DOCUMENT RESUME

ED 330 708 TM 016 276

AUTHOR Schafer, William D.; Wang, Yuh-Yin
TITLE Graphical Description of Johnson-Neyman Outcomes for Linear and Quadratic Regression Surfaces.
PUB DATE 91
NOTE 18p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Computer Software; *Graphs; Interactive Video; *Predictor Variables; *Regression (Statistics)
IDENTIFIERS *Johnson Neyman Technique; Linear Models; *Quadratic Equations; Univariate Analysis

ABSTRACT

A modification of the usual graphical representation of heterogeneous regressions is described that can aid in interpreting significant regions for linear or quadratic surfaces. The standard Johnson-Neyman graph is a bivariate plot with the criterion variable on the ordinate and the predictor variable on the abscissa. Regression surfaces are drawn for each group. If there are regions of significance, their boundaries are noted either on the graph or in the text. If there is a manageable number of cases, the cases may be plotted on the graph with different symbols for the groups. This standard style of representation often suffers from inclusion of too many cases for realistic plotting of data and from the difficulty in translating a bivariate display of data into univariate distributional characteristics of the two groups. The modification proposed here alleviates these problems. Computer programs for solutions in both linear and quadratic surfaces are included. These programs, which were written in GAUSS, differ from earlier programs in that the input required is summary information available from standard computer package output rather than raw data and is entered interactively. A 39-item list of references, three graphs, and seven pages of computer programming language are provided. (TJH)

# GRAPHICAL DESCRIPTION OF JOHNSON-NEYMAN OUTCOMES FOR LINEAR AND QUADRATIC REGRESSION SURFACES

William D. Schafer & Yuh-Yin Wang
University of Maryland at College Park

Graphical Description of Johnson-Neyman Outcomes for Linear and Quadratic Regression Surfaces

William D. Schafer and Yuh-Yin Wang
University of Maryland at College Park

The presence of heterogeneous regression slopes in an analysis of covariance design is analogous to the presence of interaction in factorial designs. It is puzzling, therefore, that current practices seem to follow conflicting recommendations in analyzing these two plans. Chosen as an example only because of recency, Maxwell and Delaney (1990) present an analysis of a two-way, fixed-effects design using, as denominators for all F tests, the mean square residual about a model that includes interaction (see Chapter 7), whereas they present an analysis of a one-way analysis of covariance design using, as the denominator of the F test for treatment effects, the mean square residual about a model with the common within-groups regression slope. Tests for heterogeneous slopes and eventual recommendations about incorporating the more complete model for all tests are treated later under the heading of an extension and using a different name for the technique (see Chapter 9). Some common computer package procedures (e.g., SPSS ANOVA) include interaction among explained sources in analyses of variance and heterogeneity among unexplained sources in analyses of covariance. These practices are contradictory. Identifying heterogeneous regressions in analysis of covariance designs makes as much sense as identifying interaction in two-way designs and on that basis should be equally frequent. Furthermore, techniques for display and significance testing exist in either context.

Interaction in factorial designs is often illustrated with the aid of interaction graphs of cell means, where the abscissa is, if one exists, a blocking factor, and the ordinate is the criterion. Contrasts, with or without familywise error rate control, can be used to compare treatment differences at levels of the blocking factor. Similarly, in analysis of covariance designs, heterogeneous slopes may be illustrated by a plot of the two or more group-specific regression equations on a graph where the abscissa is the covariate and the ordinate is the criterion. The plot may or may not display the data points. Since the scores are not grouped on the covariate, the parallel question about conditional treatment differences is usually addressed by partitioning the range of the covariate into regions where the differences are significant and where they are not. As in the case of factorial designs, this can be done with or without familywise error rate control. Known in general as the Johnson-Neyman procedure, a method to accomplish this analysis was originally developed by Johnson and Neyman (1936) for the two-group case; Pedhazur (1982) gives details of the procedure and references to extensions for more than two groups, more than one covariate, and regions where the significance interpretation is simultaneous.

Several authors have investigated the Johnson-Neyman technique and extended it beyond the one- or two-predictor cases. Potthoff (1964) described Scheffe-type solutions for simultaneous confidence intervals, multiple predictors (due to Abelson, 1953), multiple groups, and multiple criteria. Cahen and Linn (1971) have compared three methods for defining regions of significance that differ in their conservativeness. Solutions for multiple groups and multiple predictors have also been described by Forster (1971, 1975). Aitkin (1972) has developed a similar technique. Shields (1978) has found that the Johnson-Neyman technique is sensitive to heterogeneity of residual variances for equal as well as unequal group sizes, although Borich and Wunderlich (1973) did not. Hollingsworth (1977) has compared tests for homogeneous variances in contexts including the Johnson-Neyman technique and Pigache and Graham (1976) have described an extension that does not assume that variances are homogeneous. Rogosa (1980, 1981) has discussed issues surrounding the Johnson-Neyman technique, including simultaneous significance adjustments, and Wilcox (1987) has described an adaptation of Tukey-Kramer simultaneous procedures to multiple-group cases. Tsutakawa (1978) has developed a Bayesian solution and Houston (1987) as reported in Houston and Novick (1987) has demonstrated the equivalence of Bayesian solutions using noninformative priors with classical solutions, assuming equal residual variances. Borich and Wunderlich (1973) have discussed procedures for plotting results with two groups and two predictors and Wunderlich and Borich (1974) have described an extension to quadratic regressions.

Several computer programs have been prepared to yield Johnson-Neyman solutions. Carroll and Wilson (1970), Ceurvorst (1979), Karpman (1980), Kush (1986), Lautenschlager (1987), Scialfa (1987), and Strube (1988) have developed algorithms for this purpose in various programming languages and Karpman (1983, 1986) has discussed how to arrive at Johnson-Neyman solutions using SAS and SPSS transformational languages. A program that yields solutions for heterogeneous quadratic regressions can be found in Wunderlich and Borich (1974).

