# Estimating Unrestricted Population Parameters From Restricted Sample Data in Employment Testing

Michael J. Burke, New York University

Jacques Normand, U.S. Postal Service

Lucinda I. Doran, New York University

This study examined the accuracy of Alexander, Alliger, and Hanges' (1984) method for estimating unrestricted univariate predictor means and variances from sample data drawn from three populations in two personnel selection contexts: (1) where there was direct nonstrict truncation on the predictor, and (2) where there was direct strict truncation on the predictor. In addition, the accuracy of corrected (estimated unrestricted) validity coefficients based on estimated population predictor standard deviations was assessed in the nonstrict truncation condition. In general, there was inconsistency in the accuracy of the population predictor mean and standard deviation estimates obtained across the present datasets and conditions. Caution is advised in the interpretation and reporting of corrected validity coefficients in employment testing based on estimated population predictor standard deviations. *Index terms: employment testing, personnel selection, range restriction, true validity estimation, unrestricted population parameters.*

Schmidt, Hunter, and Urry (1976) demonstrated that observed validity coefficients in personnel selection are typically biased. When range restriction operates to bias these validity coefficient estimates, it is recommended that appropriate adjustments be made, when possible, to obtain a population validity coefficient estimate (Society for Industrial and Organizational Psychology, 1987). However, a problem often arises in attempting to correct ob-

served validity coefficients for direct restriction on a variable using Thorndike's (1949) Case 2 formula. That is, knowledge of the unrestricted variance of a variable in the applicant population that is needed in order to correct for range restriction using the Case 2 formula is frequently unavailable. Therefore, personnel researchers have relied on reasonable estimates of population variances obtained from literature reviews, or, in the case of range restriction corrections in meta-analytic studies, on the Schmidt and Hunter (1977) distribution of range restriction effects [i.e., ratios of restricted to unrestricted standard deviations (SDs) on the predictor].

Recently, Alexander, Alliger, and Hanges (1984) proposed a procedure for estimating the unrestricted variance of a variable from the restricted sample data. The procedure is based on the observation that it is possible to table a function of the truncated normal distribution that will allow the extent or point of truncation to be estimated (Cohen, 1959). As noted by Alexander et al. (1984), the correlation of the truncated variable with other variables may then be corrected by standard restriction-of-range formulas. The mean of the restricted variable can also be corrected. Alexander et al.'s results for both simulated data and educational data indicated that this correction for correlations may be useful when the actual unrestricted variance is unknown. In addition, Alexander, Hanges, and Alliger (1985) have extended this restriction-of-range

correction procedure to the bidimensional case. Results based on computer simulations supported the accuracy of this latter procedure in estimating unrestricted validity coefficients under conditions of strict truncation.

Although Alexander et al. (1984) applied this correction procedure to data involving strict truncation on ACT scores, the question of how well this correction procedure performs in applied personnel selection contexts remains unanswered. In a typical personnel selection situation, range restriction often does not involve strict truncation (i.e., where all persons at or above the cutoff score comprise the restricted group) on the predictor. That is, organizations frequently do not select in a top-down manner. Some high-scoring applicants are not selected due to scores on a later measure (e.g., a physical exam) or as a result of affirmative action/equal employment considerations. Other high-scoring applicants may reject offers based on factors unrelated to selection policies. Also, in employment testing there may be no prespecified cut score (Cascio, Alexander, & Barrett, 1988); the number of applicants to be selected is based on the number of position openings. The result is that not all applicants above the lowest-scoring selectee are hired.

These conditions together would be particularly difficult to model with computer simulation (monte carlo studies). Therefore, it would be informative to examine the Alexander et al. (1984) correction procedure under actual employment testing conditions in order to evaluate the accuracy of the procedure.

