SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Response

Statistical Methods in Medical Research
2020, Vol. 29(4) 991–995

DOI: 10.1177/0962280219893460
journals.sagepub.com/home/smm
SAGE

Reaction to "Sufficient statistics and insufficient explanations": Use your information

Jean-Paul Fox

1 Introduction

In the reaction of van Breukelen (2019), the claim is made that the sum score is a sufficient statistic for the latent trait parameter (θ_{ij} of person i at time point j) in a latent growth model analysis. The argumentation is given that with the sum score as the sufficient statistic, the estimated between-subject variance cannot be biased and the estimated within-subject variance is contaminated with a measurement error variance term. We agree that differences in item response patterns leading to the same sum score become irrelevant when the sum score is the sufficient statistic. However, we can show that the sum score is not the sufficient statistic for the latent trait parameter in the longitudinal IRT model (i.e. latent growth model with IRT measured longitudinal latent variables).

First, we show that additional information in the data about θ_{ij} is ignored, when considering the sum score as the sufficient statistic. Thus, the sum score is not sufficient. Second, we show that the estimated variance components (within-subject and between-subject) are contaminated with unexplained error variance, when using the sum score as the outcome variable instead of the item response data. This supports the conclusions of our paper.[1]

It is a common mistake to assume that the sum score is the sufficient statistic for the latent trait parameter, when the item responses are conditionally independently distributed given the latent trait (as in the Rasch model). This is only true when the data do not provide additional information about the latent trait. The longitudinal data consist of repeated measurements, where on each measurement occasion, a latent trait is measured and a latent growth model is assumed for the longitudinal latent traits. This latent growth model defines a distribution for the occasion-specific latent trait parameters. The data from the different measurement occasions for a subject are relevant for each occasion-specific latent trait measurement due to this distribution. Because data from the other measurement occasions provide information about each occasion-specific latent trait, the sum score is not the sufficient statistic.

The influence of the additional information on the latent trait is easily illustrated by considering the posterior expected value for θ_{ij} given the data. Consider quantitative item responses Z_{ijk} of persons i, measurement j and item k, let the Z_{ijk} be normally distributed, and we assume a linear trend for the latent trait parameter

$$\begin{aligned} Z_{ijk} &= \theta_{ij} - b_k + e_{ijk},\ e_{ijk} \sim N(0,1) \\ \theta_{ij} &= \beta_{0i} + \beta_{1i} t_{ij} + r_{ij},\ r_{ij} \sim N(0,\sigma^2) \end{aligned} \tag{1}$$

Department of Research Methodology, Measurement, and Data Analysis, Faculty of Behavioural, Management & Social Sciences, University of Twente, Enschede, The Netherlands

Corresponding author:
Jean-Paul Fox, Department of Research Methodology, Measurement, and Data Analysis, Faculty of Behavioural, Management & Social Sciences, University of Twente, Enschede, The Netherlands.
Email: j.p.fox@utwente.nl

The error distributions are simplified in comparison to the linear model of van Breukelen (equation (3)), but our complexity is sufficient to prove our point. The posterior expected value for θ_{ij} is given by (following the derivation above equation (10) of Gorter et al.[1])

$$E(\theta_{ij}|\mathbf{Z}_{ij}, \mathbf{b}, \boldsymbol{\beta}_i, \sigma^2) = \left(\frac{K}{K+\sigma^{-2}}\right)(\overline{\mathbf{Z}_{ij}+\mathbf{b}}) + \left(\frac{\sigma^{-2}}{K+\sigma^{-2}}\right)(\beta_{0i}+\beta_{1i}t_{ij})$$

The data information from occasion j is represented by the mean $\sum_{k=1}^{K}(Z_{ijk}+b_k)/K = \overline{\mathbf{Z}_{ij}+\mathbf{b}}$ and the remaining information stems from the linear trend. The posterior mean shrinks towards the mean, $\overline{\mathbf{Z}_{ij}+\mathbf{b}}$, when the number of items K increases in relation to the precision σ^{-2}. When the precision increases in relation to K, the posterior mean shrinks towards the linear trend prediction. This posterior mean estimator is based on the borrowing-strength principle, where data information from other measurement occasions is used to improve the estimator in terms of the mean squared error. Thus, the latent growth distribution of the latent trait parameters connects the parameters associated with the different measurement occasions, which makes it possible to apply the borrowing-strength principle. In latent growth modeling, the sum score should not be used as a sufficient statistic; it is a suboptimal estimator for the latent trait parameter, since it ignores data information from other measurement occasions.