The ERIC and PsychINFO document bases were searched for Johnson()Neyman in order to survey uses of the Johnson-Neyman technique. Berliner (1971) found interactions among memory aptitudes, note-taking, and test-like events. Keim-Abbott and Abbott (1977) found an interaction between instructional treatments and mental ability. McLeskey and Rieth (1982) compared reading disabled and normal children with IQ as the predictor. Dunbar and Novick (1985) found differential predictions by gender in Marine Corps training programs. Gamache and Novick (1985) have examined differential prediction of grade-point average by gender. Houston and Novick (1987) examined differential prediction by race. Reeves-Kazelskis and Kazelskis (1987) found a disordinal interaction between treatments and a prior knowledge pretest.

The purpose of this paper is to describe a modification to the usual graphical representation of heterogeneous regressions that can aid in interpreting significant regions for linear or quadratic regression surfaces. Computer programs for solutions in both cases are included. These differ from earlier programs in that the input required is summary information available from standard computer package output instead of raw data, and is entered interactively. In most applications, we feel a researcher will have run preliminary tests (for heterogeneous linear or quadratic regressions) before deciding to perform a Johnson-Neyman analysis; Wunderlich and Borich (1974) present a decision plan for making this judgment. The data are thus already prepared for analysis, likely in the format needed by a computer package, and the needed summary information is or easily can be made available, so a simple interactive program to find the Johnson-Neyman solution(s) will be more convenient to use than it would be to set up and run the analysis "from scratch" using a raw-data algorithm.

## GRAPHICAL REPRESENTATION

The standard Johnson-Neyman graph is a bivariate plot with the criterion variable on the ordinate and the predictor variable on the abscissa. Regression surfaces are drawn for each group. If there are regions of significance, their boundaries are noted either on the graph or in the text. If there are a manageable number of cases, they may be plotted on the graph with different symbols for the groups. A prototypic example can be found in Walker and Lev (1953, p. 404).

There are at least two difficulties with this style of presentation. First, the number of cases may be too large to be able realistically to plot the data. The graph would become too confusing and multiple overlapping points are difficult to represent graphically. The second difficulty is that a bivariate display of the data does not readily translate into univariate distributional characteristics of the two groups. The latter is important, particularly for the predictor variable, since the locations of significant differences are defined on the predictor and best interpreted in relation to its distribution.

A convenient way to convey information about the univariate distribution of a sample is to use a box-and-whisker plot. Although there is no general agreement about what constitutes a box-and-whisker plot, Glass and Hopkins (1984, pp. 23-24) have recommended an approach that is easy to explain and implement. It is recommended here that a box-and-whisker plot for each group be drawn on the scale of the predictor variable and placed beneath the abscissa of the graph. Also, boundaries between significant and nonsignificant regions within the range of the data can be identified by vertical lines. In this manner, it is easy to grasp the relationships among the distributions of the groups and where on these distributions significant differences are located.

An example for the linear case taken from Guthrie, Schafer and Hutchinson (1991) is included in this paper. The graph describes differential predictions by gender for document reading variety when document achievement is the predictor (both scales were constructed using the data from the National Assessment of Educational Progress 1985 study of young adult literacy). The data are for 870 blacks and 1945 whites, clearly too many subjects to plot in any meaningful way. On the artificial document achievement scale (X; an ability measure), the percentiles for blacks that formed the box-and-whisker plot were 5(-1.45), 10(-1.13), 25(-.42), 50(-.02), 75(.38), 90(.98), and 95(1.30). The corresponding percentiles for whites were 5(-.43), 10(-.28), 25(.16), 50(.84), 75(1.46), 90(2.18), and 95(2.43). The regression lines were $Y' = 8.62 + 2.17X$ for blacks and $Y' = 9.26 + 1.75$ for whites. The region of nonsignificance (confidence coefficient = .95) was between .53 and 35.12. The vertical line in the diagram corresponds to .53 on the document achievement scale; the other boundary was outside the range of the data and thus was not drawn. The figure seems to present this information in an appealing and interpretable way.

Two examples for the curvilinear case taken from Guthrie, Schafer and Wang (1991) are also included. In both examples, the criterion is reading achievement and the groups are black males and black females. The predictors are study strategies, showing nonintersecting quadratic regressions with one solution in the range of the data, and general reading activities, showing intersecting regressions with four solutions in the range of the data (both predictors are derived variables). The data were part of the 1986 NAEP reading assessment. Because both predictors were derived scales made up of variables in different partitions of the items in the balanced incomplete block design by which the items were administered to the participants, the percentiles used in the box and whisker plots were estimated assuming normality in each group instead of directly from observed distributions and are therefore only crude approximations to the actual percentiles that might otherwise have been obtained.

## COMPUTER PROGRAMS

Computer programs for Johnson-Neyman solutions for two groups were written, one for linear solutions and one for quadratic solutions. These programs were designed to be interactive and to take as input summary information available from the output of regression procedures found in standard statistical packages; our assumption is that the user will have already determined that heterogeneity of regression exists and will have found the linear or quadratic regression equations, the univariate means and standard deviations and the bivariate correlations among the variables for each group along with the group sizes.

The programs were written in GAUSS. This language was chosen because it is fast and maintains a high degree of precision. It also has available two procedures necessary for the solution: one that returns a p-value for an F distribution and one that returns the solutions to general equations of the nth degree.

The source code and a sample output are given for the linear solution. The data used were taken from an example in Walker and Lev (1953, p. 403). The confidence coefficient and the adjustments necessary for a simultaneous region (Potthoff, 1964) are easily changed where the fcv procedure is called. The solution agrees with that in Walker and Lev (1953).

The source code and a sample output are also given for the quadratic solution. This solution was developed using the two-predictor solution presented by Walker and Lev (1953, pp. 406-407). Their predictors were labeled X and Z. The solution was algebraically obtained by substituting $X^2$ for Z and simplifying. Since a quartic equations results, up to four solutions are possible, making interpretation less straightforward than for the linear case. Therefore, th program requests the domain of the predictor variable and outputs significance decisions for eleven equally spaced points in that interval. It is suggested in using the program that the observed range of the predictor is used as the end-points of its domain, effectively limiting interpretation to the interval where data are present. In order to check the program, a solution for an example was developed by hand and compared with the results.

