# Enhancing the documentation by leaving useful traces

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

MAT TRIAD-07, Bedlewo, Poland
22–24 March 2007

# Outline



## Documentation
- Enhancing
- Useful traces

## Demonstration

## Measurement
- Framework
- Model
- Scale
- Validity
- Reliability
- Estimators

## References

# Enhancing the documentation



An appropriate documentation of the scientific work process

1. advocates the **reproducibility** of the research [1]
2. improves the **data quality** [2]
3. in general, helps to avoid reinventing the wheel

but an *enhanced documentation* also

4. supports **backtracking** and following **side tracks**
5. helps to get back on the track.

Instead of formal documentation, we focus on **traces** that are left behind when working.

# *Leaving useful traces*



Leaving **useful traces** [3] enhances the documentation.

Useful traces are, for example:

- **commands**, **expressions** and **work schemes** (created primarily to utilize operations of the software)
- **free-form notes** and **comments** (written down explicitly, preferably nearby the commands etc.)

They reflect the **thought process**, and help to retrace it later (say, when revising a paper after a review process).

# Demonstration with Survo



We will demonstrate these ideas especially with the matrix operations of **Survo** (SURVO MM) software [4].

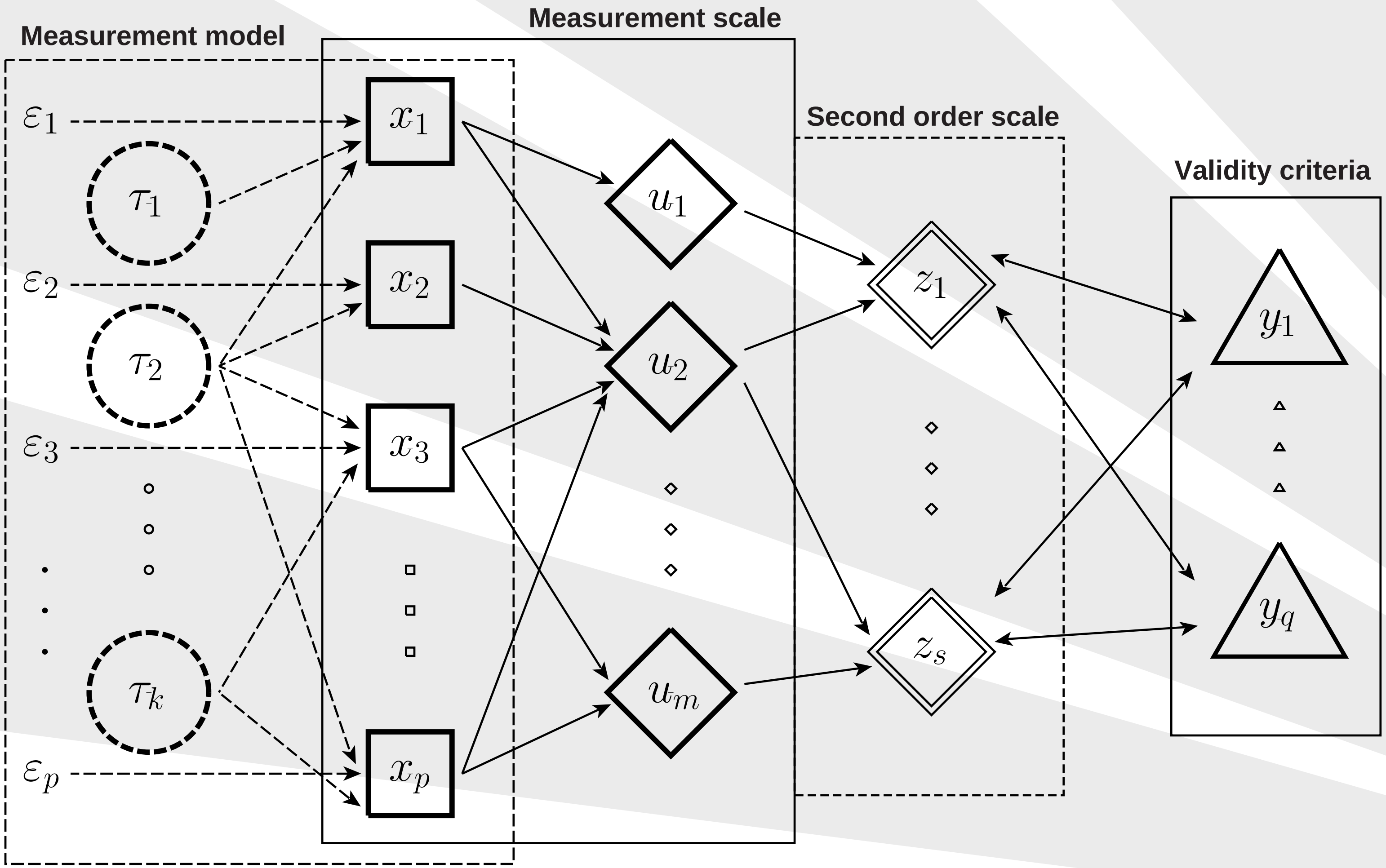
The unique *editorial interface* of **Survo** promotes building and maintaining the work process so that each step is appropriately documented.

