FAA

UNCLASSIFIED

DTIC

Technical Report

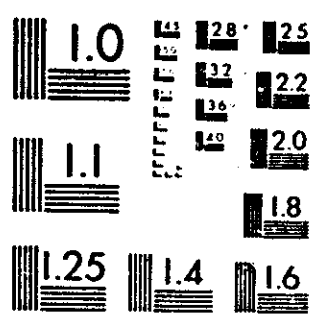
distributed by

Defense Technical Information Center
DEFENSE LOGISTICS AGENCY
Cameron Station • Alexandria, Virginia 22314

UNCLASSIFIED

1 OF 1

AD
A066132



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

LEVEL IV

(12)

FAA-AM-78-36

THE DEVELOPMENT OF THE ATC SELECTION BATTERY: A NEW PROCEDURE TO MAKE MAXIMUM USE OF AVAILABLE INFORMATION WHEN CORRECTING CORRELATIONS FOR RESTRICTION IN RANGE DUE TO SELECTION

James O. Boone
Mary A. Lewis
FAA Civil Aeromedical Institute
P.O. Box 25082
Oklahoma City, Oklahoma 73125

SEPTEMBER 1978

Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

Prepared for
U.S. DEPARTMENT OF TRANSPORTATION
Federal Aviation Administration
Office of Aviation Medicine
Washington, D.C. 20591

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FAA-AM-78-36		

4. Title and Subtitle
THE DEVELOPMENT OF THE ATC SELECTION BATTERY: A NEW PROCEDURE TO MAKE MAXIMUM USE OF AVAILABLE INFORMATION WHEN CORRECTING CORRELATIONS FOR RESTRICTION IN RANGE DUE TO SELECTION

5. Report Date
SEP[redacted]78

6. Performing Organization Code

7. Author(s)
James O. Boone ~~and~~ Mary A. Lewis

8. Performing Organization Report No.
47p.

9. Performing Organization Name and Address
FAA Civil Aeromedical Institute
P.O. Box 25082
Oklahoma City, Oklahoma 73125

10. Work Unit No. (TRAIS)

11. Contract or Grant No.

12. Sponsoring Agency Name and Address
Office of Aviation Medicine
Federal Aviation Administration
800 Independence Avenue, S.W.
Washington, D.C. 20591

13. Type of Report and Period Covered

14. Sponsoring Agency Code

15. Supplementary Notes

Work was performed under Task AM-C-78-PSY-70.

16. Abstract

A five-test selection battery was ~~is currently~~ given to select Air Traffic Controllers. Data were ~~have been~~ collected on two new tests being considered for incorporation into the battery. To determine the utility of the old and new tests, it is necessary to correlate the tests with a criterion of job success. However, since criterion information is available only on persons already selected for air traffic control work, the correlation is restricted to this upper range of persons, and is, thereby, spuriously low for prediction purposes. To properly evaluate the utility of the tests, the correlation must be corrected for this restriction in range. This paper describes a new procedure to more accurately correct correlations for restriction in range. By Monte Carlo methods the new procedure is compared with Gulliksen and Thorndike's procedures and is shown to be more accurate.

17. Key Words
Selection
Correlation
Restriction in range

18. Distribution Statement
Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	44	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

THE DEVELOPMENT OF THE ATC SELECTION BATTERY: A NEW PROCEDURE TO MAKE MAXIMUM USE OF AVAILABLE INFORMATION WHEN CORRECTING CORRELATIONS FOR RESTRICTION IN RANGE DUE TO SELECTION

Introduction.

To develop or update a test battery used for selecting personnel, two very important steps must be completed. First, the most valid tests must be chosen, and second, a weighting system must be devised which will combine these tests into a composite that yields a maximum validity coefficient. In order to do this all tests under consideration are intercorrelated with each other and correlated with a specified criterion of job success. These correlations are used to regress the test scores on the job success criterion and the coefficients from the regression analysis are then used to determine which tests should be included in or deleted from the battery and what the relative weights should be for each test. These weighted test scores are then combined to form the composite score which is used for selection.

In the selection of air traffic controllers, a five-test selection battery is currently given to applicants, each test score is weighted, and the five weighted scores are combined to form a composite which is used to select candidates for Air Traffic Control (ATC) training. This test battery is in the process of being revised, and several new selection tests have been developed which could replace part or all of the existing five-test battery. To evaluate these new tests and compare them with the existing battery, they were administered to 7,000 ATC applicants along with the existing five-test battery. The applicant scores on the five existing tests and the new tests were then correlated to see how much overlap existed between them.

In order to determine the utility of the tests, both old and new, it was necessary to correlate them with some criterion measure of job success. Unfortunately, job success measures are available only for those individuals selected to be controllers, and this selection is based on scores only on the five current selection tests. An important factor influencing the size of correlation coefficients between a test and the criterion is the range of scores available on the tests and on the criterion. Since information about the job success criterion is available only for the ATC applicants who have been selected for employment, only the upper range of scores is available on the criterion. Because of this restriction in range, the correlations between the current selection test scores and the job success criterion will be spuriously low. This situation is illustrated in Figure 1.

The new tests being considered to replace part or all of the existing test battery will have a larger range and variance in the selected group than the five tests actually used for selection. In fact, the range and variance will be restricted only to the extent that the new tests correlate with the old tests, and will be as restricted as the old tests only if this correlation is 1.0. Because of this differential restriction in range, the new tests will correlate higher with the job success criterion in the selected group than will the old tests.

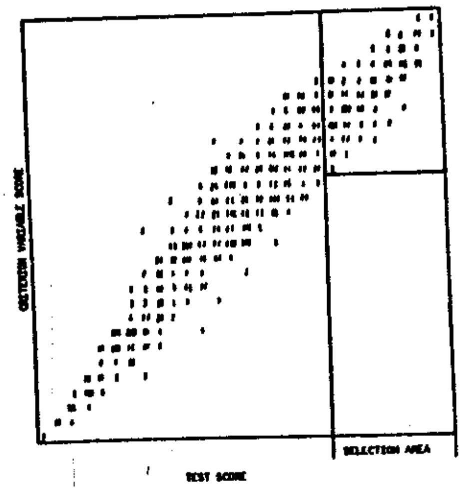


Figure 1. The effect of restricted range on a correlation coefficient. Subjects in the smaller box represent the selected group. The unrestricted correlation of the two variables is .88, and the restricted is .15.

To adjust for this spurious result, the correlations with the job success criterion must be corrected for restriction in range to assess the validity of the tests used for selection and to determine how the current tests used for selection compare with the new tests. The correction must take place prior to performance of regression analyses: otherwise, the new tests will appear superior to the current tests because of nothing more than a statistical artifact. This also means that, when corrected, the new test correlations with the criterion will generally increase less than the old test correlations.

Since a composite score is used for selection of ATC trainees, and the five tests in the existing battery are not given equal weights, some tests in the battery contribute more to the composite than others. Because of this differential contribution to the composite, some of the five tests which form the composite will be more restricted in range than others. Consequently, the correlations for some tests which form the composite, when corrected for restriction in range, will increase more than others, and the amount of increase will be proportional to the amount of restriction in the variance for each particular test.

Equal Employment Opportunity Commission (EEOC) Guidelines state that tests used for personnel selection must be demonstrated to be valid predictors of job success, and the magnitude of the validity coefficient must be both "practically and statistically significant" (3). The spuriously low correlation coefficient due to selection, then, becomes a very important legal issue in addition to its importance in assessing the value of new selection tests. Numerous litigations have occurred as a result of this problem, several of which related to the accuracy of the methods employed in correcting the validity coefficients for restriction in range (1).

There are two major statistical formulas which have been developed to correct the correlation of a test and a job success criterion. For the purposes of this study, the following notation will be used for all formulas:

x = the current selection composite score
y = the new test, or one of the five components of the current test battery
z = the job success criterion
RR = the unrestricted correlation of the variable subscripted
SS = the unrestricted standard deviation of the variable subscripted
R = the restricted correlation of the variable subscripted
S = the restricted standard deviation of the variable subscripted

Both major formulas estimate the value of RRyz based on the information available on the restricted group: Rxy, Rxz, Ryz, Sx, Sy, and Sz. They differ in their assumptions about information available on the unrestricted group.

The first formula (5), Thorndike's formula 7 case III (hereafter referred to as T7), assumes that only SSx is available for the unrestricted group and uses the ratio SSx/Sx and the restricted correlations to estimate RRxy, RRxz, SSy and SSz. These estimates in turn are used to estimate RRyz. The second major formula (4), Gulliksen's formula 37 (hereafter referred to as G37), assumes that only SSy is available on the unrestricted group and uses SSy-Sy and the restricted correlations and variances to estimate RRxy, RRxz, SSx, and SSz. These also are used to estimate RRyz, which is, of course, the desired unrestricted correlation of the test and the job success criterion.

The problem in using either of these formulas for the ATC selection situation is that both T7 and G37 require making estimates of either SSx or SSy and RRxy, when this unrestricted information is actually available from the applicant sample. The purpose of this study was to develop a procedure for correcting for restriction in range using available unrestricted values. In the two formulas already developed, estimates of SSz and RRxz only are required to estimate RRyz. In order to make maximum use of the unrestricted information, two formulas were derived by the first author of this paper. The first formula (hereafter referred to as B1) uses SSx to derive estimates of SSz and RRxz. The second formula (hereafter referred to as B2) uses SSy to derive estimates of these variables. In both formulas, the estimates, along with the actual unrestricted values of RRxy and either Sx or Sy, were used in conjunction with restricted correlations to estimate RRyz. The four formulas were compared both mathematically and by using Monte Carlo techniques to determine which can be most accurate in estimating RRyz across different selection ratios and different correlation values.

Methods.

Following Gulliksen's (4) schema for derivation of the correction formulas, three assumptions were employed, where upper case and lower case letters represent unrestricted and restricted variables respectively and x = the test used for selection, y = the new test being assessed and z = the success criterion.

Assumption 1. The slopes of the regressions of the new test and the criterion used for selection are not affected by selection.

$$Rxy \frac{Sy}{Sx} = RRxy \frac{SSy}{SSx} \qquad Rxz \frac{Sz}{Sx} = RRxz \frac{SSz}{SSx} \tag{1}$$

Assumption 2. The error made in estimating the new test scores and the criterion from the selection test scores is not affected by selection.

$$Sy^2(1 - Rxy^2) = SSy^2(1 - RRxy^2) \qquad Sz^2(1 - Rxz^2) = SSz^2(1 - RRxz^2) \tag{2}$$

Assumption 3. The partial correlation between the new test and the criterion is not affected by selection.

$$\frac{Ryz - RxyRxz}{\sqrt{(1 - Rxy^2)(1 - Rxz^2)}} = \frac{RRyz - RRxyRRxz}{\sqrt{(1 - RRxy^2)(1 - RRxz^2)}} \tag{3}$$

Based on assumptions 1 through 3, derivation of the root formulas proceed as follows.

Equation (1) is solved for RRxy,

$$RRxy = Rxy \frac{SySSx}{SSySx} \tag{4}$$

and RRxy is substituted in equation (2),

$$Sy^2(1 - Rxy^2) = SSy^2\left(1 - Rxy^2 \frac{Sy^2SSx^2}{SSy^2Sx^2}\right). \tag{5}$$

Multiplying the right side through by SSy^2,

$$Sy^2(1 - Rxy^2) = SSy^2 - Rxy^2Sy^2 \frac{SSx^2}{Sx^2} , \tag{6}$$

and solving for SSy^2,

$$SSy^2 = Sy^2\left[(1 - Rxy^2) + \left(Rxy^2 \frac{SSx^2}{Sx^2}\right)\right]. \tag{7}$$

Substituting SSy^2 in equation (4),

$$RRxy = \frac{Rxy \frac{SSx}{Sx}}{\sqrt{1 - Rxy^2 + Rxy^2\left(\frac{SSx}{Sx}\right)^2}} . \tag{8}$$

The same method can be used to derive SSz^2 and $RRxz^2$.

$$SSz^2 = Sz^2\left[1 - Rxz^2 + Rxz^2\left(\frac{SSx}{Sx}\right)^2\right] \tag{9}$$

$$RRxz = \frac{Rxz \frac{SSx}{Sx}}{\sqrt{1 - Rxz^2 + Rxz^2\left(\frac{SSx}{Sx}\right)^2}} \tag{10}$$

Solving for RRyz in equation (3), we algebraically change equation (2), dividing first by SSy^2 and taking the square root,

$$\sqrt{(1 - RRxy^2)} = \frac{Sy}{SSy}\sqrt{(1 - Rxy^2)} \tag{11}$$

and dividing by SSz^2 and taking the square root,

$$\sqrt{(1 - Rxz^2)} . \tag{12}$$

Substituting (11) and (12) in the denominator of (3)

$$\frac{Ryz - RxyRxz}{\sqrt{1 - Rxy^2}\sqrt{1 - Rxz^2}} = \frac{(RRyz - RRxyRRxz)SSySSz}{SySz\sqrt{1 - Rxy^2}\sqrt{1 - Rxz^2}} \tag{13}$$

and solving for RRyz

$$RRyz = \frac{(Ryz - RxyRxz)SySz}{SSySSz} + RRxyRRxz \quad (14)$$

The equations in assumption 1 can be algebraically combined, producing

$$RRxyRRxz = RxyRxz \frac{SySzSSx^2}{Sx^2 SSySSz} \quad (15)$$

Substituting (15) in (14) and factoring out SySz/SSySSz,

$$RRyz = \frac{SySz}{SSySSz}\left[Ryz - RxyRxz + RxyRxz \frac{SSx^2}{Sx^2} \right] \quad (16)$$

Formula (16) is the root formula for the development of the first two correction formulas, and formula (14) serves as the root formula for correction formulas (3) and (4). The first correction formula is derived on the basis that neither SSy nor SSz are available and SSy and SSz are estimated using the proportion SSx/Sx.

