

Reliability, Power, Functions, and Relations: A Reply to Humphreys

Donald W. Zimmerman, Carleton University

Richard H. Williams, University of Miami

Bruno D. Zumbo, University of Ottawa

Index terms: difference scores, error of measurement, power, significance tests, t test, test reliability, true scores.

In previous discussions of statistical power and reliability of measurement (Williams & Zimmerman, 1989; Zimmerman & Williams, 1986; Zimmerman, Williams, & Zumbo, 1993), we emphasized that controversy sometimes arises because investigators ask imprecise and inappropriate questions. After reading Humphreys' (1993) comments, we are more convinced of the importance of asking the proper questions and paying attention to underlying assumptions.

The question "How does reliability of measurement determine the power of a significance test?" implies the existence of a functional relation. Similar questions include: "How does force determine the acceleration of a moving body?" and "How does voltage determine the current in a conductor?" Calculations based on simple equations provide specific values of acceleration and of current, and the same might be true of statistical power.

As pointed out in our earlier papers (Williams & Zimmerman, 1989; Zimmerman & Williams, 1986; Zimmerman, Williams, & Zumbo, 1993), however, power is not a mathematical function of reliability unless either true score variance or error score variance is constant. Reliability is determined entirely by the relative magnitudes of true score variance and observed score variance.

APPLIED PSYCHOLOGICAL MEASUREMENT

Vol. 17, No. 1, March 1993, pp. 15-16

© Copyright 1993 Applied Psychological Measurement Inc.
0146-6216/93/010015-02\$1.35

On the other hand, statistical power is determined by the absolute magnitude of observed score variance, together with several other variables, such as sample size, significance level, and the alternative hypothesis.

For this reason, the correspondence of reliability and power has the status of a relation, not a function. Any change in reliability that is accompanied by an increase in observed score variance reduces power. A unique correspondence is lacking as long as the components of observed score variance are unrestricted. Similarly, it is not possible to describe how current depends on voltage if resistance is allowed to change at the same time voltage changes.

Humphreys (1993, p. 11) mentions the freedom of statisticians to "... obtain numerical solutions by varying independently the parameters of the power of statistical tests." The exercise of this freedom is commonplace in both statistics and test theory. If true score variance is 10 and observed score variance is 20, then the reliability coefficient, as defined in classical test theory, is .50, without further qualification. Similarly, the power of a significance test is determined by the variables listed previously, no matter how particular values arise in practice.

If a test is increased in length k times, reliability increases as a function of k according to the Spearman-Brown formula. Because research persons "... have empirical limitations placed on them" (Humphreys, 1993, p. 11), it is usually not feasible to let $k = 10$, but the formula is useful nevertheless. Practical considerations are

important and interesting, but may not be germane to derivation of equations.

The generality of our approach does not imply that we are insensitive to the practical needs of empirical researchers. On the contrary, juxtaposition of concepts in statistical power analysis and test theory has many practical consequences. The equations and the results presented in the tables in Zimmerman et al. (1993) are a step in that direction. A researcher can calculate the power to be expected under specific conditions, if desired.

Moreover, this perspective helps resolve an earlier controversy. It emphasizes the significance of the paradox originally pointed out by Overall & Woodward (1975) and in a sense makes their finding less paradoxical. Unquestionably, unreliable dependent variables can be associated with powerful significance tests. However, at the same time, there is no doubt of the importance of augmenting reliability in experimental settings whenever possible.

References

- Humphreys, L. G. (1993). Further comments on reliability and power of significance tests. *Applied Psychological Measurement, 17*, 11-14.
- Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for the measurement of change. *Psychological Bulletin, 82*, 85-86.
- Williams, R. H., & Zimmerman, D. W. (1989). Statistical power analysis and reliability of measurement. *Journal of General Psychology, 116*, 359-369.
- Zimmerman, D. W., & Williams, R. H. (1986). Note on the reliability of experimental measures and the power of significance tests. *Psychological Bulletin, 100*, 123-124.
- Zimmerman, D. W., Williams, R. H., & Zumbo, B. D. (1993). Reliability of measurement and power of significance tests based on differences. *Applied Psychological Measurement, 17*, 1-9.

Author's Address

Send requests for reprints or further information to Donald W. Zimmerman, 2738 Garber St., Berkeley CA 94705, U.S.A.; E-mail Brian D. Zumbo, ZUMBO@ACADVM1.UOTTAWA.CA.