## References


Abelson, R. P. A. (1953). A note on the Neyman-Johnson technique. *Psychometrika, 18*, 213-218.

Aitkin, M. A. (1972). *Fixed-width confidence intervals in linear regression with applications to the Johnson-Neyman technique* (Report No. RB-72-18). Princeton, NJ: Educational Testing Service.

Berliner, D. C. (1971, February). *Aptitude-treatment interactions in two studies of learning from lecture instruction.* Paper presented at the meeting of the American Educational Research Association, New York, NY.

Borich, G. D. & Wunderlich, K. W. (1973). Johnson-Neyman revisited: Determining interactions among group regressions and plotting regions of significance in the case of two groups, two predictors, and one criterion. *Educational and Psychological Measurement, 33*, 155-159.

Cahen, L. S. & Linn, R. L. (1971). Regions of significant criterion differences in aptitude-treatment-interaction research. *American Educational Research Journal, 8*, 521-530.

Carroll, J. B. & Wilson, G. F. (1970). An interactive computer program for the Johnson-Neyman technique in the case of two groups, two predictor variables, and one criterion variable. *Educational and Psychological Measurement, 30*, 121-132.

Ceurvorst, R. W. (1979). Computer program for the Johnson-Neyman technique. *Educational and Psychological Measurement, 39*, 205-207.

Dunbar, S. B. & Novick, M. R. (1985). *On predicting success in training for males and females: Marine Corps clerical specialties and ASVAB forms 6 and 7* (Report No. ONR-TR-85-2). Arlington, VA: Office of Naval Research.

Forster, F. (1971, February). *The generalized Johnson-Neyman procedures: An approach to covariate adjustment and interaction analysis.* Paper presented at the meeting of the American Educational Research Association, New York, NY.

Forster, F. (1975, April). *An alternative to ancova when group regressions are heterogeneous: The generalized Johnson-Neyman procedure.* Paper presented at the meeting of the American Educational Research Association, Washington, DC.

Gamache, L. M. & Novick, M. R. (1985). Choice variables and gender differentiated prediction within selected academic programs. *Journal of Educational Measurement, 22*, 5.-70.

Glass, G. V. & Hopkins, K. D. (1984). *Statistical methods in education and psychology* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Guthrie, J. T., Schafer, W. D. & Hutchinson, S. R. (1991). Relationships of document literacy and prose literacy to occupational and societal characteristics of young black and white adults. *Reading Research Quarterly, 26*(1), 30-48.

Guthrie, J. T., Schafer, W. D., & Wang, Y. (1991, April). *Minority reading achievement: Motivational, instructional, and familial variables for black and white males and females.* Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Hollingsworth, H. (1977). Tests for equal within-class variances for k-sample regression problems. *Journal of Experimental Education, 45*(4), 37-41.

Houston, W. M. (1987). *The equivalence of classical and Bayesian Johnson-Neyman analysis for noninformative prior distributions.* Unpublished manuscript.

Houston, W. M. & Novick, M. R. (1987). Race-based differential prediction in Air Force technical training programs. *Journal of Educational Measurement, 24*, 309-320.

Johnson, P. O. & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoires, 1*, 57-93.

Karpman, M. B. (1980). ANCOVA: A one covariate Johnson-Neyman algorithm. *Educational and Psychological Measurement, 40*, 791-793.

Karpman, M. B. (1983). The Johnson-Neyman technique using SPSS or BMDP. *Educational and Psychological Measurement, 43*, 137-147.

Karpman, M. B. (1986). Comparing two non-parallel regression lines with the parametric alternative to analysis of covariance using SPSS-X or SAS - the Johnson-Neyman technique. *Educational and Psychological Measurement, 46*, 639-644.

Keim-Abbott, S. & Abbott, R. D. (1977). Moderation of achievement prediction in an elementary school metric curriculum by trait * instructional method interactions. *Educational and Psychological Measurement, 37*, 481-486.

Kush, J. C. (1986). A FORTRAN V IBM computer program for the Johnson-Neyman technique. *Educational and Psychological Measurement, 46*, 185-187.

Lautenschlager, G. T. (1987). JOHN-NEY: An interactive program for computing the Johnson-Neyman confidence region for nonsignificant prediction differences. *Applied Psychological Measurement, 11*, 174 & 194.

Maxwell, S. E. & Delaney, H. D. (1990). *Designing experiments and analyzing data.* Belmont, CA: Wadsworth Publishing Company.

McLeskey, J. & Rieth, H. J. (1982). Controlling IQ differences between reading disabled and normal children: An empirical example. *Journal of Learning Disabilities*, 15, 481-483.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). Fort Worth, TX: Holt, Rinehart & Winston.

Pigache, R. M. & Graham, B. R. (1976). A modification of the Johnson-Neyman technique comparing two regressions applied to treatment effects dependent on baseline levels. *Biological Psychology, 4*, 213-235.

Pottnoff, R. F. (1964). On the Johnson-Neyman technique and some extensions thereof. *Psychometrika, 29*, 241-256.

Reeves-Kazelskis, C. & Kazelskis, R. (1987, November). *The effects of student-generated questions on test performance.* Paper presented at the meeting of the Mid-South Educational Research Association, Mobile, AL.

Rogosa, D. (1980). Comparing nonparallel regression lines. *Psychological Bulletin, 88*, 307-321.

Rogosa, D. (1981). On the relationship between the Johnson-Neyman region of significance and statistical tests of parallel within-group regressions. *Educational and Psychological Measurement, 41*, 73-84.

Scialfa, C. T. (1987). A BASIC program to determine regions of significance using the Johnson-Neyman technique. *Behavior Research Methods Instruments & Computers, 19*, 349-352.