Substituting the estimates for SSy (7) and SSz (9) in the root formula (16) and simplifying gives:

$$RRyz = \frac{Ryz - RxyRxz + RxyRxz \frac{SSx^2}{Sx^2}}{\sqrt{\left(1 - Rxy^2 + Rxy^2 \frac{SSx^2}{Sx^2}\right)\left(1 - Rxz^2 + Rxz^2 \frac{SSx^2}{Sx^2}\right)}} \quad (17)$$

Formula (17) is equivalent to Thorndike's T7 (and also to Gulliksen's formula 19, ref. 4, p. 149).

The second correction formula uses the information (SSy - Sy), the restriction of the variance of test y, to estimate the restriction in SSx and SSz due to selection. Proceeding on this basis, Equation (2) is solved for $RRxy^2$, giving

$$RRxy^2 = 1 - \left(\frac{Sy^2}{SSy^2}\right)(1 - Rxy^2). \quad (18)$$

Equation (1) can be expressed as

$$\frac{SSx}{Sx} = \frac{RRxySSy}{RxySy}, \quad (19)$$

and substituting (18) in (19) and solving for SSx,

$$SSx = Sx \frac{\sqrt{SSy^2 - Sy^2(1 - Rxy^2)}}{SyRxy}. \qquad (20)$$

Next, solve (2) for RRxz, yielding

$$RRxz = Rxz \frac{SzSSx}{SxSSz}. \qquad (21)$$

Substituting (21) in (2), solving for SSz^2 and simplifying produces,

$$SSz^2 = Sz \left[\frac{Sy^2 Rxy^2 - Sy^2 Rxz^2 + SSy^2 Rxz^2}{Sy^2 Rxy^2} \right]. \qquad (22)$$

Returning to the root equation (16), substituting the estimates for SSx (20) and SSz (22) and simplifying produces the second correction formula.

$$RRyz = \frac{Rxz(SSy^2 - Sy^2) + RxyRyzSSy^2}{SSy\sqrt{Rxz^2(SSy^2 - Sy^2) + Sy^2 Rxy^2}} \qquad (23)$$

Formula (23) is Gulliksen's formula G37.

The third and fourth correction formulas employ the assumptions of the first and second correction formulas, respectively, and make the additional assumptions that the new test under consideration, test y, was administered to the applicant group. Consequently, there is no need to estimate RRxy, SSy or SSx, and formula (14) can be utilized as the root formula.

Substituting estimates for SSz (9) and RRxz (10) used in deriving the first correction formula (17) in the root formula (14) and simplifying gives the third correction formula,

$$RRyz = \left[\frac{Sy(Ryz - RxyRxz)}{SSy\sqrt{(1 - Rxz^2) + \left(Rxz^2 \frac{SSx^2}{Sx^2}\right)}} \right] + \left[\frac{Rxz \frac{SSx}{Sx}}{\sqrt{(1 - Rxz^2) + \left(Rxz^2 \frac{SSx^2}{Sx^2}\right)}} \right] RRxy. \qquad (24)$$

To obtain the fourth correction formula RRxz must be derived in terms of (SSy - Sy) by first solving equation (2) for $RRxz^2$,

$$RRxz^2 = 1 - \frac{Sz^2}{SSz^2}(1 - Rxz^2). \qquad (25)$$

Substituting (22) in (25), multiplying and simplifying yields,

$$RRxz = Rxz\sqrt{\frac{SSy^2 - Sy^2 + Sy^2 Rxy^2}{SSy^2 Rxz^2 - Sy^2 Rxz^2 + Sy^2 Rxy^2}}. \qquad (26)$$

To form the fourth correction formula, (22) and (26) are substituted in the root formula (14) and simplified giving,

$$RRyz = \frac{Sy(Ryz - RxyRxz)}{SSy\sqrt{\frac{Sy^2 Rxy^2 - Sy^2 Rxz^2 + SSy^2 Rxz^2}{Sy^2 Rxy^2}}} + RRxyRxz\sqrt{\frac{(SSy^2 - Sy^2) + Sy^2 Rxy^2}{SSy^2 Rxz^2 - Sy^2 Rxz^2 + Sy^2 Rxy^2}}. \qquad (27)$$

To evaluate the effects of the selection ratio, RRxy, and RRyz on the restricted Ryz mathematically, the process employed above to obtain unrestricted parameter estimates from restricted parameters was reversed to obtain explicit restricted parameter estimates in terms of unrestricted parameters. The Ryz's were then calculated as a function of the selection ratio, RRxy, and RRyz and compared to the RRyz to determine their respective effects on restriction in RRyz.

Since the derivation of formulas for the explicit restricted parameters follows a set pattern parallel to the steps in deriving the correction formulas, the pattern will be demonstrated and the remaining formulas will simply be given. This is done for the two cases employing the assumptions: (i) (SSx/Sx) is used to estimate the amount of restriction as in correction formulas T7 and B1 (hereafter referred to as assumption A-SSx); and (ii) (SSy - Sy) is used to estimate the amount of restriction as in correction formulas G37 and B2 (hereafter referred to as assumption A-SSy).

For A-SSx, (SSx/Sx),
Equation (1) is solved for Rxy,

$$Rxy = RRxy \frac{SSySx}{SySSx}. \qquad (28)$$

Rxy is substituted into equation (2), and multiplying through and solving for Sy^2,

$$Sy^2 = SSy^2\left[(1 - RRxy^2) + RRxy^2 \frac{Sx^2}{SSx^2}\right]. \qquad (29)$$

Substituting Sy^2 (29) in equation (28),

$$Rxy = \frac{RRxy \frac{Sx}{SSx}}{\sqrt{(1 - RRxy^2) + RRxy^2 \frac{Sx^2}{SSx^2}}}. \qquad (30)$$

The pattern that parallels the development of the correction formulas can be noted by comparing (28), (29), and (30) to (4), (7), and (8). The restricted correlations in (4), (7), and (8) become unrestricted correlations in (28), (29), and (30), and the ratio of (SSx/Sx) becomes (Sx/SSx). The same pattern exists in the remaining derivations for Sz and Rxz. Consequently, these explicit equations can be given as,

$$Sz^2 = SSz^2\left[(1 - RRxz^2) + RRxz\left(\frac{Sx^2}{SSx^2}\right)\right], \text{ and} \qquad (31)$$

$$Rxz = \frac{RRxz \frac{Sx}{SSx}}{\sqrt{(1 - RRxz^2) + RRxz^2\left(\frac{Sx^2}{SSx^2}\right)}}. \qquad (32)$$

To obtain Ryz, root formula (16) is solved for Ryz,

$$Ryz = \frac{SSySSz}{SySz}\left[RRyz - RRxyRRxz + RRxyRRxz \frac{Sx^2}{SSx^2}\right], \qquad (33)$$

and (29) and (31) are substituted into (33) and simplified to produce,

$$Ryz = \frac{RRyz - RRxyRRxz + RRxyRRxz \frac{Sx^2}{SSx^2}}{\sqrt{(1 - RRxy^2) + RRxy^2\left(\frac{Sx^2}{SSx^2}\right)}\sqrt{(1 - RRxz^2) + RRxz^2\left(\frac{Sx^2}{SSx^2}\right)}}. \qquad (34)$$

For A-SSy, (SSy - Sy), equation (2) is solved for Rxy^2, giving

$$Rxy^2 = 1 - \frac{SSy^2}{Sy^2}(1 - RRxy^2). \qquad (35)$$

The equations in assumption 1 can be expressed as

$$\frac{Sx}{SSx} = \frac{RxySy}{RRxySSy}, \tag{36}$$

and substituting (35) in (36) and solving for Sx,

$$Sx = SSx \frac{\sqrt{Sy^2 - SSy^2(1 - RRxy^2)}}{SSyRRxy}. \tag{37}$$

By comparing (35), (36), and (37) to (18), (19), and (20) the pattern emerges. The restricted correlations in (18), (19), and (20) become unrestricted in (35), (36), and (37) and (Sy/SSy) and (SSy - Sy) become (SSy/Sy) and (Sy - SSy). Applying this pattern, Sz, Rxz, and Rxy is given as,

$$Sz = SSz\sqrt{\frac{SSy^2 RRxy^2 - SSy^2 RRxz^2 + Sy^2 RRxz^2}{SSy^2 RRxy^2}}, \tag{38}$$

$$Rxz = RRxz\sqrt{\frac{Sy^2 - SSy^2 + SSy^2 RRxy^2}{Sy^2 RRxz^2 - SSy^2 RRxz^2 + SSy^2 RRxy^2}}, \text{ and} \tag{39}$$

$$Rxy = \sqrt{1 - \left(\frac{SSy^2}{Sy^2}\right)(1 - RRxy^2)}. \tag{40}$$

To obtain Ryz, root formula (14) is solved for Ryz, after substituting values from formula (1),

$$Ryz = \frac{SSySSz}{SySz}\left[(RRyz - RRxyRRxz) + RRxyRRxz \frac{Sx^2}{SSx^2}\right]. \tag{41}$$

Substituting (37) and (38) in (41) and simplifying produces,

$$Ryz = \frac{RRxz(Sy^2 - SSy^2) + RRxyRRyzSSy^2}{Sy\sqrt{RRxz^2(Sy^2 - SSy^2) + SSy^2 RRxy^2}}. \tag{42}$$

To examine the effects that selection ratio, RRxy, and RRyz had on the restricted Ryz, the ratio (Sx/SSx) was assigned values of .3, .5, and .8 and Ryz was computed while varying RRxy from .01 to 1.0 at .01 intervals for RRyz values of .2, .4, and .6. To insure that each (Sx/SSx) and (SSy - Sy) represented equal selection effects, the formulas for Rxy for the (Sx/SSx) and (SSy - Sy) cases were set equal,

$$\frac{RRxy \frac{Sx}{SSx}}{\sqrt{(1 - RRxy^2) + RRxy^2 \frac{Sx^2}{SSx}}} = \sqrt{1 - \frac{SSy^2}{Sy^2}(1 - RRxy^2)}$$

and the equation solved for Sy ,

$$Sy^2 = SSy^2(1 - RRxy^2) + RRxy^2 \frac{Sx^2}{SSx^2} . \qquad (44)$$

You may notice that formula (44) is the same as formula (29) even though they were derived from different root equations. SSy^2 will be arbitrarily set at a constant 20 and Sy^2 will be solved for Sx/SSx ratios of .2, .5, and .8.

A demonstration of the characteristics of the four correction formulas in terms of more refined influences was also performed by using Monte Carlo techniques. The Monte Carlo study examined the comparative accuracy of the four correction formulas as a function of (i) the selection ratio, (ii) RRxy, and (iii) RRyz.

In order to generate data of known means, standard deviations, and intercorrelations, a program (MNRNG) (2) (see Appendix A) was modified by the authors and used. The program uses the Marsaglia's reasonably fast method to generate normally distributed variables whose covariances are those required by a specified correlation matrix input into the program. Table 1 contains the relevant portion of the correlation matrix input into this program.

Table 1. Relevant Correlations Input Into MNRNG

	1	2	3	4	5	6	7	8	9	10	11
1		0.6	0.3	0.3	0.3	0.3	0.3	0.3	0.2	0.4	0.5
2			0.1	0.2	0.3	0.4	0.5	X	X	X	X
3				0.1	0.2	0.3	0.4	0.5	0.1	0.1	0.1
4					X	X	X	X	0.2	0.2	0.2
5						X	X	X	0.3	0.3	0.3
6							X	X	0.4	0.4	0.4
7								X	0.5	0.5	0.5
8									X	X	X
9										X	X
10											X

1 The correlations denoted by X were not used in the analysis.

For the purpose of this analysis variable 1 was defined as variable x, and variables 2,3,9,10, and 11 alternated as variable y and variables 3,4,5,6,7, and 8 were used for variable z. The unrestricted correlation of x and z (RRxz) was a constant 0.30, the unrestricted correlation of x and y (RRxy) ranged from 0.2 to 0.6 in increments of 0.1, and the unrestricted correlation of y and z (RRyz) ranged from 0.1 to 0.5 in increments of 0.1 also. All possible combinations of RRxy and RRyz were generated by using the various variables from the generated data as shown in Table 2.

