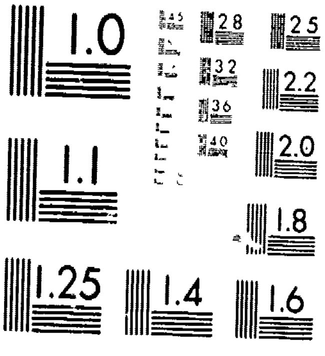

MICROCOPY RESOLUTION TEST CHART

DOCUMENT RESUME

ED 107 674 TM 004 461

AUTHOR Ary, Donald; Karabinus, Robert
TITLE The Use of Reliability Coefficient to Increase Accuracy of the Calculation of n in Power Formulas.
PUB DATE [Apr 75]
NOTE 5p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Hypothesis Testing; Mathematical Models; *Measurement Techniques; *Reliability; Sampling; *Statistical Analysis; *Tests of Significance
IDENTIFIERS Power (Statistics); Variance (Statistical)

ABSTRACT

The power of a statistical test is, in part, a function of the reliability of the dependable variable being analyzed. The substitution of sigma square divided by the reliability coefficient for sigma is proposed. This enables the researcher to incorporate dependent variable reliability information when determining the sample size required for a specified power of his statistical test. An inverse relationship between test reliability and the sample n required to maintain a given statistical power is defined. (Author)

The Use of Reliability Coefficient to Increase Accuracy of the Calculation of n in Power Formulas

Donald Ary and Robert Karabinus, Northern Illinois University

We know that the reliability of the dependent variable measure is related to power (the ability to reject a false null hypothesis). Kerlinger (1964), Helmstadter (1970), and others have reminded us that increased error variance in a dependent variable will increase the likelihood of Type II errors. However, we fail to find in the literature a formula that will enable us to systematically integrate reliability information into power computations.

The purpose of this paper is to propose a way in which knowledge of dependent variable reliability can be used in a typical two-group comparison of means analysis to enable a researcher to more accurately determine the sample size required for a specified power of his statistic.

Let us consider the formula for determining the number of subjects (n) required for rejecting H_o given an absolute difference between population means (δ), level of significance (α), proportion of replications in which H_o will be rejected (power or $1 - \beta$), and population variance (σ^2), as given by Edwards (1967):

$$n = \frac{2\sigma^2}{\delta^2}(z\alpha + z\beta)^2 \qquad \text{Formula 1}$$

Note that this formula does not incorporate the reliability of the instrument used to measure the dependent variable. The formula will yield the same n for a dependent variable measure with a variance of 100 and reliability of .40 as it does for a test with a variance of 100 and a reliability of .95.

We know that if dependent variable variance is kept constant, an investigator employing a measure with low reliability has less ability to reject a false null hypothesis (power) than an investigator employing a measure with high reliability. We also know that in making comparisons between groups we can compensate for low reliability through increasing the number in the groups.

We propose to demonstrate that we can systematically establish how much increase in n is required to compensate for a decrease in reliability by employing the definitions of true variance and error variance in such a way that as error variance increases, an appropriate compensating increase in n will be indicated.

The observed population variance can be partialled into true and error variance components using classical variance formulas from Gullikson, 1950: Observed variance = True variance + Error variance; True variance ÷ Observed variance = test reliability; therefore, True variance = Observed variance x test reliability. Whenever reliability is less than perfect, observed variance will be greater than true variance and the observed standard deviation (σ_x) will be greater than the true standard deviation (σ_{∞}).

Delta (δ) is by definition a true difference between population means ($\mu_1 - \mu_2$). If we define δ as a distance in true standard deviation units (σ_{∞}) rather than observed standard deviation units (σ_x), we can establish a systematic method for adjusting n according to reliability. We find no authority which maintains that δ must be defined in true standard deviation units. On the other hand we find no authority which maintains that it should be defined in observed standard deviation units. So we shall procede from the premise that the former is reasonable and define δ, a true difference, measured in true standard deviation units (σ_∞) rather than measured in observed standard deviation units (σ_x) which are a combination of true and error variation.

$$(\sigma_x^2 = \sigma_\infty^2 + \sigma_e^2)$$

If we define the $\mu_1 - \mu_2$ distance in σ_∞ units, the variance of $\overline{X}_1 - \overline{X}_2$ under a true H_o ($\mu_1 - \mu_2 = 0$) becomes:

$$\frac{2\sigma_\infty^2}{n}$$

and when the ratio

$$\frac{\delta}{\sqrt{\frac{2\sigma_\infty^2}{n}}} \Big/ (z_\alpha + z_\beta)$$

equals unity solving for n yields a formula identical to formula 1 except for the substitution of σ_∞^2 for σ^2.