This gain is achieved by estimating simultaneously all parameters using MCMC, which facilitates balancing the data information from the different measurement occasions. The latent growth parameters and difficulty parameters are sampled from their posterior distributions using all data and this is combined with the occasion-specific data information. Thus, when the latent growth parameters are sampled from the conditional distribution (step 3b in Appendix 1 in Gorter et al.[1]), and the difficulty parameters are sampled from their conditional distribution (step 1c in Appendix 1 in Gorter et al.[1]), then the posterior mean is updated using the sampled values for the latent growth and item difficulty parameters. Subsequently, the latent trait parameters are sampled from their posterior distribution (step 1b Appendix 1 in Gorter et al.[1]). After convergence of the MCMC algorithm, the sampled values for the latent trait parameters are distributed according to the marginal posterior distribution, $p(\boldsymbol{\theta}_{ij}|\mathbf{Z})$, which uses all data information.

The use of the sum score as a sufficient statistic in the latent growth model analysis has an influence on the variance decomposition. When the sum score is not used as a sufficient statistic, the variance decomposition is different from the one described by van Breukelen. He considered the covariance components at the level of the latent trait parameters, mainly because the sum score is defined at this level. However, to understand the variance decomposition, we partition the total sum of squares in a within-measurement component (SSW), a between-measurement (SSA, within-subject) component, and a between-subject (SSB) component. It is shown that each sum of squared errors represents different variance components, and that the item parameters affect the estimated measurement error variance. It follows that an increase in the estimated measurement error variance leads to a reduction of the within- and between-subject variance, since the total variance in the data (i.e. the total sum of squared errors) is fixed.

To prove our point, we consider a balanced design with N subjects, J measurement occasions, and K items. For balanced data, the variance components can be easily estimated by setting the sum of squared errors equal to their expectations and solving the equations.[2] This procedure is followed to examine which variance components are estimated and how the estimates influence each other. To ease the mathematical burden, we only include the random intercept of the latent growth model. The following model is considered

$$\begin{aligned} &Z_{ijk} = \theta_{ij} - b_k + e_{ijk},\ e_{ijk} \sim N(0, \delta^2_{jk}) \\ &\theta_{ij} = \beta_{0i} + r_{ij},\ r_{ij} \sim N(0, \sigma^2),\ \beta_{0i} \sim N(0, \tau_0^2) \end{aligned} \tag{2}$$

The expressions for the sum of squared errors can be obtained in closed-form. The expected value of each sum of squared errors is used to understand the variance decomposition under the model. The total sum of squared errors is partitioned according to the multilevel structure of the data

$$\begin{aligned} \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K}(Z_{ijk}-\bar{Z})^2 &= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K}(Z_{ijk}-\bar{Z}_{ij.})^2 + \sum_{i=1}^{N}\sum_{j=1}^{J}K(\bar{Z}_{ij.}-\bar{Z}_{i.})^2 \\ &\quad + \sum_{i=1}^{N}JK(\bar{Z}_{i.}-\bar{Z})^2 = SSW + SSA + SSB \end{aligned}$$

where $\bar{Z} = \sum_{i,j,k} Z_{ijk}/(NJK)$, $\bar{Z}_{i..} = \sum_{j,k} Z_{ijk}/(JK)$, and $\bar{Z}_{ij.} = \sum_{k} Z_{ijk}/K$. The SSW represents the sum of squared errors in a response pattern.