The present study examined the accuracy of Alexander et al.'s method for obtaining unrestricted estimates of univariate means and variances in two personnel selection contexts: (1) where there is direct nonstrict (incomplete) truncation on the predictor (i.e., personnel selection measure), and (2) where there is direct strict (complete) truncation on the predictor. This study also examined the accuracy of estimated unrestricted validity coefficients (based on estimated population predictor SDs from Alexander et al.'s procedure) in the nonstrict truncation condition. Because the Alexander et al. procedure assumes that the nontruncated predictor

scores are normally distributed, an unambiguous interpretation of the results will hinge on demonstrating that the predictor distributions are normal.

## Method

### Populations

Three populations were employed in examining the accuracy of Alexander et al.'s (1984) correction procedure: applicants for a distribution clerk machine job at a large government agency, sales manager applicants to a large national manufacturing company, and applicants for administrative positions at a small government agency. A distribution clerk machine worker sorts mail using a letter-sorting machine by scanning addresses and zip codes, checks the accuracy of sorted mail, and bands and loads mail in containers for distribution.

The distribution clerk machine applicant population consisted of 23,909 persons who had taken a standardized written test in 15 employment centers located throughout the U.S. during the period February 1980 to February 1981. A sample of 1,271 hirees from those 15 offices (chosen to represent all regions of the U.S.) comprised the restricted sample. The 1,271 hirees were selected based on test scores, personal reference checks, checks of criminal records, medical suitability, availability to work during unusual hours, and veteran preference policies. Of the total applicant group, 17,404 scored at or above the cutoff score on the distribution clerk machine selection test.

The sales manager applicant population consisted of 128 area sales representatives, 29 of whom had been assessed and promoted to district sales manager (the target position of the assessment program). The 29 promotees who comprised the restricted sample were selected based on assessment center scores and the organization's equal employment opportunity goals. The distribution of predictor scores of the 128 employees was based on assessment center evaluations obtained over a 7-year period (1976 to 1983).

The third applicant population consisted of 435 persons who had taken an internally developed typing test at a small government agency over a 4-year period (1982 to 1985). The applicant pool was

obtained based on recruiting efforts in nearby states and applications received at the national office. Because job offers were extended to all persons scoring above the cutoff on the predictor, the restricted sample was considered to consist of those persons scoring above this cutoff. Some persons declined the job offers (primarily due to salary considerations); however, follow-up data on actual job placements were unavailable.

## Personnel Selection Procedures

*Distribution clerk machine test.*   The distribution clerk machine test consisted of 295 items in four subtests: number series, coding, memory, and error detection. The number-correct scores were transformed to a score from 44 to 100 (using a Civil Service transformation). The organization had established a truncation point of 70 for classifying applicants as having passed or failed the selection test. The mean score on the distribution clerk machine test for the population was 75.23 (SD = 8.96).

An important assumption of the present study is that the predictor scores are normally distributed in the population. The statistics employed for examining normality were skewness and kurtosis. The skewness and kurtosis values of the nontruncated (i.e., population, $N = 23,909$) distribution-clerk machine test scores were .26 and .27, respectively. Based on measures of skewness and kurtosis, the nontruncated distribution clerk machine predictor scores were approximately normally distributed. This is especially noteworthy with respect to this test score distribution because these scores were transformed to a scale with a ceiling value of 100. Although the Kolmogorov-Smirnov goodness-of-fit test was statistically significant for the distribution clerk machine population ($D = .03$, $p < .01$), this goodness-of-fit test, like all statistical significance tests, is dependent on sample size.

A test-retest reliability coefficient of .83 was computed for the total score, based on applicants retaking the test within a time interval of approximately six months during the period 1970 to mid-1980. In addition, an observed biserial validity coefficient (see Guilford & Fruchter, 1973) of .27 was calculated between transformed distribution

clerk machine test scores and a pass/fail criterion score. The criterion was a work sample test which required the applicant to key three digits on cards at a rate of 40 cards per minute with 95% accuracy.

*Assessment center.*   The assessment center score was a weighted composite based on five dimensions: planning and organizing, decision making, stress tolerance, sensitivity, and persuasiveness. The mean score on the assessment center for the population of 128 was 138.87 (SD = 7.66). The skewness and kurtosis values of the 128 assessment center scores were .22 and $-.64$, respectively. Thus, the normality assumption was reasonably met for the manager population test score distribution. The Kolmogorov-Smirnov goodness-of-fit test for the manager population test score distribution was not statistically significant. Operational data (i.e., cost data, etc.) and criterion-related validity information on this assessment center were presented in Burke and Frederick (1986).

