MacPotthoff: Automated calculation of the Potthoff regression bias procedure

MARLEY W. WATKINS and CATHY J. HETRICK
Pennsylvania State University
University Park, Pennsylvania

Test bias, in contrast to test fairness, is best conceptualized in validity terms amenable to statistical analysis. Evidence of predictive validity may be most salient in many situations. Evaluation of predictive bias is generally operationalized via linear regression. Potthoff (1978) provided an efficient and parsimonious regression bias procedure that allows both simultaneous and separate tests of regression slopes and intercepts across groups. A Macintosh computer program, MacPotthoff, is presented for automated calculation of Potthoff regression bias statistics.

Bias in educational and psychological tests is a sensitive and controversial issue which has been addressed in both the popular and scholarly press (Gould, 1981; Jensen, 1980; Lippmann, 1923) as well as by legislation (i.e., 34 C.F.R. § 300.530). Many definitions of test bias have been presented in the literature (Flaugher, 1978). It is important, however, to distinguish between test bias as a question of psychometric adequacy and test bias as the fairness of a test's use (Messick, 1980). Psychometric adequacy can yield to statistical analysis, but fairness can only be defined within societal value systems (Reynolds & Kaiser, 1990).

In the psychometric sense, test bias "is ultimately a question of validity" (Messick, 1988, p. 40), which means that there is "*systematic* error in the estimation of some 'true' value for a group of individuals" (Reynolds & Kaiser, 1990, p. 624). Evidence of such systematic measurement error can best be revealed through analyses of empirical data (Millsap & Everson, 1993). Given the availability of item-level data and generous sample sizes, test bias can be studied via factor analytic and item response theory approaches (Reynolds, Lowe, & Saenz, 1999). However, Reynolds (1982) suggested that evidence of predictive validity is the most salient form of validity evidence in relation to test bias—especially when one is considering aptitude and intelligence tests, which typically do not generate item-level data.

Cleary, Humphreys, Kendrick, and Wesman (1975) concluded that "a test is considered fair for a particular use if the inference drawn from the test score is made with the smallest feasible random error and if there is no constant error in the inference as a function of membership in a particular group" (p. 25). Their definition of predictive bias is generally operationalized via linear regression (Reynolds, 1995), where predictions take the form of $Y = a + bX$. Within this equation, the slope of the regression line is represented by b and the intercept by a, with X as the predictor. A test without bias would predict equally well, regardless of group membership. When the slope or intercept differs significantly across groups, there is bias in prediction when the regression equation based on combined groups is used.

Evaluating regression parameters across groups by conducting separate statistical tests for slope and intercept values unnecessarily inflates the alpha level and overcomplicates the decision-making process (Reynolds, 1982, 1995; Reynolds & Kaiser, 1990; Reynolds et al., 1999). Potthoff (1978) provided a useful procedure that allows a simultaneous test of the equivalence of slopes and intercepts with a single F test. If an overall significant F results, slopes and intercepts can be tested separately to determine their independent equality.

The Potthoff technique was recommended by Reynolds (1982) as "the most efficient method for determining whether the regression equation relating any two variables is constant across groups" (p. 219). That is, it elegantly handles chance-corrected tests for both slopes and intercepts without the user's having to manually create interaction terms, correct for multiple tests, and so forth, as is necessary in general-purpose regression programs. This efficiency has been widely recognized, as is evidenced by applications of the Potthoff technique to test-bias investigations of popular intelligence tests (Bossard, Reynolds, & Gutkin, 1980; Glutting, 1986; Reynolds, Willson, & Chatman, 1985; Weiss & Prifitera, 1995; Weiss, Prifitera, & Roid, 1993).

Although practically and conceptually useful, Potthoff's (1978) regression bias procedure is not included in standard statistics packages. Consequently, it must be coded anew by each researcher. This presents logistical problems, inflates cost, and increases the risk of error. This paper presents a solution to this problem: a computer program for automated calculation of Potthoff regression bias statistics.

