Center for Advanced Studies in Measurement and Assessment

CASMA Technical Note

Number 5

Regressed Score Estimates and Inconsistencies with Classical Test Theory[1]

Robert L. Brennan[2]

December 20, 2012

[1]The research reported here was supported, in part, by a contract with the College Board. The author expresses his gratitude to Won-Chan Lee for several thoughtful comments about a previou versions of this paper.

[2]Robert L. Brennan is E. F. Lindquist Chair in Measurement and Testing and Director, Center for Advanced Studies in Measurement and Assessment (CASMA).

Center for Advanced Studies in
Measurement and Assessment (CASMA)
College of Education
University of Iowa
Iowa City, IA 52242
Tel: 319-335-5439
Web: www.education.uiowa.edu/casma

Contents

Abstract

Regressed score estimates are estimates of true scores based on assuming a linear regression of true scores on observed scores. This assumption may seem innocuous. Actually, however, it leads to some fundamental inconsistencies with certain traditional assumptions and results in classical test theory. These inconsistencies are the major focus of this paper.

Kelley's (1947) regressed score estimates (RSEs) are estimates of true scores based on assuming a linear regression of true scores (T) on observed scores (X) for the bivariate distribution of T and X. RSEs are discussed in considerable detail by Feldt and Brennan (1989, pp. 120–122) and Haertel (2006, pp. 80–82). The assumption of a linear regression of T on X may seem innocuous. Actually, however, this assumption leads to some fundamental inconsistencies with certain traditional assumptions and results in CTT. These inconsistencies are the major focus of this paper. They arise in part because CTT is not a regression model

For example, in CTT by definition the expected value of observed scores over replications of a measurement procedure is true score, which means that the regression of X on T is always linear with an intercept of 0 and a slope of 1. By contrast, the regression of T on X is unknown, and even if this regression is linear, this regression is not the inverse of the regression of X on T.

1 Regressed-score Estimates

By well-known statistical results, it is straightforward to show that the linear regression of T on X is

$$\widetilde{T} \quad = \quad (1-\rho^2_X)\overline{T} + \rho^2_X\, X, \tag{1}$$

where ρ^2_X is the reliability of X and $\widetilde{T}$ is the estimated true-score random variable based on the linear regression. Note that the word "estimated" is a bit ambiguous here. It means predicted from the linear regression; it does not mean that $\widetilde{T}$ is based on sample estimates of $\overline{X}$ and/or ρ^2. This distinction is the reason why $\widetilde{T}$ is used here rather than the more usual $\widehat{T}$. Since $\overline{T} = \overline{X}$ under CTT assumptions, Equation 1 is commonly expressed as:

$$\widetilde{T} \quad = \quad (1-\rho^2_X)\overline{X} + \rho^2_X\, X. \tag{2}$$

It is often noted that Equation 2 has a Bayesian interpretation. Specifically, as ρ^2_X increases, the magnitude of $\widetilde{T}$ is increasingly influenced by X. Conversely, as ρ^2_X decreases, the magnitude of $\widetilde{T}$ is increasingly influenced by $\overline{X}$. The limiting cases are instructive: (a) when $\rho^2_X = 0$, $\widetilde{T} = \overline{X}$ which means that all examinees receive the same RSE, namely, $\overline{X}$; and (b) when $\rho^2_X = 1$, $\widetilde{T} = X$, which means that for all examinees their RSEs are identical to their observed scores.

2 Reliability

Under CTT assumptions, there are four widely used expressions for ρ^2_X:

$$\rho_{XX'} = \rho^2_{TX} = \frac{\sigma^2_T}{\sigma^2_X} = 1 - \frac{\sigma^2_E}{\sigma^2_X}, \tag{3}$$

where $X = T + E$ and X' is a parallel version of X.[3] The first two expressions are generally regarded as definitions of reliability. Historically, $\rho_{XX'}$ was the first definition of reliability, but ρ^2_{TX} is generally considered the canonical definition. The latter is the definition of ρ^2_X in Equations 1 and 2. The last two expressions in Equation 3 can be derived from either of the first two under CTT assumptions.

Since X and $\widetilde{T}$ are estimators of the same parameter, namely, T, it is natural to consider the reliability of $\widetilde{T}$ which we denote generically as $\rho^2_{\widetilde{T}}$; i.e., $\widetilde{T}$ plays the role of X in Equation 3.

2.1 Correlational-based Definitions of Reliability

Clearly $\rho_{XX'}$ is a correlation and ρ^2_{TX} is a squared correlation. Based on standard statistical results, a correlation is unaffected by a linear transformation of either or both variables. It follows that

$$\rho_{\widetilde{T}\widetilde{T}'} = \rho_{XX'} \qquad \text{and} \qquad \rho^2_{T\widetilde{T}} = \rho^2_{TX}.$$

Therefore, under these two correlational-based definitions of reliability, $\rho^2_{\widetilde{T}}$ is identical to ρ^2_X. Since ρ^2_X is positive, the above identities imply that the rank ordering of examinees is the same for X and $\widetilde{T}$ and, consequently, percentile ranks for examinees are the same.

