

Ignorance or Negligence: Uncomfortable Truth Regarding Misuse of Confirmatory Factor Analysis

Heon-Jae Jeong¹ and Wui-Chiang Lee^{2*}

¹The Care Quality Research Group, Seoul, Korea

²Department of Medical Affairs and Planning, Taipei Veterans General Hospital, Taiwan & National Yang-Ming University School of Medicine, Taiwan

*Corresponding author: Wui-Chiang Lee, Department of Medical Affairs and Planning, Taipei Veterans General Hospital, Taiwan & National Yang-Ming University School of Medicine, 201, Section 2, Shihpai Rd, Taipei City, Taiwan 11217, Taiwan, Tel: +886-2-28757120; Fax: +886-2-28757200 E-mail: leewuichiang@gmail.com

Rec date: April 27, 2016; Acc date: April 27, 2016; Pub date: May 04, 2016

Copyright: © 2016 Heon-Jae Jeong, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

We are uncomfortable. As a scientist and biostatistician, we feel rather uncomfortable whenever people misuse an analytical method by violating necessary assumptions, which eventually leads to an inappropriate method for the situation. For example, more often than not, we use an ordinary regression model even when clustering is suspected, assuming that the sample is homogeneous. Indeed, in many cases we need to relax the assumptions of a method, because no other methods exist that are precisely suitable for the dataset. However, what if there are some methods and we are simply not aware of these better methods? What if our ignorance leads us to misuse a method? Among such disasters of the misuse of methodology, confirmatory factor analysis (CFA) would be one of the representative examples.

Brief but essential review on CFA

Since the idea related to CFA was first brought up to the psychometrics field by Spearman in the early 1900s [1,2], it has continuously evolved and played a key role as a measurement model for many uses, including survey validation [3,4]. Compared to the classic test theory (CTT), which focuses on the property of a certain test instrument as a whole, CFA's strength is that it focuses on each item of the instrument. At this point, we review the fundamentals of CFA to go further in this discussion (Figure 1).

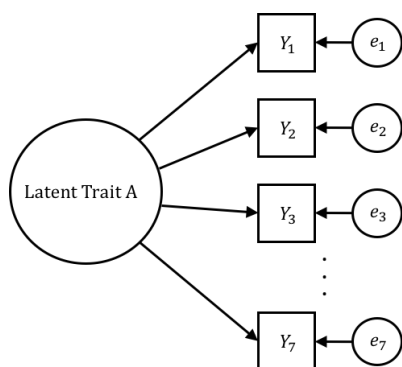


Figure 1: Confirmatory Factor Model for Seven-item Scale for Latent Trait A.

In CFA, the latent trait is assumed to account for the subject's responses. Therefore, the direction of arrows is from the latent trait to the item response, not the other way around, which is a very common

misunderstanding. In addition, this relationship is assumed to be linear, as shown in Figure 2.

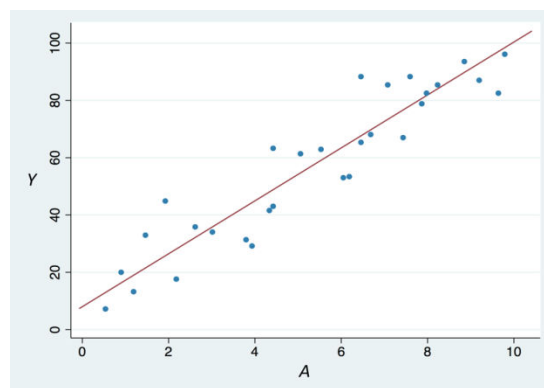


Figure 2: Graphical Presentation of CFA as a Measurement Model.

The fitted line (linear regression) is written as the following equation:

$$Y_{ij} = \mu_j + \lambda_j A_i + e_{ij}$$

where Y_{ij} is a continuous item response to item j for subject i , μ_j is the intercept of item j , and λ_j is the slope of j . In addition, e_{ij} is the error for the item and subject. Therefore, if λ_j is large (the slope of the line is steep), even a small difference in latent trait A can be captured with item j (high discrimination). In other words, items with larger λ_j are more sensitive and provide more information (precision) on the latent trait. Traditionally we call this slope, λ_j , the factor loading of item j .

As such, between the seven items and latent trait A in Figure 1, we have seven different linear regression equations in which the latent trait is the independent variable and the response of each item is the dependent variable. Thus far, everything is logical, but now we begin to analyze CFA critically.