Table 2. Variables Used as x,y, and z for Assigned Values of RRxy and RRyz

			Values of RRyz				
			0.1	0.2	0.3	0.4	0.5
			Var #	Var #	Var #	Var #	Var #
	0.2	x =	1	1	1	1	1
		y =	9	9	9	9	9
		z =	3	4	5	6	7
	0.3	x =	1	1	1	1	1
		y =	3	3	3	3	3
		z =	4	5	6	7	8
Values of RRxy	0.4	x =	1	1	1	1	1
		y =	10	10	10	10	10
		z =	3	4	5	6	7
	0.5	x =	1	1	1	1	1
		y =	11	11	11	11	11
		z =	3	4	5	6	7
	0.6	x =	1	1	1	1	1
		y =	2	2	2	2	2
		z =	3	4	5	6	7

1 Variable # used for x, y, or z.

After a sample of 1,000 subjects had been generated by using the correlation matrix specified in Table 1, the sample was sorted into descending order based on variable 1, the x variable. Using a program (REST) developed by Lewis and Boone (see Appendix A), the sample was then restricted on variable 1 using five different ratios, 10%, 20%, 30%, 40%, and 50%. For each selection ratio the four formulas for correction for restriction in range were used to estimate the value of RRyz. This was done for each selection ratio for all 25 combinations of RRxy and RRyz described in Table 2. The correlations computed from the restricted sample and the unrestricted sample were input into a subroutine (COREST) developed by Lewis and Boone (see Appendix B) which employs all four correction formulas and transformed the estimate of RRyz as well as the actual values of RRxy, RRxz, and RRyz by using the Fischer R to Z so that the values could later be averaged. This was repeated for 100 samples. A summary of the process is as follows in Table 3.

Table 3. Summary of Processes Used in Study

1. Generate 1,000 subjects with scores on 11 variables as defined by means, standard deviations, and correlations.
2. Sort sample into descending order based on scores on variable 1.
3. Restrict sample based on selection ratios of 10%, 20%, 30%, 40%, and 50%.
4. Calculate the four different estimates of RRyz for each restricted sample based on values of RRxy ranging from 0.2 to 0.6 and on values of RRyz ranging from 0.1 to 0.5.
5. Transform all correlations and estimated correlations by using Fischer R to Z transformation and for use in later averaging.

The results were then prepared in tabular and graphical form. Since the sample size was 100,000, significance tests were deemed inappropriate.

<u>Results</u>.

Figures 2 through 7 represent the calculated value of the restricted correlation, Ryz, when the unrestricted correlations, RRxz and RRyz, are equated and assigned values of .2, .4, or .6. For each figure the unrestricted correlation RRxy was allowed to vary from .01 to 1.00 by increments of .01. The ratio of the variances on the explicit selection variable was assigned values of .2, .5, and .8 and Ryz was plotted as a function of RRxy for each selection ratio. This was done for the variance assumptions of T7 and B1 (A-SSx) and also for the assumptions of G37 and B2 (A-SSy) for each assigned value of RRxz and RRyz.

The remaining figures and tables in the present study are based on the data obtained through the Monte Carlo technique described in Table 3. The actual correlation matrix obtained from the input of the matrix in Table 1 is contained in Table 4.

Table 4. Actual Correlation Matrix

	1	2	3	4	5	6	7	8	9	10	11
1	X	.60	.30	.30	.30	.30	.30	.30	.20	.41	.50
2		X	.10	.21	.31	.41	.50	X	X	X	X
3			X	.09	.20	.29	.40	.50	.10	.09	.11
4				X	X	X	X	X	.21	.21	.20
5					X	X	X	X	.29	.31	.31
6						X	X	X	.41	.40	.40
7							X	X	.49	.49	.50
8								X	X	X	X
9									X	X	X
10										X	X
11											X

In order to assess the accuracy of prediction of each correlation procedure, an error term was calculated based on the absolute value of the difference between the actual unrestricted correlation RRyz and the estimated correlation Ryz. Table 5 contains this error term, RRyz - Ryz, for each correction formula, for each selection ratio, for each value of RRxy, and for each value of RRyz. Figure 8 represents this error term as a function of selection ratio for the four correction formulas and for the actual restricted correlation Ryz. Figure 9 represents the error term as a function of RRxy for the four formulas and Ryz. Figure 10 represents the error term as a function of RRyz for the four formulas and Ryz.

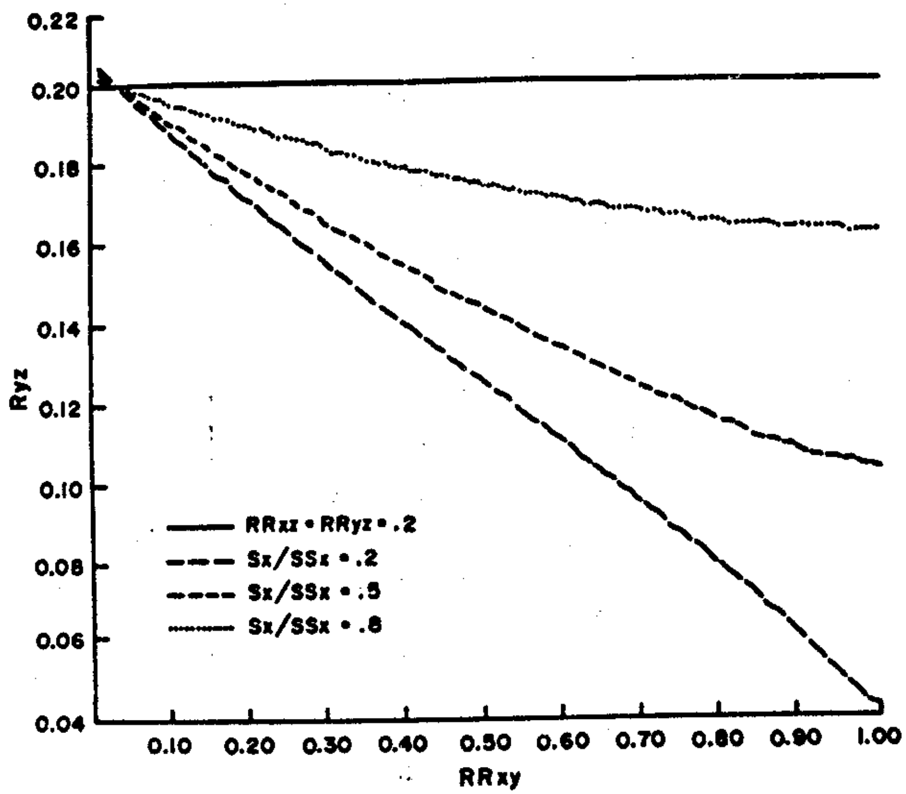


Figure 2. For the assumptions of formulas T7 and B1 when RRxz=RRyz=.2, the calculated Ryz across RRxy for values of (Sx/SSx).

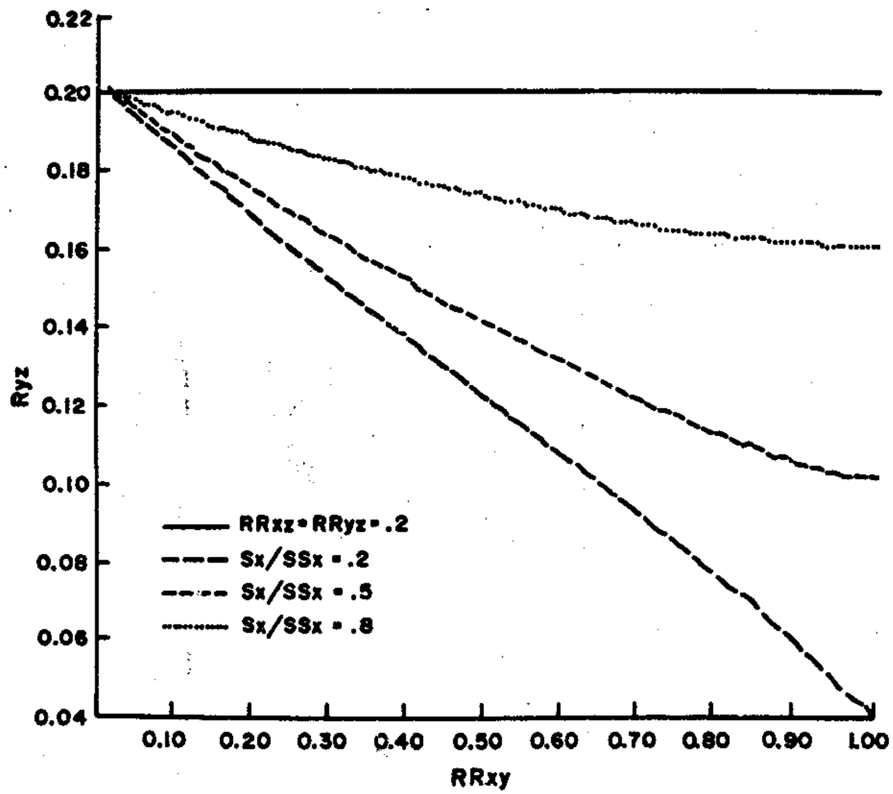


Figure 3. For the assumptions of formulas G37 and B2 when RRxz=RRyz=.2, the calculated Ryz across RRxy for values of (Sx/SSx).

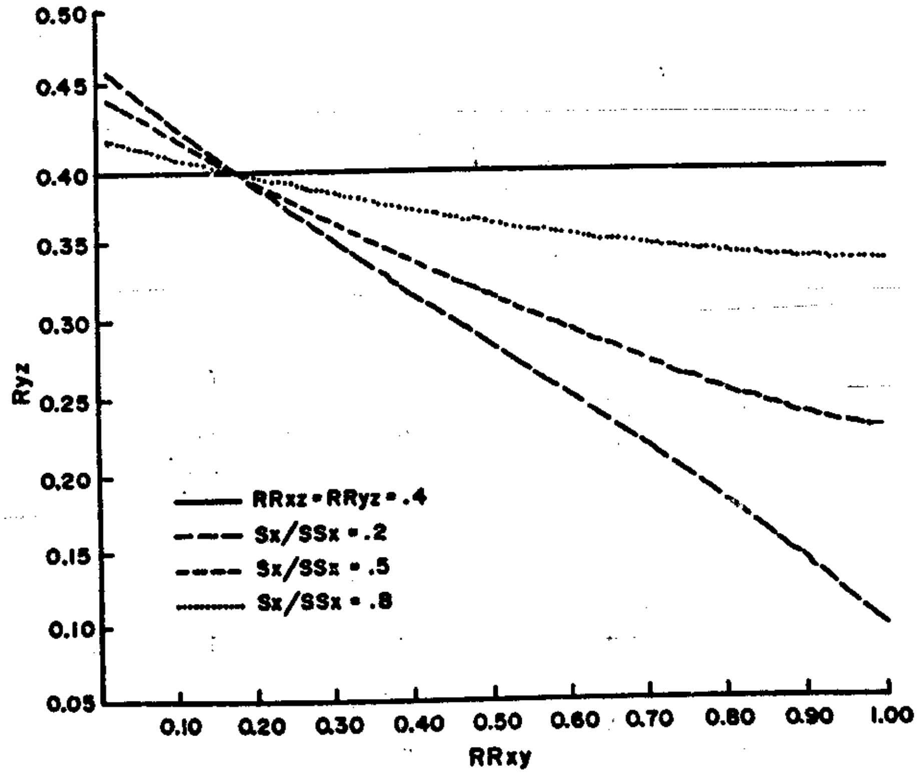


Figure 4. For the assumptions of formulas T7 and B1 when RRxz=RRyz=.4, the calculated Ryz across RRxy for values of (Sx/SSx).

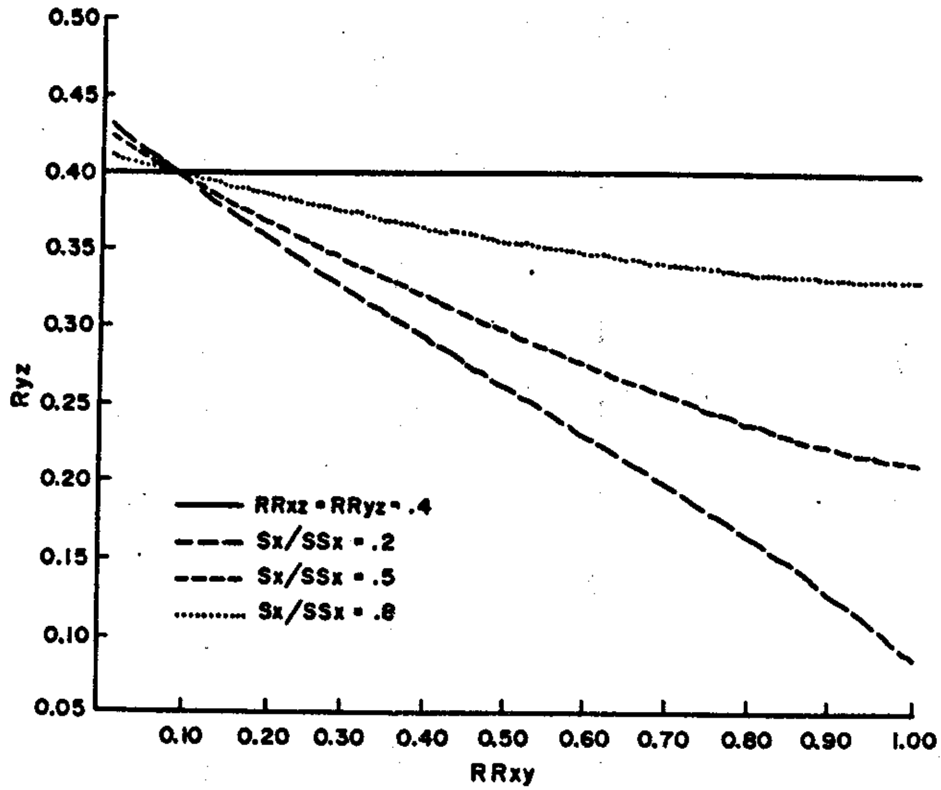


Figure 5. For the assumptions of formulas G37 and B2 when RRxz=RRyz=.4, the calculated Ryz across RRxy for values of (Sx/SSx).

