MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 14, NUMBER 1, SPRING 1986

Longitudinal Analysis of Salary Discrimination in Higher Education

Robert L. Heiny, Samuel R. Houston, and John B. Cooney

University of Northern Colorado

Abstract

Legal and statistical issues associated with the use of multiple regression models in faculty discrimination cases in higher education are presented in this paper. Faculty salary models as a function of gender, rank, tenure status, race, academic discipline, and age variables are analyzed in a longitudinal study covering three years (1982-84) at the University of Northern Colorado (UNC). Declining student enrollment during the period saw the size of the faculty drop from a high of 492 in 1982 to a low of 380 in 1984. Results of the exploratory data analysis indicate declining roles for gender, race and age variables in explaining salary differences. While the contribution of academic discipline variables in the regression models was statistically significant, results seem consistent with institutional salary policies which were in effect at each point in time.

Paper presented at the American Educational Research Association, Chicago, April, 1985

Introduction

Given the increasing frequency of litigation on matters of discrimination with regard to salaries in higher education, the courts are faced with statistical evidence that support and refute claims of discrimination at an ever increasing magnitude and complexity. The claims of discrimination are made on the basis of race, gender and age factors.

Within the past ten years, multiple regression techniques have become popular in litigation on discrimination. Two recent articles support the use of multiple regression techniques in judicial studies of race and sex discrimination, (Finkelstein, 1980; Fisher, 1980). Both researchers identify several concerns which must be addressed.

Finkelstein discusses the problems associated with the use of "tainted" variables. Predictor variables specified to reflect productivity are often affected by discriminatory practice themselves. For example, when using the variables of tenure status and rank to predict salaries, discrimination might also be present in tenure and promotion decisions (Finkelstein, 1980), thus the inclusion of the "tainted" variables may serve to mask salary discrimination if it exists.

Fisher (1980) discusses the assumptions underlying multiple regression analysis and points out the problems associated with multicollinearity and the "shotgun" approach to analyze the data. Too often, the analysis is performed with an overprescription of independent variables in an attempt to discover what may be related to the criterion variable. When many variables are included, the risk of multicollinearity is increased. As a result, the magnitude and even the sign of the coefficients in the model may be affected. Fisher warns against the "shotgun" approach. He advises the experimenter to select carefully the variables to be used and develop a rationale for inclusion which can be defended.

Recently, studies have appeared which use other statistical techniques such as canonical correlation and multiple discriminant analysis, Carter, et al. (1983). Carter applies these techniques to analyze salary equity at the University of Wisconsin at Superior for two successive years, 1981-82 and 1982-83.

The two techniques used by Carter provide an alternative to address some of the concerns expressed by Finkelstein with regard to violation of assumptions in the multiple linear regression models. Specifically, the concern about "tainted" variables can be addressed by using canonical correlation and multiple discriminant analysis. These techniques assist the experimenter in determining whether or not the variables of tenure status and rank are affected by the variables of race, age or gender. If this analysis confirms the variables in question are not "tainted", then the multiple regression model can make use of the variables to improve the fit. If, however, the analysis reveals the variables are "tainted", the regression model will exclude those variables in the model. In addition, the very fact that the variables are discovered to be tainted is important information which may be used to resolve discriminatory practices.

All three statistical procedures, multiple regression, canonical correlation, and discriminant analysis, are used in this longitudinal study of salary practices at the University of Northern Colorado (UNC). Data on all full-time faculty members at UNC for the academic years 1982-83, 1983-84, and 1984-85 are analyzed to determine the existence of salary discrimination on the basis of race, age or sex. The items collected on each faculty member include: salary, rank, tenure status, highest degree, years employed at UNC, years in each rank, years at UNC before obtaining tenure, years with the doctorate, discipline, sex, race and age.

The longitudinal data allows for an analysis of changes in salary practices as they are affected by changes in University policies. This paper relates University policy changes which occurred during the three-year period to the changes in the existence and/or extent of discrimination in UNC salaries.

The paper is subdivided into four major sections: multiple regression analysis of salaries for the three years, canonical correlation on rank and tenure status versus qualification, experience and discrimination variables, multiple discriminant analysis to determine classifications and misclassifications with regard to rank and tenure status, and a contextual analysis which compares the UNC policy changes to the state of salary patterns at UNC during the three-year period.

Variables included in the statistical analyses of salary discrimination at UNC for the years 1982-83 through 1984-85 are presented in Table 1. Before proceeding with the statistical analyses several precautions were taken to insure the internal validity of the study. First, patterns of discrimination among the predictor variables themselves were examined using discriminant analysis and canonical correlation techniques. That is to say, relationships between university status variables (e.g., tenure status, rank, rate of promotion) and the discrimination variables were carefully examined before they were included in the regression models as predictor variables. If university status variables are tainted they should be removed. Second, collinearity diagnostics were obtained on the predictor variables. Although our primary interest is in the use of R^2 values, interpretation of the regression coefficients themselves is also of interest. It can be shown that the presence of collinearity can affect both the sign and magnitude of the regression coefficients (Pedhazur, 1982). Detection of collinearity among the predictor variables would require us to re-think the specification of our model.

Inspection of the collinearity diagnostics from the regression procedure of the Statistical Analysis System (1982) indicated that the variables Longevity and Years with Doctorate were the primary sources of collinearity. Inasmuch as these variables were selected to contribute unique information to the model, the preliminary analyses indicate that these variables were already adequately represented by other predictors. Our solution to the problem was to delete Longevity and Years with the Doctorate from the set of predictor variables.

In the sections that follow, results form the canonical correlation and discriminant analyses designed to detect patterns of discrimination among the set of predictor variables are reported.

Canonical Correlation Analysis

In an attempt to ferret out potential patterns of discrimination during the past three academic years at UNC, canonical correlational analytic methods were undertaken. Canonical Analysis (CA) is a method designed to study the relations between two sets of variables, a set of predictor variables and a set of criterion variables. The set of independent or predictor variables (PV) identified in this study consisted of all the discrimination variables which included gender, race, and age. On the other hand, the set of dependent or criterion variaoles (CV) could be classified as university status

Table 1

Variables Included in the Analysis of Salary Discrimination

Variable	Description
	Rank
V1	Assistant Professor
V2	Associate Professor
V2	Professor
	Longevity
V4	Years of Service
	Degree
V5	Master's
V6	Doctorate
	Tenure Status
V7	Yes=1, No=0
	Gender
V8	Male=1, Female=0
	Race
V9	Caucasian=1, Otherwise=0
V11	Black=1, Otherwise=0
V12	Hispanic=1, Otherwise=0
	Else, Oriental, or Indian
	Time in Rank
V14	Years as Instructor
V15	Years as Assistant Professor
V16	Years as Associate Professor
V17	Years as Professor
	Time Since Receiving Doctorate
V18	Years with the Doctorate
	Time Before Receiving Tenure
V19	Years before Receiving Tenure
	Discipline
V20	School of Business=1, Otherwise=0
V21	Physical Sciences=1, Otherwise=0
V22	Social Sciences=1, Otherwise=0
V23	Humanities=1, Otherwise=0
V24	College of Performing & Visual Arts=1, Otherwise=0
V25	College of Health and Human Services=1, Otherwise=0
	Else, College of Education .
V29	Age
V30	Salary

variables. These variables included tenure, academic rank, degree earned, years spent at each level, and school or college in which the faculty member was assigned. The set of discrimination or predictor variables numbered six whereas there were 17 university status or criterion variables. Thus, the maximum number of linear combinations or composites of predictor variables and criterion variables which could be tested for a significant correlation is six.

Each of the possible six canonical correlations (Canonical R) for each of the three academic year studied at UNC was tested for statistical significance by converting Wilks' Lambda to an approximate F. In Table 2 are presented the standardized weights for the set of predictors and set of criteria associated with the three significant canonical R-values using N = 492 observations of the 1982-83 study group. All three canonical R-values are significant beyond the 0.001 level and the three canonical R-values in descending order are .76, .42, and .38. The remaining three non-significant canonical R-values and corresponding standardized weights are not reported.

The results for the 1983-84 study are presented in Table 3. It should be observed that only two of the canonical R-values were statistically significant for N = 446 observations used in the

le 2

onical Solution Using Standardized Weights for Significant Relationships for N = 492 ervations (1982-83)

dictor iables	Standardized Predictor Weights			Criterion Variables	Standardized Criterion Weights		
	PV1	PV2	PV3		CV1	CV2	CV3
der	.32	-.19	.91	Tenure	.02	-.50	-.04
casian	.16	1.67	.43	Asst. Prof.	.29	-.10	.36
ck	.00	.59	.29	Assoc. Prof.	.45	-.20	.46
panic	.02	1.10	.54	Professor	.54	-.18	.54
ental	.03	.46	.04	Masters	-.16	3.53	-.08
	.88	-.07	-.43	Doctorate	-.18	3.69	.21
				Yrs. Instr.	.02	.03	-.31
				Yrs. Asst. Prof.	.27	.38	-.15
				Yrs. Assoc. Prof.	.36	.25	.02
				Yrs. Prof.	.62	-.00	-.43
				Business	.18	-.07	-.16
				Phys. Sci.	.06	.02	-.04
				Soc. Sci.	.09	.04	-.01
				Humanities	.10	.18	-.49
				PVA	.10	-.05	-.05
				HHS	.06	.02	-.78
				Education	.18	-.07	-.58
				Canonical R	.76*	.42*	.38***

*Wilks' Lambda Significant at 0.001 when converted to an approximate F.
*Wilks' Lambda Significant at 0.001 when converted to an approximate F.
*Wilks' Lambda Significant at 0.001 when converted to an approximate F.

analysis. As is the case with Table 2 the standardized weights associated with the set of predictors and set of criteria are presented. The two significant canonical R-values are .77 and .43. Both are significant at 0.001 level.

In Table 4 results of the canonical analysis for the 1984-85 study are described for N = 380 observations. The decline in the number of observations over the three-year period is a function of declining enrollment at UNC. The first two canonical R-values (.73 and .40) are statistically significant at the 0.001 level and the corresponding standardized weights for the set of predictors and criteria are reported. The standardized weights and canonical R-values for the four non-significant relationships in 1984-85 are not presented.

Standardized canonical weights are often interpreted in a manner analogous to the interpretation of standardized regression weights in multiple linear regression. It is not surprising, therefore, to see some researchers use them as indices of the relative contribution or importance of the variables with which they are associated. Because of the multicollinearity associated

le 3

onical Solution Using Standardized Weights for Significant Relationships for N = 446 ervations (1983-84)

dictor iables	Standardized Predictor Weights		Criterion Variables	Standardized Criterion Weights	
	PV1	PV2		CV1	CV2
der	.30	.94	Tenure	-.05	.23
casian	.19	.15	Asst. Prof.	.36	.96
ack	-.02	.32	Assoc. Prof.	.65	.98
spanic	.07	.41	Professor	.82	1.03
iental	.09	-.05	Masters	-.04	-.21
e	.90	-.37	Doctorate	-.12	-.05
			Yrs. Instr.	.04	-.37
			Yrs. Asst. Prof.	.27	-.40
			Yrs. Assoc. Prof.	.28	.00
			Yrs. Prof.	.64	-.43
			Business	.18	-.26
			Phys. Sci.	.03	-.05
			Soc. Sci.	.11	-.07
			Humanities	.11	-.39
			PVA	.09	-.12
			HHS	.08	-.71
			Education	.15	-.50
			Canonical R	.77*	.43**

*Wilks' Lambda Significant at 0.001 when converted to an approximate F.
**Wilks' Lambda Significant at 0.001 when converted to an approximate F.

Table 4

Canonical Solution Using Standardized Weights for Significant Relationships for N = 38
Observations (1984-85)

Predictor Variables	Standardized Predictor Weights		Criterion Variables	Standardized Criterion Weights	
	PV1	PV2		CV1	CV2
Gender	.25	.94	Tenure	-.06.	.46
Caucasian	.18	-.08	Asst. Prof.	.39	.55
Black	-.06	.15	Assoc. Prof.	.68	.30
Hispanic	.05	.27	Professor	.73	.49
Oriental	.07	-.16	Masters	-.21	-.72
Age	.92	-.28	Doctorate	-.30	-.51
			Yrs. Instr.	.11	-.50
			Yrs. Asst. Prof.	.26	-.42
			Yrs. Assoc. Prof.	.36	.03
			Yrs. Prof.	.80	-.53
			Business	.09	-.12
			Phys. Sci.	.03	.04
			Soc. Sci.	.08	.05
			Humanities	.04	-.29
			PVA	.03	.02
			HHS	-.02	-.54
			Education	.10	-.38
			Canonical R	.73*	.40**

*Wilks' Lambda Significant at 0.001 when converted to an approximate F.
**Wilks' Lambda Significant at 0.001 when converted to an approximate F.

with the set of predictors as well as the set of criteria, the standardized canonical weights suffer from the same shortcomings as those of standardized regression coefficients. Not only the signs but the magnitude of the weights can be misleading. These limitations appeared with the results presented in Tables 2, 3, and 4. For these reasons, the investigators used structure coefficients for the purpose of interpreting and explaining the results of CA. For a further discussion of this point, see Cooley & Lohnes (1976); Thorndike & Weiss (1973).

In Tables 5, 6, and 7 are presented the corresponding structure coefficients or loadings associated with the significant canonical correlations found in the three-year study at UNC. A structure coefficient or loading in CA is the correlation of a specific variaole and a canonical variate. For example, in Table 5, we see that the age variable correlates .94 with the first predictor variate (PV1). In other words, the square of .94 changed to a percent indicates that 88.36% of the variance in the linear composite of the predicator variables (discrimination variables) can be explained by the aqe variable.

A rule of thumb is suggested by Pedhazur (1982) that structure coefficients $\geq$ .30 be considered as meaningful or useful in explaining significant canonical correlations. In Table

Table 5

Structure Loadings for Significant Canonical Correlations for N = 492 Observations (

Predictor	Structure Loadings			Criterion	Structure Loa	
~~Variables~~	Predictor Variables			~~Variables~~	~~Criterion~~ Var	
	PV1	PV2	PV3		CV1	CV2
Gender	.44	-.30	.81	Tenure	.65	-.10
Caucasian	.20	.59	-.11	Asst. Prof.	-.52	.08
Black	-.03	.03	.08	Assoc. Prof.	-.12	.04
Hispanic	-.12	.05	.28	Professor	.71	-.07
Oriental	-.03	-.30	-.08	Masters	-.31	.10
Age	.94	.01	-.28	Doctorate	.33	.11
				Yrs. Instr.	-.08	-.08
				Yrs. Asst. Prof.	.28	.23
				Yrs. Assoc. Prof.	.69	.09
				Yrs. Prof.	.78	-.07
				Business	-.11	-.09
				Phys. Sci.	.16	.05
				Soc. Sci.	-.00	.07
				Humanities	-.00	.07
				PVA	-.00	-.11
				HHS	-.21	.08
				Education	.07	-.05

le 6

ucture Loadings for Significant Canonical Correlations for N = 446 Observations (1983-84)

dictor iables	Structure Loadings Predictor Variables PV1	 PV2	Criterion Variables	Structure Loadings Criterion Variables CV1	 CV2
der	.41	.84	Tenure	.53	.00
casian	.17	-.18	Asst. Prof.	-.43	-.00
ck	-.05	.10	Assoc. Prof.	-.09	.05
panic	-.12	.26	Professor	.57	.02
ental	-.00	-.06	Masters	-.26	-.15
	.93	-.25	Doctorate	.27	.15
			Yrs. Instr.	-.05	-.19
			Yrs. Asst. Prof.	.22	-.11
			Yrs. Assoc. Prof.	.46	.04
			Yrs. Prof.	.61	-.00
			Business	-.13	.00
			Phys. Sci.	.11	.09
			Soc. Sci.	.04	.13
			Humanities	-.01	-.06
			PVA	-.01	.08
			HHS	-.10	-.25
			Education	.05	-.05

Table 7

Structure Loadings for Significant Canonical Correlations for N = 380 Observations (19

Predictor Variables	Structure Loadings Predictor Variables		Criterion Variables	Structure Loadir Criterion Variat	
	PV1	PV2		CV1	C
Gender	.35	.87	Tenure	.64	
Caucasian	.20	-.24	Asst. Prof.	-.59	-
Black	-.06	.06	Assoc. Prof.	-.14	-
Hispanic	-.14	.29	Professor	.69	
Oriental	-.02	-.07	Masters	-.19	-
Age	.94	-.19	Doctorate	.19	
			Yrs. Instr.	-.06	-
			Yrs. Asst. Prof.	.13	-
			Yrs. Assoc. Prof.	.41	
			Yrs. Prof.	.59	
			Business	-.14	
			Phys. Sci.	.14	
			Soc. Sci.	.04	
			Humanities	-.01	-
			PVA	-.05	
			HHS	-.14	-
			Education	.05	-

le 8

ful Structure Coefficients (Loadings) in Explaining Relationships between Significantly

related Canonical Variates[1]

Discrimination Variables	1982-83 (N = 492) PV1	 PV2	 PV3	1983-84 (N = 446) PV1	 PV2	1984 (N = 380) PV1	 PV2
nder	*+	*-	*+	*+	*+	*+	*+
ucasian		*+					
ack							
spanic							
iental		*-					
ge	*+			*+		*+	
University Status Variables	1982-83 (N = 492) CV1	 CV2	 CV3	1983-84 (N = 446) CV1	 CV2	1984 (N = 380) CV1	 CV2
enure	*+			*+		*+	
sst. Prof.	*-			*-		*-	
ssoc. Prof.							
rofessor	*+			*+		*+	
asters	*-		*-				
octorate	*+		*+				
rs. Instr.			*-				
rs. Asst. Prof.							
rs. Assoc. Prof.	*+			*+		*+	
rs. Prof.	*+			*+		*+	
usiness							
hys. Sci							
oc. Sci.			*+				
umanities							
/A							
HS			*-				
ducation							
anonical R	.76	.42	.38	.77	.43	.73	.40

[1] Structure coefficients $\geq$ .30 were considered as meaningful (Pehhazur's criterion). A "*+" represents a positive coefficient $\geq$ .30 and a "*-" refers to a negative structure loading $\geq$ .30.

8 the structure coefficients which are $\geq$ .30 are starred as positive or negative depending on the sign of the structure coefficient. The purpose of this table is to present the results for the three consecutive years at UNC in such a way that the significant canonical R-values might be interpreted in terms of the set of predictors and the set of criteria.

