Matrices as building blocks of measurement framework*

Kimmo Vehkalahti
Department of Mathematics and Statistics
University of Helsinki, Finland
Kimmo.Vehkalahti@helsinki.fi
http://www.helsinki.fi/people/Kimmo.Vehkalahti

McGill Matrix Wednesday: May 30, 2007

*joint work with **Simo Puntanen** and **Lauri Tarkkonen**

Outline

Sources of uncertainty

Main sources of **uncertainty** in statistical research:

- sampling (well known)
- **measurement** (too often neglected!)

1. **validity:** *are we measuring the right thing?*
   - closely connected to the substantial theory
   - only partially a statistical question
   - within the *measurement framework* we can assess:
     (a) structural validity of the measurement model
     (b) predictive validity of the measurement scale
2. **reliability:** *are we measuring accurately enough?*
   - relevant: only if validity acceptable
   - definition: ratio of true variance to total variance
   - required: estimate of measurement error variance

Estimation of reliability

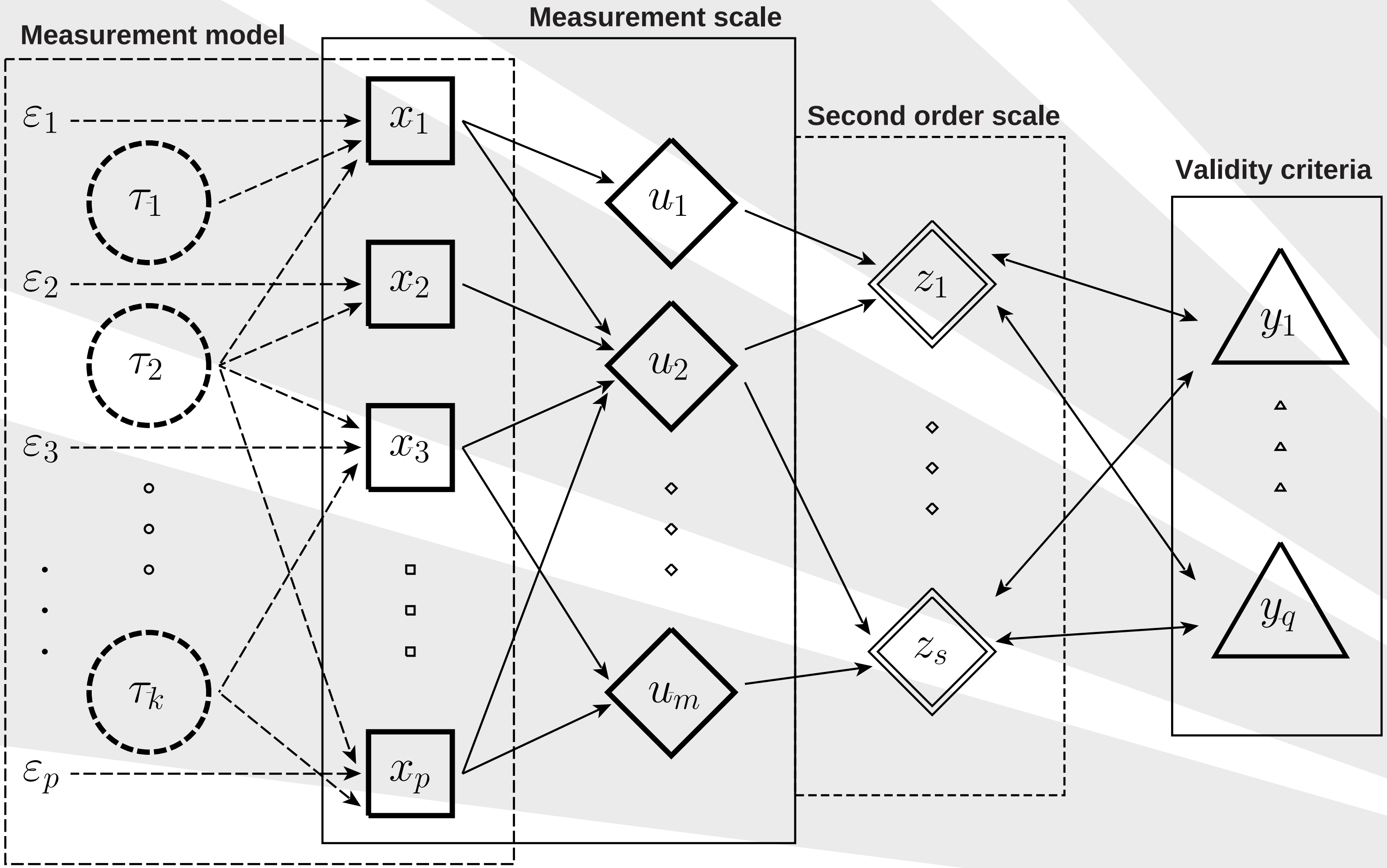
Estimation of reliability depends on the assumptions made about the **measurement model** and the **measurement scale**.