Shields, J. L. (1978). *An empirical investigation of the effect of heteroscedasticity and heterogeneity of variance on the analysis of covariance and the Johnson-Neyman technique* (Report No. ARI-TP-292). Arlington, VA: Army Research Institute for the Behavioral and Social Sciences.

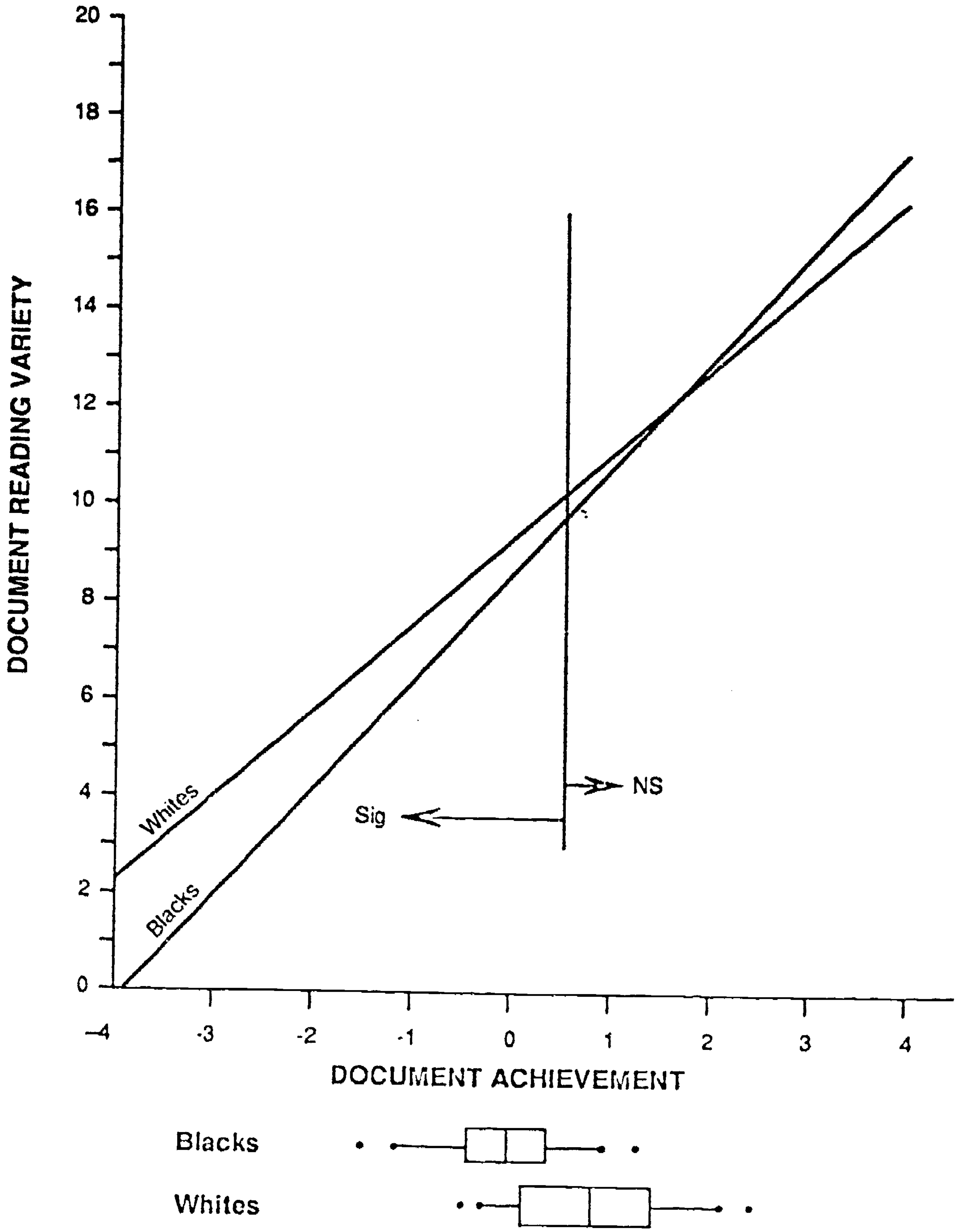
Strube, M. J. (1988). Calculation of significance regions for multiple predictors by the Johnson-Neyman technique. *Behavior Research Metnods Instruments & Computers, 20*, 510-512.

