RB 50-35

FILE COPY

RESEARCH BULLETIN

ON THE STANDARD LENGTH OF A TEST

by

Max A. Woodbury (ONR Fellow)

Institute for Advanced Study
Princeton, New Jersey

Sponsored as a Research Bulletin for ETS by Dr. Paul Horst

March 15, 1950

ON THE STANDARD LENGTH

OF A TEST*

by

Max A. Woodbury

Institute for Advanced Study
and
University of Michigan

Abstract

A new concept, that of the standard length of a test, seems useful in simplifying formulas for reliability and validity coefficients for tests of altered length. The usefulness of the concept is a consequence of the fact that it is possible to define a unit of test information, this being, in fact, the amount of information in a test of standard length. The amount of information in a test is its length L in terms of its standard length as a unit. From this one computes the reliability r of the test by the formula $r = L/(L + 1)$. For small amounts of information, the reliability and the information are approximately equal. It is readily seen that the standard length of a test is that length which would make its reliability one half. The length of a test in terms of its standard length as a unit required to produce a given reliability is given by $L = r/(1 - r)$. Both formulas reduce to $(1 + L)(1 - r) = 1$.

The correlation between two tests of length L_1 and L_2 in terms of their standard lengths as units is $r_{12} = R_{12}\ L_1L_2/(L_1 + 1)(L_2 + 1)^{1/2}$, where R_{12} is the correlation corrected for attenuation. For the case of correlation with a criterion this becomes $r_{1c} = R_{1c}\ L_1/L_1 + 1^{1/2}$ where R_{1c} is the validity coefficient corrected for attenuation.

Introduction:

In some correspondence relating to the notions advanced by Horst 1 and Taylor 2 the author introduced and used the notion of the standard length of a test to simplify markedly some formulas. This notion does not seem to be currently familiar so this note was prepared to present it to the readers of Psychometrika as an interesting and possibly useful unit of test length.

Subsequent investigations have shown some interesting relations to the theories of information developed by Fisher as a statistical tool and by Shannon in the theory of communication. The full picture in this connection is still in abeyance and only preliminary results are available. Later it is hoped to publish the results in full. A subsequent paper will cover the application of these techniques to the problem treated by Horst and Taylor mentioned above.

---

* The author wishes to acknowledge the aid of the Office of Naval Research in providing the funds for conducting the research covered in this note.

<u>Standard Length:</u>

The Spearman-Brown formula for the reliability of a test "i" as a function of its length can be written:

$$r_{ii}(t_i) = \frac{t_i r_{ii}}{a_i + (t_i - a_i) r_{ii}(a_i)} , \tag{1}$$

where a_i is its original length and $r_{ii}(a_i)$ is the reliability at length a_i and similarly for t_i and $r_{ii}(t_i)$. The conditions for which this formula is valid are not our concern here. The equation can be written in a form which will exhibit a property of the test which is independent of its length, viz.

$$t_i \;\; 1 - r_{ii}(t_i) \;\; / r_{ii}(t_i) = a_i \;\; 1 - r_{ii}(a_i) \;\; / r_{ii}(a_i).$$

This numerical property of the test has the dimension of test length, hence it has been decided to call it the <u>standard length</u> of the test "i" and denote it by the symbol T_i. Since this quantity does not depend on the actual length of the test, but only on the kinds of items in the test it can be used to compare tests whose lengths are different. The definition of the standard length gives us the formula

$$T_i = a_i \;\; 1 - r_{ii}(a_i) \; . \tag{2}$$

In terms of the new quantity we find

$$r_{ii}(t_i) = t_i/(t_i + T_i) \tag{3a}$$

and a reformulation gives

$$t_i(r) = T_i \; r/(1 - r) \tag{3b}$$

gives us the length $t_i(r)$ of the test "i" which will give a specified reliability r. It can now be easily seen that if the length of a test happens to be exactly its standard length then the reliability is one half.