In reviewing the starred variables in Table 8 it can be seen that the linear combination of predictor variables in the first canonical R for each of the three years has a positive structure loading on gender and age. Thus, PV1 might be conceptualized as a factor representing older males. If we focus on the corresponding set of university status variables (CV1) for the three years we see positive loadings on tenure, professor, years associate professor, years full professor and a negative loading on assistant professor. For 1982-83 only we see a negative loading on masters and a positive loading on doctorate. The loadings on the criterion variate for all three years suggest that CV1 reflects the factor of an experienced professional--one with tenure, higher academic rank, and more experience at the associate or full professor level. It is interesting to note that degree status (criterion set) seems unrelated to age and gender (predictor set) in the last two years of study. As one investigates the pattern that relates the predictor variate with the criterion variate in the second canonical R and in the case of 1982-83 the third canonical R, the picture becomes less clear. In 1982-83, the positive and negative loadings in PV2 suggest a factor of female Caucasian in the predictor variate whereas no significant loading was detected in the criterion variable set (CV1). From a discrimination claims point of view this might be interpreted as a positive finding. The discrimination factor in PV2 (female Caucasian) seems related to university status factor variables in no systematic way. Similarly, the PV2 seems to be a gender factor for both 1983-84 and 1984-85 but is unrelated to any university status variable in CV2 for both years. In 1982-83 a third significant canonical R was found. PV3 in this year seems to reflect a gender factor and this factor seems to show that males tended to have the doctorate, were not instructors, were social science faculty and not HHS faculty members. This gender university status pattern for 1982-83 did not show up in subsequent analyses for both 1983-84 and 1984-85 and should be considered another positive finding from a discrimination claims point of view. Finally, it should be observed that race as a discrimination variable did not exhibit a high loading in each of the three years. Race seems unrelated to the linear composite of university-status variables.

In Table 9 are presented the percent of the variance in the linear composite of the university-status variables (criterion

Table 9

Percent of Variance in Set of University Status Variables Linear Composite Explained by Discrimination Variables[1]

Discrimination	1982-83 (N = 492)			1983-84 (N = 446)		1984-198	
Variables	CV1	CV2	CV3	CV1	CV2	CV1	C
Gender	11.36	13.10	22.83	10.41	23.55	6.74	19
Caucasian	2.39	8.95	9.13	1.79	2.45	2.22	3
Black	0.06	0.08	0.18	0.19	0.38	0.25	C
Hispanic	0.95	1.01	2.17	0.97	2.25	1.17	2
Oriental	0.09	1.78	1.88	0.00	0.08	0.04	C
Age	51.74	51.75	52.94	52.45	53.65	48.66	49
Canonical R	.76	.42	.38	.77	.43	.73	

[1] Only criterion variable linear composites are presented which are associated with canonical R-values which are significant beyond the 0.001 level.

variate) that can be explained by each of the six discrimination (predictor) variables for the significant canonical R-values found. Results in this table seem to confirm that age was the dominant variable over the three years--it explained about 50% of the variance in each of the criterion variates. Gender appeared to be a much less significant factor as the percent of variance for each criterion variate explained ranged from about 7% to a high of 24%. Race as a factor was not significant as the percent of variance of the criterion variate it was able to explain ranged from a low of 0% to a high of 9%.

In summary, the results of CA seem positive from the issue of discrimination claims in higher education. While the older-male relationship with the professional-experience factor was detected in the three-year analysis, the relationship has historical roots and is less pronounced today. No other gender or race factors were found to be linked in any systematic way to any university-status factors.

Discriminant Analysis

To investigate further the possibility of discrimination patterns in tenure and promotion decisions, a statistical technique known as discriminant analysis (DA) was applied to data for the academic years 1982-83, 1983-84 and 1984-85. The DA method analyzes one variable such as tenure status by comparing it with a group of variables called independent variables or predictors. Since the tenure status variable is a binary variable, DA determines a set of weights which maximizes the criterion for group membership, called the discriminant function. This function serves as the basis for attempts to "classify" each faculty member into one of the two original groupings, tenured or nontenured. Two linear combinations of the independent variables are formed to "characterize" group membership.

After the linear combinations are determined, the values of the predictors for each individual are used to calculate discriminant scores which will indicate which of the two groups the individual's profile most closely resemble. This measure is given by posterior probabilities of group membership. After the analysis is completed for all individuals, those observations which are misclassified can be analyzed for inequities or other irregularities.

For the three academic years of interest, DA was conducted using the five variables tenure status, professor rank, associate professor, assistant professor, and instructor as the criterion variables individually. The predictors were age, gender, race, highest degree, years in rank, and discipline. Tables 10 through 12 present the linear discriminant function for each criterion variable and the resulting classifications and misclassifications for the three years.

Table 10

Discriminant Function (1982-83)

	Criteria Variable									
	Tenure		Professor		Associate		Assistant		Instructor	
Predictors	0	1	0	1	0	1	0	1	0	1
Constant	-311.99	-320.77	-313.52	-320.55	-313.56	-313.96	-315.04	-313.09	-320.36	-304.05
Doctorate	62.64	69.01	62.52	63.93	62.25	64.99	63.50	61.21	68.87	58.13
Gender	9.41	10.06	9.43	9.82	9.37	9.55	9.42	9.33	10.50	8.66
Caucasian	33.16	29.49	33.37	33.44	33.38	32.22	32.44	34.19	33.51	33.25
Black	30.93	26.91	31.12	30.98	31.16	30.13	30.12	32.07	31.72	30.77
Hispanic	32.07	27.86	32.31	32.43	32.31	31.46	31.33	33.17	33.71	31.39
Oriental	37.38	36.95	37.63	39.16	37.16	35.74	35.56	38.17	38.38	38.78
Yrs.Instr.	.40	.85	.39	.46	.37	.45	.44	.31	.48	.31
Master	-.33	.32	62.83	61.67	62.96	65.53	63.44	62.60	67.18	60.35
Yrs.Asst.	-1.26	-.63	-.38	-.44	-.37	-.31	-.44	-.30	-.02	-.59
Yrs. Assoc.	-.46	-.24	-1.24	-.90	-1.30	-1.00	-.91	-1.65	-1.24	-1.34
Yrs. Prof	494.66	493.76	-.37	.32	-.46	-.83	-.42	-.52	-.52	-.44
Business	491.53	492.14	494.90	496.21	494.72	494.02	494.64	494.76	495.27	494.35
Phys. Sci.	493.54	494.35	491.82	493.91	491.52	490.54	491.82	491.21	491.13	491.74
Soc.Sci.	496.05	497.16	493.49	493.45	493.50	493.23	493.21	493.76	493.72	493.36
Humanities	63.23	67.62	496.07	496.58	495.99	495.93	496.02	495.96	496.65	495.58
PVA	493.92	494.91	494.09	495.57	493.87	493.39	493.91	493.82	494.99	493.15
HHS	496.53	496.58	496.74	498.09	496.56	494.93	495.99	497.83	496.17	496.76
Education	492.88	493.01	493.03	494.02	492.89	492.85	492.85	493.09	493.10	492.73
Age	.88	.93	.89	.92	.88	.90	.89	.87	.96	.83

Discriminant Function (1983-84)

Criterion Variable

Predictors	Tenure		Professor		Associate		Assistant		Instructor	
	0	1	0	1	0	1	0	1	0	1
Constant	-428.52	-436.40	-431.07	-443.23	-430.27	-435.56	-436.00	-428.38	-439.13	-418.89
Doctorate	227.82	225.97	228.46	230.37	228.24	230.51	227.24	227.64	240.55	221.68
Gender	5.67	6.57	5.96	6.44	5.81	5.99	5.32	5.76	8.44	4.52
Caucasian	153.16	155.11	153.97	155.41	153.99	156.46	159.22	153.37	150.98	154.44
Black	149.88	151.09	149.40	147.99	150.81	153.92	153.29	150.02	150.04	149.99
Hispanic	152.39	155.77	153.20	154.33	153.44	156.23	157.96	152.74	152.27	152.93
Oriental	147.91	153.57	149.91	153.29	149.44	153.19	158.08	148.58	145.66	149.75
Yrs.Instr.	-.31	.52	-.24	-.27	-.18	0.00	.04	-.23	-.28	-.20
Master	226.30	222.13	225.68	225.17	226.28	227.77	223.55	225.88	235.62	221.40
Yrs.Asst.	-.18	.61	-.10	-.12	-.09	-.07	-.27	-.10	.37	-.32
Yrs. Assoc.	-1.54	-.67	-1.35	-1.10	-1.33	-.86	-.56	-1.45	-1.42	-1.47
Yrs. Prof	-.54	-.20	-.12	.77	-.63	-1.09	-.36	-.51	-.60	-.47
Business	451.57	451.05	452.55	454.96	451.15	449.72	451.52	451.52	451.61	451.48
Phys. Sci.	445.88	446.83	447.28	450.34	445.60	444.17	446.09	445.98	445.23	446.32
Soc.Sci.	445.90	447.34	446.73	448.34	445.74	444.62	445.72	446.04	446.25	445.94
Humanities	454.85	454.89	455.68	457.61	454.50	453.12	453.79	454.85	456.66	454.03
PVA	449.49	451.57	450.82	453.45	449.58	449.14	450.39	449.69	452.01	448.62
HHS	453.42	453.78	454.68	457.53	452.82	450.36	452.44	453.45	453.33	453.52
Education	445.71	446.51	446.83	449.26	445.44	444.88	445.69	445.79	446.39	445.51
Age	.82	.85	.84	.90	.83	.85	.86	.82	.92	.78

Table 12

Discriminant Function (1984-85)

	Criterion Variable									
	Tenure		Professor		Associate		Assistant		Instructor	
Predictors	0	1	0	1	0	1	0	1	0	1
Constant	-366.12	-372.00	-368.21	-379.68	-367.77	-373.13	-379.33	-366.58	-369.82	-360.54
Doctorate	404.68	399.18	404.08	406.22	403.92	406.66	410.25	405.02	404.73	398.67
Gender	10.21	11.38	10.74	11.49	10.46	10.34	10.69	10.53	10.99	8.75
Caucasian	195.52	200.25	197.14	198.60	197.17	199.71	204.31	198.50	195.97	198.98
Black	190.90	196.38	193.25	196.32	192.43	193.51	198.55	193.74	191.95	193.10
Hispanic	195.23	201.73	197.11	198.10	197.32	199.96	203.63	198.43	196.54	197.63
Oriental	186.68	194.59	189.58	192.57	189.41	193.49	201.28	191.63	187.92	190.84
Yrs.Instr.	-1.07	-.28	-.89	-.91	-.85	-.68	-.39	-.76	-1.05	-.29
Master	402.94	395.57	401.16	401.13	401.55	403.40	404.57	401.99	401.48	400.11
Yrs.Asst.	.31	.98	.46	.45	.46	.41	.18	.40	.65	-.15
Yrs. Assoc.	-1.64	-.88	-1.39	-1.18	-1.36	-.89	-.27	-1.17	-1.43	-1.54
Yrs. Prof	-.87	-.48	-.44	.55	-.88	-1.39	-.50	-.71	-.82	-.62
Business	94.33	93.57	95.16	98.14	93.66	91.27	93.26	93.93	93.99	94.71
Phys. Sci.	93.58	94.54	95.32	99.78	93.26	90.57	93.66	93.76	94.22	92.40
Soc.Sci.	93.83	95.04	94.98	97.53	93.71	91.69	92.92	93.83	94.47	92.92
Humanities	111.50	111.47	112.44	115.24	111.05	108.86	109.61	111.04	112.52	107.98
PVA	100.22	102.36	101.87	105.22	100.42	98.91	100.69	100.72	101.98	96.45
HHS	101.81	102.10	103.06	106.58	101.11	97.37	98.96	101.17	101.67	102.60
Education	94.34	95.16	95.63	98.86	94.05	91.69	93.32	94.24	95.07	92.69
Age	.88	.88	.88	.88	.89	.94	.93	.90	.93	.74

Table 13

Classifications

82-83

TENURE

TO

		0	1	T
	0	100	6	106
FROM	1	40	346	386
	T	140	352	492

PROF

TO

		0	1	T
	0	269	8	277
FROM	1	22	193	215
	T	291	201	492

ASSOC

TO

		0	1	T
	0	254	84	338
FROM	1	25	129	154
	T	279	213	492

ASST

TO

		0	1	T
	0	314	74	388
FROM	1	3	101	104
	T	317	175	492

INSTR

TO

		0	1	T
	0	426	47	473
FROM	1	1	18	19
	T	427	65	492

83-84

TENURE

TO

	0	1	T
0	95	3	98
1	21	327	348
T	116	330	446

PROF

TO

0	1		T
0	252	1	253
1	19	174	193
T	271	175	446

ASSOC

TO

	0	1	T
0	262	42	304
1	21	121	142
T	283	163	446

ASST

TO

	0	1	T
0	295	50	351
1	1	94	95
T	296	150	446

INSTR

TO

	0	1	T
0	398	32	430
1	0	16	16
T	398	48	446

84-85

TENURE

TO

	0	1	T
0	70	2	7
1	26	282	30
T	96	284	38

PROF

TO

	0	1	
0	208	0	2
1	19	153	1
T	227	153	

ASSOC

TO

	0	1
0	228	28
1	12	112
T	240	140

ASST

TO

	0	1
0	273	29
1	1	77
T	274	106

INSTR

TO

	0	1
0	356	18
1	0	6
T	356	24

0 - indicates individual does not belong to class
1 - indicates individual does belong to class

FROM - is ACTUAL STATUS
TO - is PREDICTED STATUS

Table 14

R^2 Values for Full and Restricted Models for 1982-83 through 1984-85

	Academic Year		
Model	1982-83	1983-84	1984-85
Full Model (FM)	.8630 A	.8691 A	.9006 A
FM - Discrimination Set	.8510 8	.8616 8	.8990 A
FM - Gender	.8580 8	.8651 8	.8995 A
FM - Race	.8626 A	.8680 A	.9002 A
FM - Age	.8480 B	.8659 8	.9005 A

Note: R^2 values in a column with the same letter as the full model are not significantly different from each other. All P's$\leq$.01.

As seen in the above tables, the number of misclassifications in all five analyses decrease from 1982-83 to 1984-85. Several policy changes within the institution provide possible explanations for this pattern. These relationships will be discussed in the section entitled Contextual Analysis.

Upon examination of the individual cases identified by DA as misclassified, the majority were explained by rational, nondiscriminatory factors or by historical factors due to evolving standards at UNC. For example, in the year 1984-85, UNC has 72 faculty members who are not tenured. The DA method indicates two of these individuals possess values for the predictors which more closely resemble the individuals who are tenured.

The first faculty member is a male who has a special seven year agreement with the Board of Trustees in lieu of tenure. The second faculty member is a male who is hired annually on state grant money through the Colorado State Vocational Education Program. Even though he has excellent credentials, he is on soft money and is therefore not tenured.

The majority of the 26 faculty members who are tenured but more closely resemble the nontenured group are faculty members who do not possess the doctorate. These faculty members were tenured in the period from 1965-1975 when the availability of qualified faculty and the standards for obtaining tenure were quite different from the period since 1975.

Similar analyses were performed for the misclassifications for each rank. Few individual cases were identified which required further attention. In no instance was there any pattern of cases which would indicate systemic discrimination by the University on the basis of gender, age or race.

For 1984-85, the ranks of associate professor and assistant professor had a number of misclassifications from 0 to 1 (See Table 13). Upon further study, most of the misclassifications of this nature were situations in which a faculty member possessed a higher rank than the DA method predicted for the individual. The DA method consistently misclassified such individuals in all ranks for each year. These individuals had been promoted prior to 1976 when standards for promotion began to change at the institution.

This technique is an excellent tool for identifying general patterns as well as individual faculty members who may have been treated differentially. Certainly this method cannot be treated in isolation; however, it provides additional information to the institution in an attempt to correct whatever inequities which may exist. Both the canonical correlation and discriminant analyses show the variables of tenure status and rank are not "tainted" with respect to the discrimintation variables. Therefore, the variables of tenure status and rank may be used in the multiple regression analysis of salaries to improve the overall predictive efficiency.

Multiple Regression Analysis

Multiple regression (MR) analyses were performed to examine the relationship between salaries of full-time faculty and a set of discrimination variables, i.e., gender, race and age, for the years 1982-83 through 1984-85. Predictor vectors were coded for the MR analyses to reflect an individual's gender, race, age, qualifications, academic discipline, rank, tenure status, years spent in each rank and years before receiving tenure. Justification for including variables related to a faculty member's status within the institution was provided by the results of the canonical correlation analysis. Recall that there was no relationship between the academic status variables and the discrimination variables of gender and race. That is to say, no evidence was found that rank, tenure status, time in rank and time before receiving tenure were the result of discriminatory practices.

For each of the three years under consideration, salaries were regressed on the variables listed in Table 1 (the full model). Subsequently, salaries were regressed on a model containing all of the variables in the full model except for the set of discrimination variables: gender, race and age (the restricted model). Differences in R^2 values for the full and restricted models were tested by means of the F-distribution (Pedhazur, 1982). If the set of discrimination variables was found to account for a significant proportion of variance in salaries, the variables were examined one at a time to identify the specific source(s) of discrimination. Diagnostics were also performed to determine if the collinearity assumption had been violated. R^2 values of the full and restricted models for each of the three years are presented in Table 14.

Results of the MR analyses for the 1982-83 year show that the full model accounted for 86% of the variance in faculty salaries, F(22,469) = 133.93, $p<.01$, while the restricted model accounted for 85% of the variability in salaries, F(17,474) = 159.42, $p<.01$. Although the difference in R^2 values for the two models was small, it was statistically significant, F(5,469) = 8.21, $p<.01$. Further analyses of the 1982-83 data found that gender, F(1,469) = 17.11, $p<.01$ and age, F(1,469) = 51.35, $p<.01$, accounted for a significant proportion of the variance in faculty salaries. There was a tendency for males to earn higher salaries than females and the relationship between age and salary was found to be positive. No evidence of discrimination on the basis of race was detected by the analysis, F(3,469)<1.

A pattern similar to that found in 1982-83 emerged from the 1983-84 salary data. The squared multiple correlation coefficient for the full model was .87, F(22,423) = 127.61, $p<.01$, while the R^2 value of the restricted model was .86, F(17,428) = 156.79, $p<.01$. Again deleting the set of discrimination variables from the full model produced a statistically significant decrease in R^2, F(5,423) = 4.83, $p<.01$. Subsequent analyses show once again that gender and age accounted for a significant proportion of the variance in salaries, $F(1,423) = 12.90$, $p < 01$; $F(1,423) = 10.32$, $p < .01$, respectively. The increment in the proportion of variance in salaries attributable to race was not significant, $F(3,423) = 1.18$.