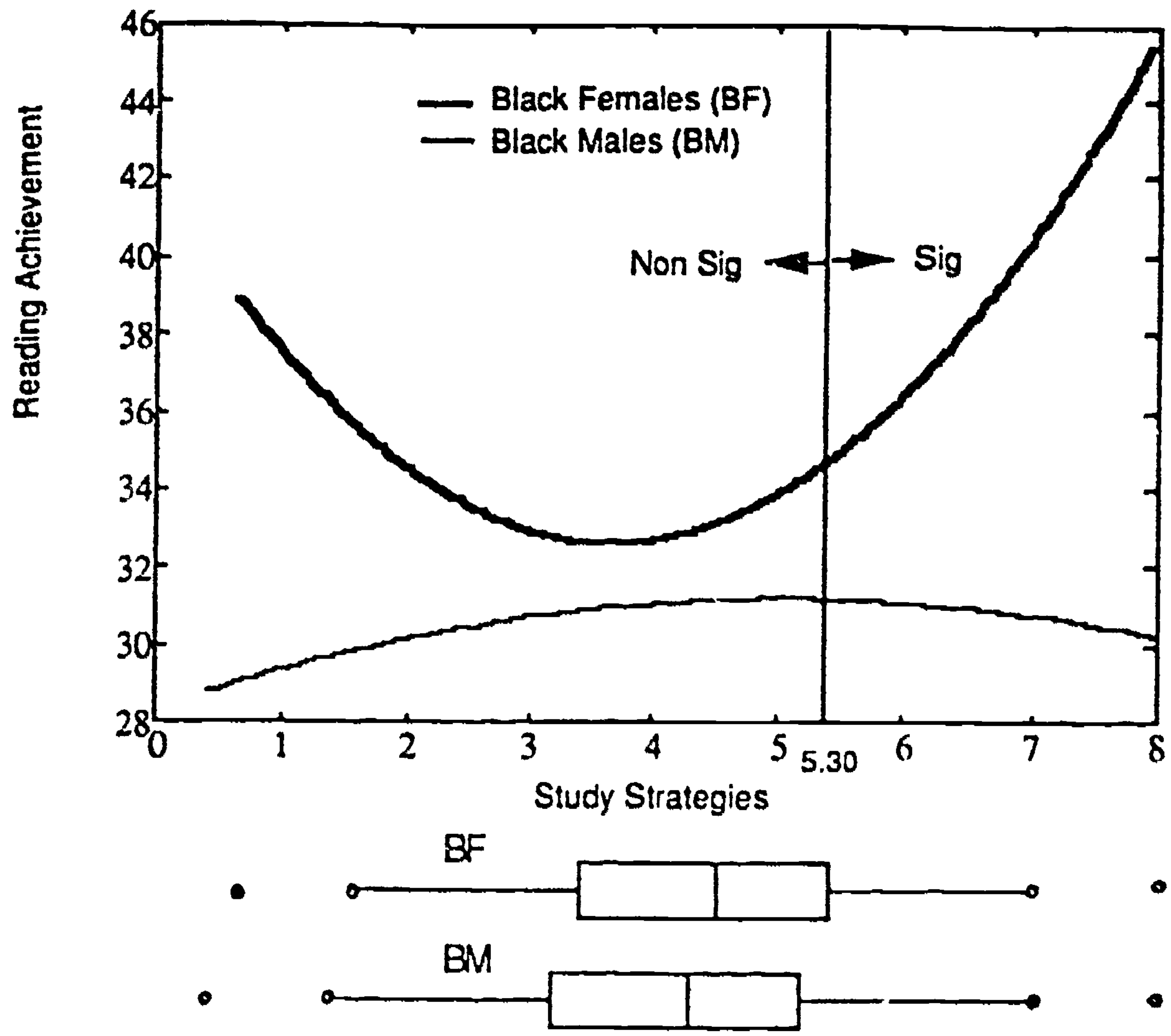
Tsutakawa, R. K. (1978). Bayesian comparison of two regression lines. *Journal of Educational Statistics, 3*, 179-188.

Walker, H. M. & Lev, J. (1953). *Statistical inference.* New York, NY: Holt, Rinehart & Winston.

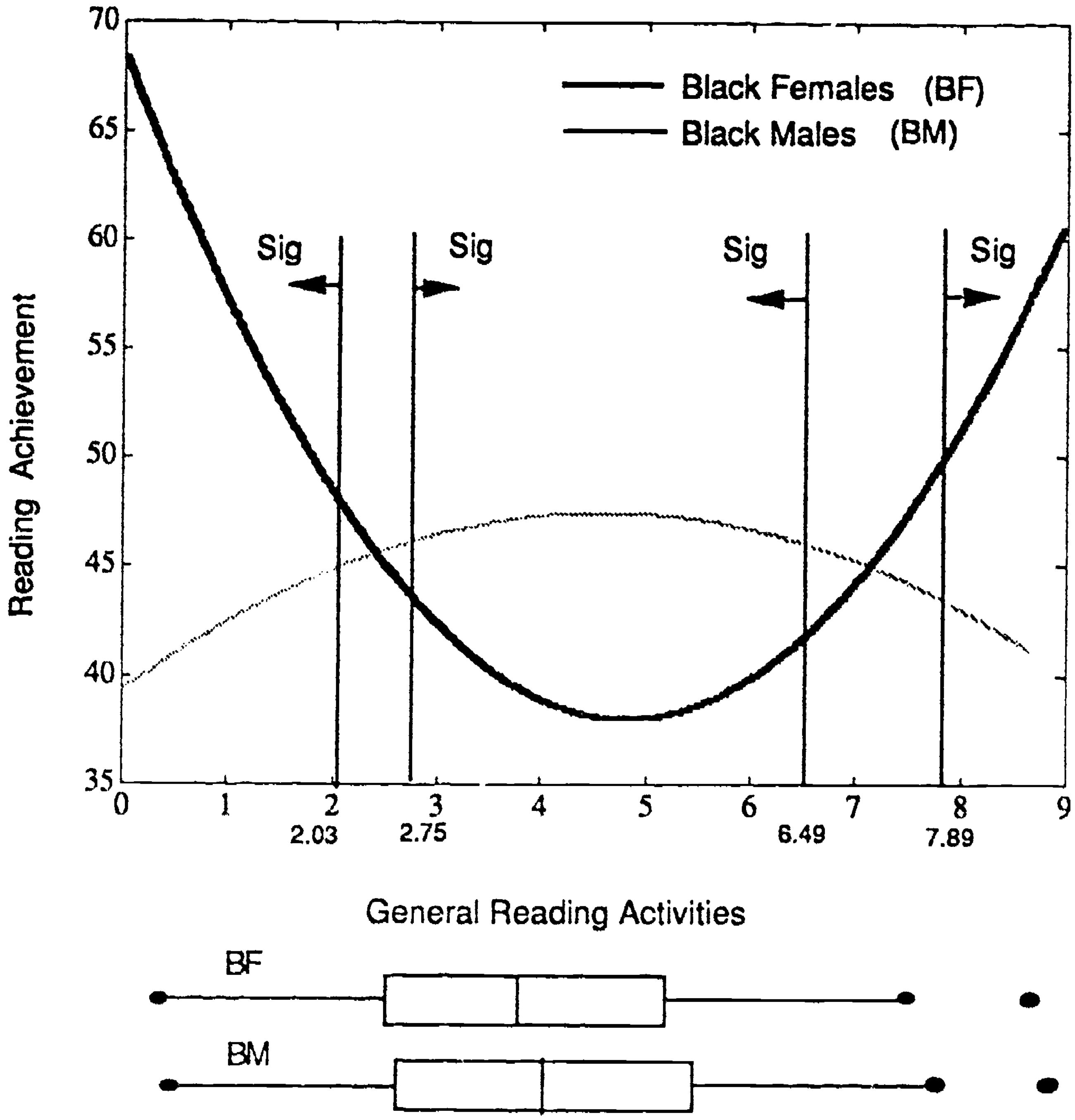
Wilcox, R. R. (1987). Pairwise comparisons of j independent regression lines over a finite interval, simultaneous pairwise comparison of their parameters, and the Johnson-Neyman procedure. *British Journal of Mathematical and Statistical Psychology, 40*, 80-93.

