Classical Test Theory
Main concepts

Christophe Lalanne
ch.lalanne@gmail.com

November, 2009

Summary

❝ abstract here. . . ❞

© 2009, www.aliquote.org

Outline

1.
2.
3.
4.

Various functions used throughout this chapter were collated in the package **Psychomisc**.

© 2009, www.aliquote.org

[1]

© 2009, www.aliquote.org

The True score model

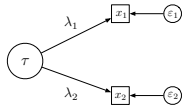
© 2009, www.aliquote.org

Types of test

Several tests are said to be *congenerically* equivalent if all tests may be expressed in terms of one factor and a residual error.

Parallel tests are the special case where (usually two) tests have equal factor loadings ($\lambda_1 = \lambda_2$ in the diagram below).

Tau equivalent tests have equal factor loadings but may have unequal errors. Congeneric tests may differ in both factor loading and error variances.



© 2009, www.aliquote.org

Formalisation

Let's write the assumptions used to define these models:

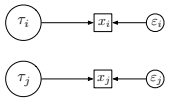
(a_1) τ-equivalence, $\tau_i = \tau_j$

(a_2) essential τ-equivalence, $\tau_i = \tau_j + \lambda_{ij}, \quad \lambda_{ij} \in \mathbb{R}$

(a_3) τ-congenerity, $\tau_i = \lambda_{ij0} + \lambda_{ij1}\tau_j, \quad \lambda_{ij0}, \lambda_{ij1} \in \mathbb{R}, \lambda_{ij1} > 0$

(b) uncorrelated errors, $\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$

(c) equal error variances, $\mathbb{V}(\varepsilon_i) = \mathbb{V}(\varepsilon_j)$



© 2009, www.aliquote.org

Formalisation (Con't)

Now, we see that:

- Parallel tests are defined by Assumptions (a_1), (b) and (c),
- Essentially τ-equivalent tests are defined by Assumptions (a_2) and (b),
- Congeneric tests are defined by Assumptions (a_3) and (b).

© 2009, www.aliquote.org

Congeneric Test

```
require(psych)
test <- sim.congeneric(short=FALSE, N=100)
round(cor(test$observed),2)
```

Raw correlation between sum scores and latent scores is 0.85, and R^2 of scores with factors is 0.87.

The first principal axis accounts for 45% of the variance in sum scores.

	I1	I2	I3	I4	PA	h^2
I1	—				0.91	0.83
I2	0.64	—			0.71	0.50
I3	0.48	0.39	—		0.53	0.28
I4	0.42	0.30	0.22	—	0.44	0.20

© 2009, www.aliquote.org

circumplex structure

In summary

Pros:

- provides useful guidelines for item writing/checking/revision and facilitates field-testing of the instrument using small samples
- few but simple mathematical formulation, with straightforward estimation of model parameters; this yields scoring rules of practical interest
- a class of 'weak' model because its assumptions are easily met

In summary

Cons:

- At the item level, CTT is mostly organized around the difficulty and discrimination parameters which are sample dependent: higher item difficulty would be obtained from an examinee samples of lower-average knowledge, and higher item discrimination would occur from an heterogeneous examinee sample.
- At the test level, difficulty directly affects the test scores, and the true score model does not allow specific pattern response (sum score is a sufficient statistic): As a consequence, it is not possible to predict one's performance on another test.

References

[1] P M Fayers and D Machin. *Quality of life. The assessment, analysis and interpretation of patient-reported outcomes*. Wiley, 2000.

– Typeset with FoilTEX (version 2)