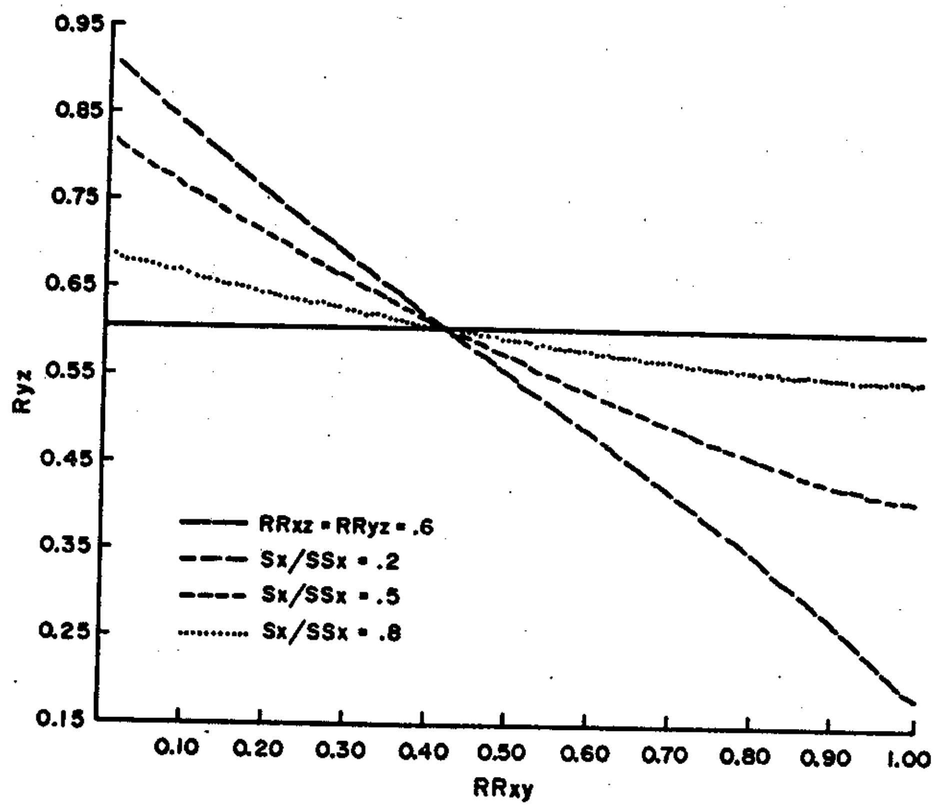


Figure 6. For the assumptions of formulas T7 and B1 when RRxz=RRyz=.6, the calculated Ryz across RRxy for values of (Sx/SSx).

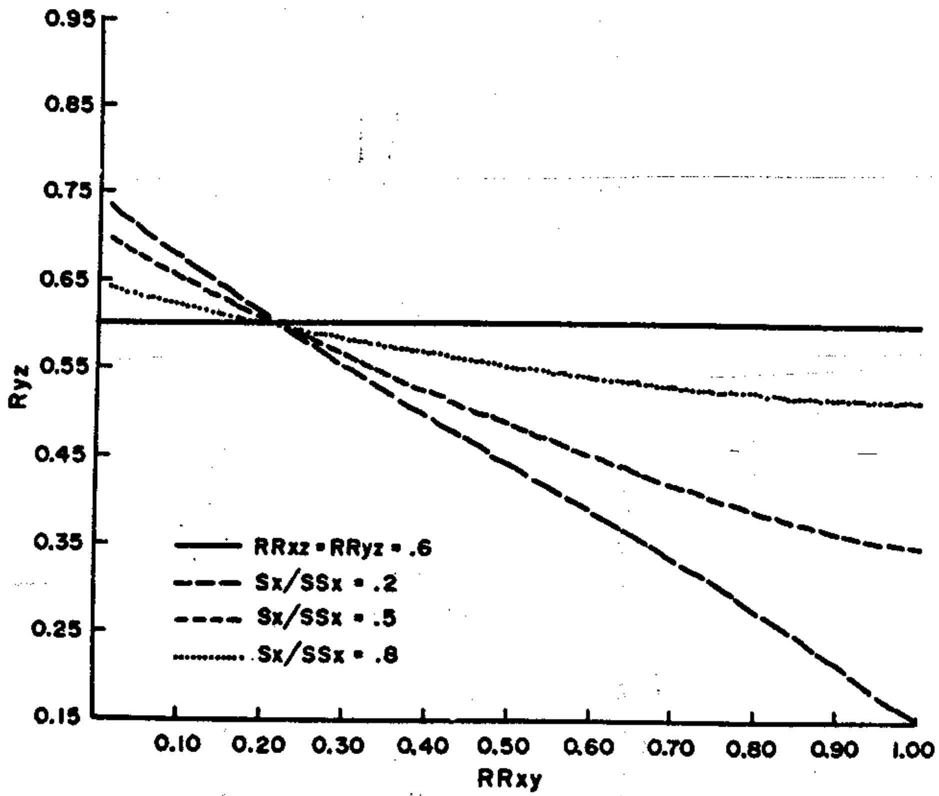


Figure 7. For the assumptions of formulas G37 and B2 when RRxz=RRyz=.6, the calculated Ryz across RRxy for values of (Sx/SSx).

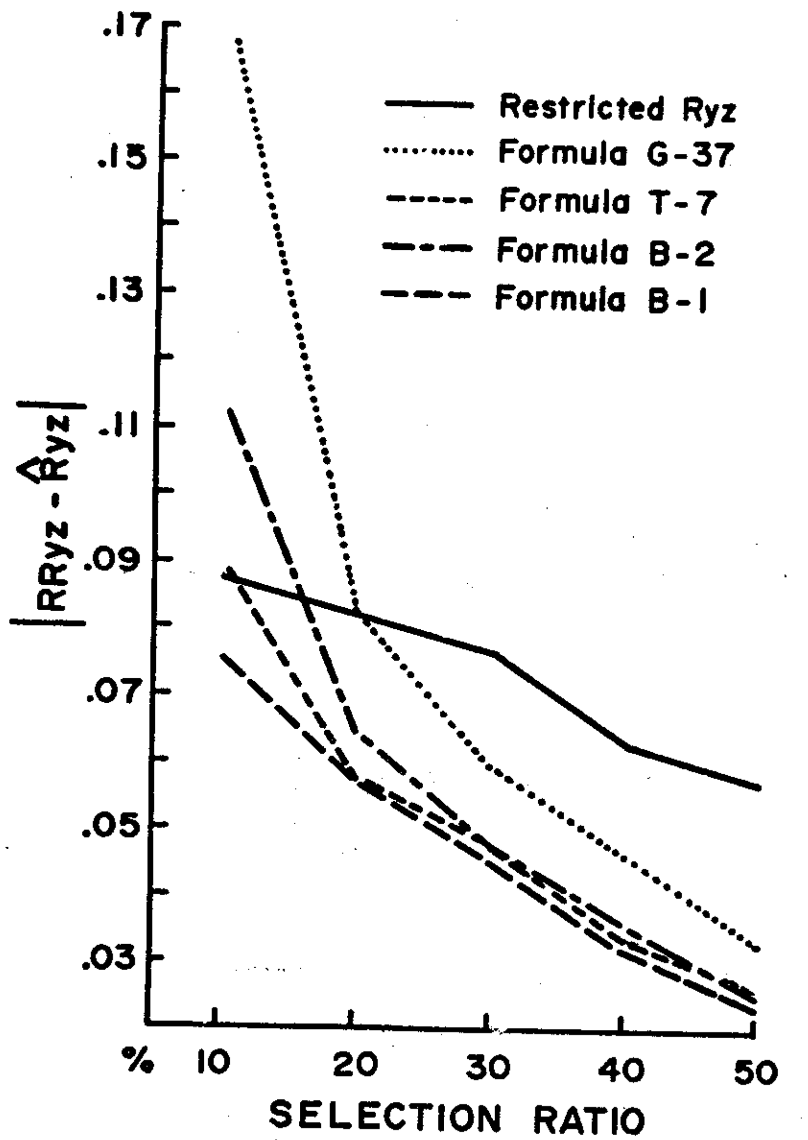


Figure 8. Error by selection ratio for the four correction formulas and the actual restricted value of Ryz.

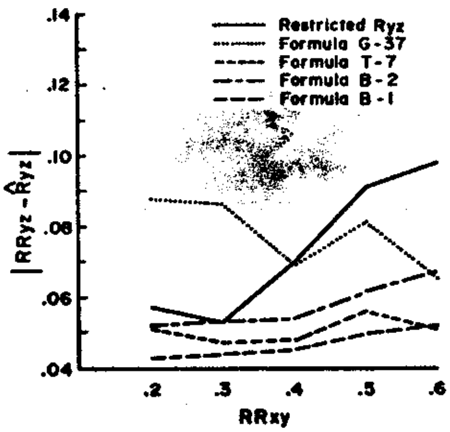


Figure 9. Error by values of RRxy for the four correction formulas and the actual restricted value of Ryz.

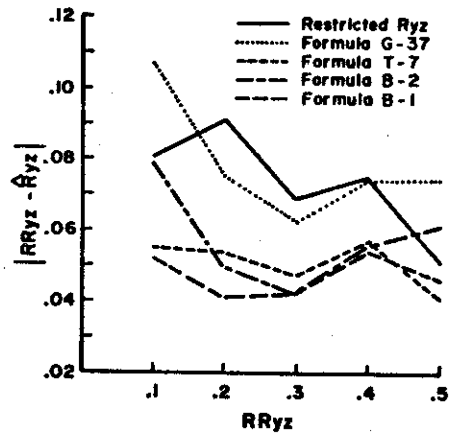


Figure 10. Error values of RRyz for the four correction formulas and the actual restricted value of Ryz.

Table 5. Average Error in Estimation of RRyz

Error by Formula

	B1	T7	G37	B2
Means =	0.047	0.051	0.078	0.058
Stds =	0.04	0.05	0.11	0.07

Error by Selection Ratio

	10%	20%	30%	40%	50%
Means =	0.112	0.065	0.050	0.037	0.028
Stds =	0.26	0.14	0.11	0.08	0.06

Error by RRxy

	.60	.20	.30	.40	.50
Means =	0.059	0.059	0.058	0.054	0.062
Stds =	0.16	0.14	0.14	0.14	0.18

Error by RRyz

	.10	.20	.30	.40	.50
Means =	0.073	0.055	0.048	0.060	0.056
Stds =	0.17	0.13	0.11	0.15	0.18

Table 6 contains the average Ryz for formula B1 for each value of RRyz by values of RRxy by selection ratio. To average the correlations, they were transformed using the Fischer R to Z transformation, averaged, and then transformed back to a correlation. Table 7 contains the same information for formula T7, Table 8 contains the information for formula G37, and Table 9 contains the information for formula B2.