*Typing test.*   The typing test consisted of practice copy and straight-copy typing. The practice copy was a 2-minute warm-up, which was given to each examinee prior to administering the straight-copy typing test. The straight-copy typing test was a timed 5-minute test consisting of double-spaced narrative text (i.e., 690 words with no numbers or symbols). This test assessed an examinee's basic keystroke speed and typing accuracy.

A net typing speed score was calculated for the straight-copy typing test by subtracting two times the number of typing errors (i.e., the number of wrong letters, punctuation mistakes, strike-overs, etc.) from the number of words per minute. The net typing speed score was the score used to categorize applicants as having passed or failed the typing test. The government agency had, prior to the present study, selected a truncation point of 45. This truncation point was approximately at the mean test score for the 435 examinees over the period of this study (M = 44.36, SD = 16.81). The skewness and kurtosis values of the 435 test scores were .25 and .25, respectively. These latter values indicated that the administrative population test score distribution was approximately normal. In addition, the goodness-of-fit test for this test score distribution was statistically nonsignificant.

A test-retest reliability coefficient of .80 was calculated based on 36 examinees who were tested twice. The time lapse between the initial test and the retest varied from 1 day to 3 months. Criterion-related validity data were not available for the test.

## Data Analysis

Several range restriction conditions were examined in the three applicant populations. For each condition, estimates of the unrestricted mean and SD on the predictor were calculated from the restricted sample data. These estimates were based on formulas presented in Alexander et al. (1984) and discussed below. The estimated population means and SDs were then compared to the actual unrestricted predictor means and SDs. In addition, estimates of the unrestricted validity coefficient for the assessment center predictions and distribution clerk machine test predictions were made. These validity coefficients were computed by initially estimating the unrestricted predictor SD in each sample, obtaining the observed validity data noted above, and then incorporating these data into Thorndike's (1949) Case 2 formula. The estimated unrestricted validity coefficients were then compared to unrestricted validity coefficients based on the applicant population predictor SDs.

For the distribution clerk machine population, three range restriction conditions on the predictor were examined: (1) actual selection, where the predictor scores of the 1,271 hirees comprised the restricted sample; (2) top-down selection, where the top 1,271 scorers on the distribution clerk machine test comprised the restricted sample; and (3) selection above the truncation point (i.e., the lowest observed predictor score of an employee), where predictor scores of all examinees above the truncation point comprised the restricted sample.

In the sales manager population, the three range restriction conditions on the predictor were (1) actual selection, where the predictor scores of the 29 promoted employees comprised the restricted sample; (2) top-down selection, where the top 29 predictor scores (i.e., actually 30 due to a tie between scores 29 and 30) based on a selection ratio of .22 comprised the restricted sample; and (3) selection above the truncation point (i.e., the lowest observed predictor score for a promoted employee), where predictor scores of all examinees above the truncation point comprised the restricted sample.

In the administrative population, the range restriction condition on the predictor, based on actual selection (comparable to Condition 3 for distribution clerk machine and sales manager) was a restricted sample consisting of all persons scoring above the truncation point between 1982 and 1985 ($N = 204$).

*Estimating unrestricted means, SDs, and validity coefficients.* Initially, Cohen's ratio $SD_{obs}^2/(\overline{X}_{obs} - X_c)^2$, the observed sample variance divided by the squared difference between the sample mean and point of truncation, was calculated for each of the above conditions. Then, using Table 1 of Alexander et al. (Cohen's ratio and corresponding restricted SD of $z$ score for truncation point), the SD in a normal distribution truncated at the $z$ cut (i.e., truncation point) was obtained for each condition. Finally, for each condition, the observed sample SD ($SD_{obs}$) was corrected using the tabled SD ($SD_{tab}$, i.e., from Table 1 of Alexander et al.) by

$$\hat{SD} = \frac{SD_{obs}}{SD_{tab}} \quad . \tag{1}$$

Essentially, Equation 1 increases the sample SD proportionately to the amount of restriction estimated by Cohen's ratio.