Implementation of the new University salary model for 1984-85 virtually eliminated discrimination in salaries on the basis of gender, race or age. For the full model $R^2 = .90$ while the restricted model resulted in an $R^2 = .89$. The difference in R^2 values for the full and restricted models was not statistically significant, $F(5,357) = 1.14$, $p < 05$.

In summary, evidence was found that males earned higher salaries than females from 1982-83 to 1983-84; however, the difference between male and female salaries was eliminated after the implementation of a new salary model. There was also a tendency for older faculty members to earn higher salaries than younger faculty members during the same period. Similarly, the relationship between age and salary was eliminated in 1984-85. There was no evidence of salary discrimination on the basis of race during any of the three years under consideration.

Contextual Analysis

Before discussing the results, a brief history of UNC is required in order to understand the context within which the results occurred. UNC is a former normal school which was founded in 1889. The institution evolved from the normal school to a teacher's college (1935), to a state college (1957), to a university (1970) as have many other similar institutions in the country. However, UNC differed in one significant aspect. During the 1920-1940 period, UNC embarked on a unique path of offering many graduate programs particularly at the doctoral level. Instead of developing the programs from a solid base of bachelor degree programs to a broadly based masters degree program to the doctoral level, UNC jumped immediately to the doctoral level. This lack of breadth eventually caused serious problems of enrollment and quality of doctoral work in the late 1970's.

To further compound problems, the institution engaged in the practice of hiring its own graduates, particularly in the late 1950's and 1960's. These faculty members were tenured and promoted rapidly under standards which were less rigorous than those that exist now at UNC. Tenure was nearly automatic after three years of service and promotions were granted every four years. Thus a faculty member would normally become a tenured full professor after eight years of service. Many did not possess the credentials which would justify a similar rank or status at another institution of higher education. Thus the faculty member was "trapped" at UNC unless the faculty member was willing to take a lower rank at a different institution. All these factors resulted in an older faculty that was not mobile in the market place.

In addition, enrollment began to decline in 1977 and with one exception continued to decline in the 1980's. The institution's enrollment has fallen from a peak of 11,770 in 1977-78 to 8,800 in 1984-85.

All these factors have led to numerous policy changes which are important to place the analysis in context. Prior to 1982-83, tenure and promotion decisions were made by a process which called for departmental recommendations to be passed to the council of deans who made a strong recommendation rarely overturned by the vice president or president. Little was known of the criteria or method of decision used by deans. Beginning in 1982-83, the council of deans was replaced by a committee of faculty members and the criteria for tenure and promotion were more stringent and clearly defined. This change was the final step in a movement towards higher tenure and promotion standards initiated in the late 1970's.

As a result, obtaining tenure and/or promotion is considerably more difficult now than at any time before. In fact there are numerous instances in which faculty members possess a rank for which they would no longer be qualified under the new policies. These tougher standards which have been used for faculty members hired since 1976 cause numerous misclassifications in the DA analysis presented in the previous section.

With the enrollment decline came the need to reduce staff, faculty and the budget. In 1982, the decline culminated in a major reduction in force which led to the termination of 47 faculty members, 38 of whom were tenured. From 1977-78 to

1984-85, the University lost 155 faculty positions or 24% of the faculty positions it employed in 1977-78. The faculty in 1984-85 is considerably younger than its counterpart which existed in 1982-83.

In 1983-84, the institution initiated an early retirement plan to encourage faculty members to retire. Forty-two (42) faculty members accepted the offer and retired at the conclusion of the 1983-84 academic year.

These two events, the reduction in force and the early retirement plan, help explain the dramatic improvement in the results of both the regression analysis and the discriminant analysis classification analysis over the three-year period. UNC lost approximately 90 of its older faculty members during this period and was able to hire a significant number of new faculty members. Thus a substantial change in the demographics of the remaining faculty has occurred. The improving pattern of rank and tenure classifications is to be expected as fewer faculty members who were tenured or promoted under past policies are employed at UNC.

Finally, in an effort to improve the salaries of its faculty and to correct individual inequities, UNC developed a new faculty salary model which was implemented for the 1984-85 year. This new model called for a survey of 29 peer institutions to be selected on the basis of similar role, mission, programs, enrollment and budget to that of UNC. At the same time the institution developed a comprehensive evaluation system which was used to help determine salaries. Therefore, a faculty member's salary was determined by the rank, discipline, time in rank and the evaluation rating for the previous year.

This new salary model led to a substantial redistribution of salary dollars among the faculty. No salary was reduced; however, a number of faculty members had their salary frozen. In contrast, a number of faculty members received salary raises of between $6,000-$9,000 or an increase of 20% to 30%.

Any faculty member who received an unsatisfactory evaluation received no salary raise. There were approximately 20% of the faculty who fell into this category for 1984-85 salary determinations. Thus the salary patterns which had existed in 1982-83 and 1983-84 changed dramatically for 1984-85. The purpose for the change was two-fold as mentioned above: (a) to improve salaries of the faculty at UNC relative to peer institutions and (b) to base salary decisions on rational factors such as qualifications and evaluations rather than historical factors or inconsistent policies of the past.

The results of the regression analysis clearly demonstrate the success of the new salary model in neutralizing the gender factor in salaries. The effects of the reduction in force effective in 1983 and the early retirement plans effective in 1984 are clearly seen in the analysis of the age factor over the three years. These factors combined with the new salary model have

1984-85, the University lost 155 faculty positions or 24% of the faculty positions it employed in 1977-78. The faculty in 1984-85 is considerably younger than its counterpart which existed in 1982-83.

In 1983-84, the institution initiated an early retirement plan to encourage faculty members to retire. Forty-two (42) faculty members accepted the offer and retired at the conclusion of the 1983-84 academic year.

These two events, the reduction in force and the early retirement plan, help explain the dramatic improvement in the results of both the regression analysis and the discriminant analysis classification analysis over the three-year period. UNC lost approximately 90 of its older faculty members during this period and was able to hire a significant number of new faculty members. Thus a substantial change in the demographics of the remaining faculty has occurred. The improving pattern of rank and tenure classifications is to be expected as fewer faculty members who were tenured or promoted under past policies are employed at UNC.

Finally, in an effort to improve the salaries of its faculty and to correct individual inequities, UNC developed a new faculty salary model which was implemented for the 1984-85 year. This new model called for a survey of 29 peer institutions to be selected on the basis of similar role, mission, programs, enrollment and budget to that of UNC. At the same time the institution developed a comprehensive evaluation system which was used to help determine salaries. Therefore, a faculty member's salary was determined by the rank, discipline, time in rank and the evaluation rating for the previous year.

This new salary model led to a substantial redistribution of salary dollars among the faculty. No salary was reduced; however, a number of faculty members had their salary frozen. In contrast, a number of faculty members received salary raises of between $6,000-$9,000 or an increase of 20% to 30%.

Any faculty member who received an unsatisfactory evaluation received no salary raise. There were approximately 20% of the faculty who fell into this category for 1984-85 salary determinations. Thus the salary patterns which had existed in 1982-83 and 1983-84 changed dramatically for 1984-85. The purpose for the change was two-fold as mentioned above: (a) to improve salaries of the faculty at UNC relative to peer institutions and (b) to base salary decisions on rational factors such as qualifications and evaluations rather than historical factors or inconsistent policies of the past.

The results of the regression analysis clearly demonstrate the success of the new salary model in neutralizing the gender factor in salaries. The effects of the reduction in force effective in 1983 and the early retirement plans effective in 1984 are clearly seen in the analysis of the age factor over the three years. These factors combined with the new salary model have produced a salary structure which has no indication of age dependency.

The race factor was not significant in any of the three years analyzed in this study. UNC has undergone significant changes both externally imposed and internally imposed. The statistical techniques used to assess the status of salaries and tenure and promotion decisions confirm the changes have improved the consistency of these decisions. When analyzed within the context of evolving institutional policies, these statistical tools can provide valuable insight into the status of decisions made with regard to salaries, tenure or promotion.

References

Carter, R. D., et al. (1983). "Multivariate Alternatives to Regression Analysis in the Evaulation of Salary Equity-Parity." Annual Forum of Association for Institutional Research, Toronto, Ontario.

Cooley, W.W. & Lohnes, P.R. (1976). *Evaluation research in education.* New York: Wiley.

Finkelstein, M.O. (1980). "The Judicial Reception of Multiple Regression Studies in Race and Sex Discrimination Cases," *80 Columbia Law Review,* 737-754.

Fisher, F.M. (1980). Multiple Regression in Legal Proceedings, *80 Columbia Law Review,* 702-736.

Pedhazur, E.J. (1982). *Multiple regression in behavioral research* (2nd ed.). New York: Holt, Rinehart and Winston.

Thorndike, R.J. & Weiss, D.J. (1973). A study of the stability of canonical correlations and canonical components. *Educational and Psychological Measurement, 33,* 123-134.

Multiple Comparisons Via Multiple Linear Regression: Learning the Obvious Takes Time

John D. Williams

The University of North Dakota

Perhaps a best starting point is at the beginning--the beginning of my involvement in multiple linear regression ala Ward, Bottenberg and Jennings. A presession to the AERA annual meeting in New York in 1967 was my first exposure to this type of analysis. I must admit something less than being fully enthralled with their ideas at the time. Despite computer accessibility for the five day workshop, I didn't actually run any programs. To me it was just a new fad. When getting back to Grand Forks (N.D.) I did feel some pangs of conscience and tried running a simple ANOVA by regression. The problem was a three group situation; I was trying to run:

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + e_1 \qquad [1]$$

where

X_1 = 1 if a member of group 1, 0 otherwise,

X_2 = 1 if a member of group 2, 0 otherwise,

X_3 = 1 if a member of group 3, 0 otherwise,

b_1, b_2, b_3 are regression coefficients,

Y = the criterion score, and

e_1 = the error in prediction with this model.

The program used at the presession was DATRAN, a forerunner of LINEAR (which of course, I didn't actually use). The program available to me back

Paper presented at the American Educational Research Association, Chicago, April, 1985

in North Dakota was a stock IBM program; in retrospect, such stock programs typically have automatic inclusion of a unit vector (or constant). Well, what happened next is both a descriptor of something about my personality (stubborn) or possibly lack of intelligence (slow). On a daily basis for seven weeks, (that's 35 times) I unsuccessfully tried running the program exactly as shown in equation 1 without any change. I thought possibly there was something wrong with the computer or the program; <u>never</u> did it cross my mind that I might have made a conceptual error. Finally, I started monkeying with the input (I was convinced the stuff in Bottenberg and Ward, 1963, was wrong). Well, I finally made the right mistake, and the program actually worked correctly. One form of that mistake is as follows:

$$Y = b_0 + b_1X_1 + b_2X_2 + e_1. \qquad [2]$$

The difference between equation 2 and equation 1 ostensibly is the exclusion of b_0 in equation 1 and the exclusion of b_3X_3 in equation 2.

Also, I now know that equations 1 and 2 are reparameterizations of one another. There are also some other "obvious" things about equation 2; it took me only four years to discover some of the obvious.

Equation 2 can allow not only a simple ANOVA, but also describes some important aspects of Dunnett's (1955) test (Williams, 1971); b_0 is not just a constant, but is equal to $\bar{Y}_3$, the so-called left out group. Also, $b_1 = \bar{Y}_1 - \bar{Y}_3$ and $b_2 = \bar{Y}_2 - \bar{Y}_3$. Equation 2 could be rewritten as:

$$Y = \bar{Y}_3 + (\bar{Y}_1 - \bar{Y}_3)X_1 + (\bar{Y}_2 - \bar{Y}_3)X_2 + e_1. \qquad [3]$$

The tests of the regression coefficients $b_1 = \bar{Y}_1 - \bar{Y}_3$ and $b_2 = \bar{Y}_2 - \bar{Y}_3$ are identically equal to the t values in Dunnett's test.

In addition to an ANOVA, other simple designs can be shown in a regression lay-out, such as the analysis of covariance, the t test, and treatments x subjects designs. The use of equations such as equation 2 to complete these designs was shown in Williams (1970). As usual, I had no idea at the time of the relationship to multiple comparisons. In some ways, the relationships are so simple and direct that it gives me cause for some degree of humility to remember how long it took me to discern the obvious again.

Through the use of full and restricted models, a process to test comparisons equivalent to Tukey's (1953) test was shown (Williams, 1974a). With three groups, beginning with equation 1, $Y = b_1X_1 + b_2X_2 + b_3X_3 + e_1$. Now suppose the test of $\overline{Y}_2 = \overline{Y}_3$ is of interest. In terms of the regression coefficients $b_2 = b_3$ is the appropriate restriction. Then $Y = b_1X_1 + b_2X_2 + b_2X_3 + e_2$ or

$$Y = b_1X_2 + b_2(X_2 + X_3) + e_2.$$

Let $V_1 = X_2 + X_3$; then

$$Y = b_1X_1 + b_2V_1 + e_2. \qquad [4]$$

Equation 4 can be reparameterized so that the unit vector (constant term) is reintroduced by excluding either X_1 or V_1. Excluding X_1 yields:

$$Y = b_0 + b_2V_2 + e_2. \qquad [5]$$

Testing $t = \sqrt{F} = \sqrt{\dfrac{(R_2^2 - R_5^2)/1}{(1 - R_2^2)/(N - K)}}$ yields a t appropriate to testing $\overline{Y}_2$ to $\overline{Y}_3$.

On the other hand, there is an easy way to run Tukey's test by regression. All that is necessary is the <u>set</u> of reparameterizations of equation 1:

$$Y = b_0 + b_1X_1 + b_2X_2 + e_1, \qquad [2]$$

$$Y = b_0 + b_1X_1 + b_3X_3 + e_1, \qquad [6]$$

and $$Y = b_0 + b_2X_2 + b_3X_3 + e_1. \qquad [7]$$

Here, the test of the computed t values is identical to a similar test for Tukey's test. (It took a full three years after doing the same thing with Dunnett's test to realize that Tukey's test could be accomplished through successive psuedo-Dunnett's tests). One complication is that most published studentized range tables are in terms of q, rather than in terms of testing the regression coefficients for significance. A table showing a direct solution using tests on the (partial) regression weights is given in Williams (1976, 1980).

In that I routinely would find all simple reparameterizations of an equation for an ANOVA solution, taking seven years to discover the obvious says something.

Two-Way Disproportionate ANOVAs

The two-way analysis of variance with disproportionate cell frequencies has been discussed in many different publications; Bottenberg and Ward (1963) showed a regression solution for the general case, and Jennings (1967) concentrated on the disproportionate situation. To be honest, I had a lot of trouble understanding the Jennings article, so I tried to go about doing what I could understand from the original Bottenberg and Ward presentation. One aspect of Bottenberg, Ward and Jennings in their various writings is a concern for explicitly stating exactly the hypothesis being tested through the use of a restriction on the regression coefficients. This aspect has been both a blessing and a curse; it is a blessing in the sense that the approach allows a precise methodology. It is a curse in that users are often at a disadvantage because of the cognitive completixity and relative mathematical sophistication required in comparison to traditional analysis of variance methodologies. It could be argued that a middle ground can be attempted; to some degree, that middle ground was something I tried to do (Williams, 1974b).

As an example of a two-way ANOVA with disproportionate cell frequencies the following data set was originally published in Williams (1972):

Data for Disproportionate Two-Way Analysis of Variance

Effect	Effect		
	B_1	B_2	B_3
A_1	8	1	6
	6	1	2
	4		
A_2	10	7	10
		5	9
		4	7
		4	5
		3	4

The solution given (1972) that was meant to simplify the process was to form four models:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + e_3; \qquad [8]$$

where

X_1 = 1 if from an individual in cell 1 (row 1, column 1), 0 otherwise;

X_2 = 1 if from an individual in cell 2 (row 1, column 2), 0 otherwise;

X_3 = 1 if from an individual in cell 3 (row 1, column 3), 0 otherwise;

X_4 = 1 if from an individual in cell 4 (row 2, column 1), 0 otherwise;

X_5 = 1 if from an individual in cell 5 (row 2, column 2), 0 otherwise;

and b_0 to b_5 are regression coefficients for this model.

$$Y = b_6 + b_7X_7 + e_4; \qquad [9]$$

where

X_7 = 1 from an individual in row 1, 0 otherwise and

b_6, b_7 are regression coefficients for this model.

$$Y = b_8 + b_9X_9 + b_{10}X_{10} + e_5; \qquad [10]$$

where

X_9 = 1 if from an individual in column 1, 0 otherwise;

X_{10} = 1 if from an individual in column 2, 0 otherwise; and

b_8, b_9 and b_{10} are regression coefficients for this model.

$$Y = b_{11} + b_{12}X_7 + b_{13}X_9 + b_{14}X_{10} + e_6. \qquad [11]$$

Now a solution in terms of sums of squares can be given as follows:

From: equation 8, $SS_{ATTRIBUTABLE} = 80.80$;

$SS_{DEVIATION} = 51.20$;

equation 9, $SS_{ATTRIBUTABLE} = 20.36$;

equation 10, $SS_{ATTRIBUTABLE} = 37.43$ and

equation 11, $SS_{ATTRIBUTABLE} = 80.25$.

This information could be used to construct a fitting contants solution or a hierarchical solution (Cohen, 1968) or the solution described by Jennings (1967); although Jennings laboriously goes through the process of testing hypotheses through restrictions on a reparameterization of the full model:

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + e_3. \qquad [12]$$

This model corresponds to equation 8, except that the unit vector is omitted (b_0) and the sixth cell is represented through b_6X_6. Because my solution, while it coincides with Jennings, can be addressed without adjusting the sums of squares as must be done for a fitting constants solution or a hierarchical solution, I called this solution the "unadjusted main effects" solution--in retrospect, a poor choice of names. It was called this because of the means of extracting the sums of squares--but its usefulness is because it corresponds to the Jennings solution. That by the way, is another story--I spent an hour and a half convincing Earl that my solution gave the same results as his; at first he was skeptical. Finally, he accepted that, "computationally, their respective sums of squares was the same," but thought only people such as myself who understand both approaches and used my approach as a computational short cut should use it; if you didn't know what hypotheses were being tested, you probably shouldn't use it. I thought Earl was being a little harsh back in 1972, but today I'm coming closer to agreement with that position.

In particular, it could be noted that the so-called "full rank model" as described by Timm and Carlson (1975), and which in fact they describe using my (1972) data set, has no better claim to being a full rank model solution than Jennings (1967); the hypotheses tested by these and other approaches are considered in Williams (1977a). It is unfortunate that the Timm and Carlson (1975) solution might be seen by some as "standard practice" or "state of the art". The issue really is, which hypotheses are of greatest interest? If the Timm and Carlson hypotheses are truly of the greatest interest, they can be addressed via the Bottenberg and Ward approach.