Wunderlich, K. W. & Borich, G. D. (1974, April). *Curvilinear extensions to Johnson-Neyman regions of significance and some applications to educational research.* Paper presented at the meeting of the American Educational Research Association, Chicago, IL.


20
18
16
14
12
10
8
6
4
2
0
DOCUMENT READING VARIETY
Whites
Blacks
Sig
NS
-4
-3
-2
-1
0
1
2
3
4
DOCUMENT ACHIEVEMENT
Blacks
Whites



Reading Achievement
46
44
42
40
38
36
34
32
30
28
Black Females (BF)
Black Males (BM)
Non Sig
Sig
0
1
2
3
4
5
5.30
6
7
8
Study Strategies
BF
BM


Black Females (BF)
Black Males (BM)
Sig
Sig
Sig
Sig
Reading Achievement
70
65
60
55
50
45
40
35
0
1
2
3
4
5
6
7
8
9
2.03
2.75
6.49
7.89
General Reading Activities
BF
BM

```
new;
proc fcv(ndf,ddf,alpha);
local f,fl,fh,ndf,ddf,alpha,a;
a=0;
fl=0;
fh=200;
f=1;
do until abs(a-alpha)<.000001;
   a = cdffc(f,ndf,ddf);
      if a < alpha;
        fh = f;
        f = (fl + f) / 2;
      else;
        fl = f;
        f = (fh + f) / 2;
      endif;
 endo;
retp(f);
endp;
print "What is the size of group 1?";
n1 = con(1,1);
lprint "The size of group 1 is " n1;
print "What is the mean of group 1 on the predictor?";
mx1 = con(1,1);
lprint "The mean of group 1 on the predictor is " mx1;
print "What is the standard deviation of group 1 on the predictor?";
sdx1 = con(1,1);
lprint "The standard deviation of group 1 on the predictor is " sdx1;
vx1 = sdx1 ^ 2;
print "What is the standard deviation of group 1 on the criterion?";
sdy1 = con(1,1);
lprint "The standard deviation of group 1 on the criterion is " sdy1;
vy1 = sdy1 ^ 2;
print "What is the intercept for group 1?";
a1 = con(1,1);
lprint "The intercept for group 1 is " a1;
print "What is the slope for group 1?";
b1 = con(1,1);
lprint "The slope for group 1 is " b1;
print "What is the size of group 2?";
n2 = con(1,1);
lprint "The size of group 2 is " n2;
print "What is the mean of group 2 on the predictor?";
mx2 = con(1,1);
lprint "The mean of group 2 on the predictor is " mx2;
print "What is the standard deviation of group 2 on the predictor?";
sdx2 = con(1,1);
lprint "The standard deviation of group 2 on the predictor is " sdx2;
vx2 = sdx2 ^ 2;
print "What is the standard deviation of group 2 on the criterion?";
sdy2 = con(1,1);
lprint "The standard deviation of group 2 on the criterion is " sdy2;
vy2 = sdy2 ^ 2;
print "What is the intercept for group 2?";
a2 = con(1,1);
lprint "The intercept for group 2 is " a2;
print "What is the slope for group 2?";
b2 = con(1,1);
lprint "The slope for group 2 is " b2;
r1 = b1 * sdx1/sdy1;
```

```
lprint "The correlation for group 1 is " r1;
r2 = b2 * sdx2/sdy2;
lprint "The correlation for group 2 is " r2;
df = n1 + n2 - 4;
f = fcv(1,df,.05);
cyy1 = (n1-1) * vy1;
cyy2 = (n2-1) * vy2;
cxx1 = (n1-1) * vx1;
cxx2 = (n2-1) * vx2;
cxy1 = r1 * sqrt(cxx1 * cyy1);
cxy2 = r2 * sqrt(cxx2 * cyy2);
t = (cyy1-cxy1^2/cxx1+cyy2-cxy2^2/cxx2);
a = -f/df * t * (1/cxx1+1/cxx2)+(b1-b2)^2;
a = a/2;
b = f/df * t * (mx1/cxx1 + mx2/cxx2) + (a1-a2) * (b1-b2);
c = -f/df * t * ((n1+n2)/n1/n2 + mx1^2/cxx1 + mx2^2/cxx2) + (a1-a2)^2;
c = c/2;
x = polyroot (a|b|c);
print "Below are the two solutions at which the difference between the";
lprint "Below are the two solutions at which the difference between the";
print "means of the two groups is exactly significantly different from zero.";
lprint "means of the two groups is exactly significantly different from zero.";
print "In the first column are values of X, the predictor,";
lprint "In the first column are values of X, the predictor,";
print "but only those solutions having zero in the second column are real.";
lprint "but only those solutions having zero in the second column are real.";
print x;
lprint x;
```

```
The size of group 1 is          8.0000000
The mean of group 1 on the predictor is       0.031250000
The standard deviation of group 1 on the predictor is        0.24486120
The standard deviation of group 1 on the criterion is        0.56109620
The intercept for group 1 is          0.96750000
The slope for group 1 is         1.4400000
The size of group 2 is          10.000000
The mean of group 2 on the predictor is       -0.19000000
The standard deviation of group 2 on the predictor is        0.27448860
The standard deviation of group 2 on the criterion is        0.14854970
The intercept for group 2 is          0.20630000
The slope for group 2 is      -0.088100000
The correlation for group 1 is         0.62841297
The correlation for group 2 is        -0.16279027
Below are the two solutions at which the difference between the
means of the two groups is exactly significantly different from zero.
In the first column are values of X, the predictor,
but only those solutions having zero in the second column are real.

      -5.0798438        0.00000000
     -0.21572336        0.00000000
```

```
proc fcv(ndf,ddf,alpha);
local f,fl,fh,ndf,ddf,alpha,a;
a=0;
fl=0;
fh=200;
f=1;
do until abs(a-alpha)<.000001;
   a = cdffc(f,ndf,ddf);
      if a < alpha;
        fh = f;
        f = (fl + f) / 2;
      else;
        fl = f;
        f = (fh + f) / 2;
      endif;
 endo;
retp(f);
endp;
print "What is the size of group 1?";
n1 = con(1,1);
lprint "The size of group 1 is " n1;
print "What is the mean of group 1 on the predictor?";
mx1 = con(1,1);
lprint "The mean of group 1 on the predictor is " mx1;
print "What is the mean of group 1 on the squared predictor?";
mz1 = con(1,1);
lprint "The mean of group 1 on the squared predictor is: " mz1;
print "What is the standard deviation of group 1 on the predictor?";
sdx1 = con(1,1);
vx1 = sdx1 ^ 2;
lprint "The standard deviation of group 1 on the predictor is: " sdx1;