If the length of a test is expressed in terms of the standard length as a unit, $L_i = t_i/T_i$ we find the relation between L_i and r_{ii} to be

$$r_{ii} = L_i/(L_i + 1), \tag{3c}$$

and

$$L_i = r_{ii}/(1 - r_{ii}). \tag{3d}$$

The quantity L_i which we shall call the "total information" in the test "i" at the given length is determined by the reliability uniquely and conversely. However it has the advantage that tests can be compared validly in terms of this quantity and their length while reliabilities can not. Thus if test "1" has reliability 0.7 and length 35 minutes and test "2" has reliability 0.8 and length 40 minutes one might presume that they are equally "reliable" per unit length since the reliabilities are proportional to the length. However the total information in test "1" is 2.33 units and in test "2" the total information is 4.00 units which are not in proportion to their length, the test "2" having the greatest amount of information per unit of time. Thus in an hour test "1" would have 4.00 units and test "2" would have 6.00 units of information and thus has 50% more information per unit of time.

Correlation and Validity:

Computations involving the correlation between two tests as a function of their lengths can be handled simply in terms of the standard lengths of the tests and the correlation between the tests corrected for attenuation. Let "i" and "j" be the two tests, r_{ii} and r_{jj} be their observed reliabilities at the respective lengths of a_i and a_j and let r_{ij} be their observed correlation at these lengths. Then another well known formula gives their correlation when their lengths are t_i and t_j:

(4) $$r_{ij}(t_i,t_j) = r_{ij} \; {t_i t_j}^{1/2} / \; {a_i+(t_i-a_i)r_{ii}}^{1/2} \; {a_j+(t_j-a_j)r_{jj}}^{1/2},$$

which reduces to (1) when $i = j$, $a_i = a_j$ and $t_i = t_j$. This can be written as

(5a) $$r_{ij}(t_i,t_j) = R_{ij} \; {t_i t_j}^{1/2} \; {(t_i+T_i)(t_j+T_j)}^{1/2},$$

or also as

(5b) $$r_{ij}(L_i,L_j) = R_{ij} \; {L_i L_j/(L_i+1)(L_j+1)}^{1/2},$$

where

(6) $$R_{ij} = r_{ij}/ \; {r_{ii} r_{jj}}^{1/2}.$$

The value of the correlation corrected for attenuation, like the standard length is not affected by changes of length of the two tests. If both tests are of standard length then the correlation between them is just half of R_{ij}.

The case of correlation with a criterion (validity) scarcely needs separate treatment. Presumably the criterion has unit reliability, but this assumption need not be made since this will not affect the formulas below. Let "c" be the criterion and $r_{ic}(t_i)$ the validity of the test "i" at length t_i and let R_{ic} be the validity corrected for attenuation and one has

(7) $$r_{ic}(t_i) = R_{ic} \quad t_i/t_i+T_i \quad ^{1/2} = R_{ic} \quad L_i/L_i+1 \quad ^{1/2}$$

where

(8) $$R_{ic} = \quad r_{ic}/ \quad r_{ii} \quad ^{1/2}.$$

From (7) we can find the equation for the length of the test which will give a specified validity: Note that only validities smaller in absolute value than the validity corrected for attenuation can be obtained and that the sign of the validity is unchanged by lengthening the test. Let r_{ic} be the desired validity, let $t(r_{ic})$ be the length of the test that will give this validity and we have:

(9) $$t(r_{ic}) = \frac{T_i \quad r_{ic}/R_{ic} \quad ^2}{1 - \quad r_{ic}/R_{ic} \quad ^2}.$$

Bibliography:

1 Horst, Paul
Determination of Optimal Test Length to Maximize the Multiple Correlation Psychometrika, 14 (1949), 79-88.

2 Taylor, Calvin
Maximizing Predictive Efficiency for a Fixed Total Testing Time Psychometrika, (Forthcoming)