Several estimators suggested, we focus on two of them:

- new alternative: **Tarkkonen's rho**
  - based on measurement framework approach [1, 2, 3]
  - realistic assumptions, well applicable in practice
  - multidimensionality now stressed in psychology [4, 5]
- most widely used: **Cronbach's alpha**
  - based on Spearman's one-factor model (>100 years ago)
  - routinely used for >50 years (despite of criticism)
  - problem: underestimation (too strict assumptions)

Measurement framework

- *guidelines of the study from the plans to the analyses*
- basis for a consistent assessment of **measurement quality**



Measurement model

Let $\boldsymbol{x} = (x_1, \ldots, x_p)'$ measure k **(important here: $k < p$)** unobservable **true scores** $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)'$ with unobservable **measurement errors** $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)'$.

Assume $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{cov}(\boldsymbol{\tau}, \boldsymbol{\varepsilon}) = \mathbf{0}$. The measurement model is

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{B}\boldsymbol{\tau} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{B} \in \mathbb{R}^{p \times k}$ specifies the relationship between $\boldsymbol{x}$ and $\boldsymbol{\tau}$.

Denoting $\mathrm{cov}(\boldsymbol{\tau}) = \boldsymbol{\Phi}$ and $\mathrm{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}$ we have

$$\mathrm{cov}(\boldsymbol{x}) = \boldsymbol{\Sigma} = \boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}' + \boldsymbol{\Psi}, \tag{2}$$

where it is assumed that $\boldsymbol{\Sigma} > \mathbf{0}$ and $\boldsymbol{B}$ has full column rank.

Model: Estimation of parameters

The **parameters** are the $pk + k(k+1)/2 + p(p+1)/2$ (unique) elements of the matrices $\boldsymbol{B}$, $\boldsymbol{\Phi}$, and $\boldsymbol{\Psi}$. In general, there are too many, since $\boldsymbol{\Sigma}$ has only $p(p+1)/2$ elements.

- Identifiability is obtained by imposing assumptions on the true scores and the measurement errors.
- **Typical:** assume that $\mathrm{cov}(\boldsymbol{\tau}) = \boldsymbol{I}_k$, an identity matrix of order k, and $\mathrm{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Psi}_d = \mathrm{diag}(\psi_1^2, \ldots, \psi_p^2)$.
- With these the model conforms with the orthogonal factor analysis model where the *common factors are directly associated with the true scores* and the *specific factors are interpreted as measurement errors*.

Assuming multinormality the parameters can be estimated using e.g., **the maximum likelihood** method of factor analysis.

Model: Structural validity

Structural validity is a property of the measurement model.

- Important, as the model forms the core of the framework and hence affects the quality of all scales created.
- Lack of structural validity revealed by testing hypotheses
  - on the dimension of $\boldsymbol{\tau}$
  - on the effects of $\boldsymbol{\tau}$ on $\boldsymbol{x}$ (matrix $\boldsymbol{B}$)
- Whole approach could be called *semi-confirmatory*.
- Also: appropriate factor rotation and residuals of the model.