Table 6

Average Ryz for Formula B1
by RRxy and Selection Ratio for Each RRyz

RRyz = 0.1

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.082	0.121	0.097	0.162	0.089
20% Selection	0.115	0.133	0.059	0.097	0.057
30% Selection	0.103	0.119	0.058	0.119	0.098
40% Selection	0.111	0.110	0.103	0.139	0.145
50% Selection	0.092	0.094	0.124	0.137	0.143

RRyz = 0.2

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.207	0.180	0.244	0.215	0.263
20% Selection	0.184	0.144	0.232	0.212	0.223
30% Selection	0.173	0.157	0.212	0.197	0.204
40% Selection	0.185	0.180	0.203	0.213	0.224
50% Selection	0.194	0.189	0.183	0.191	0.207

RRyz = 0.3

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.272	0.249	0.302	0.277	0.275
20% Selection	0.251	0.247	0.290	0.267	0.291
30% Selection	0.258	0.292	0.284	0.269	0.287
40% Selection	0.279	0.310	0.306	0.290	0.301
50% Selection	0.281	0.307	0.306	0.296	0.312

RRyz = 0.4

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.302	0.296	0.388	0.355	0.388
20% Selection	0.357	0.343	0.383	0.338	0.359
30% Selection	0.368	0.380	0.396	0.344	0.385
40% Selection	0.375	0.388	0.405	0.358	0.404
50% Selection	0.396	0.391	0.404	0.376	0.409

RRyz = 0.5

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.453	0.463	0.385	0.423	0.389
20% Selection	0.453	0.491	0.451	0.473	0.465
30% Selection	0.466	0.509	0.453	0.469	0.481
40% Selection	0.475	0.511	0.470	0.482	0.506
50% Selection	0.477	0.500	0.482	0.491	0.504

Table 7

Average Ryz for Formula T7
by RRxy and Selection Ratio for Each RRyz

RRyz = 0.1

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.071	0.107	0.123	0.122	0.039
20% Selection	0.096	0.107	0.044	0.094	0.039
30% Selection	0.100	0.106	0.051	0.122	0.093
40% Selection	0.104	0.120	0.099	0.148	0.142
50% Selection	0.087	0.105	0.115	0.137	0.145

RRyz = 0.2

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.223	0.190	0.264	0.146	0.211
20% Selection	0.169	0.131	0.234	0.205	0.217
30% Selection	0.178	0.159	0.213	0.195	0.197
40% Selection	0.185	0.192	0.207	0.219	0.225
50% Selection	0.191	0.201	0.183	0.191	0.210

RRyz = 0.3

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.254	0.256	0.346	0.285	0.273
20% Selection	0.243	0.242	0.309	0.269	0.294
30% Selection	0.267	0.292	0.291	0.276	0.287
40% Selection	0.278	0.321	0.313	0.299	0.303
50% Selection	0.279	0.319	0.309	0.297	0.315

RRyz = 0.4

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.330	0.336	0.424	0.352	0.392
20% Selection	0.348	0.340	0.401	0.340	0.364
30% Selection	0.376	0.377	0.406	0.351	0.385
40% Selection	0.378	0.397	0.414	0.367	0.407
50% Selection	0.396	0.403	0.408	0.377	0.411

RRyz = 0.5

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.484	0.517	0.458	0.500	0.441
20% Selection	0.453	0.496	0.470	0.474	0.476
30% Selection	0.482	0.512	0.467	0.480	0.484
40% Selection	0.479	0.524	0.483	0.492	0.511
50% Selection	0.479	0.511	0.489	0.492	0.507

Table 8

Average Ryz for Formula G37 by RRxy and Selection Ratio for Each RRyz

RRyz = 0.1

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.235	0.294	0.154	0.260	0.198
20% Selection	0.235	0.264	0.135	0.106	0.085
30% Selection	0.218	0.183	0.105	0.141	0.104
40% Selection	0.204	0.130	0.145	0.160	0.158
50% Selection	0.137	0.106	0.164	0.141	0.146

RRyz = 0.2

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.257	0.223	0.346	0.343	0.334
20% Selection	0.265	0.245	0.309	0.211	0.237
30% Selection	0.268	0.216	0.257	0.222	0.219
40% Selection	0.261	0.192	0.243	0.233	0.233
50% Selection	0.231	0.198	0.221	0.195	0.211

RRyz = 0.3

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.248	0.247	0.355	0.247	0.326
20% Selection	0.323	0.332	0.331	0.267	0.295
30% Selection	0.351	0.341	0.317	0.286	0.294
40% Selection	0.352	0.327	0.338	0.309	0.310
50% Selection	0.315	0.315	0.339	0.301	0.315

RRyz = 0.4

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.228	0.261	0.440	0.328	0.409
20% Selection	0.375	0.429	0.416	0.334	0.362
30% Selection	0.443	0.412	0.431	0.362	0.394
40% Selection	0.436	0.402	0.438	0.375	0.412
50% Selection	0.424	0.400	0.437	0.379	0.413

RRyz = 0.5

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.359	0.445	0.370	0.333	0.342
20% Selection	0.462	0.559	0.489	0.469	0.472
30% Selection	0.528	0.529	0.486	0.485	0.491
40% Selection	0.519	0.525	0.504	0.501	0.514
50% Selection	0.500	0.507	0.512	0.495	0.507

Table 9

Average Ryz for Formula B2
by RRxy and Selection Ratio for Each RRyz

RRyz = 0.1

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.137	0.175	0.116	0.221	0.216
20% Selection	0.152	0.199	0.106	0.104	0.085
30% Selection	0.134	0.149	0.086	0.129	0.105
40% Selection	0.140	0.114	0.124	0.145	0.154
50% Selection	0.110	0.093	0.149	0.139	0.144

RRyz = 0.2

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.239	0.205	0.290	0.298	0.361
20% Selection	0.193	0.198	0.273	0.216	0.237
30% Selection	0.199	0.186	0.235	0.211	0.216
40% Selection	0.208	0.179	0.221	0.220	0.229
50% Selection	0.207	0.186	0.204	0.193	0.207

RRyz = 0.3

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.265	0.245	0.314	0.316	0.330
20% Selection	0.258	0.288	0.309	0.267	0.295
30% Selection	0.279	0.311	0.300	0.277	0.293
40% Selection	0.292	0.311	0.320	0.296	0.306
50% Selection	0.290	0.303	0.324	0.298	0.312

RRyz = 0.4

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.305	0.279	0.399	0.381	0.407
20% Selection	0.307	0.375	0.399	0.336	0.362
30% Selection	0.369	0.382	0.412	0.353	0.392
40% Selection	0.375	0.387	0.419	0.363	0.408
50% Selection	0.399	0.388	0.420	0.378	0.409

RRyz = 0.5

RRxy =	0.2	0.3	0.4	0.5	0.6
10% Selection	0.412	0.439	0.332	0.395	0.295
20% Selection	0.386	0.502	0.468	0.472	0.469
30% Selection	0.450	0.501	0.467	0.476	0.488
40% Selection	0.462	0.510	0.483	0.487	0.509
50% Selection	0.477	0.497	0.497	0.493	0.504

Figures 11 through 30 graphically represent the estimated RRyz (Ryz) as a function of assigned values of RRxy. Each assigned value of RRyz is graphed separately for each correction formula, and each graph contains a line representing each of the five selection ratios. The actual value of RRyz is represented as a straight line. Figures 11 through 15 represent Ryz as a function of RRxy by selection ratio for formula B1. Figures 16 through 20 represent Ryz as a function of RRxy by selection ratio for formula T7. Figures 21 through 25 represent Ryz as a function of RRxy by selection ratio for formula B2, and Figures 26 through 30 represent Ryz as a function of RRxy by selection ratio for formula G37.

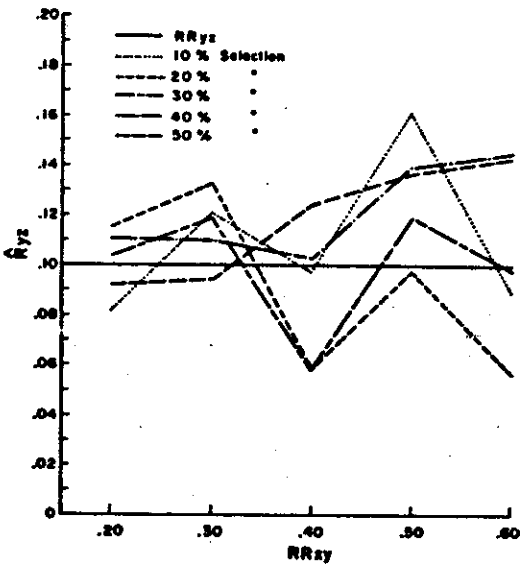


Figure 11. The estimated unrestricted correlations for formula B1, RRyz = .1.

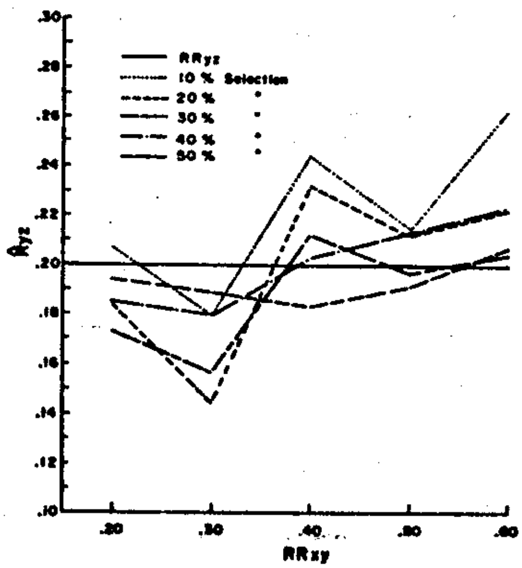


Figure 12. The estimated unrestricted correlations for formula B1, RRyz = .2.

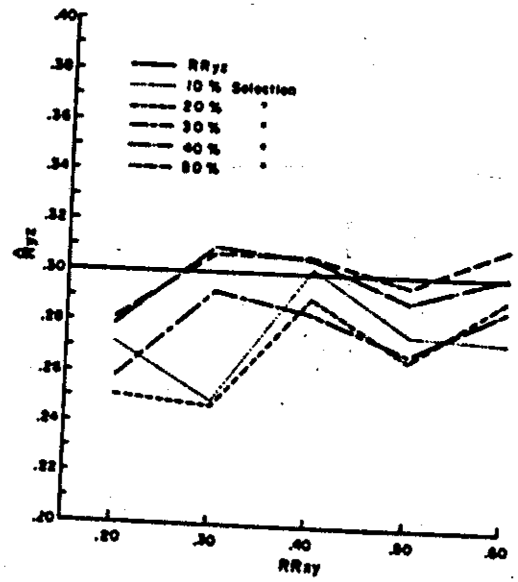


Figure 13. The estimated unrestricted correlations for formula B1, RRyz = .3.

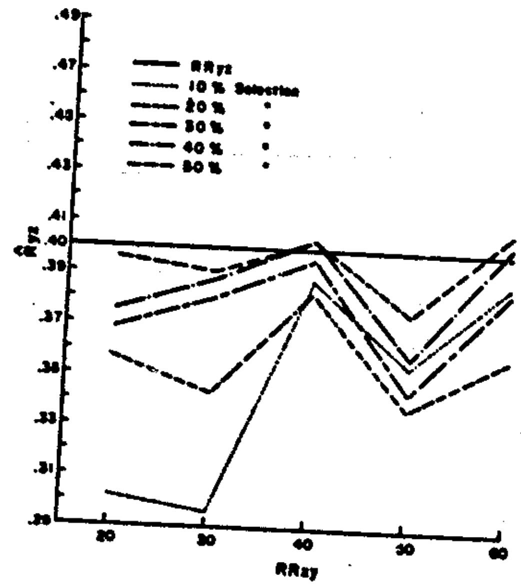


Figure 14. The estimated unrestricted correlations for formula B1, RRyz = .4.

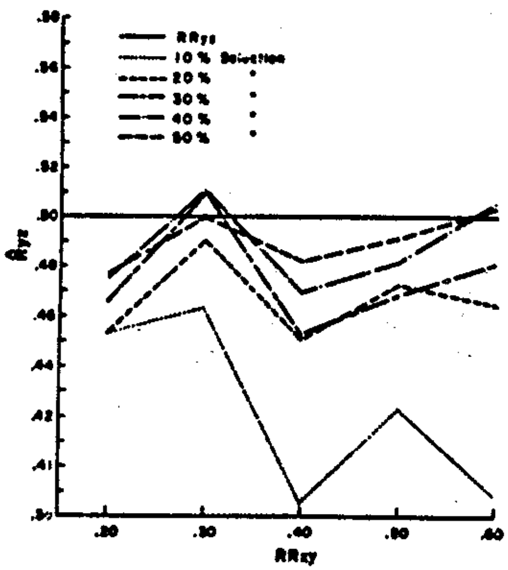


Figure 15. The estimated unrestricted correlations for formula B1, RRyz = .5.

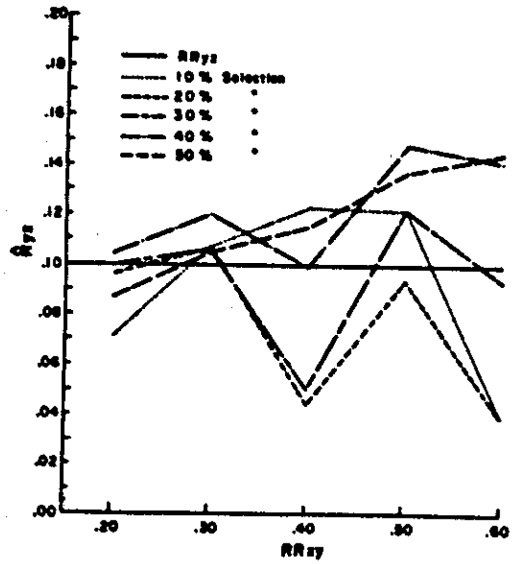


Figure 16. The estimated unrestricted correlations for formula T7, RRyz = .1.