Program

MacPotthoff is a FutureBASIC program that operates on Macintosh computers under Mac OS Versions 7 and 8. Data are input via a text file where group membership and univariate predictor and criterion scores are tab delimited. MacPotthoff will conduct analyses with two to nine groups and up to 10,000 cases.

Output from MacPotthoff is both textual and graphical. Textual output includes a partial listing of the input file to ensure data integrity, as well as simultaneous slope and intercept F values and their associated degrees of freedom and approximate p values. Graphical output consists of a scatterplot of the data, color coded by group membership,

---

Correspondence should be addressed to M. W. Watkins, Department of Educational and School Psychology and Special Education, Pennsylvania State University, 227 CEDAR Building, University Park, PA 16802-3109 (e-mail: mwwl0@psu.edu).

overlaid with color-coded common and group regression lines.

Availability

MacPotthoff is available from the first author upon receipt of an initialized Macintosh disk accompanied by appropriate first-class return postage. Alternatively, MacPotthoff can be downloaded from http://espse.ed.psu.edu/spsy/Watkins/SPSY-Watkins.ssi.

REFERENCES

BOSSARD, M., REYNOLDS, C. R., & GUTKIN, T. B. (1980). A regression analysis of test bias on the Stanford–Binet Intelligence Scale. *Journal of Clinical Child Psychology*, **9**, 52-54.

CLEARY, T. A., HUMPHREYS, L. G., KENDRICK, S. A., & WESMAN, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist*, **30**, 15-41.

FLAUGHER, R. L. (1978). The many definitions of test bias. *American Psychologist*, **33**, 671-679.

GLUTTING, J. J. (1986). Potthoff bias analysis of K–ABC and nonverbal scale IQs among Anglo, Black, and Puerto Rican kindergarten children. *Professional School Psychology*, **1**, 225-234.

GOULD, S. J. (1981). *The mismeasure of man*. New York: Norton.

JENSEN, A. R. (1980). *Bias in mental testing*. New York: Free Press.

LIPPMANN, W. (1923). A judgment of the tests. *New Republic*, **34**, 322-323.

MESSICK, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, **35**, 1012-1027.

MESSICK, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33-45). Hillsdale, NJ: Erlbaum.

MILLSAP, R. E., & EVERSON, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, **17**, 297-334.

POTTHOFF, R. F. (1978). *Statistical aspects of the problem of biases in psychological tests* (Institute of Statistics Mimeo Series No. 479). Chapel Hill: University of North Carolina, Department of Statistics.

REYNOLDS, C. R. (1982). Methods for detecting construct and predictive bias. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias* (pp. 199-227). Baltimore: Johns Hopkins University Press.

REYNOLDS, C. R. (1995). Test bias and the assessment of intelligence and personality. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence* (pp. 545-573). New York: Plenum.

REYNOLDS, C. R., & KAISER, S. M. (1990). Bias in assessment of aptitude. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement* (pp. 611-653). New York: Guilford.

REYNOLDS, C. R., LOWE, P. A., & SAENZ, A. L. (1999). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *Handbook of school psychology* (3rd ed., pp. 549-595). New York: Wiley.

REYNOLDS, C. R., WILLSON, V. L., & CHATMAN, S. P. (1985). Regression analysis of bias on the K–ABC. *Journal of School Psychology*, **23**, 195-204.

WEISS, L. G., & PRIFITERA, A. (1995). An evaluation of differential prediction of WAIT achievement scores from WISC–III FSIQ across ethnic and gender groups. *Journal of School Psychology*, **33**, 297-304.

WEISS, L. G., PRIFITERA, A., & ROID, G. (1993). The WISC–III and the fairness of predicting achievement across ethnic and gender groups. In B. A. Bracken & R. S. McCallum (Eds.), *Wechsler Intelligence Scale for Children* (Advances in Psychoeducational Assessment, Journal of Psychoeducational Assessment Monograph Series, pp. 35-42). Cordova, TN: Psychoeducational Corporation.