A summary table that computationally tests hypotheses proportional to cell frequencies such as proposed by Jennings can easily be formed from the information from equations 8, 9, 10 and 11:

$SS_{ROWS} = 20.36$; $SS_{COLS} = 37.43$;

$SS_{RC} = 80.80 - 80.25 = .55$;

$SS_{within} = 51.20$. The summary table is as follows:

Table 1

Summary Table for Two-Way
Disproportionate Cell Frequencies

Source of Variation	df	SS	MS	F
Rows	1	20.36	20.36	4.77
Columns	2	37.43	18.72	4.38
R X C	2	.55	.28	.07
Within	12	51.20	4.27	

In regard to multiple comparisons in a two-way layout, equation 12 is an appropriate starting point. The number and type of comparisons (contrasts) would be important for deciding on the type of test (Dunnett's, Tukey's, Scheffe's, 1959, and Dunn's, 1961). As an example of constructing a contrast to test a hypothesis of interest, suppose the researcher wants to compare column 1 to column 2, weighing the cells by their size, the hypothesis, in terms of sample means, is:

$$\frac{3\bar{Y}_1 + 1\bar{Y}_4}{4} = \frac{2\bar{Y}_2 + 5\bar{Y}_5}{7} .$$

In terms of the regression coefficients,

$$\frac{3b_1 + b_4}{4} = \frac{2b_2 + 5b_2}{7}$$

Unraveling and solving for b_1 yields: $b_1 = 8/21b_2 + 20/21b_5 - 7/21b_4$.

Substituting this restriction into equation 12 yields:

$$Y = (8/21b_2 + 20/21b_5 - 7/21b_4)X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + e_7; \qquad [13]$$

or

$$Y = b_2(X_2 + 8/21X_1) + b_3X_3 + b_4(X_4 - 7/21X_1) + b_5(X_5 + 20/21X_1) + b_6X_6 + e_7. \qquad [14]$$

Reparameterization with $b_6 = 0$ yields:

$$Y = b_0 + b_2(X_2 + 8/21X_1) + b_3X_3 + b_4(X_4 - 7/21X_1) + b_5(X_5 + 20/21X_1) + e_7. \qquad [15]$$

Equation 14 can be used in programs where unit vectors can be ommitted. Its reparameterization, equation 15, is useful when a unit vector is automatically incorporated into a regression solution. Equations 8 and 12 (full models) yield $R_F^2 = .61212$. Equations 14 and 15 (restricted models) yield $R_F^2 = .38544$. Then:

$$t = \sqrt{F} = \sqrt{\frac{(R_F^2 - R_R^2)/1}{(1 - R_F^2)/12}} = 2.648.$$

This t value should be tested against an appropriate table depending upon the type and number of total comparisons considered by the researcher.

This approach to multiple comparisons is probably much closer to the approach of Jennings and Bottenberg and Ward than I would have considered 10 to 15 years ago. Additional considerations regarding multiple comparisons in the two-way analysis of variance ban be found in Williams (1980).

Multiple Comparisons in the Analysis of Covariance

Students would often ask questions such as, "How do you do multiple comparisons on adjusted means in the analysis of covariance?" I've often been impressed with questions students ask; I'm sure they've been less impressed with at least some of my answers. Well, for several years, I didn't have any good answer to the aforementioned question (other than, "That's a good question.") and as the answer finally came to me, there was far more embarrassment than awe. The "answer" had been on the printouts that I'd been using for years. In a nutshell, it was simply the test of signifiance for the group partial regression weights in a full model. An example of a solution for this problem was taken from Williams (1979).

Table 2 is taken from Williams (1974b, p. 104 and 109). In Table 2, X_1 is a binary variable for membership in group 1, X_2 is a binary variable for membership in group 2 and X_3 is similarly a binary variable for membership in group 3 and X_4 represents a pretest score; the Y value represents a posttest score.

Table 2

Data for the Analysis of Covariance

Y	X_1	X_2	X_3	X_4
35	1	0	0	12
27	1	0	0	17
32	1	0	0	13
29	1	0	0	10
27	1	0	0	8
38	0	1	0	29
25	0	1	0	12
36	0	1	0	17
25	0	1	0	22
31	0	1	0	15
27	0	0	1	17
35	0	0	1	22
19	0	0	1	10
17	0	0	1	8
32	0	0	1	13

Under the assumption of a single regression line on the covariate (the pretest, X_4) an analysis of covariance can be accomplished with two linear models:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_4X_4 + e_8, \qquad [16]$$

and

$$Y = b_0 + b_4X_4 + e_9. \qquad [17]$$

In that a large part of the print-out regarding equation 16 is useful, the print-out is reproduced in Table 3.

The usual analysis of covariance can be completed by using:

$$F = \frac{(R_F^2 - R_R^2)/(g - 1)}{(1 - R_2^2)/(N - C - g)} = \frac{(.61959 - .47476)/2}{(1 - .61959)/11} = 2.09,$$

which for df = 2, 11, $p > .05$.

In equation 16 the X_3 variable has been omitted. Thus $b_1 = \overline{Y}_1$adj - $\overline{Y}_3$adj and $b_2 = \overline{Y}_2$adj - $\overline{Y}_3$adj. To find the adjusted means, the following equations can be used:

$\overline{Y}_3\text{adj} = b_0 + b_4X_4 = 15.36 + .76(15) = 26.76$;

$\overline{Y}_1\text{adj} = b_1 + Y_3\text{adj} = 5.52 + 26.76 = 32.28$; and

$\overline{Y}_2\text{adj} = b_2 + Y_3\text{adj} = 3.20 + 27.76 = 29.96.$

The adjusted values agree with those originally given by Williams (1974b, p. 106), though the method shown here is simplified somewhat.

More importantly, the standard error of the regression coefficients corresponding to X_1 and X_2 are respectively equal to the standard errors for comparing $\overline{Y}_1$adj to $\overline{Y}_3$adj and $\overline{Y}_2$adj to $\overline{Y}_3$adj. Thus, the computed t values given in Table 3 are directly usable in whichever multiple comparison procedure the researcher prefers. The use of Dunnett's (1955), Tukey's (1953), Dunn's (1961) and Scheffe's (1959) tests are described in a regression format using computed t values in Williams (1976, 1980). Were there interest in comparing $\overline{Y}_1$adj to $\overline{Y}_2$adj, a model of the form:

$$Y = b_0 + b_1X_1 + b_3X_3 + b_4X_4 + e_8 \qquad [18]$$

could be used, with focus on the computed t value for the X_1 variable.

Table 3

Print-Out for Equation 16

Variable	Mean	Standard Deviation	Correlation X vs Y	Regression Coefficient	Std. Error of Reg. Coef.	Computed T Value
4	15.00	5.85	0.689	0.76	0.22783	3.33582
1	0.33	0.48	0.039	5.52	2.73396	2.01905
2	0.33	0.48	0.398	3.20	2.92653	1.09345
Dependent Y	29.66	6.12				
INTERCEPT	15.36					

MULTIPLE CORRELATION 0.78714

STD. ERROR OF ESTIMATE 4.26230

MULTIPLE CORRELATION SQUARED 0.61959

ONE MINUS MULTIPLE CORRELATION SQD. 0.38041

Analysis of Variance for the Regression

Source of Variation	Degrees Of Freedom	Sum of Squares	Mean Squares	F Value
Attributable to Regression	3	325.49	108.497	5.972
Deviation from Regression	11	199.84	18.167	
Total	14	525.33		

f course, multiple covariates and/or more complex comparisons can be ncorporated; multiple covariates can be incorporated without adding too uch complexity to the solution. The remarkable thing is that the soluion to multiple comparisons for the analysis of covariance is easily ichieved.

Multiple Comparisons in Repeated Measure Designs

Again, the impetus (to me) for interest in multiple comparisons in epeated measures designs in general, and treatments x subjects designs n particular comes from students. Students would ask, "O.K., so now e can do a treatments x subjects design by regression. How do we run nultiple comparisons?" Since they asked the question long before I had any suitable answer, a question might be asked, "What answer did I give?" To quote both the famous and infamous (e.g. Steve Martin and John Mitchell), "I forgot." Considering that that answer can be as simple as, "It's right there on your printout," I won't dwell anymore on why it took so long.

Multiple Comparisons for Treatments X Subjects Designs

To consider multiple comparisons for treatments x subjects designs (or repeated measure designs) an example taken from Chapter 7 of Williams (1974b, p. 56) is used; see Table 4.

Table 4

Three Treatment Methods of Paired-Associate Learning with Educable Mentally Retarded Subjects

Subject	Treatment One	Treatment Two	Treatment Three
1	18	27	15
2	17	24	14
3	14	13	12
4	5	8	6
5	11	14	10
6	9	12	8
7	14	16	15
8	12	17	9
9	22	21	16
10	10	18	15

The information in Table 4 can be placed in a tabular form suitable for use in regression format; see Table 5.

Table 5

Illustration of Design Matrix for Treatments X Subjects Designs

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
3	1	0	0	1	0	0	0	0	0	0	0	0	0	60
7	0	1	0	1	0	0	0	0	0	0	0	0	0	60
5	0	0	1	1	0	0	0	0	0	0	0	0	0	60
7	1	0	0	0	1	0	0	0	0	0	0	0	0	55
4	0	1	0	0	1	0	0	0	0	0	0	0	0	55
4	0	0	1	0	1	0	0	0	0	0	0	0	0	55
14	1	0	0	0	0	1	0	0	0	0	0	0	0	39
13	0	1	0	0	0	1	0	0	0	0	0	0	0	39
12	0	0	1	0	0	1	0	0	0	0	0	0	0	39
5	1	0	0	0	0	0	1	0	0	0	0	0	0	19
8	0	1	0	0	0	0	1	0	0	0	0	0	0	19
6	0	0	1	0	0	0	1	0	0	0	0	0	0	19
11	1	0	0	0	0	0	0	1	0	0	0	0	0	35
14	0	1	0	0	0	0	0	1	0	0	0	0	0	35
10	0	0	1	0	0	0	0	1	0	0	0	0	0	35
9	1	0	0	0	0	0	0	0	1	0	0	0	0	29
12	0	1	0	0	0	0	0	0	1	0	0	0	0	29
8	0	0	1	0	0	0	0	0	1	0	0	0	0	29
14	1	0	0	0	0	0	0	0	0	1	0	0	0	45
16	0	1	0	0	0	0	0	0	0	1	0	0	0	45
15	0	0	1	0	0	0	0	0	0	1	0	0	0	45
12	1	0	0	0	0	0	0	0	0	0	1	0	0	38
17	0	1	0	0	0	0	0	0	0	0	1	0	0	38
9	0	0	1	0	0	0	0	0	0	0	1	0	0	38
22	1	0	0	0	0	0	0	0	0	0	0	1	0	59
21	0	1	0	0	0	0	0	0	0	0	0	1	0	59
16	0	0	1	0	0	0	0	0	0	0	0	1	0	59
10	1	0	0	0	0	0	0	0	0	0	0	0	1	43
18	0	1	0	0	0	0	0	0	0	0	0	0	1	43
15	0	0	1	0	0	0	0	0	0	0	0	0	1	43

The values in Table 5 are defined as follows:

Y = the criterion score;

X_1 = 1 if the score corresponds to Treatment 1, 0 otherwise;

X_2 = 1 if the score corresponds to Treatment 2, 0 otherwise;

X_3 = 1 if the score corresponds to Treatment 3, 0 otherwise;

X_4 = 1 if the score is obtained from Subject 1, 0 otherwise;

$X_5 = 1$ if the score is obtained from Subject 2, 0 otherwise;

$X_6 = 1$ if the score is obtained from Subject 3, 0 otherwise;

$X_7 = 1$ if the score is obtained from Subject 4, 0 otherwise;

$X_8 = 1$ if the score is obtained from Subject 5, 0 otherwise;

$X_9 = 1$ if the score is obtained from Subject 6, 0 otherwise;

$X_{10} = 1$ if the score if obtained from Subject 7, 0 otherwise;

$X_{11} = 1$ if the score is obtained from Subject 8, 0 otherwise;

$X_{12} = 1$ if the score is obtained from Subject 9, 0 otherwise;

$X_{13} = 1$ if the score is obtained from Subject 10, 0 otherwise; and

X_{14} = the sum of the criterion scores for each subject separately.

A full model for this data could be given as:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + b_{12}X_{12} + e_{10}; \qquad [19]$$

an alternative model would be:

$$Y = b_0 + b_1X_1 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7 + b_8X_8 + b_9X_9 + b_{10}X_{10} + b_{11}X_{11} + b_{12}X_{12} + e_{10}. \qquad [20]$$

See Table 6 for a printout using equation 19.

From Table 6, it can be seen that $t_1 = 1.10362$ and $t_2 = 4.59846$; that t values are respectively the tests regarding comparing $\bar{Y}_1$ to $\bar{Y}_3$ and $\bar{Y}_2$ to $\bar{Y}_3$, taking into account that the subjects serve as their own controls. A similar printout could be generated using a model corresponding to equation 20. Values from this printout show $t_1 = -3.49484$, $t_3 = -4.59847$; these t values correspond to comparing $\bar{Y}_1$ to $\bar{Y}_2$ and $\bar{Y}_3$ to $\bar{Y}_2$. Also, the corresponding means are $\bar{Y}_1 = 13.20$, $\bar{Y}_2 = 17.00$ and $\bar{Y}_3 = 12.00$. These computed t values should be compared to an appropriate multiple comparison table for significance.

Table 6

Output of Full Model for Treatments X Subjects Design

Variable No.	Mean	Standard Deviation	Correlation X vs Y	Regression Coefficient	Std. Error Of Reg. Coef.	Computed T Value	Beta
1	0.33333	0.47946	-0.12145	1.19998	1.08732	1.10362	0.11210
2	0.33333	0.47946	0.41105	4.99997	1.08732	4.59846	0.46710
4	0.10000	0.30513	0.39195	5.66663	1.98515	2.85451	0.33690
5	0.10000	0.30513	0.28185	4.00001	1.98515	2.01496	0.23781
6	0.10000	0.30513	-0.07046	-1.33331	1.98515	-0.67164	-0.07927
7	0.10000	0.30153	-0.51085	-7.99992	1.98515	-4.12987	-0.47562
8	0.10000	0.30153	-0.15854	-2.66665	1.98515	-1.34329	-0.15854
9	0.10000	0.30153	-0.29066	-4.66664	1.98515	-2.35077	-0.27745
10	0.10000	0.30153	0.06166	0.66668	1.98515	0.33583	0.03964
11	0.10000	0.30153	-0.09248	-1.66665	1.98515	-0.83956	-0.09909
12	0.10000	0.30153	0.36993	5.33332	1.98515	2.68661	0.31708
Dependent Y	14.06667	5.13226					

INTERCEPT 12.26667

MULTIPLE CORRELATION 0.92774

STD. ERROR OF ESTIMATE 2.43131

MULTIPLE CORRELATION SQUARED 0.86070

ONE MINUS MULTIPLE CORRELATION SQD. 0.13930

Analysis of Variance for the Regression

Source of Variation	Degrees Of Freedom	Sum of Squares	Mean Squares	F Value
Attributable to Regression	11	657.46021	59.76910	10.11102
Deviation from Regression	18	106.40308	5.91128	
Total	29	763.86329		

Using the Shortcut Method

The solution just given in the last section presumed that each subject (except one) is separately coded using a binary coding scheme. Clearly, if the number of subjects is at all large, the coding procedure described in Williams (1977b) and using:

$$Y = b_0 + b_1X_1 + b_2X_2 + b_{14} + e_{10} \qquad [21]$$

might be preferrable. However, one difficulty with using this shortcut procedure is that the standard error of the regression coefficients for X_1 and X_2 are too small due to the degrees of freedom, as generated by the computer program, not being accurate for deviation from regression. These t values could be adjusted by multiplying by an appropriate constant. The appropriate constant is: $c = \sqrt{\frac{MS_{w_{21}}}{MS_{w_{19}}}}$ where $MS_{w_{21}}$ is the mean square within (or deviation from regression) for equation 21 and $MS_{w_{19}}$ is the mean square within for equation 19. The $MS_{w_{21}}$ is 4.09225 and $MS_{w_{19}}$ is 5.91125. Thus, c = .83203. The values generated by equation 21 for t_1 and t_2 (comparing $\overline{Y}_1$ to $\overline{Y}_3$ and $\overline{Y}_2$ to $\overline{Y}_3$) are $t_1 = 1.32641$ and $t_2 = 5.52678$. Multiplying t_1 and t_2 by c yields corrected $t_1 = 1.10361$ and corrected $t_2 = 4.59845$, within rounding error of the values found earlier. Of course, $MS_{w_{19}}$ would not be available were the researcher using the shortcut method. However, $MS_{w_{19}} = \frac{SS_w}{N-S-g+1}$ where N is the total number of scores, S is the number of subjects and g is the number of groups. The denominator can also be found as (S-1)(g-1).

Repeated Measures Designs

Multiple comparisons also can be relatively routinized for large data sets involving repeated measures. Williams and Williams (1984) showed research application of a hypotheses testing process for k groups ıasured at three times for large N. More recently, they showed in press) the same solutions to the problem done earlier in Williams 1980); a 3 x 4 repeated measure design with five entries per cell was ıade to show a problem that was not solvable in a regression format; ·ortunately (or unfortunately) a solution was found, so the chapter vas entitled, "Problems less amenable to a regression solution." In ıpplying this solution to the larger data set, two progressively easier solutions were found; the preferred solution (i.e., easiest to accomplish) is embarrassingly close to a simple Bottenberg and Ward/Ward and Jennings (1973) solution.

Perhaps the point of all of this is to give some comfort to those who have struggled within the use of regression as a technique to address research questions, particularly as they look over their shoulders and think they may never master the process. Insofar as I might be seen as one who has mastered this process, let me point out, I'm still learning!

References

Bottenberg, R. A., & Ward, J. H. (1963). *Applied multiple linear regression.* Lackland Air Force Base, Texas: Personnel Research Laboratory PRL-TDR-63-6.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70,* 526-543.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association, 56,* 52-64.

Dunnett, C. W. (1955). A multiple comparison procedure for comparing several treatments with a control. *Journal of the American Statistical Association, 50,* 1086-1121.

Jennings, E. (1967). Fixed effects analysis of variance by regression analysis. *Multivariate Behavioral Research, 2*, 95-108.

Scheffe, H. (1959). *The analysis of variance.* New York: Wiley.

Timm, N. H., & Carlson, J. E. (1975). Analysis of variance through full rank models. *Multivariate Behavioral Research: Monograph,* 75-1.

Tukey, J. W. (1953). The problem of multiple comparisons. Dittoed, Princeton University.