These conclusions assume, however, that $\widetilde{T}$ in Equation 2 is based on using $\overline{X}$ for the entire population. By contrast, for example, if males are regressed to their mean and females are regressed to their mean, it is highly unlikely that the rank ordering or percentile ranks for the entire population will be the same based on X and $\widetilde{T}$.

2.2 Variance-ratio Expressions for Reliability

Replacing X with $\widetilde{T}$ in σ^2_T/σ^2_X (third expression in Equation 3) gives

$$\rho^2_{\widetilde{T}} = \frac{\sigma^2_T}{\sigma^2_{\widetilde{T}}} = \frac{\sigma^2_T}{(\rho^2_X)^2\sigma^2_X} = \frac{\rho^2_X}{(\rho^2_X)^2} = \frac{1}{\rho^2_X}, \tag{4}$$

which states that $\rho^2_{\widetilde{T}}$ is the inverse of ρ^2_X. Clearly, whenever $\rho^2_X \neq 1$, this is a nonsensical expression for the reliability of $\widetilde{T}$. For example, whenever $\rho^2_X < 1$, $\rho^2_{\widetilde{T}} > 1$, and as $\rho^2_X \to 0$, $\rho^2_{\widetilde{T}} \to \infty$ according to Equation 4.

Replacing X with $\widetilde{T}$ in $1 - \sigma^2_E/\sigma^2_X$ (fourth expression in Equation 3) gives

$$\rho^2_{\widetilde{T}} = 1 - \frac{\sigma^2_E}{\sigma^2_{\widetilde{T}}} = 1 - \frac{\sigma^2_E}{(\rho^2_X)^2\sigma^2_X},$$

[3] It is common to assume that X and X' are classically parallel; it is sufficient to assume here that X and X' are congeneric (see Feldt & Brennan, 1989, pp. 113–116, or Haertel, 2006, pp.74–76).

where

$$\frac{\sigma^2_E}{\sigma^2_X} = \frac{\sigma^2_X - \sigma^2_T}{\sigma^2_X} = 1 - \rho^2_X.$$

It follows that

$$\rho^2_{\widetilde{T}} = 1 - \frac{1-\rho^2_X}{(\rho^2_X)^2} = \frac{(\rho^2_X)^2 + \rho^2_X - 1}{(\rho^2_X)^2}. \tag{5}$$

This, too, is a nonsensical result for the reliability of $\widetilde{T}$ whenever $\rho^2_X \neq 1$. For example, as $\rho^2_X \to 0$, $\rho^2_{\widetilde{T}} \to -\infty$.

3 Covariances Involving Error

The results in Equations 4 and 5 are nonsensical because their derivations begin with expressions for reliability (third and fourth expressions in Equation 3) that make a CTT assumption that does not hold for RSEs. This issue is examined next from two perspectives.

3.1 Zero Covariance for True and Error Scores in CTT

Letting $E = X - T$, often CTT results are derived assuming that $\sigma_{T,E} = 0$, which means that T and E are uncorrelated. For RSEs, however, the error random variable is $\widetilde{E} = \widetilde{T} - T$, and $\sigma_{T,\widetilde{E}} \neq 0$, as indicated next:[4]

$$\begin{aligned}
\sigma(T,\widetilde{E}) &= \sigma(T,[(1-\rho^2_X)\overline{X} + \rho^2_X\, X] - T) \\
&= \sigma(T,\rho^2_X\, X - T) \\
&= \sigma(T,\rho^2_X\, X) - \sigma(T,T) \\
&= \rho^2_X \sigma(T,X) - \sigma^2(T) \\
&= \rho^2_X \sigma(T,T+E) - \sigma^2(T) \\
&= \rho^2_X \sigma^2(T) + \rho^2_X \sigma(T,E) - \sigma^2(T) \\
&= \rho^2_X \sigma^2(T) - \sigma^2(T) \\
&= -\sigma^2(T)(1-\rho^2_X) && (6) \\
&= -\sigma^2(\widetilde{E}). && (7)
\end{aligned}$$

The equivalence of Equations 6 and 7 follows from the well-known result that the square of the so-called "standard errors of estimate" (sometimes called the "standard errors of estimation") for the linear regression of T on X (i.e., RSEs) is $\sigma^2_{\widetilde{E}} = \sigma^2_T(1-\rho^2_X)$, which is usually expressed in the measurement literature as

$$\sigma^2_{\widetilde{E}} = \sigma^2_X\, \rho^2_X\, (1-\rho^2_X). \tag{8}$$

[4]In this proof, for ease of readability, most variables are put in parentheses rather than used as subscripts.

3.2 Zero Expectation for Errors for Persons in CTT

As noted previously, most developments of CTT make the assumption that $\sigma_{T,E} = 0$. Feldt and Brennan (1989, p. 108) and Haertel (2006, p. 69), however, replace this assumption with the following

> *expectation assumption*: The expected value of the errors (E) for any subpopulation of examinees is 0 under the *restriction* that the examinees are *not* selected based on their values for X.