The corrected unrestricted mean was estimated by

$$\hat{\overline{X}} = X_c - z\,\hat{SD} \quad , \tag{2}$$

where $z$ is the truncation point in standard score form and the other terms are as defined above.

For the sales manager and distribution clerk machine samples, the corrected validity coefficients ($\hat{r}$s) were calculated using the $\hat{SD}$s from Equation 1 (for Condition 1, nonstrict truncation) and Thorndike's Case 2 formula. A value of .41 was used as the observed validity coefficient between assessment ratings and supervisory performance ratings. This value was obtained from a validity study involving the 29 promoted employees (see Burke & Frederick, 1986). As noted above, the observed validity coefficient for the distribution clerk ma-

chine test was .27. The corrected $r$s were then compared with the actual unrestricted validity coefficients ($\hat{\rho}$s). The actual unrestricted validity coefficients were based on correction with the respective applicant population predictor SDs.

## Results and Discussion

As noted above, the normality assumption was reasonably met in all three populations based on skewness and kurtosis measures. Therefore the results, in particular for the nonstrict truncation conditions where population validity coefficients were estimated, can be interpreted with respect to violations of strict (complete) truncation.

Presented in Table 1 are the observed sample statistics and corrected statistics for strict and nonstrict range restriction conditions, and the actual population parameters. Although the population parameters are referred to as actual values for presentation purposes, in reality they are also estimates of the true population values. For the nonstrict truncation condition for sales managers (i.e., actual selection), the mean predictor score was accurately estimated. The population mean score for distribution clerk machine was overestimated. The predictor SDs were also overestimated to a fair degree in each nonstrict truncation condition. That is, the estimated predictor SDs were 123% and 120% of the population SDs for distribution clerk machine and sales manager, respectively. This overestimation of the variance on the predictor resulted in the actual unrestricted validity coefficient being substantially overestimated by .08 ($\hat{r} = .52$, $\hat{\rho} = .44$) for sales managers and .04 ($\hat{r} = .34$, $\hat{\rho} = .30$) for distribution clerks.

The degree of nonstrict truncation in these two populations was substantial. This was due in part to the high quality of the applicant pool at the manufacturing company (i.e., currently employed area sales representatives) as well as to organizational hiring/promotion practices that attempted to increase minority, female, and veteran representation in the workforce at both the manufacturing company and government agency. It is doubtful that these selected samples represent extreme cases of nonstrict truncation in applied personnel selec-

Table 1
Comparison of Observed Sample Statistics,
Corrected Statistics, and Population Parameters for Direct
Strict and Nonstrict Truncation

| Population and Type of Truncation | N | Observed Mean | Observed SD | Corrected Mean | Corrected SD | Population $\mu$ | Population $\sigma$ |
|---|---|---|---|---|---|---|---|
| Distribution Clerk Machine | 23,909 | | | | | 75.23 | 8.96 |
|   Nonstrict | | | | | | | |
|     Actual Selection | 1,271 | 83.15 | 8.43 | 79.28 | 11.05 | | |
|   Strict | | | | | | | |
|     Top-Down Selection | 1,271 | 95.54 | 3.35 | 80.11 | 3.37 | | |
|     Selection Above $X_c$ | 17,404 | 79.14 | 6.85 | 70.56 | 11.19 | | |
| Sales Managerial | 128 | | | | | 138.87 | 7.66 |
|   Nonstrict | | | | | | | |
|     Actual Selection | 29 | 142.72 | 6.74 | 138.64 | 9.22 | | |
|   Strict | | | | | | | |
|     Top-Down Selection | 30[a] | 149.30 | 3.45 | 141.52 | 6.71 | | |
|     Selection Above $X_c$ | 98 | 141.90 | 6.00 | 139.08 | 7.90 | | |
| Administrative[b] | 378 | | | | | 44.36 | 16.81 |
|   Actual Selection | 204 | 56.75 | 8.64 | 47.25 | 13.65 | | |

[a]Because a tie existed between scores 29 and 30, the figures were based on 30 scores.