Fundamental assumption that we violate while using CFA

We mentioned that CFA is a linear regression model that predicts continuous response variables from the latent trait as an explanatory variable [5]. This means that responses can be plugged into a CFA legitimately only when they are continuous. However, how often does a survey instrument include a true continuous variable as a response option? Probably not often. More often than not, we use five- or seven-point Likert scales, which are unfortunately not a continuous response

by definition. Therefore, using CFA to validate an instrument with a dichotomous or categorical outcome variable actually violates the fundamental idea of CFA—that is, the linearity—although we are used to this violation.

Let us keep our eyes closed to this issue as another common oversight to instrument utilization exists. We frequently calculate the simple mean of items to determine the level of latent trait. Using the previously discussed seven-item model as an example, the mean of item responses is said to be the latent trait level. However, this approach works only on the assumption that all items are equally important. In other words, items' factor loadings are the same or at least similar, which is called τ ("tau") equivalence [6]. However, such τ equivalence is rarely the case; therefore, we have to appropriately place more weight on items with larger loading and less weight on the items with smaller loading. Otherwise, we will wrongfully under- or overvalue items. We call the result of this approach "factor score." Whenever τ equivalence is not satisfied, the factor score should be considered instead of the simple mean of the item responses.

Now we are trapped in a dilemma. To determine the factor score and test τ equivalence, we need factor loading. To get factor loading, we have to run a CFA. However, if item responses are not continuous we cannot (or, more precisely, should not) use CFA. At this point, we must take a leap: We often decide to regard responses from the Likert scale as a continuous variable and then run a CFA. However although a really well-designed Likert scale can have symmetry and balance as its properties, this does not mean the responses are continuous. At this point, you have every right to ask what methodology other than CFA we can or should use for a categorical response-like Likert scale. The answer we offer as a potential alternative for CFA is the graded response model (GRM) of item response theory (IRT).

IRT GRM as an alternative to CFA for categorical responses

Simply put, IRT GRM is a CFA designed for categorical variables. This model was proposed by Samejima [7,8]. Through GRM we calculate parameters equivalent to μ and λ in CFA, but in a different form involving probability.

In addition, the τ equivalence issue in CFA can be effectively handled with IRT GRM. For example, we can estimate a respondent's latent trait level by calculating the empirical Bayesian mean of the responses of items by considering item parameters [9]. The weighting based on factor loading (in CFA terminology) is naturally taken into account in this process. Although computationally demanding, most statistical software packages have found their own ways to perform this efficiently, so it would not be a statistician's problem any longer.

Conclusion

Time to get over the ignorance and negligence

We began this discussion because of the overwhelming number of studies that wrongfully use CFA for categorical responses, and we suggested IRT GRM as an alternative option. Contrary to our expectation, not every statistician or researcher is aware of the continuous assumption of observed response variable in CFA. For these statisticians and researchers, the misuse arises from ignorance, so this article may help correct their inaccurate beliefs.

Although this article mainly discussed the suspected misuse of CFA, such a debate can and should be expanded to all the methodologies and their possibilities. We often dream of a one-size-fits-all method that works with all types of data, such as continuous, dichotomized, and categorical data. Maybe it is because choosing the correct method and meticulously checking assumptions are cumbersome. However, such onerousness cannot justify our negligence, which might jeopardize the entire study. All our efforts in scientific research should be based on a rock-solid foundation of the correct methodology; otherwise, it is just like holding a superstitious ritual for rain.

Whenever we find ourselves feeling pesky when reviewing methodology, imagine seeing one of our close colleagues running a linear regression model for a dichotomous outcome variable. What would we say to our beloved colleague? The answer is loud and clear.

References

1. Spearman C (1904) General Intelligence, objectively determined and measured. *The American Journal of Psychology* 15: 201-292.
2. Spearman C (1904) The proof and measurement of association between two things. *The American Journal of Psychology* 15: 72-101.
3. Hoyle RH (2000) Confirmatory factor analysis. *Handbook of applied multivariate statistics and mathematical modeling*.
4. Heinen T (1996) Latent class and discrete latent trait models: Similarities and differences. Sage Publications.
5. Thompson B (2004) Exploratory and confirmatory factor analysis: Understanding concepts and applications. American Psychological Association.
6. Acock AC (2013) Discovering structural equation modeling using Stata, Revised Edition. Stata Press.
7. Samejima F (1969) Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*.
8. Fayers P (2004) Item response theory for psychologists. *Quality of Life Research* 13: 715-716.
9. Baker FB, Kim S-H (2004) Item response theory: Parameter estimation techniques. CRC Press.