Similarly with other questions of validity, knowledge of the theory and practice of the application needed.

Measurement scale

In further analyses, the variables $\boldsymbol{x}$ are best used by creating **multivariate measurement scales** $\boldsymbol{u} = \boldsymbol{A}'\boldsymbol{x}$, where $\boldsymbol{A} \in \mathbb{R}^{p \times m}$ is a matrix of the weights. Using (2) we obtain

$$\operatorname{cov}(\boldsymbol{u}) = \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{A}'\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}'\boldsymbol{A} + \boldsymbol{A}'\boldsymbol{\Psi}\boldsymbol{A}, \tag{3}$$

the (co)variances generated by the **true scores** and
the (co)variances generated by the **measurement errors**.

Some examples of measurement scales: factor scores, psychological test scales, or any other linear combinations of the observed variables. The weights of the scale may also be predetermined values according to a theory.

Scale: Predictive validity

Predictive validity is a property of the measurement scale.

- Assessed by the correlation(s) between the (second order) scale and an *external criterion*.
- In general, a second order scale is denoted by $\boldsymbol{z} = \boldsymbol{W}'\boldsymbol{u} = \boldsymbol{W}'\boldsymbol{A}'\boldsymbol{x}$, where $\boldsymbol{W} \in \mathbb{R}^{m \times s}$ is a weight matrix and a criterion is denoted by $\boldsymbol{y} = (y_1, \ldots, y_q)'$.
- Often, these scales are produced by regression analysis, discriminant analysis, or other multivariate statistical methods.

In the most general case, the predictive validity would be assessed by the **canonical correlations** between $\boldsymbol{z}$ and $\boldsymbol{y}$.

Scale: Predictive validity, example

Example: consider the regression model $y = \beta_0 + \boldsymbol{\beta}'\boldsymbol{u} + \delta$, where y is the response variable, β_0 is the intercept, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$ is the vector of the regression coefficients, $\boldsymbol{u}$ is the vector of the predictors (e.g., factor scores), and δ is a model error.

Now, the criterion y is a scalar, and the second order scale is given by the prediction scale $z = \hat{\boldsymbol{\beta}}'\boldsymbol{u}$, where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_m)'$. Hence the predictive validity is equal to ρ_{zy}, the multiple correlation of the regression model.

Monte Carlo simulations carried out using **SURVO MM** [**6**] indicate that the factor scores offer the most stable method for predictor selection in the regression model. See [**1**] for details.

Tarkkonen's rho

According to the definition of reliability, Tarkkonen's rho is obtained as a **ratio of the variances**, i.e., the diagonal elements of the matrices in (3). Hence we have [**1**, **2**, **3**]

$$\boldsymbol{\rho_u} = \operatorname{diag}\left(\frac{\boldsymbol{a}_1'\boldsymbol{B\Phi B}'\boldsymbol{a}_1}{\boldsymbol{a}_1'\boldsymbol{\Sigma a}_1}, \ldots, \frac{\boldsymbol{a}_m'\boldsymbol{B\Phi B}'\boldsymbol{a}_m}{\boldsymbol{a}_m'\boldsymbol{\Sigma a}_m}\right)$$

$$= (\boldsymbol{A}'\boldsymbol{B\Phi B}'\boldsymbol{A})_d \times [(\boldsymbol{A}'\boldsymbol{\Sigma A})_d]^{-1}$$

or, in a form where the matrix $\boldsymbol{\Psi}$ is explicitly present:

$$\boldsymbol{\rho_u} = \operatorname{diag}\left(\left[1 + \frac{\boldsymbol{a}_1'\boldsymbol{\Psi a}_1}{\boldsymbol{a}_1'\boldsymbol{B\Phi B}'\boldsymbol{a}_1}\right]^{-1}, \ldots, \left[1 + \frac{\boldsymbol{a}_m'\boldsymbol{\Psi a}_m}{\boldsymbol{a}_m'\boldsymbol{B\Phi B}'\boldsymbol{a}_m}\right]^{-1}\right)$$

$$= \{\boldsymbol{I}_m + (\boldsymbol{A}'\boldsymbol{\Psi A})_d \times [(\boldsymbol{A}'\boldsymbol{B\Phi B}'\boldsymbol{A})_d]^{-1}\}^{-1}$$

Special cases

Many models, scales, and reliability coefficients established in the test theory of psychometrics are special cases of the framework.