[b]Actual selection was based on a top-down selection strategy.

tion contexts. Even if this were the case, the results nevertheless suggest that caution be used in interpreting corrected correlations calculated from estimated population SDs on the predictor (using Alexander et al.'s procedure) when there is nonstrict truncation.

Caution is advised in the use of the Alexander et al. (1984) procedure, primarily as a result of the overestimation of the population SDs on the predictor in the nonstrict truncation conditions and the resulting overestimation of the unrestricted population validity coefficients. A confidence interval based on the estimated population SD from the smaller sample (sales manager) would not include the population SD. Clearly, more research relating to the performance of the Alexander et al. correction procedure under nonstrict (incomplete) truncation with large populations in employment testing would be worthwhile.

With respect to the top-down selection conditions in the distribution clerk machine, sales, and administrative populations, the population means were overestimated whereas the population SDs were underestimated. This latter result is somewhat encouraging for organizations which knowingly have an approximate top-down selection program and desire to correct observed validity coefficients with the use of estimated population SDs. That is, the corrected correlation (based on an estimated population SD on the predictor) is likely to be an underestimate of the (unrestricted) population validity coefficient, although not as severe an underestimate as the observed validity coefficient under such a selection context.

Alexander et al. obtained results which were similar to this suggestion regarding possible undercorrection of validity coefficients when there was strict truncation in one of their educational datasets. In another educational dataset, they obtained slightly inconsistent findings, that is, where the corrected correlation overestimated the population value. This study's estimated population predictor SDs for a type of top-down selection (i.e., complete selection above the cutoff score, the lowest-scoring selectee, shown in Table 1) were mixed in terms of accuracy of estimates.

Further research in other applied settings, especially those involving large samples (and populations) and direct nonstrict truncation, would provide additional data for judging the degree of confidence in this procedure. Due to the overestimation of the population predictor SD and consequent overestimation of the validity coefficients in the actual nonstrict truncation conditions, researchers should report the uncorrected correlation coefficient along with the corrected coefficient when using estimated population SDs for such corrections in employment testing.

## References

Alexander, R. A., Alliger, G. M., & Hanges, P. J. (1984). Correcting for range restriction when the population variance is unknown. *Applied Psychological Measurement, 8*, 431–437.

Alexander, R. A., Hanges, P. J., & Alliger, G. M. (1985). Correcting for restriction of range in both $X$ and $Y$ when the unrestricted variances are unknown. *Applied Psychological Measurement, 9*, 317–323.

Burke, M. J., & Frederick, J. T. (1986). A comparison of economic utility estimates for alternative $SD_y$ estimation procedures. *Journal of Applied Psychology, 71*, 334–339.

Cascio, W. F., Alexander, R. A., & Barrett, G. V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. *Personnel Psychology, 41*, 1–24.

Cohen, A. C. (1959). Simplified estimators for the normal distribution when samples are singly censored or truncated. *Technometrics, 1*, 217–237.

Guilford, J. P., & Fruchter, B. (1973). *Fundamental statistics in psychology and education* (5th ed.). New York: McGraw-Hill.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529–540.

Schmidt, F. L., Hunter, J. E., & Urry, V. W. (1976). Statistical power in criterion-related validation studies. *Journal of Applied Psychology, 61*, 473–485.

Society for Industrial and Organizational Psychology. (1987). *Principles for the validation and use of personnel selection procedures* (3rd ed.). College Park MD: Author.

Thorndike, R. L. (1949). *Personnel selection*. New York: Wiley.

## Author's Address

Send requests for reprints or further information to Michael J. Burke, Graduate Division, Leonard N. Stern School of Business, New York University, Room 506 Merrill Hall, 90 Trinity Pl., New York NY 10006, U.S.A.