The expected value of the SSW is considered to obtain the components that explain the variance at this level of observations. The expected value is derived by plugging in the linear model (equation (2)) and taking the expected value of the random components. It follows that

$$\begin{aligned} E(SSW) &= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K} E\Big((\theta_{ij} - b_k + e_{ijk}) - (\theta_{ij} - \bar{b}_{.} - \bar{e}_{ij.})\Big)^2 \\ &= \sum_{i=1}^{N}\sum_{j=1}^{J}\sum_{k=1}^{K} (b_k - \bar{b}_{.})^2 + E(e_{ijk} - \bar{e}_{ij.})^2 \\ &= NJ\sum_{k=1}^{K}(b_k - \bar{b}_{.})^2 + N\sum_{j=1}^{J}\sum_{k=1}^{K}(\delta_{jk}^2 + \bar{\delta}_{j.}^2/K) \end{aligned} \tag{3}$$

We have used that the errors are assumed to be independently distributed from each other, and the inner product of the binomial products is zero, since the expected value of the random errors is zero. The difficulty parameters are assumed to be fixed. It follows directly that the unexplained variance in the SSW is reduced due to the variance explained by the item difficulty parameters. When items differ more in item difficulty, the more variance is explained by the item difficulty parameters. This reduces the amount of unexplained variance which is captured by the measurement error variance δ_{jk}^2. This is not possible under the CTT model, which does not account for item difficulty differences.

The SSA represents the information about the within-subject variance, σ^2. The expected value of the SSA is derived in a similar manner

$$\begin{aligned} E(SSA) &= \sum_{i=1}^{N}\sum_{j=1}^{J} KE\Big((\beta_{0i} + r_{ij} - \bar{b}_{.} + \bar{e}_{ij.}) - (\beta_{0i} + \bar{r}_{i.} - \bar{b}_{.} + \bar{e}_{i..})\Big)^2 \\ &= \sum_{i=1}^{N}\sum_{j=1}^{J} K\Big(E(r_{ij} - r_{i.})^2 + E(\bar{e}_{ij.} - \bar{e}_{i..})^2\Big) \\ &= NK(J-1)\sigma^2 + N\sum_{j}(\bar{\delta}_{j.}^2 + \bar{\delta}_{..}^2/J) \end{aligned} \tag{4}$$

For the observed SSA, the estimated within-subject variance, σ^2, is contaminated with the average measurement error variance, when not accounting for it. This bias in the within-subject variance estimate under CTT was also remarked by van Breukelen. However, this contamination of the within-subject variance estimate contains more components. The explained within-subject variance, $(r_{ij} - r_{i.})^2$ is lower under CTT than under IRT, since differences in response patterns leading to the same sum score are ignored. So, the explained within-subject variance is lower under CTT and a lower reduction in unexplained (measurement error) variance. Differences in item difficulties are ignored under CTT, which leads to a higher amount of unexplained (measurement error) variance. In conclusion, the overestimation of the within-subject variance with CTT is caused by the unexplained measurement error variance, which includes the contamination of the average measurement error variance.

Finally, the SSB represents the information about the between-subject variance, τ_0^2. The expected value of the SSB is given by

$$\begin{aligned} E(SSB) &= JK\sum_{i=1}^{N} E\Big((\beta_{0i} + \bar{r}_{i.} - \bar{b}_{.} + \bar{e}_{i..}) - (\bar{\beta}_{0.} + \bar{r}_{..} - \bar{b}_{.} + \bar{e}_{...})\Big)^2 \\ &= JK\sum_{i=1}^{N}\sum_{j=1}^{J} E(\beta_{0i} - \bar{\beta}_{0.})^2 + E(\bar{r}_{i.} - r_{..})^2 + E(\bar{e}_{i..} - \bar{e}_{...})^2 \\ &= JK(N-1)\tau_0^2 + K(N-1)\sigma^2 + (N-1)\bar{\delta}_{..}^2 \end{aligned} \tag{5}$$

This shows that the between-subject variance estimate is also contaminated by the measurement error variance under CTT. Furthermore, under CTT, the component $(\beta_{0i} - \bar{\beta}_{0.})^2$ explains less variance than under IRT, since again different response patterns leading to the same sum score are ignored. This leads to an increase of the unexplained error variance. When overestimating the within-subject variance (according to the SSA in equation (4)), the between-subject variance is underestimated, since the SSB represents the total variance between subjects. The estimated τ_0^2 is contaminated, when not accounting for the average measurement error variance, but not as large as the contamination in the estimated σ^2. Note that the repeated measurements, θ_{ij}, are nested within the subject, β_{0i}. This nested random-effect structure introduces a dependence where the higher-level variance estimate is influenced by the lower-level variance estimate. Proust-Lima et al.[3] did not detect biased Type-I errors of regression effects, but they also did not have this nested random effect structure in their multivariate latent variable model.