Example: $\boldsymbol{x} = \boldsymbol{\mu} + \mathbf{1}\tau + \boldsymbol{\varepsilon}$ and $u = \mathbf{1}'\boldsymbol{x}$ (unweighted sum). Now, $\boldsymbol{\Sigma} = \sigma_\tau^2\mathbf{1}\mathbf{1}' + \boldsymbol{\Psi}_d$ and $\sigma_u^2 = \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} = p^2\sigma_\tau^2 + \mathrm{tr}(\boldsymbol{\Psi}_d)$.

$$\rho_{uu} = \frac{p^2\sigma_\tau^2}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}} = \frac{p}{p-1}\left(\frac{p^2\sigma_\tau^2 - p\sigma_\tau^2}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right)$$

$$= \frac{p}{p-1}\left(\frac{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1} - \mathrm{tr}(\boldsymbol{\Psi}_d) - \mathrm{tr}(\boldsymbol{\Sigma}) + \mathrm{tr}(\boldsymbol{\Psi}_d)}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right)$$

$$= \frac{p}{p-1}\left(1 - \frac{\mathrm{tr}(\boldsymbol{\Sigma})}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right) = \frac{p}{p-1}\left(1 - \frac{\sum_{i=1}^{p}\sigma_{x_i}^2}{\sigma_u^2}\right),$$

which is the original form of Cronbach's alpha [**7**].

Propositions for further research

- **developing** means for the correction for attenuation in various statistics (e.g. regression coefficients) (see [**1**])
- **specifying** the connections between the measurement framework and multivariate statistical methods (discriminant analysis, canonical correlations, correspondence analysis etc.) (see [**1**])
- **examining** the connections between the measurement framework and generalizability theory (see [**8**])
- **studying** the statistical properties of Tarkkonen's rho (sampling distribution etc.)
- **modifying** t-test for the measurement error variances
- **building** confidence intervals using the standard error of measurement
- **determining** the scales that maximize the reliability
- ...

Thank you for your attention!

1 Kimmo Vehkalahti, Simo Puntanen, and Lauri Tarkkonen. Effects of measurement errors in predictor selection of linear regression model. *Computational Statistics & Data Analysis*, 2007. `http://dx.doi.org/10.1016/j.csda.2007.05.005`.

2 Kimmo Vehkalahti, Simo Puntanen, and Lauri Tarkkonen. Estimation of reliability: a better alternative for Cronbach's alpha. Reports on Mathematics 430, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland, 2006. `http://mathstat.helsinki.fi/reports/Preprint430.pdf`.

3 L. Tarkkonen and K. Vehkalahti. Measurement errors in multivariate measurement scales. *Journal of Multivariate Analysis*, 96:172–189, 2005.

4 Joseph F. Lucke. The α and the ω of congeneric test theory: An extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement*, 29:65–81, 2005.

5 Jos M. F. Ten Berge and Gregor Soĉan. The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, 69:611–623, 2004.

6 Seppo Mustonen. SURVO MM: Computing environment for creative processing of text and numerical data. `http://www.survo.fi/mm/english.html`, 2001.

7 L. J. Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16:297–334, 1951.

8 Maria Valaste, Kimmo Vehkalahti, and Lauri Tarkkonen. Generalizability: Reliability or Validity? Paper to be presented, International Meeting of the Psychometric Society, Tokyo, Japan, July 2007.