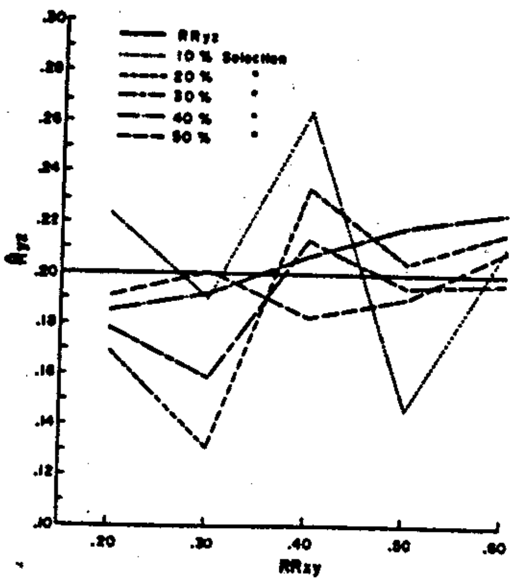


Figure 17. The estimated unrestricted correlations for formula T7, RRyz = .2.

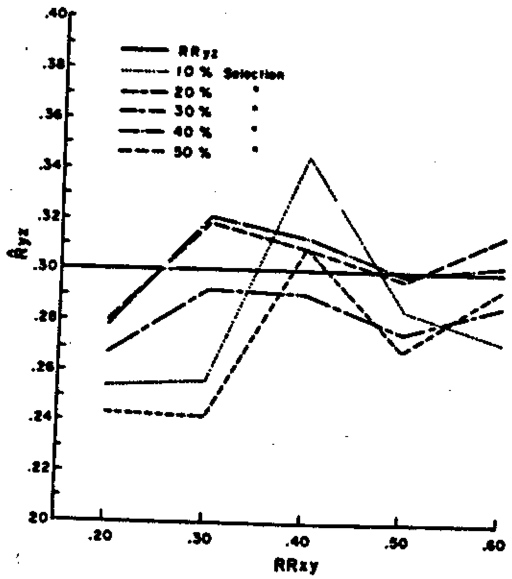


Figure 18. The estimated unrestricted correlations for formula T7, RRyz = .3.

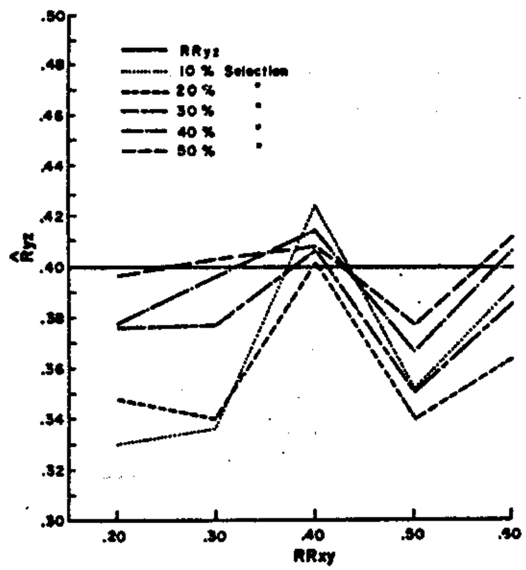


Figure 19. The estimated unrestricted correlations for formula T7, RRyz = .4.

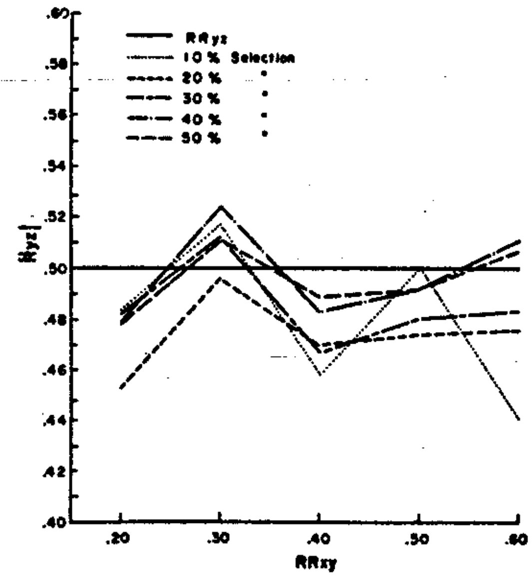


Figure 20. The estimated unrestricted correlations for formula T7, RRyz = .5.

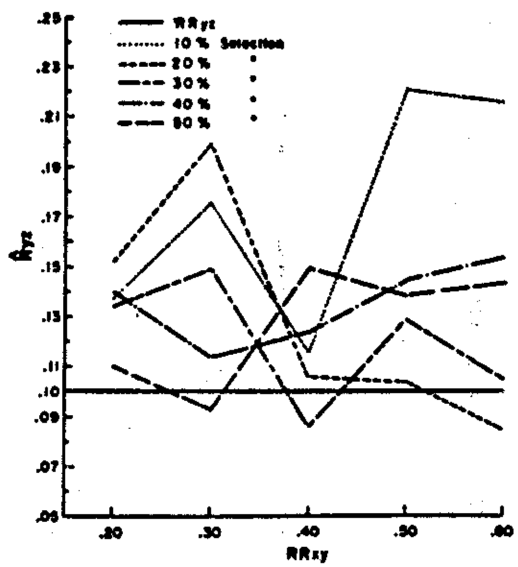


Figure 21. The estimated unrestricted correlations for formula B2, RRyz = .1.

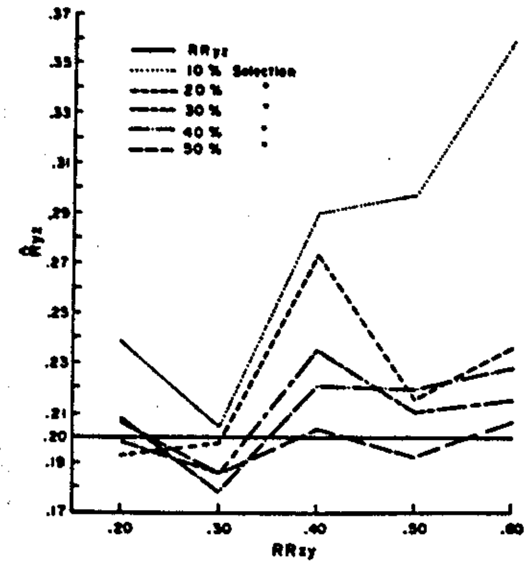


Figure 22. The estimated unrestricted correlations for formula B2, RRyz = .2.

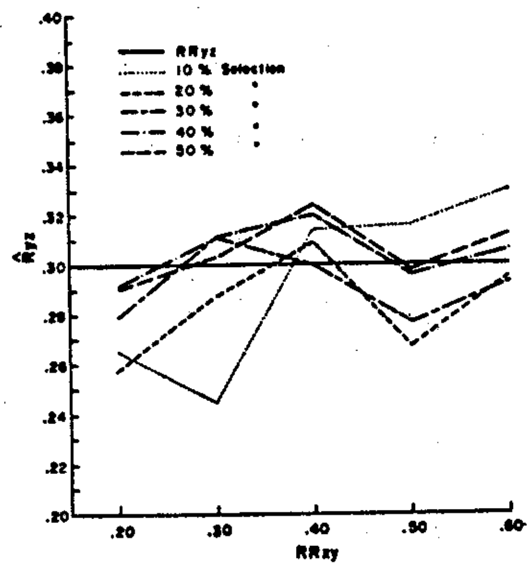


Figure 23. The estimated unrestricted correlations for formula B2, RRyz = .3.

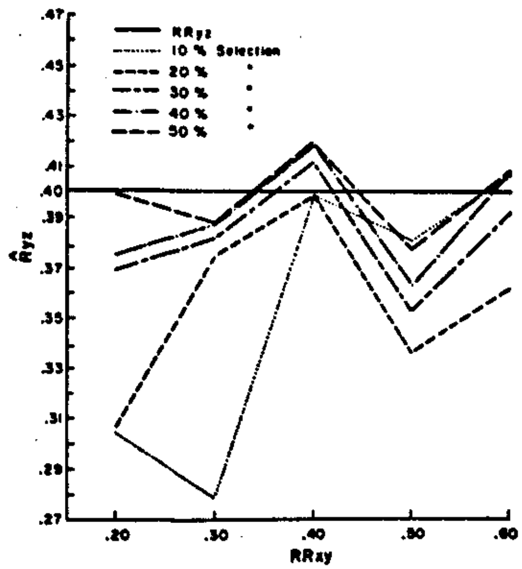


Figure 24. The estimated unrestricted correlations for formula B2, RRyz = .4.

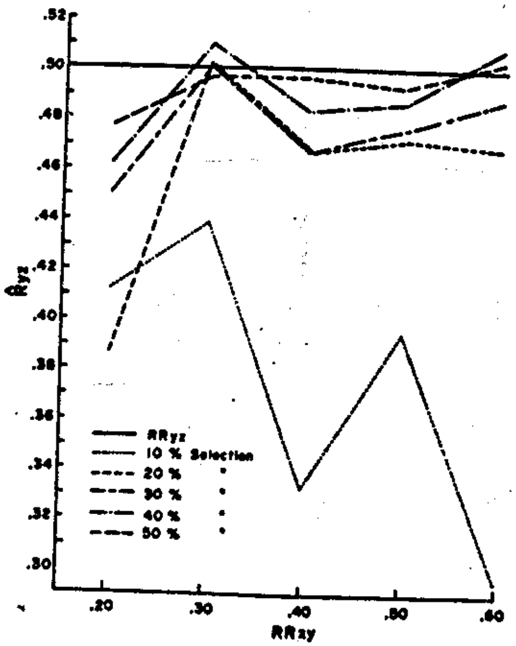


Figure 25. The estimated unrestricted correlations for formula b2, RRyz = .5.

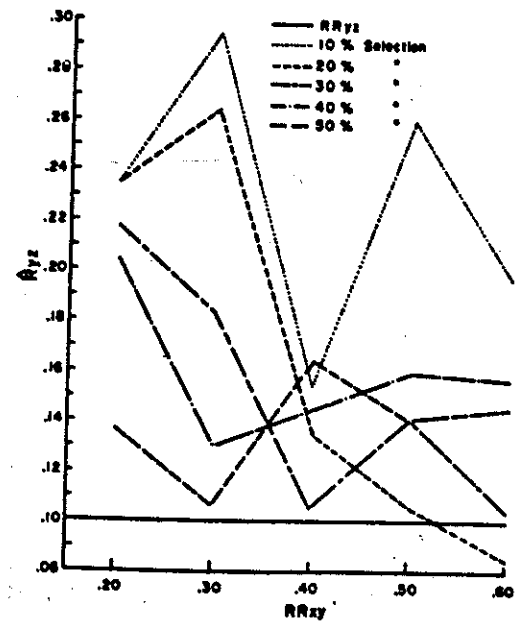


Figure 26. The estimated unrestricted correlations for formula G37, RRyz = .1.

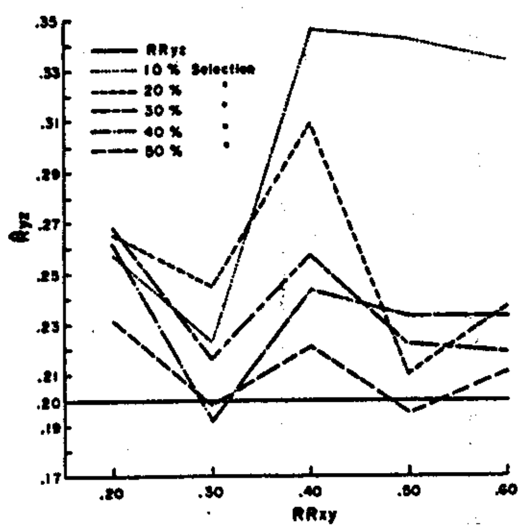


Figure 27. The estimated unrestricted correlations for formula G37, RRyz = .2.

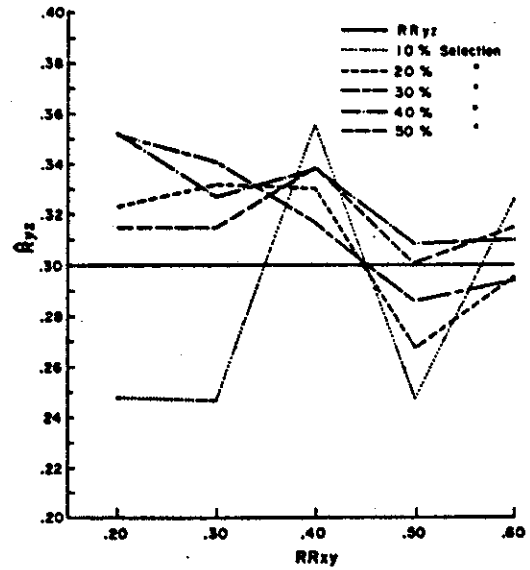


Figure 28. The estimated unrestricted correlations for formula G37, RRyz = .3.

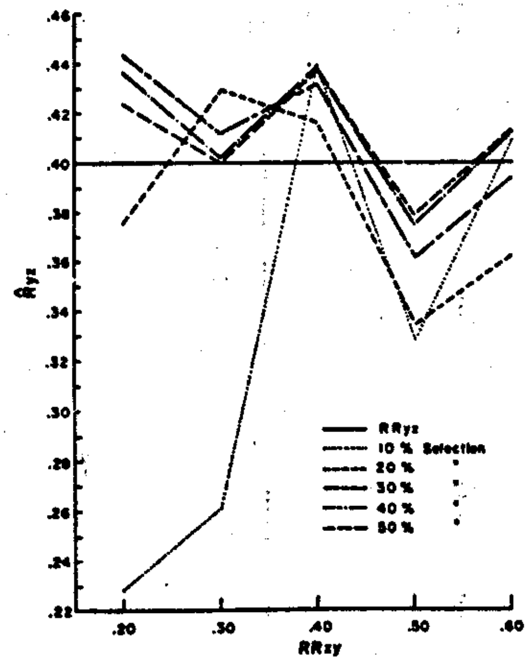


Figure 29. The estimated unrestricted correlations for formula G37, RRyz = .4.