For binary data, the latent response formulation is used, which also introduces an error term. This error term represents the measurement error. Consider the probability of a positive response under the IRT model

$$\begin{aligned} P(Y_{ijk} = 1|\theta_{ij}, b_{kj}) &= F(\theta_{ij} - b_k) \\ &= P(Z_{ijk} < \theta_{ij} - b_k) \\ &= P(\theta_{ij} - b_k + e_{ijk} > 0) \end{aligned} \tag{6}$$

where F is a cumulative distribution function, and Z_{ijk} is normally distributed with mean zero and variance one (probit model) or logistically distributed with location zero and scale parameter one (logistic model). As a result, the error term e_{ijk} is normally (probit model) or logistically (logit model) distributed. It can be seen that the model in equation (6) corresponds to the model in equation (1), only that a link function is used to map the linear term to the binary observations. The (latent) error term, e_{ijk}, represents the measurement error, which can be computed, and the measurement error variance is restricted to one to identify the model. This error variance represents the δ^2 and shows that IRT also accounts for the measurement error. This is necessary; otherwise, the within-subject and between-subject variance would be incorrectly estimated and contaminated with unexplained error variance.

The re-scaling of the plausible values cannot introduce additional bias, since it only translates the plausible values to a scale on which the latent growth parameters are estimated. The plausible values can be contaminated with measurement errors (unexplained errors) and they are together translated to another scale. However, the chosen location and variance of the (latent) scale cannot lead to a further discrimination in potential bias, since the scale parameters determine the general location and total variance of the scale.

In conclusion, when using all data information, the sum score is not the sufficient statistic, and differences in response patterns remain relevant for estimating the within-subject as well as the between-subject variance. The total amount of variance is restricted at each level, which is represented by the sums of squared errors. When overestimating the within-subject variance, the between-subject variance will be underestimated (see equation (5)). The within-subject variance will be overestimated, when not accounting for unexplained error variance(s). The IRT latent variables, θ_{ij} and β_{0i}, explain more variability, since they are based on differences in item response patterns in comparison to differences based on sum scores. In the two-level CTT-based approach, the within-subject variance is contaminated with (increased) unexplained error variance. When considering binary data, the same conclusions can be drawn. In the IRT-based approach, there is still a measurement error term except that the error variance is restricted to identify the model.

In reaction to the strong assumptions that are made under the IRT-based latent growth model (e.g. item clustering, longitudinal covariance patterns). There can be other types of dependences that need to be taken into account. Recently, Fox et al.[4] and Klotzke and Fox[5,6] developed Bayesian covariance structure models that are very flexible in dealing simultaneously with multiple types of clusters. In their Bayesian approach, the IRT-based latent growth model can be extended through a multivariate extension in which additional dependences are directly modeled through a structured covariance matrix. Thus, further generalizations are possible within the IRT-modeling framework, while keeping the desirable psychometric properties of the IRT model.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Gorter R, Fox JP, Riet GT, et al. Latent growth modeling of IRT versus CTT measured longitudinal latent variables. *Stat Meth Med Res* 2019. doi: 10.1177/0962280219856375.
2. Searle SR. *Linear models*. New York: Wiley, 1971.
3. Proust-Lima C, Philipps V, Dartigues JF, et al. Are latent variable models preferable to composite score approaches when assessing risk factors of change? Evaluation of type-I error and statistical power in longitudinal cognitive studies. *Stat Meth Med Res* 2019; **28**: 1942–1957.
4. Fox J-P, Mulder J and Sinharay S. Bayes factor covariance testing in item response models. *Psychometrika* 2017; **82**: 979–1006.
5. Klotzke K and Fox J-P. Bayesian covariance structure modelling of responses and process data. *Front Psychol* 2019; **10**: 1675.
6. Klotzke K and Fox J-P. Modeling dependence structures for response times in a Bayesian framework. *Psychometrika* 2019; **84**: 649–672. doi: 10.1007/s11336-019-09671-8.
7. Van Breukelen GJP. Sufficient statistics and insufficient explanations. *Stat Meth Med Res* 2019.