Ward, J. W., & Jennings, E. E. (1973). *Introduction to linear models.* Englewood Cliffs, New Jersey: Prentice-Hall.

Williams, J. D. (1970). A regression approach to experimental design. *Journal of Experimental Education, 39*(1), 83-90.

Williams, J. D. (1971). A multiple regression approach to multiple comparisons for comparing several treatments with a control. *Journal of Experimental Education, 39*(3), 93-96.

Williams, J.D. (1972). Two way fixed effects analysis of variance with disaproportionate cell frequencies. *Multivariate Behavioral Research, 7*, 67-83.

Williams, J. D. (1974a) A simplified regression formulation of Tukey's test. *Journal of Experimental Education, 42*(4), 80-82.

Williams, J. D. (1974b) *Regression analysis in educational research.* New York: MSS Information Corporation.

Williams, J. D. (1976). Multiple comparisons by multiple linear regression. *Multiple Linear Regression Viewpoints Monograph Series-2, 7*(1).

Williams, J. D. (1977a). Full rank and non-full rank models with contrast and binary coding systems for two-way disproportionate cell frequency analyses. *Multiple Linear Regression Viewpoints, 8*(1), 1-18.

ms, J. D. (1977b). A note on coding the subjects effects in treat-
ents x subjects designs. Multiple Linear Regression Viewpoints, 8(1),
32-35.

ms, J. D. (1979). Contrasts with unequal N by multiple linear
regression. Multiple Linear Regression Viewpoints, 9(3), 1-7.

ams, J. D. (1980). Multiple comparisons in higher dimensional designs.
Multiple Linear Regression Viewpoints Monograph Series #5, 10(3).

ams, J. D., & Williams, J. A. (1984). Testing hypotheses in a repeated
measures design on employees attitudes with large samples. Paper
presented at the Annual Meeting of the American Educational Research
Association, New Orleans, April.

iams, J. D., & Williams, J. A. (in press). Testing hypotheses in a
repeated measures design-An example. Multiple Linear Regression Viewpoints.

'LE LINEAR REGRESSION VIEWPOINTS
E 14, NUMBER 1, SPRING 1985

The Effect of the Violation of the Assumption of Independence When Combining Correlation Coefficients in a Meta-Analysis

Susan M. Tracz

California State University, Fresno

Patricia B. Elmore

Southern Illinois University, Carbondale

Meta-analysis is a technique for combining the summary statistics from viously conducted research studies. Pioneered by Gene V Glass (1976) a-analysis gives not only an indication of the direction of the results of : studies, but provides an index of the magnitude of the effect as well. ta-analyses are reported in terms of mean effect size, $\overline{ES}$. There are two)es of effect sizes. An experimental effect size is the mean of the experintal group minus the mean of the control group divided by the standard viation,

$$ES = \frac{\bar{X}_E - \bar{X}_C}{s_X},$$

ile a correlational effect size is simply a correlation coefficient,

$$ES = r.$$

Paper presented at the American Educational Research Association, Chicago, April 1985

Meta-analysis has been further refined by Hedges (1983), who has been developing techniques for using effect sizes as data points and then fitting regression models. The focus of this paper, however, will be the use of correlation coefficients in meta-analyses and the effect of the violation of the assumption of independence in these analyses.

Independence

A necessary assumption for the results of statistical analyses to be tenabl is independence. All inferential statistical techniques require independence of observations. By independence is meant that the probability of including one subject or data point will in no way affect the probability of including any oth subject or data point. Another way of defining independence is to say that the value of a variable for a subject is not predictable from the value of a variabl for any other subject.

So far independence has been defined in reference to primary studies perfor by researchers who draw a random sample of subjects, measure the subjects on variables of interest, and calculate statistics from the measured data using their hypothesized models. The meta-analysts, on the other hand, draw a sample of studies usually from journal articles, record the numerous statistics reported in each study, and calculate a statistic based on effect sizes or a meta-statistic from a data set of simple statistics. When jumping from the level of individual studies to combinatory techniques, studies parallel subjects and simple statistics parallel observations on variables. In the framework of combinatory methodology, then, independence means that the value of any statist which is included should in no way be predictable from the value of any other included statistic.

The typical study which is chosen for inclusion in a meta-analysis, however will yield more than one effect size or simple statistic. When the meta-analyst uses all the statistics available in a particular study to calculate the mean size, the assumption of independence is violated. Landman and Dawes (1982) ፡ five ways in which the assumption of independence can be violated in meta-ɘs. These five types of violations are as follows:

> "1) Multiple measures from the same subjects, . . .
> 2) Measures taken at multiple points in time from the same subjects, . . .
> 3) Nonindependence of scores within a single outcome measure, . . .
> 4) Nonindependence of studies within a single article, . . . and
> 5) Nonindependent samples across articles" (pp. 506-507).

Kraemer (1983) specifically provides the caveat that "only one effect size tudy can be used to ensure independence" (p. 99) in meta-analyses. This that the ratio of effect sizes to studies in a meta-analysis should be n order to avoid violating this assumption. However, even a cursory review ıblished meta-analyses reveals that the assumption of independence is, in , seldom met.

ɔose

The purpose of this study was to determine the effect of the violation the assumption of independence on the distribution of r and the distribution Fisher's Z. In this Monte Carlo simulation the following four parameters were ɛd with the values specified:

N - the sample size within a study (20, 50, 100),

p - the number of predictors (1, 2, 3, 5),

rho(i) - the population intercorrelation among predictors (0, .3, .7),

rho(p) - the population correlation between predictors and criterion (0, .3, .7).

Predictor and criterion variables were generated to conform to all possible combinations of the parameters specified above and then correlated. The main parameter of interest was rho(i), since it was the index of nonindependence when it assumed a nonzero value in the multiple predictor cases. When only one predictor was used or when the intercorrelation among predictors, rho(i), equaled zero, then the assumption of independence was not violated.

Method

In this study dependent and independent correlations were generated between criterion and predictor variables. The values of the parameter p, the number of predictors, were one, two, three, and five, and path diagrams for each case appear in Figures 1 through 4 respectively. In these diagrams the G variables are the common generating variables used along with error to form the X variables or predictors, which are in turn combined along with error to produce the Y or criterion variables. The arrows between variables indicate the relationship among the endogenous variables. The associated lower case letters are the standardized regression coefficients for path analysis. The arrows which are not

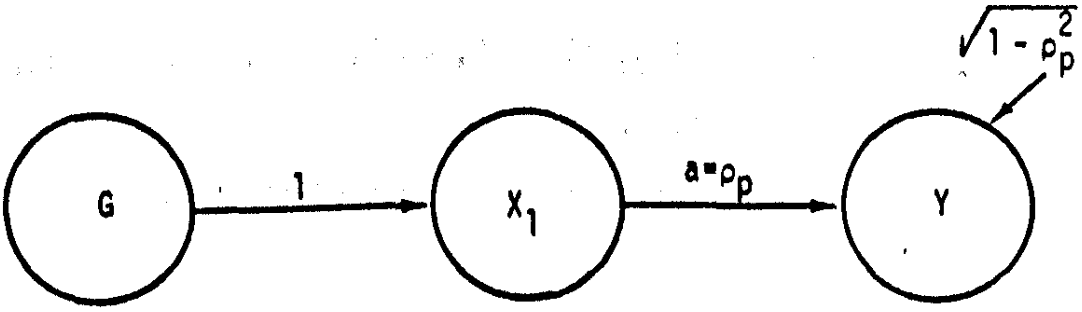
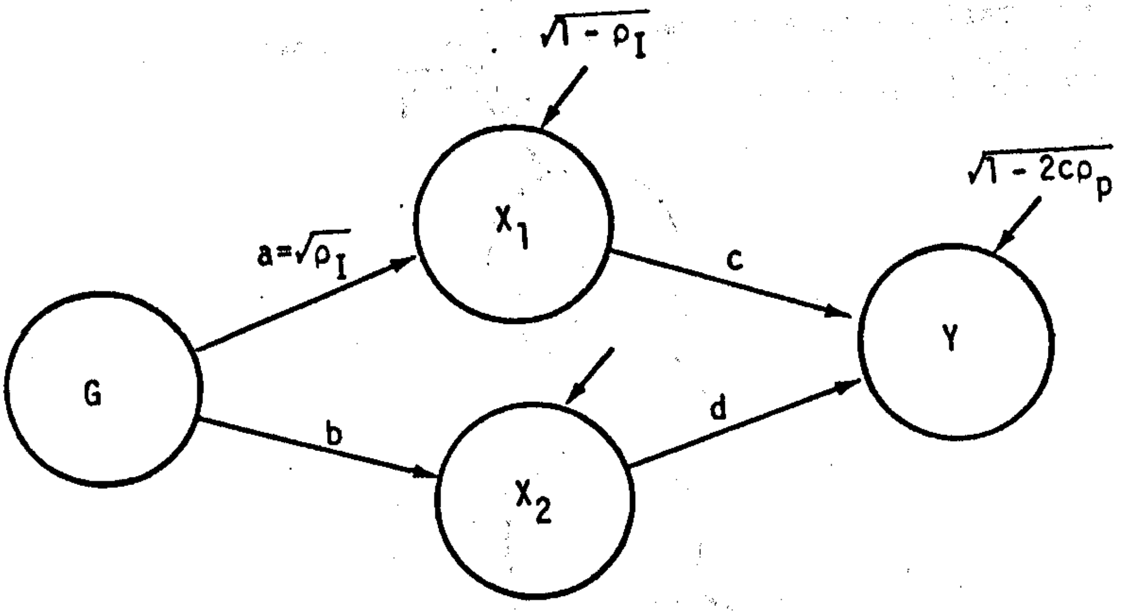
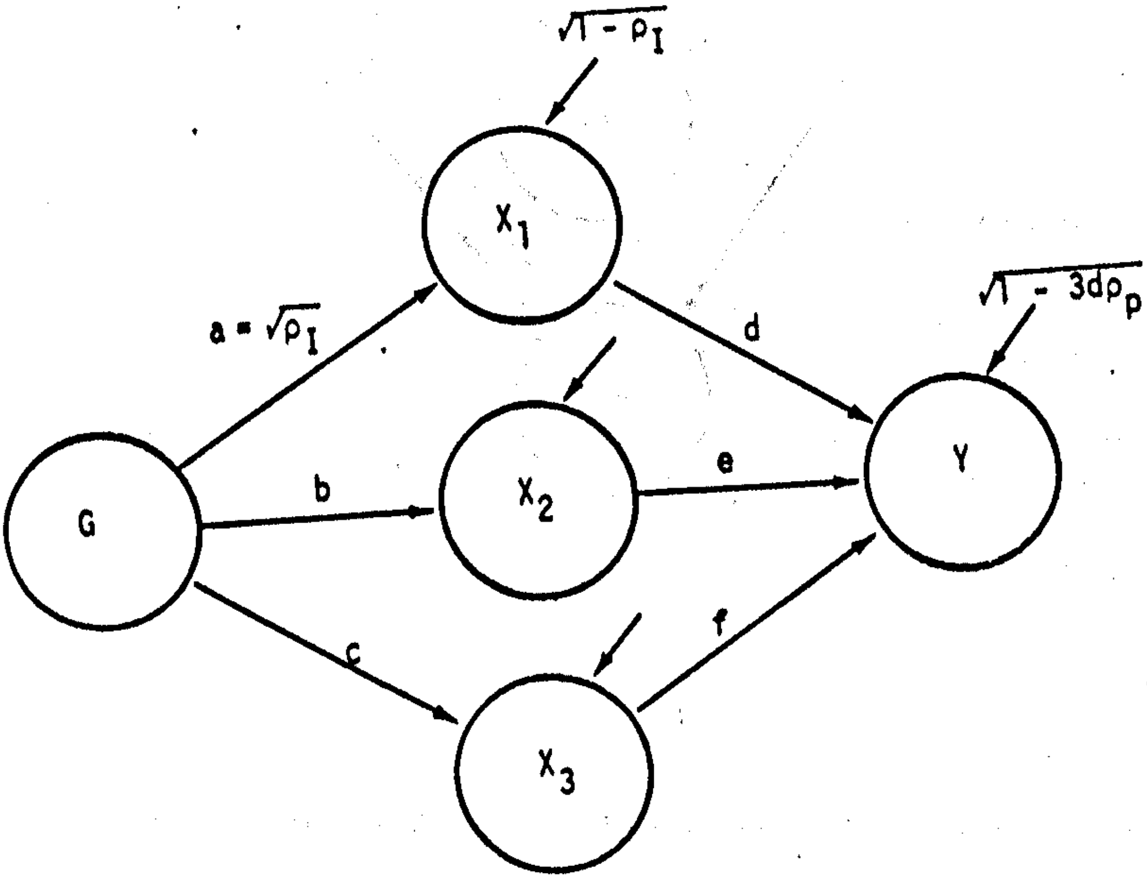


Figure 1. Path diagram for the one predictor case.



<u>Figure 2</u>. Path diagram for the two predictor case.



<u>Figure 3.</u> Path diagram for the three predictor case.

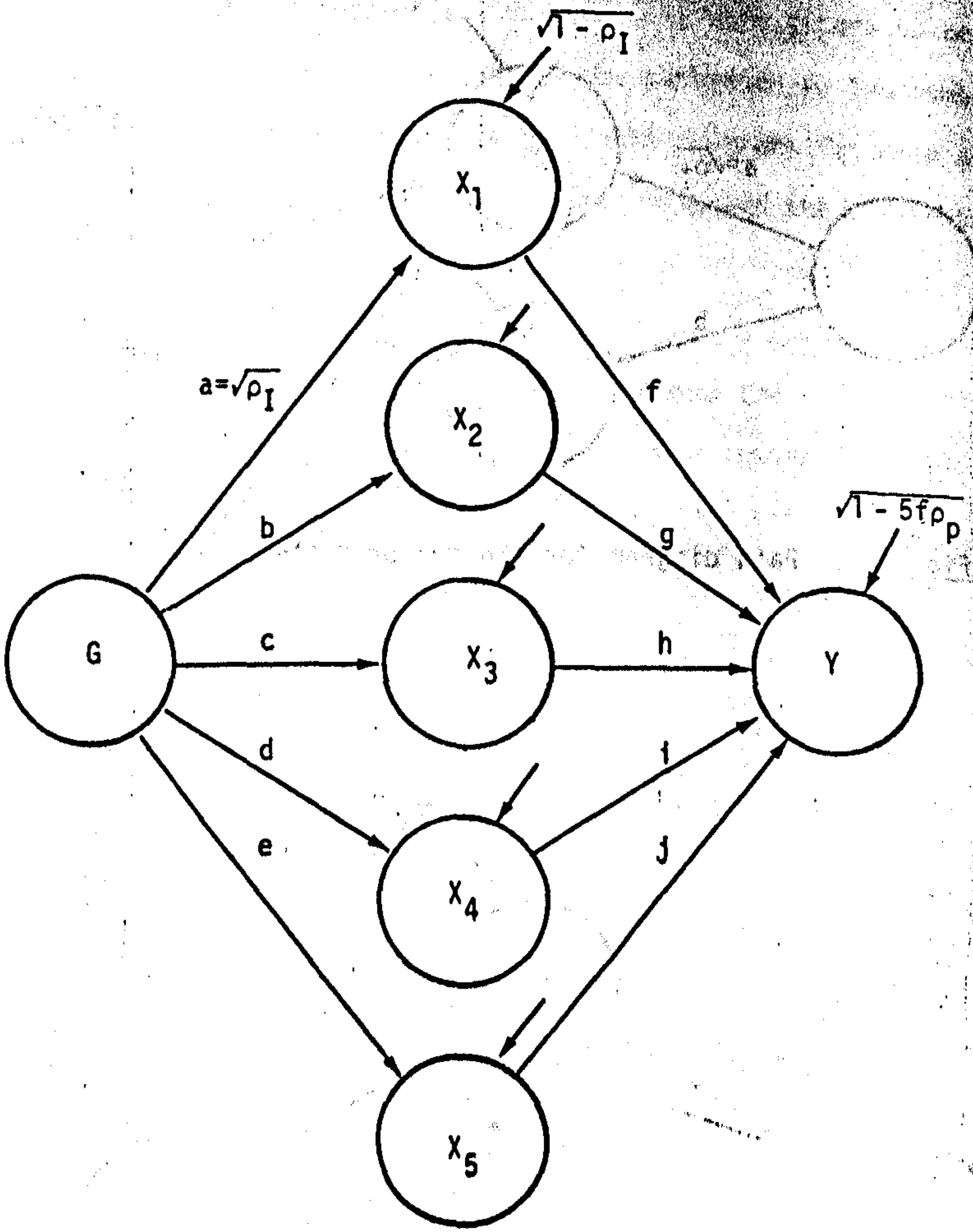


Figure 4. Path diagram for the five predictor case.

nected indicate exogenous variation, and those coefficients are given as well.

The following algorith derived by Knapp and Swoyer (1967) was used to erate correlated vectors of numbers:

$$Y = aX + \sqrt{1 - a^2}Z$$

ere X = a vector of randomly chosen numbers from the standard normal distribution,

Z = another vector of randomly chosen numbers from the standard normal distribution, and

a = the desired correlation between X and Y.

In the unique one predictor case, the intercorrelation among predictors uld not be varied since only one predictor was present. Therefore, independence xists in this case. Here the X1 vector was set equal to G, a vector of randomly hosen standard normal deviates, so the path coefficient between G and X1 is one. he path coefficient between X1 and Y, a, was set equal to the population correlation etween predictors and criterion, rho(p). Since a = rho(p), the error coefficient for Y was $\sqrt{1 - a^2}$ or $\sqrt{1 - \text{rho}(p)^2}$. The Y vector was then created as follows:

$$Y = aX1 + \sqrt{1 - a^2}Z$$

where Z = a vector of randomly chosen numbers from the standard normal distribution. The vectors for X1 and Y were then correlated.

