Z_i^* + \varepsilon_i$.

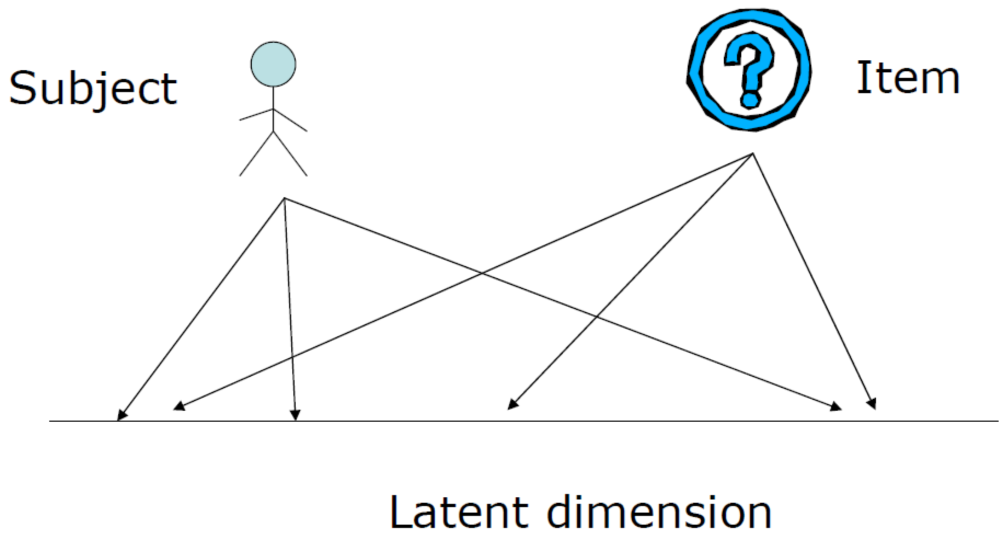
Adding random measurement error to X biases β_1 downwards towards 0 and in this case β_2 upwards. In general β_2 will be driven away from its conditional value towards its marginal value


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Esimated marginal slope for Z
Estimated Slope coefficients
beta_1 rho=0.2
beta_2 rho=0.2
marginal beta_2
beta_1 rho=0.5
beta_2 rho=0.5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Proportion of random error variance in X

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

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

UNIVERSITY OF OXFORD

Four item Guttman scale with errors

	Response Pattern			
*	+	+	+	+
*	+	+	+	-
	+	+	-	+
*	+	+	-	-
	+	-	+	+
	+	-	+	-
	+	-	-	+
*	+	-	-	-
	-	+	+	+
	-	+	+	-
	-	+	-	+
	-	+	-	-
	-	-	+	+
	-	-	+	-
	-	-	-	+
*	-	-	-	-

* 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

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)