print "What is the st. dev. of group 1 on the squared predictor?";
sdz1 = con(1,1);
vz1 = sdz1 ^ 2;
lprint "The st. dev. of group 1 on the squared predictor is: " sdz1;
print "What is the st. dev. of group 1 on the criterion?";
sdy1 = con(1,1);
vy1 = sdy1 ^ 2;
lprint "The st. dev. of group 1 on the criterion is: " sdy1;
print "What is the X-Y correlation for group 1?";
rxy1 = con(1,1);
lprint "The X-Y correlation for group 1 is: " rxy1;
print "What is the Xsquared-Y correlation for group 1?";
ryz1 = con(1,1);
lprint "The Xsquared-Y correlation for group 1 is: " ryz1;
print "What is the X-Xsquared correlation for group 1?";
rxz1 = Con(1,1);
lprint "The X-Xsquared correlation for group 1 is: " rxz1;
print "What is the intercept for group 1?";
a1 = con(1,1);
lprint "The intercept for group 1 is: " a1;
print "What is the slope on X for group 1?";
bx1 = con(1,1);
lprint "The slope on X for group 1 is: " bx1;
print "What is the slope on Xsquared for group 1?";
bz1 = con(1,1);
lprint "The slope on Xsquared for group 1 is: " bz1;
print "What is the size of group 2?";
n2 = con(1,1);
lprint "The size of group 2 is " n2;
```

```
print "What is the mean of group 2 on the predictor?";
mx2 = con(1,1);
lprint "The mean of group 2 on the predictor is: " mx2;
print "What is the mean of group 2 on the squared predictor?";
mz2 = con(1,1);
lprint "The mean of group 2 on the squared predictor is: " mz2;
print "What is the standard deviation of group 2 on the predictor?";
sdx2 = con(1,1);
vx2 = sdx2 ^ 2;
lprint "The standard deviation of group 2 on the predictor is: " sdx2;
print "What is the st. dev. of group 2 on the squared predictor?";
sdz2 = con(1,1);
vz2 = sdz2 ^ 2;
lprint "The st. dev. of group 2 on the squared predictor is: " sdz2;
print "What is the st. dev. of group 2 on the criterion?";
sdy2 = con(1,1);
vy2 = sdy2 ^ 2;
lprint "The st. dev. of group 2 on the criterion is: " sdy2;
print "What is the X-Y correlation for group 2?";
rxy2 = Con(1,1);
lprint "The X-Y correlation for group 2 is: " rxy2;
print "What is the Xsquared-Y correlation for group 2?";
ryz2 = con(1,1);
lprint "The Xsquared-Y correlation for group 2 is: " ryz2;
print "What is the X-Xsquared correlation for group 2?";
rxz2 = con(1,1);
lprint "The X-Xsquared correlation for group 2 is: " rxz2;
print "What is the intercept for group 2?";
a2 = con(1,1);
lprint "The intercept for group 2 is: " a2;
print "What is the slope on X for group 2?";
bx2 = con(1,1);
lprint "The slope on X for group 2 is: " bx2;
print "What is the slope on Xsquared for group 2?";
bz2 = con(1,1);
lprint "The slope on Xsquared for group 2 is: " bz2;
print "What is the lower limit of the domain of X?";
lx = con(1,1);
lprint "The lower limit of the domain of X is: " lx;
print "What is the upper limit of the domain of X?";
ux = con(1,1);
lprint "The upper limit of the domain of X is: " ux;
diff = (ux-lx)/10;
df = n1 + n2 - 6;
f = fcv(1,df,.05);
cyy1 = (n1-1) * vy1;
cyy2 = (n2-1) * vy2;
cxx1 = (n1-1) * vx1;
czz1 = (n1-1) * vz1;
cxx2 = (n2-1) * vx2;
czz2 = (n2-1) * vz2;
cxy1 = rxy1 * sqrt(cxx1 * cyy1);
cxy2 = rxy2 * sqrt(cxx2 * cyy2);
cyz1 = ryz1 * sqrt(czz1 * cyy1);
cyz2 = ryz2 * sqrt(czz2 * cyy2);
cxz1 = rxz1 * sqrt(cxx1 * czz1);
cxz2 = rxz2 * sqrt(cxx2 * czz2);
p = (cyy1-cxy1*bx1-cyz1*bz1) + (cyy2-cxy2*bx2-cyz2*bz2);
t = f*p/(n1+n2-6);
a = czz1/(cxx1*czz1-cxz1^2) + czz2/(cxx2*czz2-cxz2^2);
```

```
a = (bx1-bx2)^2 - t * a;
b = cxz1/(cxx1*czz1-cxz1^2) + cxz2/(cxx2*czz2-cxz2^2);
b = (bx1-bx2)*(bz1-bz2) + t * b;
c = cxx1/(cxx1*czz1-cxz1^2) + cxx2/(cxx2*czz2-cxz2^2);
c = (bz1-bz2)^2 - t * c;
e=(mx1*czz1-mz1*cxz1)/(cxx1*czz1-cxz1^2)+(mx2*czz2-mz2*cxz2)/(cxx2*czz2-cxz2^2);
e = (a1-a2)*(bx1-bx2) + t * e;
g=(mz1*cxx1-mx1*cxz1)/(cxx1*czz1-cxz1^2)+(mz2*cxx2-mx2*cxz2)/(cxx2*czz2-cxz2^2);
g = (a1-a2)*(bz1-bz2) + t * g;
h1a = cxx1*czz1/(cxx1*czz1-cxz1^2);
h2a = cxx2*czz2/(cxx2*czz2-cxz2^2);
h1b = mx1^2/cxx1-2*mx1*mz1*cxz1/cxx1/czz1+mz1^2/czz1;
h2b = mx2^2/cxx2-2*mx2*mz2*cxz2/cxx2/czz2+mz2^2/czz2;
h = (a1-a2)^2 - t * (h1a*h1b + h2a*h2b + (n1+n2)/n1/n2);
f1 = c;
f2 = 2 * b;
f3 = a + 2*g;
f4 = 2 * e;
f5 = h;
x = polyroot (f1|f2|f3|f4|f5);
print "Below are the four solutions at which the difference between the";
lprint "Below are the four solutions at which the difference between the";
print "means of the two groups is exactly significantly different from zero.";
lprint "means of the two groups is exactly significantly different from zero.";
print "In the first column are values of X, the predictor,";
lprint "In the first column are values of X, the predictor,";
print "but only those solutions having zero in the second column are real.";
lprint "but only those solutions having zero in the second column are real.";
print x;
lprint x;
x = lx-diff;
print "The critical value of F is " f " at 1 & " df " degrees of freedom";
lprint "The critical value of F is " f " at 1 & " df " degrees of freedom";
do while x < ux;
 x = x + diff;
 d = (a1-a2) + (bx1-bx2)*x + (bz1-bz2)*x^2;
 q1 = h1a*((x-mx1)^2/cxx1+(x^2-mz1)^2/czz1-2*cxz1*(x-mx1)*(x^2-mz1)/cxx1/czz1);
 q2 = h2a*((x-mx2)^2/cxx2+(x^2-mz2)^2/czz2-2*cxz2*(x-mx2)*(x^2-mz2)/cxx2/czz2);
 q = (n1+n2)/n1/n2 + q1 + q2;
 den = p * q / (n1 + n2 - 6);
 t2 = d^2/den;
 print "At x = " x " the value of F is " t2;
 lprint "At x = " x " the value of F is " t2;
 if t2 < f;
  print "The difference between the two groups is not significant.";
  lprint "The difference between the two groups is not significant.";
  else;
  print "The difference between the two groups is significant.";
  lprint "The difference between the two groups is significant.";
  endif;
 endo;
```