A different procedure was used for data generation in the multiple predictor cases. In Figure 2, path coefficients a = b and c = d. In Figure 3, a = b = c and d = e = f. In Figure 4, a = b = c = d = e and f = g = h = i = j. In these three diagrams the correlations between any two predictors is equal to the product of the path coefficients connecting those two predictors with the generating variable or the quantity, a^2, since all the coefficients between generating variables and predictors are equal. For the correlation between two predictors to equal rho(i), the path coefficient, a, was set equal to $\sqrt{\text{rho}(i)}$. Then all the X vectors were generated as follows:

$$X(i) = \sqrt{a}G + \sqrt{1 - a}Z(i)$$

Where X(i) = a vector of values for a predictor and i assumes incremental values for vectors from one to p, the number of predictors,

a = rho(i) = the population intercorrelation among predictors,

Z(i) = a vector of randomly chosen standard normal deviates and i assumes incremental values for vectors from one to p, the number of predictor

The following points concern the generation of the Y vectors. First it should be noted that each Y is a linear combination of the p predictors plus error. The weight of that combination is c in Figure 2, d in Figure 3, and f in Figure 4. Second, it should be noted that correlation coefficients can be reconstructed from the standardized regression coefficients in a path diagram. In Figure 2, the correlations between the two predictors and the criterion can be reconstructed as follows:

$$r_{yx_1} = c + abd,$$

$$r_{yx_2} = d + bac,$$

but since c = d, and a = b = $\sqrt{rho(i)}$, the correlation between Y and any predictor X(i), can be written as follows:

$$r_{yx_i} = c + \rho(i)c = c(1 + \rho(i)).$$

Also since r_{yx_i} is an estimate of rho(p), that value can be substituted into the equation so that it can be solved for c as follows:

$$\rho(p) = c(1 + \rho(i))$$

$$c = \frac{\rho(p)}{1 + \rho(i)} .$$

In Figure 3 in parallel fashion, the correlations between the three predic and the criterion can be reconstructed as follows:

$$r_{yx_1} = d + abe + acf,$$

$$r_{yx_2} = e + bcf + bad,$$

$$r_{yx_3} = f + cbe + cad,$$

since $a = b = c = \sqrt{rho(i)}$, and $d = e = f$, the correlation between Y and any dictor, X(i), can be written as follows:

$$r_{yx_i} = d + \rho(i)d + \rho(i)d = d(1 + 2\rho(i)).$$

so since r_{yx_i} is an estimate of rho(p), that value can be substituted into the uation so that it can be solved for d as follows:

$$\rho(p) = d(1 + 2\rho(i)),$$

$$d = \frac{\rho(p)}{1 + 2\rho(i)}.$$

In Figure 4 the last obvious parallel exists. The correlations between the ive predictors and the criterion can be reconstructed as follows:

$$r_{yx_1} = f + abg + ach + adi + aej,$$

$$r_{yx_2} = g + baf + bch + bdi + bej,$$

$$r_{yx_3} = h + caf + cbg + cdi + cej,$$

$$r_{yx_4} = i + daf + dbg + dch + dej,$$

$$r_{yx_5} = j + eaf + ebg + ech + edi,$$

but since $a = b = c = d = e = \sqrt{rho(i)}$, and $f = g = h = i = j$, the correlation between Y and any predictor, X(i), can be written as follows:

$$r_{yx_i} = f + \rho(i)f + \rho(i)f + \rho(i)f + \rho(i)f = f(1 + 4\rho(i)).$$

Again r_{yx_i} estimates rho(p) so with the appropriate substitutions the solution for f is as follows:

$$\rho(p) = f(1 + 4\rho(i)),$$

$$f = \frac{\rho(p)}{1 + 4\rho(i)} .$$

So far in generating the Y variables in the two, three, and five predictor cases, the weights of the combinations, c, d, and f, respectively, have solutio But in each case a weight for the error term is needed. In the Knapp and Swoye algorith, the value a^2can be viewed as r^2, the amount of variance accounted for so $1 - a^2$ is the amount of variance not accounted for and $\sqrt{1 - a^2}$ is the weight the error vector, Z.

In the three multiple predictor cases studied here, formulas for the R^2 val are given below:

$$R^2_{y \cdot 12} = c\rho_{yx_1} + c\rho_{yx_2} = 2c\rho(p),$$

$$R^2_{y \cdot 123} = d\rho_{yx_1} + d\rho_{yx_2} + d\rho_{yx_3} = 3d\rho(p),$$

$$R^2_{y \cdot 12345} = f\rho_{yx_1} + f\rho_{yx_2} + f\rho_{yx_3} + f\rho_{yx_4} + f\rho_{yx_5} = 5f\rho(p).$$

The Y variables were generated as follows:

$$Y = c(X1 + X2) + \sqrt{1 - 2c\rho(p)}Z,$$

$$Y = d(X1 + X2 + X3) + \sqrt{1 - 3d\rho(p)}Z,$$

$$Y = f(X1 + X2 + X3 + X4 + X5) + \sqrt{1 - 5f\rho(p)}Z.$$

Correlations between the criterion variables and each of the predictors were the calculated in the multiple predictor cases

The number of replications was chosen by solving for n_r in the formula for the standard error of the mean of the correlation coefficient given below:

$$\sigma_{\bar{r}} = \frac{\sqrt{\frac{(1 - \rho^2)^2}{n_s}}}{\sqrt{n_r}} .$$

lue for $\sigma_{\bar{r}}$ was arbitrarily set at .01, which was deemed sufficiently for precision in this study. In this formula, ρ is the population lation, rho(p), and was set equal to zero. The symbol, n_s, is the sample and was set equal to 20. Substituting these values into the equation ed n_r, the number of replications, to assume the largest value that would ssible among the values for parameters, rho(p) and n_s, that were chosen for study. The solution for n_r, the number of replications, was 500.

For each combination of N, p, rho(i), and rho(p) and for all r and Z ributions, the means, medians, and standard deviations were calculated.

lts

The means, medians, and standard deviations of the correlation coefficients all values of rho(i), rho(p), and the number of predictors, p, when N=20 ar in Table 1. The same information when N = 50 and N = 100 appears in les 2 and 3 respectively.

The means, medians, and standard deviations of the Fisher's Z transformation the correlation coefficients for all values of rho(i), rho(p), and the ber of predictors, p, when n = 20 appear in Table 4. The same information n N = 50 and N = 100 appears in Tables 5 and 6 respectively.

Inspection of these tables shows that when the population correlation efficient, rho(p), equals zero both the mean of r and the median of r hover ound that value and neither is consistently higher or lower than the other. wever, when rho(p) assumes a nonzero value the median of r is usually larger an mean r. This is because r is a biased statistic and its distribution is gatively skewed when rho(p) is positive. This ordering of the mean and the dian when rho(p) is not zero does not occur in the Fisher's Z distribution.

As N increases both the mean of r and the mean of Z are better estimators f the parameter rho(p). This follows from the Central Limit Theorem. Both he median of r and the median of Z tend to be better estimators of the population

Table 1

Means, Medians, and Standard Deviations for Correlation Coefficients When N = 20

		rho(i)								
		0			.3			.7		
p	rho(p)	$\bar{r}$	Md_r	SD_r	$\bar{r}$	Md_r	SD_r	$\bar{r}$	Md_r	SD_r
1[a]	0	.015	.007	.230						
	.3	.294	.322	.206						
	.7	.690	.706	.126						
2	0	.002	.011	.225	-.004	-.007	.223	.002	-.004	.234
	.3	.300	.316	.214	.296	.299	.208	.297	.311	.209
	.7	.683	.698	.129	.692	.714	.125	.695	.710	.117
3	0	.001	.003	.230	-.009	-.013	.233	.002	-.007	.228
	.3	.295	.313	.213	.289	.305	.214	.295	.316	.211
	.7	b			.686	.703	.126	.687	.703	.126
5	0	-.002	-.004	.233	.008	.007	.227	.004	.000	.221
	.3	.293	.309	.216	.307	.320	.208	.292	.303	.202
	.7	b			b			.694	.714	.120

[a] With one predictor nonzero rho(i) values are undefined.

[b] This combination would generate data which are undefined.

Table 2

Means, Medians, and Standard Deviations for Correlation Coefficients When N = 50

		rho(i)								
		0			.3			.7		
p	rho(p)	$\bar{r}$	Md_r	SD_r	$\bar{r}$	Md_r	SD_r	$\bar{r}$	Md_r	SD_r
1[a]	0	.001	-.001	.141						
	.3	.303	.305	.128						
	.7	.697	.705	.073						
2	0	.005	.000	.142	-.001	-.003	.140	.004	.005	.149
	.3	.294	.307	.132	.300	.305	.131	.304	.305	.130
	.7	.697	.705	.075	.694	.703	.076	.696	.703	.069
3	0	.002	.001	.139	.007	.003	.145	.001	-.002	.142
	.3	.294	.301	.130	.295	.300	.130	.295	.300	.136
	.7	b			.696	.703	.075	.694	.700	.076
5	0	-.002	-.001	.143	-.006	-.009	.144	-.005	-.007	.141
	.3	.299	.303	.129	.300	.305	.129	.295	.300	.128
	.7	b			b			.699	.705	.071

[a] With one predictor nonzero rho(i) values are undefined.

[b] This combination would generate data which are undefined.

Table 3

Means, Medians, and Standard Deviations for Correlation Coefficients When N = 100

		rho(i)								
		0			.3			.7		
p	rho(p)	$\bar{r}$	Md_r	SD_r	$\bar{r}$	Md_r	SD_r	$\bar{r}$	Md_r	SD_r
1[a]	0	.008	.005	.108						
	.3	.299	.303	.091						
	.7	.698	.701	.053						
2	0	.004	.003	.099	-.008	-.009	.101	.009	.012	.097
	.3	.297	.303	.091	.304	.308	.091	.303	.303	.088
	.7	.700	.704	.051	.699	.703	.053	.699	.703	.048
3	0	-.005	-.009	.098	.002	.002	.102	-.001	.000	.097
	.3	.301	.305	.092	.302	.305	.092	.300	.302	.088
	.7	b			.698	.701	.050	.695	.699	.050
5	0	-.002	-.002	.099	.003	.001	.100	-.003	-.002	.100
	.3	.295	.298	.093	.296	.302	.093	.302	.306	.094
	.7	b			b			.699	.702	.051

[a] With one predictor nonzero rho(i) values are undefined.

[b] This combination would generate data which are undefined.

Table 4

Means, Medians, and Standard Deviations for Fisher's Z Transformation of the Correlation Coefficients When N = 20

		rho(i)								
		0			.3			.7		
p	rho(p)	$\overline{Z}$	Md_z	SD_z	$\overline{Z}$	Md_z	SD_z	$\overline{Z}$	Md_z	SD_z
1[a]	0	.016	.007	.243						
	.3	.317	.334	.233						
	.7	.885	.879	.237						
2	0	.002	.011	.238	-.004	-.007	.235	.002	-.004	.247
	.3	.327	.327	.246	.321	.309	.240	.323	.321	.242
	.7	.873	.864	.242	.890	.895	.241	.893	.887	.230
3	0	.001	.003	.244	-.009	-.013	.246	.002	-.007	.241
	.3	.321	.324	.244	.313	.315	.244	.321	.327	.242
	.7	b			.879	.874	.242	.880	.873	.241
5	0	-.002	-.004	.246	.009	.007	.240	.004	-.001	.233
	.3	.319	.319	.248	.334	.331	.240	.316	.313	.231
	.7	b			b			.891	.895	.229

[a] With one predictor nonzero rho(i) values are undefined.

[b] This combination would generate data which are undefined.

Table 5

Means, Medians, and Standard Deviations for Fisher's Z Transformation of the Correlation Coefficients When N = 50

		rho(i)								
		0			.3			.7		
p	rho(p)	$\bar{Z}$	Md$_z$	SD$_z$	$\bar{Z}$	Md$_z$	SD$_z$	$\bar{Z}$	Md$_z$	SD$_z$
1[a]	0	.001	-.001	.144						
	.3	.319	.315	.144						
	.7	.876	.877	.144						
2	0	.005	.000	.145	-.001	-.003	.142	.004	.005	.152
	.3	.309	.317	.146	.316	.315	.147	.320	.315	.146
	.7	.877	.877	.145	.870	.873	.147	.873	.873	.136
3	0	.002	.001	.141	.007	.003	.148	.001	-.002	.145
	.3	.309	.310	.146	.310	.310	.145	.311	.309	.152
	.7	b			.874	.874	.145	.870	.867	.149
5	0	-.002	-.001	.146	-.006	-.009	.147	-.005	-.007	.144
	.3	.315	.313	.145	.316	.315	.145	.310	.310	.143
	.7	b			b			.878	.877	.141

[a]With one predictor nonzero rho(i) values are undefined.

[b]This combination would generate data which are undefined.

Table 6

Means, Medians, and Standard Deviations for Fisher's Z Transformation of the Correlation Coefficients When N = 100

		rho(i)								
		0			.3			.7		
p	rho(p)	$\bar{Z}$	Md_Z	SD_Z	$\bar{Z}$	Md_Z	SD_Z	$\bar{Z}$	Md_Z	SD_Z
1[a]	0	.008	.005	.110						
	.3	.311	.313	.101						
	.7	.870	.869	.102						
2	0	.004	.003	.101	-.008	-.009	.102	.009	.012	.098
	.3	.309	.312	.100	.317	.318	.101	.316	.313	.098
	.7	.874	.875	.100	.873	.872	.104	.872	.874	.094
3	0	-.005	-.009	.099	.002	.002	.103	-.001	.000	.098
	.3	.313	.315	.102	.315	.315	.103	.313	.312	.097
	.7	b			.870	.869	.097	.863	.865	.097
5	0	-.002	-.002	.100	.003	.001	.101	-.003	-.002	.101
	.3	.308	.308	.103	.309	.311	.102	.315	.316	.105
	.7	b			b			.871	.872	.100

[a] With one predictor nonzero rho(i) values are undefined.

[b] This combination would generate data which are undefined.

parameter, rho(p), as N increases as well. Both the mean and the median are consistent estimators. It should be remembered here that when r equals zero, Fishe Z also equals zero. However, when r is .3, Z is .31; and when r is .7, Z is .86.

Inspection of the tables shows that there is no discernible trend in mean r, mean Z, median r, and median Z over levels of rho(i) or levels of p. This seems to indicate that nonindependence of the data does not affect the estimation of the population parameter, rho(p). This is, of course, only for the case when the same parameter is being estimated by all the data.

When evaluating the standard deviations they should be referenced to the known expected values in the cases when independence is not violated. For the r distribution, the standard error of r can be found by substituting the values for the parameters used in this study into the following formula:

$$\sigma_r = \sqrt{\frac{(1 - \rho(p)^2)^2}{n}} .$$

Therefore, the standard error of r when rho(p) is 0 and N is 20 is approximately .224. The standard error of r when rho(p) is .3 and N is 20 is approximately .204 The standard error of r when rho(p) is .7 and N is 20 is approximately .114. When rho(p) is 0 and N is 50 the standard error of r is approximately .141. When rho(p is .3 and N is 50 the standard error of r is approximately .129. When rho(p) is . and N is 50 the standard deviation is approximately .072. The standard error of r when rho(p) is 0 and N is 100 is .1. The standard error of r when rho(p) is .3 and N is 100 is approximately .091. Finally, the standard error of r when rho (p) .7 and N is 100 is approximately .051.

Inspection of Tables 1, 2, and 3 shows that all the standard deviations are close to their expected values. The largest deviation of the standard deviation from its expected value was .015 and that was in an independent case. This deviation is of no practical concern. There is some improvement as N increases

;e standard deviations are consistent estimators, but there are no apparent es over levels of rho(i) or p.

For the Fisher's Z distribution, the values of the standard deviations can und by substituting the values for the parameter used in this study into 'ollowing formula:

$$\sigma_{Z_r} = \frac{1}{\sqrt{N - 3}} .$$

əfore, the standard error of Z when N is 20 is approximately .243. The dard error of Z when N is 50 is approximately .146. Finally, the standard r of Z when N is 100 is approximately .102.

Again inspection of Tables 4, 5, and 6 shows that all the standard deviations very close to their expected values. There is some improvement in the estimates I increases, but there are no apparent changes over either levels of rho(i) or p.

:lusion

The general conclusion, then, is that nonindependence does not affect the imation of either the measures of central tendency or the standard deviations correlation coefficients and for Fisher's Z transformation of the correlation ·fficients when the same population paremeter is being estimated.

REFERENCES

Glass, G. V. (1977). Integrating findings: The meta-analysis of research. In L. S. Shulman (Ed.), *Review of research in education* (Vol. 5, pp. 351-37

Glass, G. V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher, 5*, 3-8.

Glass, G. V. McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social researc* Beverly Hills, CA: Sage.

Hedges, L. V. (1981). Distribution theory for Glass' estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107-128.

Hedges, L. V., & Olkin, I. (1983). Regression models in research synthesis. *The American Statistician, 37(2),* 137-140.

Kendall, M. G., & Stuart, A. (1952). *The advanced theory of statistics* (Vol. 1, 5th ed.). New York: Hafner.

Knapp, T. R., & Swoyer, V. H. (1967). Some empirical results concerning the power of Bartlett's test of the sign of a correlation matrix. *American Educational Research Journal, 4*, 13-17.

Kraemer, H. C. (1983). Theory of estimation and testing of effect sizes: Use in meta-analysis. *Journal of Educational Statistics, 8*(2), 93-101.

Landman, J. R., & Dawes, R. M. (1982). Psychotherapy outcome: Smith and Glass conclusions stand up under scrutiny. *American Psychologist, 37*(5), 504-516

SAS Institute Inc. (1982a). *SAS user's guide: Basics, 1982 edition.* Cary, NC: SAS Institute.

SAS Institute Inc. (1982b). *SAS user's guide: Statistics, 1982 edition.* Cary, NC: SAS Institute.

Time Series Arima Models of Undergraduate Grade-Point Average

Bruce R. Rogers

University of Northern Iowa

Abstract

The Box-Jenkins approach to time series analysis, a regression method analyzing sequential dependent observations, was used to select the appropriate stochastic model for describing undergraduate grade point ages. The technique, applied to approximately a half century of from two universities, suggested that the moving average model ided the optimal fit. Suggestions were made for further exploration PA data.

per presented at the American Educational Research Association, Chicago, April, 85

Whenever a phenomena is observed over time, it is often useful to search for temporal patterns within the data. Economists have studied stock market prices, sociologists have examined population levels, and psychologists have investigated changes in the incidence of depression. For such purposes, a variety of time series analysis procedures have been developed, derived primari from the theory of multiple regression. These techniques require data gathered from at least fifty time periods (McCleary and Hay, 1980, p. 20). Since arcnival data covering this many time periods is not as commonly collected in education as in some other fields, these mathematical approaches are not as widely used in educational research. It is the purpose of this paper to illustrate such an application, using undergraduate grade point averages.

Although educational institutions evaluate their students each term, a single group of pupils is not often evaluated fifty times on the same variable, as would be required for a time series analysis. However, a meaningful time series can be realized by obtaining the average grades given during each of the grading periods across a lengthy time span. For about the last half century, many universities and colleges have adopted a 5-point grading scale, using either the letters A through E or the numbers 1 through 5. Some of the institutions calculated, at each grading period, the average of grades awarded to their students, with the intent of maintaining reasonable consistency in their grading standards both among their departments and across time. Approximately fifteen years ago, reports began appearing that a conspicious increase was occuring each year in the grading patterns at many institutions (Birnbaum, 1977). Although that pattern appears to have abated during the past few years (Suslow, 1977), grades remain at a noticably higher level than to the increase.