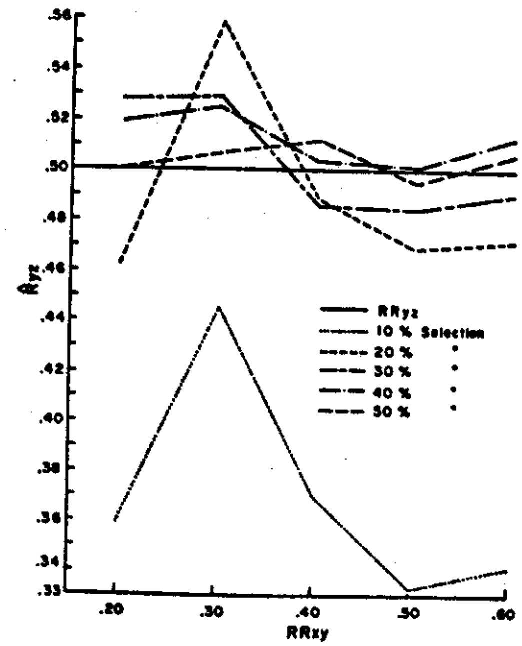


Figure 30. The estimated unrestricted correlations for formula G37, RRyz = .5.

Discussion.

In any Monte Carlo study a decision must be made concerning which components are to be varied and what the range of their variation will be. The components selected for variation and their range in this study were established subjectively based on values the authors considered representative of practical situations. Consequently, the discussion of the results is more a comparison of the practical utility of each formula rather than a strict mathematical comparison.

Restricted correlations. Figures 2 through 7 present the estimated restricted correlations by selection ratios (Sx/SSx) of .2, .5, and .8 for RRyz values of .2, .4, and .6 across values of RRxy ranging from .01 to 1.00 in .01 increments. As presented in the figures, the estimated Ryz's converge with the actual RRyz as a function of RRxy without regard to selection ratio. As the values of RRxy increase beyond the convergent point, Ryz becomes more an underestimate of RRyz, and as RRxy is decreased below the convergent point, Ryz becomes more and more an overestimate of RRyz. Further, the degree of error in Ryz sharply increases as the selection ratio decreases. It is also apparent from the figures that as RRxz and RRyz increase, the point of convergence for Ryz and RRyz on RRxy also increases, indicating the point of convergence is related to the various intercorrelations of RRxz, RRyz, and RRxy, but not to the selection ratio. The only difference between the estimates of Ryz when using the two different sets of assumptions to derive the formulas (A-SSy and A-SSx) is the point of convergence. The A-SSy assumptions result in convergence lower on RRxy than the A-SSx assumptions.

The general practical conclusion to be drawn from this portion of the study is that unless the situation being studied contains the exact and particular interrelationship of RRxz, RRyz, and RRxy necessary for Ryz to converge with RRyz, the restricted Ryz will consistently be either an underestimate or overestimate of the unrestricted RRyz, with the amount of error increasing sharply as the selection ratio becomes more and more extreme. Consequently, correcting the restricted correlations is almost always warranted.

Main effects. Table 5 demonstrates the overall accuracy of each of the four formulas in terms of the average amount of error each incurred in estimating RRyz. Their rank order from least to most error is: B1, T7, B2, and G37. The first three formulas are not remarkably different; however, G37 is far less accurate than B1, T7, and B2. The clearest effect on error is produced by the selection ratio (Table 5). As the selection ratio becomes more extreme, the amount of error increases, with the increase becoming larger and larger with each step down in the selection ratio. Table 5 shows little fluctuation in error for RRxy and no systematic pattern. The effects of RRyz in Table 5 show a pattern that was found consistently throughout the analyses. When RRyz = RRxz, the error component is at a minimum. RRxz was held at a constant .30 for this study and, as can be noted in Table 5, the error increases as RRyz moves in either direction from .30.

Practical conclusions related to main effects include the following. If sufficient information is available, the B1 formula produces the most accurate estimate for RRyz. In order to have sufficient information to use B1, the new test being evaluated would need to be administered to the applicant group at the same time the old selection test is administered. Then RRxy and SSy are available for use in B1. If the new test being evaluated was not administered to the applicant group, then the most accurate correction formula would be T7 which does not require RRxy and SSy.

The selection ratio, it appears, has the largest impact on errors in estimating RRyz. If selection is extreme, 10 percent or less, the formulas for estimating RRyz are unstable and highly inaccurate. This is a difficult practical situation to resolve. A general advertisement for applicants without sufficient specific qualification statements results in a larger number of unqualified candidates and more extreme selection. However, with a highly specific advertisement self-selection becomes a secondary selection process, and the statistics computed on the applicant group are already restricted producing spuriously low validity correlations. One strategy would be to administer the selection tests to a random sample in the general population, stratifying by race and sex in order to meet Equal Employment Opportunity Commission requirements. This would yield unrestricted variances without the influence of any selection procedure.

Since RRyz is not known and RRxy is computed after the test administration, little practical guidance can be offered related to these parameters. The usual advice is clearly applicable, viz, choose a test or construct a test for selection that parallels the actual job tasks as closely as possible.

<u>Interaction effects</u>. As seen in Figure 8, when error in prediction is examined by selection ratio for each formula and for the actual restricted correlation of Ryz, there is a tremendous amount of error for the 10-percent selection ratio, with formula B1 doing a much better job than either T7, B2, or G37 in estimating RRyz. As the selection ratio increases beyond moderate selection (30 percent), the formulas tend to perform similarly in estimating RRyz, with the exception of G37 which consistently has more error than the other three formulas across all selection ratios.

Figure 9 demonstrates that formula B1 again is consistently the better estimator of RRyz across values of RRxy. It can also be noted from Figure 9 that as the value of RRxy increases, Ryz rapidly becomes a poorer estimator of RRyz, particularly after it passes the point at which RRyz equals RRxz (.30). Once again, G37 is a much less accurate estimator of RRyz than the other three formulas.

When RRyz is less than .30, as shown in Figure 10, B1 is the better estimator of RRyz. All formulas converge when RRyz equals RRxz (.30) and T7 is the best estimator for higher values of RRyz although the differences are small. Once again formula G37 is clearly the least accurate estimator of RRyz.

In terms of the selection ratio by RRxy by RRyz interaction, Figures 11 through 30 revealed the following. Error in this case is mainly influenced by the selection ratio with some minor influence on error added by RRxy and RRyz. With a low selection ratio and low RRxy and RRyz, the error component is relatively large. When the selection ratio is low and RRxy and RRyz are high, the errors are moderate to large. With a high selection ratio and low RRxy and RRyz, the errors are moderate to small. When the selection ratio and RRxy and RRyz are all high, the comparative error is minimal.

The four-way interactions given in Tables 6 through 9 and Figures 11 through 30 for selection ratio/RRxy/RRyz/formula show that when the selection ratio is small to moderate (10 to 30 percent) across all values of RRxy and Ryz, formula B1 results in the least amount of error. When there is a high selection ratio and high RRxy and RRyz, all the formulas tend to be about the same in accuracy.

The practical implications for the interaction effects can be stated briefly. The selection ratio has such an overwhelming effect that generally the interaction effects are primarily due to the selection ratio. When the selection ratio is small to moderate (10 to 30 percent), formula B1 is clearly the most accurate estimator and should be used regardless of RRxy and RRyz. When the selection ratio goes above 30 percent, B1, T7, and B2 are practically equivalent. Formula G37 is the least desirable correction formula across conditions. Thus, overall, B1 results in the most accurate estimates of RRyz, especially when the selection ratio is 30 percent or less, regardless of the values of RRxy or RRyz.

ACKNOWLEDGMENT

We gratefully acknowledge the assistance of Barbara Rizzuti, Carol Young, and Jo Ann Steen during the formulation and preparation of this manuscript.

References

1. Boldt, R. F.: Robustness of Range Restriction in Court. Paper presented at the annual meeting of the American Psychological Association: Toronto, 1978.

2. Boone, J. O.: Documentation for Fortran IV Programming That Produces Random Observations From a Multivariately, Normally Distributed Population Where the Means, Standard Deviations and Intercorrelations Among Variables Can Be Specified. Journal of Florida Education Association, May 1977.

3. Equal Employment Opportunity Commission. Uniform Guidelines on Employee Selection Procedures. Federal Register, Dec. 30, 1977, 42, 251, 65517.