When an instrument with less than perfect reliability is employed, error variance is added to produce observed variance equal to the true variance divided by reliability.

$$\sigma_x^2 = \frac{\sigma_\infty^2}{r}$$

The expected observed variance of $\overline{X}_1 - \overline{X}_2$ becomes:

$$\frac{2(\sigma_\infty + \sigma_e)^2}{n} \quad \text{or} \quad \frac{2\sigma_\infty^2}{nr}$$

and a null hypotheses at α level will be rejected β proportion of trials when the ratio:

$$\frac{\delta}{\sqrt{\frac{2\sigma_\infty^2}{nr}}} \Big/ (z_\alpha + z_\beta)$$

equals unity. Solving this equation for n we have

$$n = \frac{2\sigma_\infty^2}{\delta^2 r} (z_\alpha + z_\beta)^2 \qquad \text{Formula 2}$$

As a result of the amendment in the numerator of the formula, the size of $\underline{n}$ will be larger when claculated by Formula 2 than when calculated with Formula 1, except when reliability is perfect ($r_{xx} = 1.0$). To illustrate this modification, consider an example first computed with the original Formula 1, then with the amended Formula 2, using two different reliability coefficients.

An investigator has a treatment which is hypothesized to increase WISC full scale IQ scores. He can randomly assign subjects to E and C groups. He wants a true mean difference of one-third the standard deviation to be statistically significant at the .05 level in 90% of replications of the experiment. Given the WISC standard deviation of 15, he computes numbers of subjects requ'red using Formula 1 as follows:

$$\underline{n} = \frac{2(15)^2}{5^2} (1.96 + 1.28)^2 \quad \frac{450}{25} (10.5) = \underline{189.0}$$

Employing Formula 1 the investigator would conclude that he needs 189 subjects in each of his two groups to obtain the power desired. However, if we define δ in true standard deviation units this would be the case only if the reliability of the WISC is unity (which it is not).

Now assume the experiment is to be done with 10½ year olds for whom the Wechsler (1949) manual reports a reliability of .95. Using the amended Formula 2, we have the following:

$$\underline{n} = \frac{2(15)^2}{5^2(.95)} (1.96 + 1.28)^2 = \frac{450}{23.75} (10.5) = \underline{198.9}$$

Even with this very high reliability, the number required in each of the two groups is about 10 more than the number derived with the standard Formula 1.

A further increase in $\underline{n}$ will be required if the experiment is done with 7½ year olds, for whom the Wechsler (1949) manual reports a reliability of .92.

Again, using Formula 2, we have the following:

$$\underline{n} = \frac{2(15)^2}{5^2(.92)} (1.96 + 1.28)^2 = \frac{450}{23} (10.5) = \underline{205.4}$$

If we define δ in true standard deviation units rather than observed standard deviation units the value $\sigma_x\sqrt{r}$ substituted for σ enables us to incorporate information concerning the reliability of the dependent variable measure into the calculation of the sample $\underline{n}$ required to maintain a given statistical power.

REFERENCES

Edwards, A.L. Statistical Methods, second edition. New York: Holt, Rinehart & Winston, Inc., 1967.

Gulliksen, H. Theory of Mental Tests. New York: John Wiley & Sons, Inc., 1950.

Helmstadter, G.C. Research Concepts in Human Behavior, Education, Psychology, Sociology. New York: Appleton - Century - Crofts, Inc., 1970.

Kerlinger, F.N. Foundations of Behavioral Research, Educational and Psychological Inquiry. 2nd Ed. New York: Holt, Rinehart and Winston, Inc., 1973.

Wechsler, D. Wechsler Intelligence Scale for Children. New York: The Psychological Corp., 1949.