A variety of factors have been suggested to explain the phenomena of tutional grade average fluctuation (Birnbaum, 1977), but there has been k of data that support the proposed explanations. Rogers (1983) ned several independent variables (demographic and economic) for the ibility of explaining temporal variation over an extended time frame, found each of them lacking in explanatory power.

Any "explanation" of a pnenomena implies that the phenomena can be uately described. Mathematical models, and regression models in particular, appropriate for such a description, but an examination of the literature ests that most authors rely solely on visual graphs rather than employing ematical modeling. It was the purpose of this study to use a stochastic series approach to generate mathematical models that might appropriately ribe the entire sequence of grade point data.

Method

ple

Grade point average data were collected from two midwestern universities about a fifty year span. For the first, hereafter called University A, a was collected for each year from 1929 through 1982. This data is plotted a time series plot in Figure 1. For the second institution, hereater called versity B, data was collected each year from 1932 to 1982, except for the rs 1943 through 1946, when no data was available. This data is plotted in jure 2.

ocedure

These data were analyzed with the time series analysis procedures ought together in 1970 by George E. P. Box and Gwilyn M. Jenkins, in their lume entitled Time Series Analysis: Forecasting and Control (revised

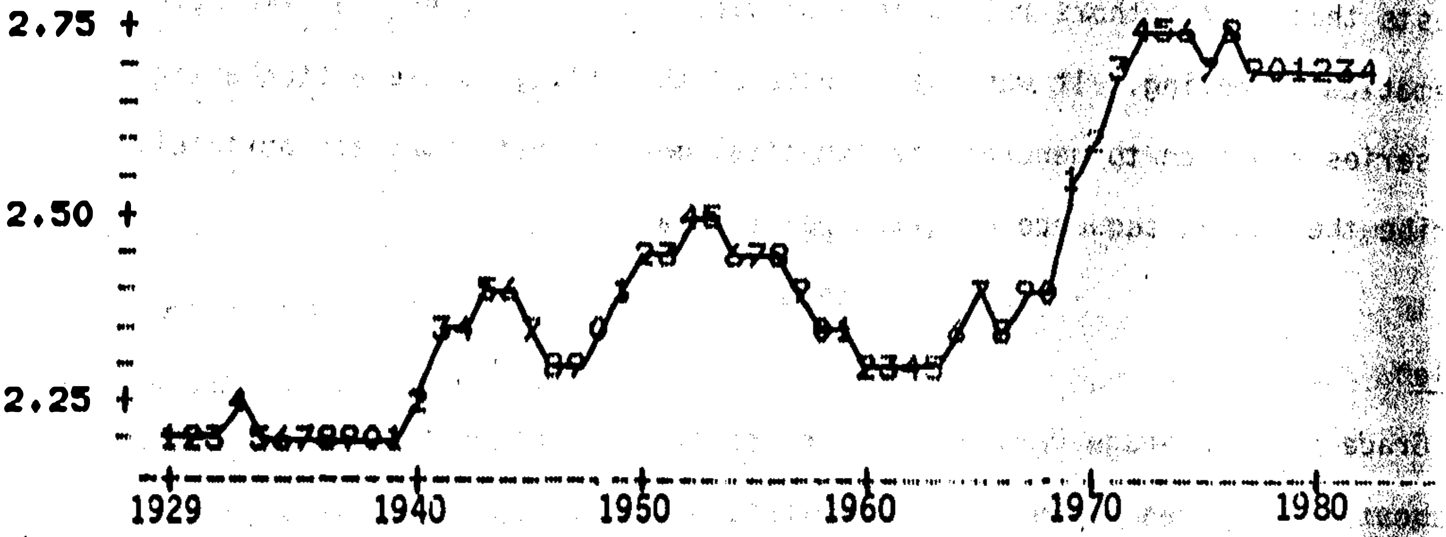


Figure 1. Grade Point Average (GPA) at University A, by year, from 1929 to 1982. (Prior to 1944 the data is for the whole year; afterward it is for fall term.)

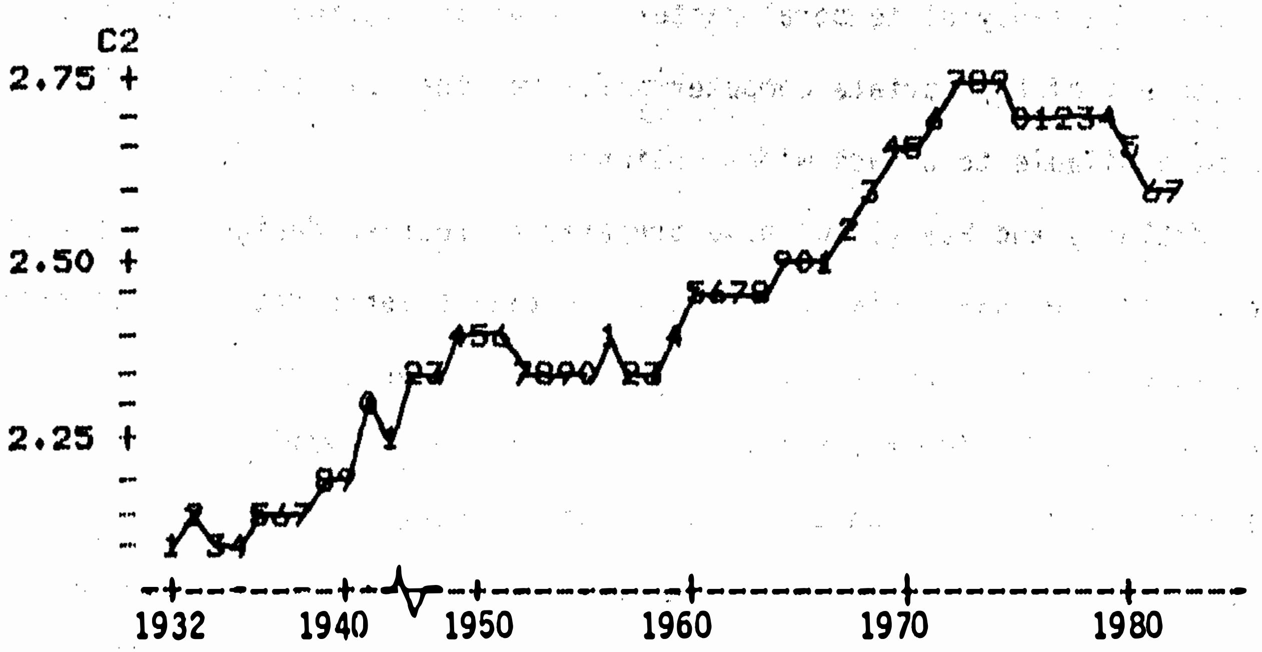


Figure 2. GPA at University 8, by year, from 1932 to 1982 (fall term). Far 1943-1946, data are not available.

edition 1976). These Auto-Regressive Integrated Moving Average (ARIMA) models (often referred to as "Box-Jenkins" models) require a large amount of data. However, when data are collected over an extended time period, as in this study, there is the possibility that the social meaning of the data could change over time. Thus, it becomes difficult to assign the same interpretation to the data at the beginning and end of the series. Nonetheless, the study of temporal patterns is an intriguing one, and with the development of appropriate computer software, the Box-Jenkins methods have become available to a much wider audience.

McCleary and Hay (1980) have prepared a treatise designed to encourage the use of the Box-Jenkins analysis for social science data, and to explicat strategies for both analyzing the data on the computer and presenting the computer output. Their strategies undergird the analysis in this study. The data was processed on a Harris computer, using MINITAB (Ryan, et al., 1982). Other approaches and other computer programs could have been used, but this was the one available for this project. The reader will need to interpret the methodological procedure of this study in that light.

The empirical identification procedures recommended by Box and Jenkins require an analysis of the autocorrelation function (ACF) and the partial autocorrelation (PACF) of the time series. The graphed ACF and PACF for both of the University time series are shown in Figures 3 and 4. The ACF is a set of correlations, each one of which represents the correlation between the original sequence and itself when lagged

k units. For observations close together, e.g., 1 or 2 lags, we most often find a higher correlation than for observations further apart, as is typifie in Figures 1 and 2, where the correlations are slowly dying out as the lags

```
                          Autocorrelations

           -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
             +----+----+----+----+----+----+----+----+----+----+
  1   0.931                           XXXXXXXXXXXXXXXXXXXXXXXX
  2   0.849                           XXXXXXXXXXXXXXXXXXXXXX
  3   0.753                           XXXXXXXXXXXXXXXXXXXX
  4   0.666                           XXXXXXXXXXXXXXXXXX
  5   0.567                           XXXXXXXXXXXXXXX
  6   0.484                           XXXXXXXXXXXXX
  7   0.392                           XXXXXXXXXXX
  8   0.312                           XXXXXXXXX
  9   0.221                           XXXXXXX
 10   0.119                           XXXX
 11   0.026                           XX
 12  -0.042                          XX
 13  -0.084                         XXX
 14  -0.112                        XXXX
 15  -0.101                        XXXX
 16  -0.083                         XXX
 17  -0.060                         XXX


                      Partial Autocorrelations

           -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
             +----+----+----+----+----+----+----+----+----+----+
  1   0.931                           XXXXXXXXXXXXXXXXXXXXXXXX
  2  -0.140                       XXXXX
  3  -0.135                        XXXX
  4   0.028                           XX
  5  -0.145                       XXXXX
  6   0.064                           XXX
  7  -0.137                        XXXX
  8   0.017                           X
  9  -0.134                        XXXX
 10  -0.185                     XXXXXX
 11   0.051                           XX
 12   0.066                           XXX
 13   0.126                           XXXX
 14  -0.008
 15   0.237                           XXXXXXX
 16  -0.010
 17  -0.034                           X
```

Figure 3. Estimated ACF and PACF for GPA. University A.

increase. This dying out phenomena is a consequence of the fundamental tenent of the ARIMA model, namely that the effect of any given input to the system declines over time. (Note that this is just the opposite of a time series of a bank savings account where, assuming a constant interest rate, the compounded interest from the first dollar invested is always larger than that from any subsequent dollar invested.) When the data is properly modeled, the residuals (errors resulting from the model) should be randomly distributed, and thus yield an ACF with with values that are all statistically non-significant. The goal of the Box-Jenkins approach is to find such a model.

The Box-Jenkins approach is a three stage procedure to build a model, consisting of Identification, Estimation, and Diagnosis. Each of these will be illustrated in the following analysis. The cycle iterates until an interpretable solution is found.

University A

Identification.

An examination of the ACF of the raw data (Figure 3) shows that the ACF falls to zero slowly, indicating that there is a strong systematic trend in the data. The most common method for removing this trend is to transform the data by replacing each observation with the difference between it and the preceding observation. When this differencing transformation is complete, the ACF is again computed. Figure 5 shows the ACF for the differences. The values are much smaller, indicating almost random data. However, there are some spikes, which may be due to sampling error or to some systematic process, so further analysis is required.

The PACF is interpreted similar to the ACF, except that each value is the correlation between observations <u>k</u> units apart <u>after</u> the correlation at

Autocorrelations

```
     -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
       +----+----+----+----+----+----+----+----+----+----+
0.939                           XXXXXXXXXXXXXXXXXXXXXXXX
0.899                           XXXXXXXXXXXXXXXXXXXXXXX
0.827                           XXXXXXXXXXXXXXXXXXXXXX
0.749                           XXXXXXXXXXXXXXXXXXXX
0.673                           XXXXXXXXXXXXXXXXXX
0.595                           XXXXXXXXXXXXXXXX
0.511                           XXXXXXXXXXXXXX
0.436                           XXXXXXXXXXXX
0.356                           XXXXXXXXXX
0.291                           XXXXXXXX
0.220                           XXXXXXX
0.174                           XXXXX
0.129                           XXXX
0.093                           XXX
0.060                           XX
0.032                           XX
```

Partial Autocorrelations

```
     -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
       +----+----+----+----+----+----+----+----+----+----+
 0.939                          XXXXXXXXXXXXXXXXXXXXXXXX
 0.142                          XXXXX
-0.263                   XXXXXXXX
-0.180                      XXXXX
 0.003                          X
 0.000                          X
-0.113                       XXXX
-0.002                          X
-0.042                         XX
 0.043                          XX
-0.069                        XXX
 0.116                          XXXX
 0.039                          XX
-0.040                         XX
-0.060                         XX
-0.006                          X
```

Figure 4. Estimated ACF and PACF for GPA. University B.

```
                              Autocorrelations

         -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1
           +----+----+----+----+----+----+----+----+----+----+
  1  0.128                          XXXX
  2  0.141                          XXXXX
  3  0.046                          XX
  4  0.029                          XX
  5 -0.322                  XXXXXXXXX
  6 -0.008                          X
  7 -0.111                       XXXX
  8 -0.065                        XXX
  9  0.086                          XXX
 10 -0.166                      XXXXX
 11 -0.157                      XXXXX
 12  0.008                          X
 13 -0.104                       XXXX
 14 -0.265                   XXXXXXXX
 15  0.025                          XX
 16 -0.085                        XXX
 17 -0.035                         XX


                         Partial Autocorrelations

         -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.
           +----+----+----+----+----+----+----+----+----+----+
  1  0.128                          XXXX
  2  0.126                          XXXX
  3  0.015                          X
  4  0.004                          X
  5 -0.345                 XXXXXXXXXX
  6  0.071                          XXX
  7 -0.038                         XX
  8 -0.025                         XX
  9  0.158                          XXXXX
 10 -0.354                 XXXXXXXXXX
 11 -0.100                       XXXX
 12  0.068                          XXX
 13 -0.145                      XXXXX
 14 -0.133                       XXXX
 15 -0.081                        XXX
 16 -0.162                      XXXXX
 17  0.043                          XX
```

Figure 5. Estimated ACF and PACF for first differences. University A.

intermediate lags has been controlled or "partialled out". The PACF in Figure 3 shows a single spike, which may be the result of what is called a moving average (MA) component. This moving average component can be conceptualized as a random "shock" which is added to each observation to obtain the predicted value for the next observation.

The distinguishing characteristic of a moving average process is the finite duration of the shock. The shock persists for q observations and then is completely suppressed (McCleary and Hay, p. 61). Such a "shock" might be the result of the new grades that are added each term for each particular student. Since the majority of students will leave the institution after four years, the impact of any particular student will vanish when that individual leaves.

From the ACF and PACF we can now tentatively "identify" the model as an ARIMA (0, 1, 1). The zero indicates that there is no auto regressive (AR) term, the middle 1 indicates that differencing is to be used (this is the Integrative (I) term), and the last 1 indicates a moving average (MA) term.

Estimation.

When the estimates of the parameters were computed, it was found that the (0, 1, 1) model produced a t-value of only 1.23 for the MA term. Since this value was not statistically significant at the .05 level (nor anywhere near there), the model was rejected, and the procedure returned to the identification stage.

Identification.

It might be useful at this point to emphasize that since the estimated ACF and PACF are based on very small samples, they are subject to relatively large sampling errors. Consequently, any identification is very tentative.

Because the ACF and PACF for first differences appeared rough, it seeme appropriate to take second differences, i.e., differences between the difference scores. Figure 6 shows the resulting ACF and PACF. They appear more interpretable, suggesting a (0, 2, 1) model. An examination of Figure also suggested that the variance was not constant across time. To attempt to correct this, a logarithmic transformation of the data was performed.

Estimation.

Table 1 shows the results of estimating the (0, 2, 1) model. The movir average parameter of .9767 satisfies the stationarity requirement that its absolute value be less than 1.0, and is also statistically significant at less than the .05 level.

Diagnosis.

The simplest diagnostic procedure is to compare the results of the give model and alternative models. In this way, it can be shown that a particula model is optimal in that neither a simpler nor a more complex model will suffice. The simpler model (0, 1, 1) was already shown to be inadequate. The more complex model (0, 2, 2) yielded a statistically insignificant secon MA term, so it was rejected. The (1, 2, 1) model was also tested, but the AR term was insignificant. Thus, the ARIMA (0, 2, 1) model was accepted as the "best" fit.

The equation generated by this procedure can be conveniently written in the following form: $(1-B)^2 y_t = (1-.9767B)a_t$ where B is the backshift operator, and a_t is the random-shock element (McCleary and Hay, (1980), p. 4(64). The backshift operator is defined as $By_t = y_{t-1}$ and follows the usual algebraic rules. The operator (1-B) represents first differences and $(1-B)^2$ represents second differences.

```
                          Autocorrelations

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
  1  -0.505             XXXXXXXXXXXXXX
  2   0.073                          XXX
  3  -0.063                        XXX
  4   0.196                          XXXXXX
  5  -0.387                XXXXXXXXXXX
  6   0.251                          XXXXXXX
  7  -0.097                        XXX
  8  -0.057                         XX
  9   0.237                          XXXXXXX
 10  -0.139                       XXXX
 11  -0.085                        XXX
 12   0.143                          XXXXX
 13   0.037                          XX
 14  -0.265                   XXXXXXXX
 15   0.217                          XXXXXX
 16  -0.097                        XXX
 17   0.036                          XX
```

```
                       Partial Autocorrelations

          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
            +----+----+----+----+----+----+----+----+----+----+
  1  -0.505             XXXXXXXXXXXXXX
  2  -0.244                    XXXXXXX
  3  -0.201                     XXXXXX
  4   0.118                          XXXX
  5  -0.318                  XXXXXXXXX
  6  -0.139                       XXXX
  7  -0.131                       XXXX
  8  -0.287                   XXXXXXXX
  9   0.236                          XXXXXXX
 10  -0.073                        XXX
 11  -0.179                      XXXXX
 12   0.039                          XX
 13  -0.010                            X
 14  -0.072                        XXX
 15  -0.028                         XX
 16  -0.206                     XXXXXX
 17   0.021                          XX
```

Figure 6. Estimated ACF and PACF for second differences. University A.