The size of group 1 is 383.00000
The mean of group 1 on the predictor is 2.4500000
The mean of group 1 on the squared predictor is: 7.3400000
The standard deviation of group 1 on the predictor is: 1.2083000
The st. dev. of group 1 on the squared predictor is: 6.3812000
The st. dev. of group 1 on the criterion is: 9.0277000
The X-Y correlation for group 1 is: 0.053170747
The Xsquared-Y correlation for group 1 is: 0.074295255
The X-Xsquared correlation for group 1 is: 0.96621986
The intercept for group 1 is: 55.900000
The slope on X for group 1 is: 2.1500000
The slope on Xsquared for group 1 is: -0.38000000
The size of group 2 is 409.00000
The mean of group 2 on the predictor is: 2.7100000
The mean of group 2 on the squared predictor is: 9.2200000
The standard deviation of group 2 on the predictor is: 1.3527700
The st. dev. of group 2 on the squared predictor is: 7.5953900
The st. dev. of group 2 on the criterion is: 9.9769700
The X-Y correlation for group 2 is: 0.15559475
The Xsquared-Y correlation for group 2 is: 0.062946167
The X-Xsquared correlation for group 2 is: 0.97227624
The intercept for group 2 is: 49.300000
The slope on X for group 2 is: 3.2400000
The slope on Xsquared for group 2 is: -0.38000000
The lower limit of the domain of X is: 0.00000000
The upper limit of the domain of X is: 5.0000000
Below are the four solutions at which the difference between the means of the two groups is exactly significantly different from zero.
In the first column are values of X, the predictor,
but only those solutions having zero in the second column are real.

| | |
|---|---|
| 3.4551557 | 1.8045932 |
| 3.4551557 | -1.8045932 |
| 4.0886800 | 0.00000000 |
| -0.47620005 | 0.00000000 |

The critical value of F is 3.8531229 at 1 & 786.00000 degrees of freedom
At x = 0.00000000 the value of F is 6.4847231
The difference between the two groups is significant.
At x = 0.50000000 the value of F is 12.214326
The difference between the two groups is significant.
At x = 1.0000000 the value of F is 23.092253
The difference between the two groups is significant.
At x = 1.5000000 the value of F is 31.537363
The difference between the two groups is significant.
At x = 2.0000000 the value of F is 25.703420
The difference between the two groups is significant.
At x = 2.5000000 the value of F is 17.870088
The difference between the two groups is significant.
At x = 3.0000000 the value of F is 13.120948
The difference between the two groups is significant.
At x = 3.5000000 the value of F is 9.3430146
The difference between the two groups is significant.
At x = 4.0000000 the value of F is 4.6310035
The difference between the two groups is significant.
At x = 4.5000000 the value of F is 1.3514992
The difference between the two groups is not significant.
At x = 5.0000000 the value of F is 0.28075711
The difference between the two groups is not significant.