

ON THE STANDARD LENGTH OF A TEST*

MAX A. WOODBURY

INSTITUTE FOR ADVANCED STUDY
AND
UNIVERSITY OF MICHIGAN

(1) A new descriptive parameter for tests, the standard length, is defined and related to reliability, correlation, and validity by means of simplified versions of known formulas. (2) The standard error of measurement is found to be related in simple fashion to the amount of information in a test in the sense of R. A. Fisher. The amount of information is computable as the test length divided by the standard length of the test. (3) The invariant properties of the standard length of a test under changes in length are discussed and proved. Similar results for the correlation coefficient corrected for attenuation and the index of validity are indicated.

Introduction

In connection with another study the notion of the *standard length* of a test turned out to be a useful means of simplifying notation and clarifying proofs. This brief note is presented to introduce this new and possibly valuable notion. The standard length is related to the *information* of R. A. Fisher† through the variance of the errors of measurement. There is an indirect relation to the type of information considered by Shannon‡ and later by Wiener.

It has long been recognized that the reliability of a test can be used (under certain restrictions which do not concern us here) to obtain the reliability of the test after it has been lengthened. Similar relations hold for the correlation between tests or the correlation of a test with a criterion (validity coefficient). This leads to the notion of the reliability and the validity as mathematical functions of the length of the test and the correlation between two tests as functions of the test lengths.

*The research covered by this note was supported by the Office of Naval Research.

†Fisher, R. A. *Statistical Methods for Research Workers*, 10th Edition. London: Oliver and Boyd, 1946, p. 346.

‡Shannon, C. E. A mathematical theory of communication. *Bell System Technical Journal*, 1948, 27, 379-423; 623-656.

Standard Length

The functional dependence of reliability upon test length is of a rather special algebraic character and involves only one parameter. It will be to the advantage of all if this parameter is chosen to simplify the formula. In the usual form of the relation the parameter is the reliability at a given (observed) length and gives the reliability for a test of e_i times the original length. The well known expression

$$\frac{e_i r_{ii}}{1 + (e_i - 1) r_{ii}} \quad (1)$$

expresses this relation.

If we rather arbitrarily define (however see the comment following (4)) the standard length of the test i as

$$r_i = \frac{t_i(1 - r_{ii})}{r_{ii}}, \quad (2)$$

where t_i is the observed test length and r_{ii} the observed reliability, then we find that the standard length computed for a test after it has been altered in length is the same as when computed for the original length. Specifically we find, since the new length is $e_i t_i$ and the new reliability is given in (1), that the new standard length is

$$\frac{(e_i t_i) \left[1 - \frac{e_i r_{ii}}{1 + (e_i - 1) r_{ii}} \right]}{\frac{e_i r_{ii}}{1 + (e_i - 1) r_{ii}}} = \frac{t_i(1 - r_{ii})}{r_{ii}}$$

which is the same as before. It should be noted in passing that any other invariant of the test must be a function of the standard length, where by an invariant of the test we mean any parameter of the test which does not depend on the length. It is clear that any invariant describes the contents of the test, not the accidental feature of its length. The standard length of a test together with its length determines the reliability. The formula for this purpose is

$$r_{ii} = \frac{t_i}{t_i + r_i}, \quad (3)$$

where t_i is the length of the test. From this it is easy to see that when a test has a length equal to its standard length it has a reli-

ability of one-half, when it has a length of twice its standard length it has a reliability of two-thirds, etc. As a matter of convenience we note that in order to obtain a reliability of r_{ii} the length of the test must be given by the relation

$$\frac{t_i}{\tau_i} = \frac{r_{ii}}{1 - r_{ii}} \tag{4}$$

Other definitions of τ_i in (2) would lead to less simple formulas for (3) and (4) so that this may be considered as justification for the particular choice for τ_i .

Fisher has used a concept of information which gives the variance of errors as the reciprocal of the amount of information. This concept can be related to the reliability through the easily derived formula

$$r_{ii} = \frac{1}{1 + \sigma_i^2}, \tag{5}$$

where σ_i is the standard error of measurement and the standard deviation of the true scores is taken as a unit. Combining this equation with (3) we see that the amount of information is

$$J_i = \frac{1}{\sigma_i^2} = \frac{r_{ii}}{1 - r_{ii}} = \frac{t_i}{\tau_i}, \tag{6}$$

i.e., the length of the test measured in terms of its standard length as a unit. Thus a unit of information is the amount of information in a test of standard length.

Correlation and Validity

The formula analogous to (1) for computing the correlation between the tests i and j after each has been lengthened is*

$$r_{ij} \sqrt{\left(\frac{e_i}{1 + (e_i - 1)r_{ii}} \right) \left(\frac{e_j}{1 + (e_j - 1)r_{jj}} \right)} \tag{7}$$

where e_i and e_j are the ratios of the lengths of the lengthened tests to the original tests, and r_{ii} , r_{jj} and r_{ij} are the original reliabilities and correlation. By noting the relationship of (7) to (1) and (3) one can write down immediately the equation for the correlation as

*See Peters, C. C., and Van Voorhis, W. R. *Statistical Procedures and their Mathematical Bases*. N. Y.: McGraw-Hill Book Co., 1940, Eq. 111, p. 193.

a function of the lengths of the tests, viz.

$$r_{ij} = r_{i\infty, j\infty} \sqrt{\left(\frac{t_i}{t_i + \tau_i}\right)\left(\frac{t_j}{t_j + \tau_j}\right)} \quad (8)$$

where

$$r_{i\infty, j\infty} = \frac{r_{ij}}{\sqrt{r_{ii} r_{jj}}} \quad (9)$$

is the correlation coefficient corrected for attenuation and where the other symbols are defined as in (3). It should be noted that the correlation coefficient corrected for attenuation is the same for the lengthened tests as for the original tests so that it, like the standard lengths, is invariant under changes of length and describes a property of the content of the tests only and not of their lengths. To prove this, substitute in (9) from (7) and (1) to obtain the new coefficient corrected for attenuation. Further, any other invariant of the two tests must be a function of the three already described, viz., the standard lengths and the correlation coefficient corrected for attenuation.

The case of correlation with a criterion (validity) scarcely needs separate treatment. Let c denote the criterion and r_{ic} the validity coefficient of test i at length t_i and we have

$$r_{ic} = r_{i\infty, c} \sqrt{\frac{t_i}{t_i + \tau_i}} \quad (10)$$

where $r_{i\infty, c}$ is the index of validity, computable from the formula

$$r_{i\infty, c} = \frac{r_{ic}}{\sqrt{r_{ii}}} \quad (11)$$

The index of validity, like the standard length and the correlation corrected for attenuation, is invariant under changes in the length of the test i . From (10) we can find the length of the test which will give a specified validity: Note that only validities smaller in absolute value than the index of validity can be obtained and that the sign of the validity is unchanged by lengthening the test. Let r_{ic} be the desired validity, let t_i be the length of the test which will give this validity and we have

$$t_i = \frac{\tau_i r_{ic}^2}{r_{i\infty, c}^2 - r_{ic}^2} \quad (12)$$

Manuscript received 6/5/50

Revised Manuscript received 8/8/50