Letting $\boldsymbol{E}$ be the expectation operator, this means formally that $\boldsymbol{E}_p\, E_p = 0$ provided examinees are not selected based on X. To illustrate that the restriction is necessary, we consider the special case in which all examinees have the same observed score. In this special case, $\sigma_{X,E} = 0$, and

$$\sigma_{T,E} = \sigma_{(X-E,E)} = \sigma_{X,E} - \sigma_E^2 = -\sigma_E^2, \tag{9}$$

which is clearly not zero.[5]

By contrast, suppose persons are selected on the basis of T. Since $\boldsymbol{E}_p\, E_p = 0$, the regression of E on T is a horizontal line passing through $E = 0$. It follows that $\sigma_{T,E} = 0$.[6] In short, the *expectation assumption* implies that $\sigma_{T,E} = 0$, which leads to the conclusion in Equation 7 that $\sigma_{T,\widetilde{E}} \neq 0$. This means that T and $\widetilde{E}$ are correlated, whereas T and E are not correlated under CTT assumptions.

From the perspective of CTT, conditioning on a specific observed score, x, leads to a violation of a crucial CTT assumption. By contrast, the very definition of RSEs is based on conditioning on observed scores—i.e., for any given x, the RSE is $\boldsymbol{E}(T|x)$. Clearly, RSEs are not simply an addendum to CTT.

4 Concluding Comments

Another inconsistency between CTT and RSEs concerns bias. Let x and τ designate realizations of X and T, respectively. In CTT, $\boldsymbol{E}\, x = \tau$, where the expectation is taken over persons with the same observed score,[7] which means that x is an unbiased estimate of τ. By contrast, for RSEs the expected value of true scores for persons with the same observed score is

$$\boldsymbol{E}(T|x) = \widetilde{\tau} = (1 - \rho_X^2)\overline{X} + \rho_X^2\, x,$$

[5] Comparing Equations 7 and 9 leads to the conclusion that, when the selected examinees have the same observed score, not only is a crucial CTT assumption violated, but also $\sigma_{T,E} = \sigma_{T,\widetilde{E}} = -\sigma_E^2$.

[6] Note that the *expectation assumption* is more general than assuming that $\sigma_{T,E} = 0$, since the selection of examinees can be based on anything other than X.

[7] Alternatively, in CTT the expectation could be over replications of the measurement procedure for the same person.

which means that a RSE for an examinee is a biased estimate of τ except when $x = \overline{X}$.[8] In general, for $x < \overline{X}$, RSEs tend to be greater than true scores, and for $x > \overline{X}$, RSEs tend to be less than true scores. That is, RSEs exhibit regression to the mean.

The error variance in CTT is $\sigma^2_E = \sigma^2_X(1 - \rho^2_X)$, whereas for RSEs the variance of the errors of estimate (or estimation) is $\sigma^2_{\widetilde{E}} = \sigma^2_X \, \rho^2_X (1 - \rho^2_X)$. Obviously, $\sigma^2_{\widetilde{E}} < \sigma^2_E$, which is perhaps the most frequently cited rationale for preferring RSEs to X. This characteristic of RSEs is purchased at a substantial price, however. For example,

- RSEs are biased estimates,
- the usual variance-ratio expressions for reliability do not apply to RSEs,
- RSEs are correlated with true scores, and
- theoretically, any examinee has as many different RSEs as there are subgroups to which the examinee belongs.

It is also worth noting that E is measurement error in CTT, whereas $\widetilde{E}$ is prediction error for RSEs. This implies that an estimate of the variance of E and an estimate of the variance of $\widetilde{E}$ are not estimating the same parameter, which raises serious questions about the merits of preferring RSEs to X based solely on the fact that $\sigma^2_{\widetilde{E}} < \sigma^2_E$. In short, RSEs are not an integral part of CTT, especially since RSEs have characteristics that are inconsistent with certain CTT assumptions and results. RSEs have an internally consistent logic, but that logic is not isomorphic with CTT, and investigators need to exercise considerable care in interpreting results that mix CTT and RSEs.

5 References

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). New York: American Council on Education and Macmillan. (Currently published by Oryx).

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 65–110). Westport, CT: American Council on Education/Praeger.

Kelley, T. L. (1947). *Fundamentals of statistics.* Cambridge, MA: Harvard University Press.

[8] There is a degree of ambiguity in this explanation. Specifically, for an infinite population of persons, there is a distribution of true scores for any subpopulation of persons all of whom have the same x. For the vast majority of these persons the RSE, $\widetilde{\tau}$, will *not* be τ, but there will be some persons for whom $\widetilde{\tau} = \tau$ (at least in the limit). This statement is based, in part, on the fact that T is continuous since it is defined as an expectation over observed scores, and the population is assumed to be infinite.