RELIABILITY OF A GUTTMAN SCALE SCORE *

Charles H. Proctor, North Carolina State University

A probability model that seems to reflect the spirit of the Guttman scale model was used in an earlier paper [Proctor, 1970] to furnish an analysis of item response data. By using the estimates for the probability model provided there and by assigning integer scores to the true types, it becomes possible, as this note will describe, to calculate a reliability for these scale scores. The calculation involves the use of both the estimates of the proportions of the underlying true types as well as of the, so called, misclassification parameter. This is a conditional probability with the following meaning. If a subject belongs to a given true type then his responses to each item can be anticipated and the misclassification parameter is the probability that his response to a given item is opposite from that anticipated. By assigning equal probabilities to all true types an alternative reliability, called "flat" reliability, can be calculated. It is quite a bit simpler to associate a standard error to the flat reliability than to the scale reliability, and the quantity may also be more intrinsically interesting.

The assignment of integer scores to the true types, to some extent runs counter to the spirit of the ordered category, rather than numerical, nature of the true types. In data handling practice, scores may be preferred to just the category assignment, particularily for use in a regression computation, consequently some measure of reliability would be welcome for correcting regression coefficients and multiple correlation coefficients for attenuation [Cochran, 1970].

The scoring may be described as follows. A response pattern will be represented, as usual, as a string of plusses and minuses. The integer scores 0, 1, ..., K will correspond to the true type patterns (namely -- ...-, -- ...+, ..., -+...+, ++...+) with that number of plus responses. A non-scale response pattern will be scored for that true type which maximizes its (the non-scale response pattern's) posterior probability. The prior or underlying probabilities of the true types (namely, θ_0, θ_1, ..., θ_K), will be estimated by the maximum likelihood scoring method along with the probability of misclassification, α. These estimators were described in the earlier paper [Proctor, 1970]. The posterior probability of the ℓ^{th} true type for a given, say the i^{th}, response pattern, is obtained by multiplying $\hat{\theta}_\ell$ by $\hat{\alpha}^{D_{i\ell}}(1-\hat{\alpha})^{K-D_{i\ell}}$, where "$D_{i\ell}$ is the number of item responses that need to be changed to modify true type ℓ into observed response pattern i." The hat notation signals the use of estimates.

The true type scores may be denoted by τ and the response pattern scores by X. The correlation between these two random variables could reasonably be referred to as the index of Guttman scale score reliability while its square will be called the scale reliability and written SR. [See Lord and Novick, 1968, p. 61, for the definition of reliability.] The joint distribution of τ and X is fully specified by the parameters θ_0, ..., θ_K, and α and by the scheme for scoring. Having point estimates of these parameters, it is a routine matter to calculate the scale reliability ($\widehat{SR}$, say) as if these were the parameter values. This produces a consistent estimate of scale reliability.

In some respects it is unfortunate that the reliability of a scale should depend on the underlying distribution of the true scores. If as a standard distribution of true scores one takes the uniform (each τ-value has probability 1/K+1) and if the response patterns are scored by the number of plusses, then the correlation between true score and observed score depends only on the misclassification parameter. The square of this correlation will be called flat reliability--"flat" in honor of the uniform distribution of τ. The following results point out how the formula for flat reliability is derived.

The observed score X is now the sum of item zero-one scores, say $X = X_1 + X_2 + \cdots + X_K$. Here $X_1 = 1$ whenever a true type $\tau = 1$ appears (which appearance has probability 1/K+1) and the response is a consistent one, or when other true types appear (each with probability 1/K+1) and the response is not consistent. Upon recalling that the quantity α is the probability of an inconsistent (or "misclassified") response while $1-\alpha$ is the chance of a response consistent with the true type one obtains:

$$E(X_1) = [K\alpha + (1-\alpha)]/(K+1).$$

Similarily,

$$E(X_2) = [(K-1)\alpha + 2(1-\alpha)]/(K+1).$$

Finally,

$$(1) \qquad E(X) = K/2.$$

By slightly heavier algebra one can find:

$$(2) \qquad E(X^2) = K/2 + K(K-1)[1 - \alpha(1-\alpha)]/3,$$

and

$$(3) \qquad V(X) = E(X^2) - [E(X)]^2$$

$$= \frac{K(K+2)}{12}\left[1 - 4\alpha(1-\alpha)\frac{K-1}{K+2}\right].$$

Incidentally, the fact is often used that the sum of the squares of first K integers is $K(K+1)(2K+1)/6$. Their sum is $K(K+1)/2$. Formula (1) was obtained by squaring $X_1 + X_2 + \dots + X_K$, using the fact that $E(X_i^2) = E(X_i)$ for these indicator variates, and then by finding $E(X_iX_j)$ as the sum of 3 parts--one from those true types where $X_i = 1$ and $X_j = 1$ arises from two consistent responses [with probability $(1-\alpha)^2$], another where $X_iX_j = 1$ is produced by two inconsistent responses [with probability α^2] and the third case where $X_i = 1$ and $X_j = 1$ reflects one inconsistent and the other a consistent response [with probability $\alpha(1-\alpha)$].