4. Gulliksen, H.: Theory of Mental Tests. Wiley and Sons: New York, 1950.

5. Thorndike, R. L.: Personnel Selection. Wiley and Sons: New York, 1949.

Appendix A

```
      PROGRAM MNRNG
      DIMENSION R(11,11),XMM(11),STD(11),CP(30,30),C(30,30),X(11),
     + Z(11),TV(30,30),TDS(30),XBAR(11),XM(11),SDEV(11),STP(11),Y(11),
     + XM2(11),VAR(30,30),RC(30,30),VAT(11)
      BYTE XMAT(72),YMAT(72)
      OPEN(UNIT=1,NAME='MONTY.DAT',TYPE='NEW')
      OPEN(UNIT=2,NAME='COR.DAT',TYPE='OLD',READONLY)
      OPEN(UNIT=3,NAME='CORR.DAT')
      OPEN(UNIT=4,NAME='CNTL.DAT',TYPE='OLD',READONLY)
      ICAL=0
      READ(4,901,END=99)NV
901   FORMAT(20I6)
      READ(4,902)XMAT
902   FORMAT(72A1)
      READ(4,902)YMAT
      READ(4,901,END=99)NOS
      READ(2,XMAT)(XMM(J),J=1,NV)
      READ(2,XMAT)(STD(J),J=1,NV)
21    FORMAT(F6.3)
      DO 30 J=1,NV
      XM(J)=0.
      XM2(J)=0.
      DO 30 K=1,NV
      CP(J,K)=0.
30    RC(J,K)=0.
      DO 49 I=1,NV
      READ(2,YMAT)(R(I,J),J=1,NV)
78    FORMAT(<NV>F3.0)
      R(I,I)=0.0
49    CONTINUE
      DO 22 I=1,NOS
      CALL MSCORE (R,NV,XMM,STD,NV,X,1,NV,1,NV,3,ICAL)
      DO 20 J=1,NV
      XM(J)=XM(J)+X(J)
      XM2(J)=XM2(J)+X(J)**2
      DO 20 K=1,NV
20    CP(J,K)=CP(J,K)+X(J)*X(K)
      WRITE(1,908)(X(J),J=1,NV)
908   FORMAT(<NV>F11.6)
22    CONTINUE
99    DO 110 J=1,NV
      XBAR(J)=XM(J)/NOS
      SDEV(J)=SQRT(((NOS*XM2(J))-(XM(J)**2))/(NOS*(NOS-1.)))
      DO 100 K=1,NV
      IF(J.EQ.K)RC(J,K)=1.
      IF(J.EQ.K)GOTO100
      RC(J,K)=((NOS*CP(J,K))-(XM(J)*XM(K)))/SQRT((NOS*XM2(J)-
     *XM(J)**2)*(NOS*XM2(K)-XM(K)**2))
100   CONTINUE
110   WRITE(3,980)SDEV(J),(RC(J,K),K=1,NV)
980   FORMAT(F7.2,<NV>F6.3)
      STOP
      END
      SUBROUTINE MSCORE (R,NRR,XMM,STD,NDRR,X,NRRX,NCCX,
     +NDRRX,NDCCX,NCOUNT,ICAL)
C  PROGRAMMED BY PETER TAM
C THIS SUBROUTINE GENERATES MULTIDEMENSIONAL SCORES, THE USER
C SPECIFIES THE NO. OF PERSONS AND SCORES FOR EACH PERSON.  BOTH PEOPLE
C AND SCORES CAN BE MADE UNLIMITED BY ADJUSTING THE DIMENSION STATEMENTS
C
C R= THE INPUT R MATRIX OF INTERCORRELATIONS, THE DIOGANAL AND UPER DIAGONAL
C  ELEMENTS SHOULD BE 0.0, TYPE IN THE WHOLE MATRIX OF R
C  NRR= THE NO. OF ROWS ACTUALLY IN R MATRIX
C  XMM= THE VECTOR OF MEANS FOR THE VARIABLES
C  STD = THE VECTOR OF STANDARD DEVIATIONS FOR THE VARIABLES.
C  NDRR= THE NO. OF ROWS DIMENSIONED FOR R MATRIX IN THE CALLING PROGRAM
C X= THE MATRIX OF SCORES GENERATED.  ROWS REPRESENT PERSONS, COLUMNS THE SCORES
C  NRRX= THE NO. OF SUBJECTS (ROWS) NEEDED
C  NCCX= THE NO. OF VARIABLES (COLS) NEEDED FOR EACH PERSON
C  NDRRX= THE NO. OF ROWS DIMENSIONED FOR X IN THE CALLING PROGRAM
C  NDCCX= THE NUMBER OF COLUMNS DIMENSIONED FOR X IN THE CALLING PROGRAM
```

```
C NCOUNT..THE PRINT CHOICE.  IF VALUE LESS THAN OR EQ. 3, NO  PRINT.
C  IF NCOUNT GT 3 WILL PRINT ON DESIGNATED NCOUNT.
      DIMENSION R(NDRR,NDRR),XMM(NDRR),STD(NDRR),VAR(30,30),C(30,30),
     + X(NDRRX,NDCCX),TV(30,30),TDS(30)
      DO 10 I=1,NRR
      VAR(I,I)=STD(I)**2
   10 CONTINUE
      NNR=NRR-1
      DO 30 J=1,NNR
      K=J+1
      DO 20 I=K,NRR
      VAR(I,J)=R(I,J)*STD(J)*STD(I)
   20 VAR(J,I)=VAR(I,J)
   30 CONTINUE
      IF(NCOUNT.GT.2)GOTO34
      WRITE(NCOUNT,31)
   31 FORMAT(1H1,///5X,'VAR-COV MATRIX FROM INPUT CORREL. MATRIX')
      DO 32 I=1,NRR
   32 WRITE(NCOUNT,33)(VAR(I,J),J=1,NRR)
   33 FORMAT((1H0,5X,6(F15.2,2X)))
   34 DO 35 I=1,NRR
      DO 35 J=1,NRR
   35 C(I,J)=0.0
      DO 40 I=1,NRR
   40 C(I,1)=VAR(I,1)/STD(1)
      DO 100 I=2,NRR
      IM1=I-1
      A=0.0
      DO 50 J=1,IM1
   50 A=A+C(I,J)**2
  997 FORMAT(I3,2F10.3)
      C(I,I)=SQRT(VAR(I,I)-A)
      IP1=I+1
      IF(IP1.GT.NRR)GOTO150
      DO 90 L=IP1,NRR
      B=0.0
      DO 80 K=1,IM1
   80 B=B+C(L,K)*C(I,K)
      C(L,I)=(VAR(L,I)-B)/C(I,I)
   90 CONTINUE
  100 CONTINUE
  150 DO 180 I=1,NRRX
      DO 180 J=1,NCCX
  180 X(I,J)=0.0
      DO 230  I=1,NRRX
      DO 170 L=1,NCCX
      CALL MIX (ZW,ICAL)
  170 TDS(L)=ZW
      DO 220 J=1,NCCX
      DO 210 K=1,J
  210 X(I,J)=X(I,J)+(C(J,K)*TDS(K))
      X(I,J)=X(I,J)+XMM(J)
  220 CONTINUE
  230 CONTINUE
      IF(NCOUNT.GE.3)RETURN
      WRITE(NCOUNT,250)
  250 FORMAT(///5X,'TRANSFORMATION MATRIX C',//)
      DO 260 I=1,NRR
  260 WRITE(NCOUNT,33)(C(I,J),J=1,NRR)
      WRITE(NCOUNT,500)
  500 FORMAT(//5X,'RECHECK PARAMETER INFORMATION',//5X,
     1 'ORIGINAL PARAMETER R MATRIX')
      DO 520 I=1,NRR
  520 WRITE(NCOUNT,33)(R(I,J),J=1,NRR)
      WRITE(NCOUNT,540)(XMM(I),I=1,NRR)
  540 FORMAT(//5X,'ORIGINAL VECTOR OF MEANS',/
     1 11X,6(F8.2,2X))
      WRITE(NCOUNT,300)
```

```
300    FORMAT(1H1,///20X,'MULTIDIMENSIONAL SCORES',/15X,
     1 'NOTE. ROWS ARE SUBJECTS, COLS ARE VARIABLES')
       RETURN
       END
       SUBROUTINE MIX(AA,IC)
       DOUBLE PRECISION ILT
       BYTE NUM(10),SEC(8)
C
C RANDOM NORMAL DEVIATES BY MARSAGLIA-S REASONABLY FAST METHOD.
C  PROGRAMMED BY ONE OF MEETER-S STUDENTS.  1970
C
       DATA A/.8638/,B/.1107/,C/.0228002039/,D/.0026997961/
       DATA A0/17.49731196/,A1/2.36785163/,A2/2.15797544/
       DATA NUM/'0','1','2','3','4','5','6','7','8','9'/
       IC=IC+1
       IF(IC.NE.1)GOTO1
       ILT=0
       CALL TIME (SEC)
       DO 56 J=1,10
       IF(SEC(7).EQ.NUM(J))ILT=ILT+(J-1)*10
       IF(SEC(8).EQ.NUM(J))ILT=ILT+J-1
       IF(SEC(5).EQ.NUM(J))ILT=ILT+(J-1)*60
56     CONTINUE
       I1=0
       I2=0
       DO 98 K=1,ILT
98     XYZ=RAN(I1,I2)
1      U1=RAN(I1,I2)
       IF(U1.GT.A)GOTO2
       AA=2.*(U1/A+RAN(I1,I2)+RAN(I1,I2)-1.5)
       RETURN
2      IF(U1.GT.A+B)GOTO4
       AA=1.5*((U1-A)/B+RAN(I1,I2)-1.)
       RETURN
4      IF(U1.GT.A+B+C)GOTO6
       X=6.*(U1-A-B)/C-3.
21     Y=.358*RAN(I1,I2)
       XA=ABS(X)
       IF(XA.GT.1.)GOTO7
19     G3=A0*EXP(-(X*X)/2.)-A1*(3.-XA)**2
       GOTO13
7      IF(XA.GT.1.5)GOTO9
8      G3=A0*EXP(-(X*X)/2.)-A1*(3.-XA)**2-A2*(1.5-XA)
       GOTO13
9      IF(XA.GT.3.)GOTO12
11     G3=A0*EXP(-(X*X)/2.)-A1*(3.-XA)**2
       GOTO13
12     G3=0.
13     IF(Y.LT.G3)GOTO14
20     X=6.*RAN(I1,I2)-3
       GOTO21
14     AA=X
       RETURN
6      V1=2.*((U1-A-B-C)/D)-1.
23     V2=2.*RAN(I1,I2)-1.
       R=V1*V1+V2*V2
       IF(R5.GT.1.)GOTO22
15     Z=SQRT((9.-2.*ALOG(R))/R)
       X1=V1*Z
       X2=V2*Z
       X1A=ABS(X1)
       IF(X1A.GT.3.)GOTO16
17     X2A=ABS(X2)
       IF(X2A.GT.3.)GOTO22
18     AA=X2
       RETURN
22     V1=2.*RAN(I1,I2)-1.
       GOTO23
16     AA=X1
       RETURN
       END
```

```
      PROGRAM REST
C  THIS PROGRAM TAKES SORTED OUTPUT FROM FILE NORM.DAT
C  AND CORRELATIONS AND SD'S FROM COR.DAT AND RESTRICTS
C  THE FILE BY INCREMENTS OF 10%
C  THE PROGRAM CALLS A SUBROUTINE (COREST) WHICH
C  CORRECTS FOR RESTRICTION IN RANGE USING FORMULAS
C  B-1, E-2, G-37, AND T-7
      DIMENSION X(11),S(11),R(11,11),SX(11),SX2(11),SXY(11,11),C(4),
     * RR(11,11),SS(11),IV(5),IR(5)
      OPEN(UNIT=1,NAME='NORM.DAT',TYPE='OLD',READONLY)
C  THIS FILE CONTAINS THE DATA GENERATED BY THE PROGRAM MHRNG
      OPEN(UNIT=2,NAME='CORR.DAT',TYPE='OLD',READONLY)
C  THIS FILE CONTAINS THE UNRESTRICTED CORRELATIONS AND SD'S FROM NORM
      OPEN(UNIT=3,NAME='RNEW.DAT')
      XT=0.
      IR(1)=2
      IR(2)=9
      IR(3)=3
      IR(4)=10
      IR(5)=11
      IV(1)=3
      IV(2)=4
      IV(3)=4
      IV(4)=3
      IV(5)=3
      DO 3 J=1,11
    3 READ(2,910)SS(J),(RR(J,K),K=1,11)
  910 FORMAT(F7.2,11F6.3)
      DO 1 J=1,11
      SX(J)=0.
      S(J)=0.
      SX2(J)=0.
      DO 1 K=1,11
      R(J,K)=0.
    1 SXY(J,K)=0.
      DO 60 IRS=1,5
      DO 10 MR=1,100
      READ(1,920)(X(J),J=1,11)
  920 FORMAT(11F11.6)
      XT=XT+1.
      DO 10 J=1,11
      SX(J)=SX(J)+X(J)
      SX2(J)=SX2(J)+X(J)**2
      DO 10 K=1,11
   10 SXY(J,K)=SXY(J,K)+X(J)*X(K)
      DO 20 J=1,11
      S(J)=SQRT(((XT*SX2(J))-(SX(J)**2))/(XT*(XT-1.)))
      DO 19 K=1,11
      R(J,K)=((XT*SXY(J,K))-(SX(J)*SX(K)))/
     * SQRT((XT*SX2(J)-SX(J)**2)*(XT*SX2(K)-SX(K)**2))
   19 CONTINUE
   20 CONTINUE
      J=1
      DO 25 IM=1,5
      K=IR(IM)
      L=IV(IM)
      MN=L+4
      DO 25 N=L,MN
      YZ=RR(K,N)
      XY=RR(J,K)
      OR=S(N)
      CALL COREST(YZ,XY,SS(J),SS(K),OR,S(J),S(K),R(J,K),R(J,N),R(K,N),XT,C)
      ZB=R(K,N)
   25 WRITE(3,925)C,YZ,RR(J,N),XY,SS(J),S(J),SS(K),S(K),ZB,
     * SS(N),S(N),R(J,K),R(J,N)
  925 FORMAT(16F7.3)
   60 CONTINUE
   99 STOP
      END
```

Appendix B

```
      SUBROUTINE COREST
      SUBROUTINE COREST(RRYZ,RRXY,SXX,SYY,SZ,SX,SY,RXY,RXZ,RYZ,XT,C)
C
C
C  THIS SUBROUTINE TAKES INPUT FROM THE PROGRAM REST AND
C  CALCULATES ESTIMATED RRYZ USING FORMULAS B-1, B-2, AND
C  G-37, AND T-7
C  RRYZ,RRXY= UNRESTRICTED CORRELATIONS; SXX,SYY=UNRESTRICTED SD'S
C  RXY,RXZ,RYZ= RESTRICTED CORRELATIONS, SX,SY,SZ=RESTRICTED SD'S
C  X= EXPLICIT SELECTION VARIABLE, Y= IMPLICIT SELECTION VARIABLE
C  Z= CRITERION VARIABLE
C  THE ESTIMATES ARE TRANSFORMED USING FISCHER R TO Z SO THAT
C  THEY CAN LATER BE AVERAGED AND RETURNED TO PROGRAM REST
C
C
      DIMENSION C(4)
  910 FORMAT(10F7.3)
      DO 478 J=1,4
      C(J)=0.
478   CONTINUE
      XX=(SXX**2)/(SX**2)
      XXX=SQRT(XX)
      FIR=(SY*(RYZ-(RXY*RXZ)))/(SYY*SQRT((1-RXZ**2)+
     1 (RXZ**2*XX)))
      SEC=(RXZ*XXX)/SQRT((1-RXZ**2)+(RXZ**2*XX))
      B1=FIR+SEC*RRXY
      DET=RYZ-RXY*RXZ+RXY*RXZ*XX
      RM=RXY**2
      XNUM=SQRT(((1.-RM)+(RM*XX))*((1.-RXZ**2)+(RXZ**2*XX)))
      T7=DET/XNUM
      SB=SY**2
      DEN=RXZ*(SYY**2-SB)+RXY*RYZ*SB
      XNUM=SYY*SQRT((RXZ**2*(SYY**2-SB))+SB*(RXY**2))
      OX=SYY**2
      G37=DEN/XNUM
      FIR=(RYZ-RXY*RXZ)*SY*SZ
      SEC=SYY*SQRT(SZ**2*(((SY**2*RXY**2)-(SY**2*RXZ**2)+
     * (SYY**2*RXZ**2))/(SY**2*RXY**2)))
      THIR=FIR/SEC
      FOUR=RXZ*SQRT((SYY**2-SY**2+(SY**2*RXY**2))/((SYY**2*
     * RXZ**2)-(SY**2*RXZ**2)+(SY**2*RXY**2)))
      B2=THIR+FOUR*RRXY
      C(3)=(ALOG((1.+G37)/(1.-G37)))/2.
      C(4)=(ALOG((1.+B2)/(1.-B2)))/2.
      C(1)=(ALOG((1.+B1)/(1.-B1)))/2.
      C(2)=(ALOG((1.+T7)/(1.-T7)))/2.
   99 RETURN
      END
```

* U. S. GOVERNMENT PRINTING OFFICE : 1970—279-109/6580