Our examples come from certain new developments in the area of multivariate statistical modeling with **measurement errors**. The following slides offer a brief introduction to those issues.

# Measurement framework [5, 6]



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



# Measurement model



Formally: Let $\boldsymbol{x} = (x_1, \ldots, x_p)'$ measure $k$ $(< p)$ 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.

# Measurement scale



The variables $\boldsymbol{x}$ are used in further analyses by creating **multivariate measurement scales**

$$\boldsymbol{u} = \boldsymbol{A}'\boldsymbol{x}, \tag{3}$$

where $\boldsymbol{A} \in \mathbb{R}^{p \times m}$ is a matrix of the weights, e.g., factor score coefficients or predetermined values according to a theory.

Using (2) we obtain

$$\operatorname{cov}(\boldsymbol{u}) = \boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A} = \boldsymbol{A}'\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}'\boldsymbol{A} + \boldsymbol{A}'\boldsymbol{\Psi}\boldsymbol{A}, \tag{4}$$

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

# Validity and reliability



In general, the sources of **uncertainty** in statistical research are:

- **sampling** (thoroughly known and handled)
- **measurement** (often neglected in statistics!)

1. **validity:** *are we measuring the right thing?*
   - closely connected to the substantial theory
   - 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

# Reliability estimation



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

Several estimators suggested:

- most widely used: **Cronbach's alpha** [7]
  - based on Spearman's one-factor model ($>$100 years ago)
  - routinely used for $>$50 years (despite of criticism)
  - problem: underestimation (too strict assumptions)
- new, better alternative: **Tarkkonen's rho** [8, 5, 6]
  - based on measurement framework approach
  - realistic assumptions, well applicable in practice
  - also other research supports multidimensionality [9]

# Tarkkonen's rho & Cronbach's alpha



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 (4). Hence we have

$$\begin{aligned} \boldsymbol{\rho_u} &= (\boldsymbol{A}'\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}'\boldsymbol{A})_d \times (\boldsymbol{A}'\boldsymbol{\Sigma}\boldsymbol{A})_d && (5) \\ &= \{\boldsymbol{I}_m + (\boldsymbol{A}'\boldsymbol{\Psi}\boldsymbol{A})_d \times [(\boldsymbol{A}'\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{B}'\boldsymbol{A})_d]^{-1}\}^{-1}. && (6) \end{aligned}$$

Cronbach's alpha is a special case of Tarkkonen's rho under a simple model $\boldsymbol{x} = \mathbf{1}\tau + \boldsymbol{\varepsilon}$ and with a simple scale $u = \mathbf{1}'\boldsymbol{x}$. It is easy to show that in this case, (5) or (6) lead to

$$\alpha = \frac{p}{p-1}\left(1 - \frac{\mathrm{tr}(\boldsymbol{\Sigma})}{\mathbf{1}'\boldsymbol{\Sigma}\mathbf{1}}\right) = \frac{p}{p-1}\left(1 - \frac{\sum_{i=1}^{p}\sigma_{x_i}^2}{\sigma_u^2}\right), \qquad (7)$$

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

# References



[1] Robert Gentleman and Duncan Temple Lang. Statistical analyses and reproducible research. *Bioconductor Project Working Papers*, 2, 2004. http://www.bepress.com/bioconductor/paper2.

[2] T. Dasu and T. Johnson. *Exploratory Data Mining and Data Cleaning*. Wiley, Hoboken, New Jersey, 2003.

[3] Kimmo Vehkalahti. Leaving useful traces when working with matrices. *Research Letters in the Information and Mathematical Sciences*, 8:143–154, 2005. Proceedings of the 14th International Workshop on Matrices and Statistics. (Paul S.P. Cowpertwait, ed.) Massey University, Auckland, New Zealand, March 29 – April 1, 2005. http://iims.massey.ac.nz/research/letters/volume8/.

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

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

[6] Kimmo Vehkalahti, Simo Puntanen, and Lauri Tarkkonen. Effects of measurement errors in predictor selection of linear regression model. Reports on Mathematics 439, Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland, 2006. http://mathstat.helsinki.fi/reports/Preprint439.pdf.

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

[8] 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.

[9] Joseph F. Lucke. The $\alpha$ and the $\omega$ of congeneric test theory: An extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement*, 29:65–81, 2005.