By similar steps one finds that:

$$(4) \qquad E(\tau) = K/2$$

and

$$(5) \qquad E(\tau^2) = (2K^2 + K)/6 ,$$

while

$$(6) \qquad V(\tau) = \frac{K(K+2)}{12} .$$

$$(7) \qquad E(\tau X) = [K(2K+1) - \alpha K(K+2)]/6$$

and

$$(8) \qquad Cov(\tau, X) = K(K+2)(1-2\alpha)/12.$$

The final step is to square $Cov(X,\tau)$ and then divide by $V(X)$ and by $V(\tau)$ to get:

$$(9) \qquad FR = (1 - 2\alpha)^2/\left[1 - 4\alpha(1-\alpha)\frac{K-1}{K+2}\right] ,$$

the formula for flat reliability of a Guttman scale. It is apparent that a probability of misclassification of over 0.5 will cause reliability to go below zero. That is, "guessing" will cause reliability to decrease.

As an estimate of FR one would replace α by $\hat{\alpha}$ as obtained from the maximum likelihood estimation calculations. Since that computational routine also provides a standard error for $\hat{\alpha}$ an approximate one can be furnished for $\widehat{FR}$. By taking the derivative of FR with respect to α, substituting $\hat{\alpha}$ for α in that expression and then multiplying its absolute value into the standard error of $\hat{\alpha}$ this standard error becomes:

$$(10) \qquad S.E.(\widehat{FR}) = \frac{12(1 - 2\hat{\alpha})[S.E.(\hat{\alpha})]}{(K+2)\left[1 - 4\hat{\alpha}(1-\hat{\alpha})\frac{K-1}{K+2}\right]^2} .$$

A set of data that showed $\hat{\alpha} = .0780$ with $S.E.(\hat{\alpha}) = .00764$ would thus show $\widehat{FR} = .852$ with $S.E.(\widehat{FR}) = .016$. This set of data is No. IV in Table 1. The scale reliability was estimated as $\widehat{SR} = .872$ for those data. As might be expected there is quite close numerical agreement between SR and FR over various sets of data (See Table 1). There would seem to be some advantages to FR since it is standardized for the distribution of true types, and this uniform distribution could be seen as a somewhat ideal distribution for Guttman scaling purposes. In cases where the underlying distribution is not close to uniform (as for data Set I in Table 1), there SR may be different from FR (in fact $\widehat{FR} = .66$ while $\widehat{SR} = .72$ for those data). Perhaps the full range of integer scores would not be entirely appropriate for these data.

FOOTNOTES

* Computer time was made available for this work from a National Science Foundation grant to the Triangle Universities Computer Center.

REFERENCES

[1] Cochran, W. G. Some effects of errors of measurement or multiple correlation. Journal of the American Statistical Association, 1970, 65, 22-34.

[2] Hayes, D. G. and Borgatta, E. F. An empirical comparison of restricted and general latent distance analysis. Psychometrika, 1954, 19, 271-279.

[3] Proctor, C. H. A probabilistic formulation and statistical analysis of Guttman scaling. Psychometrika, 1970, 35, 73-78.

TABLE 1. Guttman Scale Reliability and Flat Reliability for Six Sets of Data *

Set	True Type Proportions θ_0	θ_1	θ_2	θ_3	θ_4	θ_5	Misclassification $\hat{\alpha}$	Scale Reliability $\hat{SR}$	Flat Reliability $\hat{FR} \pm S.E.(\hat{FR})$
I	.17	.01	.08	.09	.07	.59	.163 ± .014	.720	.661 ± .035
II	.05	.15	.23	.25	.13	.19	.038 ± .006	.925	.932 ± .012
III	.11	.12	.27	.26	.20	.03	.074 ± .009	.818	.861 ± .018
IV	.24	.15	.22	.13	.08	.18	.078 ± .008	.872	.852 ± .016
V	.14	.10	.14	.30	.21	.11	.028 ± .005	.951	.951 ± .010
VI	.33	.11	.14	.13	.17	.13	.051 ± .010	.932	.907 ± .019

* Source [Hayes and Borgatta, 1954].