Table 1. Parameter estimates for the ARIMA (0, 2, 1) model. University A.

```
FINAL ESTIMATES OF PARAMETERS
NUMBER      TYPE       ESTIMATE        ST. DEV.    T-RATIO
   1       MA  1         0.9767         0.0439       22.26

DIFFERENCING. 2 REGULAR

RESIDUALS.         SS =      0.0191286   (BACKFORECASTS EXCLUDED)
                   DF =      51    MS =   0.0003751
NO. OF OBS.        ORIGINAL  SERIES       54      AFTER DIFFERENCING   52
```

Table 2. Parameter estimates for the ARIMA (0, 2, 2) model. University B.

```
FINAL ESTIMATES OF PARAMETERS
NUMBER      TYPE      ESTIMATE        ST. DEV.    T-RATIO
   1       MA  1        1.1475         0.1224        9.38
   2       MA  2       -0.5302         0.1220       -4.35

DIFFERENCING. 2 REGULAR
RESIDUALS.         SS =     0.0429018   (BACKFORECASTS EXCLUDED)
                   DF =     43    MS =   0.0009977
NO. OF OBS.        ORIGINAL SERIES      47      AFTER DIFFERENCING    45
```

The random shock element a_t is the stochastic component in the equation. the ARIMA model this moving average component can be shown to be mathe-ically equivalent to the exponentially weighted average of all previous ervations (Pankratz, 1983, p. 49, 109; McCleary and Hay, (1980), p. 63).

University B

ntification.

An examination of the estimated ACF and PACF of the raw data (Figure 4) gests that this data is also non-stationary and needs to be differenced. single spike on the PACF suggests a (0, 1, 1) model.

imation.

The (0, 1, 1) model produced an estimate of the Moving Average parameter n a t-value of .23. Since this was far from statistical significance, ifications needed to be made. Second differences were used, since the a appeared to approximate a quadratic trend. The (0, 2, 1) model produced arameter with a t-value of 11.12, which was highly significant.

gnosis.

The model was first diagnosed by comparing it with a more complex model. ordingly, a (0, 2, 2) model was tested. It produced significant t-values both MA parameters, as shown in Table 1. To compare the two models, the n squares of the residuals was computed. The (0, 2, 1) model yielded = .0011274, while the (0, 2, 2) model yielded MSR = .0009977. Finally, 1, 2, 2) model (yet more complex) was tested, but it yielded MSR = 11641. Consequently, the (0, 2, 2) model was favored, since it yielded smallest MSR.

The ACF and PACF for the Residuals of model (0, 2, 2) are shown in ure 7. No spikes are shown at lag 1 or any other lags. The residuals ear to meet the diagnostic criteria, so the model is accepted.

```
                          Autocorrelations

         -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
           +----+----+----+----+----+----+----+----+----+--
 1   0.157                         XXXXX
 2  -0.056                        XX
 3  -0.098                       XXX
 4  -0.023                        XX
 5  -0.122                      XXXX
 6  -0.094                       XXX
 7  -0.131                      XXXX
 8  -0.065                       XXX
 9  -0.055                        XX
10  -0.094                       XXX
11  -0.012                         X
12   0.125                         XXXX
13   0.152                         XXXXX
14   0.064                         XXX
15  -0.086                       XXX
16  -0.013                         X

                Partial Autocorrelations

         -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8
           +----+----+----+----+----+----+----+----+----+--
 1   0.157                         XXXXX
 2  -0.082                       XXX
 3  -0.078                       XXX
 4   0.001                         X
 5  -0.135                      XXXX
 6  -0.065                       XXX
 7  -0.131                      XXXX
 8  -0.065                       XXX
 9  -0.079                       XXX
10  -0.142                     XXXXX
11  -0.034                        XX
12   0.059                         XX
13   0.065                         XXX
14   0.003                         X
15  -0.123                      XXXX
16   0.002                         X
```

Figure 7. Estimated ACF and PACF for residuals from Arima (0, 2, 2) model. University B.

The model can be conveniently written as $(1-B)^2 y_t = (1 - 1.1475B + .5302B^2)a_t$.

Conclusion

This paper has suggested that meaningful mathematical models can be created to describe the time series of changes in the yearly grade point average at a university. The models are very tentative, partly because of the small number of available observations and also because of their relative complexity.

While this paper has not answered the questions about the so-called "grade inflation," it has indicated that a mathematical description of the time series of grades is sufficiently complex to suggest that no simple answer may suffice. The data is unstationary, as shown by the need for differencing. It further appears to be best modeled by an approach that postulates random shocks that persist for only a finite time, yet each of which can be represented as an exponentially weighted average of all previous observations. This perhaps reflects both the influx of new students and the persistent effects of traditional grading practices.

Data for this study was available for only two institutions of higher education, so the generalizability of the results is limited. Studies with data from other institutions would serve to indicate the existence of general patterns across institutions.

References

Birnbaum, R. (1977). Factors relating to university grade inflation. Journal of Higher Education, 48, 519-539.

Box, G. E. P. and Jenkins, G. M. (1976). Time series analysis: Forecasting and control. San Francisco, Holden-Day.

McCleary, R. and Hay, R. A. (1980). Applied time series analysis for the social sciences. Beverly Hills: Sage Publications.

Pankratz, A. (1983). Forecasting with univariate Box-Jenkins models. New York: John Wiley.

Rogers, B. R. (1983). A time series approach to the longitudinal study of undergraduate grades. Paper presented at the annual meeting of the National Council of Measurement in Education, April 13, 1983, Montreal. (ERIC No. ED 235 228)

Ryan, T. A., Joiner, B. L., and Ryan, B. F. (1982). Minitab reference manual University Park, Pa: Pennsylvania State University.

MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 14, NUMBER 1, SPRING 1985

Notes and Tables to Accompany the Presentation: Multiple Regression Analysis with Dichotomous Outcome Variables: Issues and Examples*

Ric Brown

California State University, Fresno

OVERVIEW

The purpose of this 'applied' presentation is to demonstrate the use of multiple regression analysis in situations where the outcome variable is dichotomous and the predictor variables are intervally scaled. The more common procedure in this situation is discriminant function analysis. However, Cohen and Cohen (1975) state:

> "A few moments of reflection will make it apparent that for the special case where two groups are to be discriminated.... the analysis reduces to a single MRC for a single dichotomous Y (which can be coded 1 - 0, or with any other pair of different values). The MRC analysis is mathematically and statistically identical with a CA when p=1; hence, it is identical with a DA for 2 groups. $R_Y^2.12...k$ equals the (sole) $R_c^2(=R_{Y_1}^2)$ and the multiple regression equation is proportional to the discriminant function and hence perfectly correlated with it(p.442)."

Mathematical formulations can be found in Tatsouka (1975).

Issues regarding the use of the general linear model (discriminant function or multiple regression) with qualitative variables is beyond the scope of this presentation. Press and Wilson (1978) argue that logistic regression is preferable to discriminant function analysis when one or more of the discriminating variables is qualitative. However, they also state a preference for discriminant analysis estimators "if the populations are normal with identical covariance matrices."

*Note: Also see Myers, M., Templer, D., and Brown, R. (1984). Coping ability of women who become victims of rape. Journal of Consulting and Clinical Psychology, 52 (1), 73-78.

Paper presented at the American Educational Research Association, Chicago, April, 198

EXAMPLE 1

The research sought to investigate the coping skills of rape victims to determine if some women may be more vulnerable to rape than others. The study investigated five domains: psychosocial competency, mental health, alcohol and drug use, cognitive resources, and physical ability. Seventy-two rape victims and 72 control women were administered psychometric instruments and a biographi inventory. Information was also obtained from significant others. The stronge domain of prediction was psychosocial competency, with the rape victim scoring lower on measures of social presence, dominance, and assertiveness, and higher on external/social locus of control. A past history of alcohol or drug abuse added to the rape-vulnerability profile. Rape victims were more likely to have a past history of psychiatric hospitalization and suicidal thoughts. They did not differ from control women on the Vocabulary subtest of the Wechsler Adult Intelligence Scale-Revised, but they scored lower on the Achievement via Independence Scale of the California Psychological Inventory. Physical ability attributes were not associated with rape vulnerability (see article).

Points:

1) choice of the stepwise model
2) acceptability of the regression approach to journals
3) presentation of the data

EXAMPLE 2

The problem of unwed adolescent pregnancy has been studied in the past primarily as a symptom of individual psychopathology. These studies yielded equivocal results. Gradually, the broader social context of pregnant teenagers began to be studied. Past research pointed to the importance of the family contributing to the problem.

The objectives of this study were to investigate whether family variables could discriminate between the families of unwed pregnant and non-pregnant teens All teen subjects met the research criteria of being unwed, under eighteen years of age, enrolled in local high schools, and living with their families of origin Thirty-one pregnant teen families and 28 non-pregnant teen families comprised the study sample. Each subject completed the Moos Family Environment Scale (FE In addition, each parent completed a questionnaire which included a problem checklist, demographic information, questions about the teen's dating behavior and recent family structural changes.

The hypothesis that incongruence of perception and other family adjustment variables could differentiate the two groups was explored. Pregnant teens were found to have longer boyfriend relationships and fewer problems as rated by the parents. Their family's perceptions were more congruent regarding cohesion and mother/daughter interaction, but less congruent in terms of family conflict (tables 1 and 2).

Points:

1) choice of full model

TABLE 1

Means of Variables by Pregnant/Non-Preqnant Groups

Variable	Group	
	Pregnant (1)	Non-Pregnant (2)
Length of Boyfriend Relationship(mos.)	10.20	3.20
Conflict Incongruence	3.93	2.56
Number of Problems	.8	1.70
Control Incongruence	2.6	2.18
Cohesion Incongruence	3.26	4.0
Organization Incongruence	3.6	3.25
Mother/Daughter Incongruence	28.23	33.0
Family Changes	1.63	1.56
Independence Incongruence	2.93	3.06

TABLE 2

Summary Table of the Regression Analysis with Incongruence of Perception and Other Family Variables

Independent Variables	Beta
Length of Boyfriend Relationship	-.43
Conflict Incongruence	-.28
Total Number of Problems	.20
Control Incongruence	-.09
Cohesion Incongruence	.13
Organization Incongruence	-.08
Mother/Daughter Incongruence	.21
Number of Family Changes	-.02
Independence Incongruence	.03

$R = .67$ $p < .01$

EXAMPLE 3

This study examined the effects of acculturation on adolescent development, specifically focusing on daydreaming as one aspect of coping and adaptation. An investigation of two samples of acculturating (Hispanic and Native American) and acculturated (Caucasian) adolescents revealed two variables that, in combination, significantly differentiated the two groups. These two variables, fear of failure daydreams and distractibility, suggested that acculturating adolescents were more likely to report guilty and fearful daydreaming themes and less likely to report concentration difficulties than their acculturated coparts (tables 3,4 and 5).

Points

1) choice of the stepwise model

Table 3

Point Biserial Correlations of Daydreaming Variables with Acculturation Index

Variables	Correlation
Frequency	.06
Absorption in Daydreaming	.01
Acceptance of Daydreaming	.16
Positive Reactions	-.14
Frightened Reactions	.04
Visual Imagery	.03
Problem-Solving Daydreams	.02
Future in Daydreams	.06
Bizarre and Improbable Daydreams	.04
Mind Wandering	-.16
Achievement-Oriented Daydreams	.07
Hallucinatory-Vividness	.08
Fear of Failure Daydreams	.33
Hostile Daydreams	.01
Guilt Daydreams	.27
Boredom	-.05
Distractability	-.12

Table 4

Summary Table of the Stepwise Multiple Regression Analysis with Acculturation as the Dependent Variable

Independent Variables	Multiple R	R Square	Change in R Square	Simp
Fear of Failure Daydreams (DM)	.33	.11	.11	.
Distractibility (DQ)	.42	.17	.06*	-.

*Variables beyond this point did not significantly account for additional between group variability (PC<05).

Table 5

Acculturating vs. Acculturated Group Means
on the Independent Variables

Variable	Means	
	Acculturating(1)	Acculturated(2)
Daydreaming		
Frequency	35.38	36.54
Absorption in Daydreaming	52.67	52.86
Acceptance of Daydreaming	30.82	28.66
Positive Reactions	30.59	28.16
Frightened Reactions	38.88	39.64
Visual Imagery	32.76	33.38
Problem-Solving Daydreams	30.03	30.34
Future in Daydreams	30.71	31.96
Bizarre & Improbable Daydreams	41.38	41.98
Mind Wandering	32.32	30.24
Achievement-Oriented Daydreams	37.44	38.86
Hallucinatory-Vividness	40.68	42.30
Fear of Failure Daydreams	34.68	39.48
Hostile Daydreams	34.15	39.24
Guilt Daydreams	41.85	46.18
Boredom	41.32	40.60
Distractibility	36.26	34.66

NOTE: A high score on each daydreaming scale means that respondents disagreed with the scale's major theme. For example, a high score on Fear of Failure Daydreams means that the subject reports few fear of failure daydreams.

REFERENCES

Berkeley, J. (1982). The Role of Daydreaming in Acculturating and Acculturated Adolescent Adaptation. Unpublished doctoral dissertation, California School of Professional Psychology, Fresno, California.

Cohen, J. Cohen, P. (1975). Applied Multiple Regression/Correlation Analysis. New York: John Wiley and Sons.

Honeyman, B. (1981). A Study of Unwed Pregnant and Non-pregnant Adolescents. Unpublished doctoral dissertation, California School of Profession Psychology, Fresno, California.

Myers, M.B.; Templer, D.I. and Brown, Ric (1984). Coping Abilities of Women who became Victims of Rape, Journal of Consulting and Clinical Psychology, 52 (1), 73-78.

Press, S.J. and Wilson, S. (1978). Choosing between Logistic Regression and Discriminant Analysis, Journal of the American Statistical Association, 73 (364), 699-705.

Tatsuoka, M.M. (1975). The General Linear Model: Selected Topics in Advanced Statistics, 7, Champaign, IL: Institute for Personality and Ability Testing.

Significant Interaction: I Got What I Needed

Keith A. McNeil and Gail Smith

Dallas Independent School District

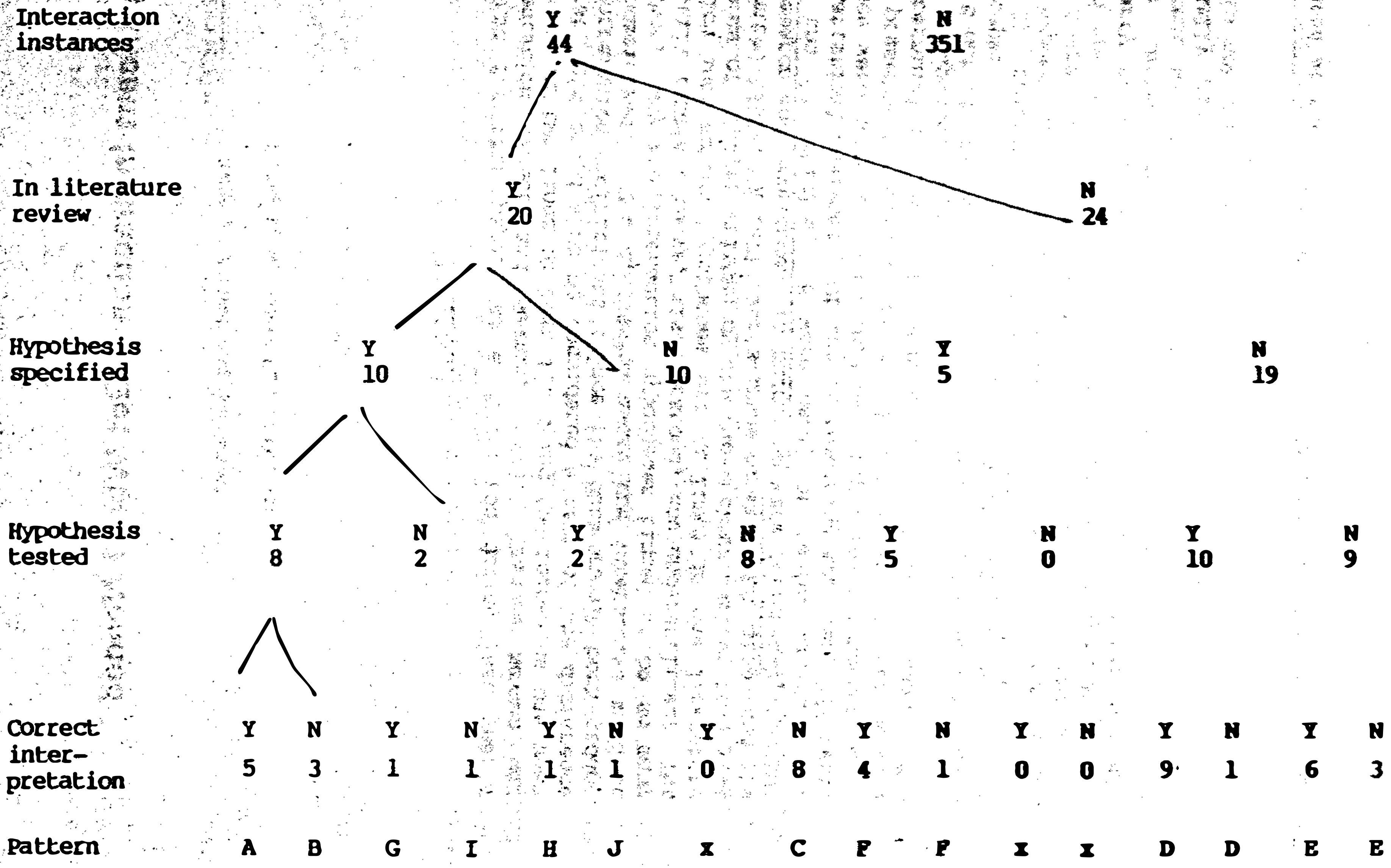
Background

The impetus for this paper was a discussion during last year's
'/SIG presentation (Hoedt and Newman, 1984). Isadore Newman was
cussing a test of two lines of best fit being considered as one when
alluded that this could also be considered a test of the difference
ween two correlation coefficients (since the data within both groups
l been standaridized.) The discussion awoke the interactive mind of
: first author. Why are interaction hypotheses hinted at on so many
nts, but still remain elusive, misunderstood, and underutilized? The
ent to which interaction hypotheses are utilized in the literature
came the focus of a paper written by the two authors earlier this year
:Neil and Smith, 1985). A full year's issue of Urban Education and the
irnal of Research and Development in Education were reviewed by the two
hors. Of the 57 articles, 38 were essay or review articles not
ntaining statistical analyses. Of the 19 remaining articles, 386 tests
significance were computed, with only 44 interaction hypotheses being
ted. The presence (Y) or absence (N) of each aspect of four crucial
ps was determinwd for each of these 44 interaction instances. The
ttern of Y/N responses is presented in Table 1.

In only 5 out of the 44 instances (Pattern A) did the author follow
e four steps: 1) identify the interaction hypothesis in the
terature, 2) specify the interaction hypothesis, 3) test the
teraction hypothesis, and 4) correctly interpret the interaction
pothesis. There were 8 instances of Pattern C, wherein the author
entified in the review of literature juicy interaction hypotheses, but
iled to carry through. Pattern D represents the computer society,
erein the canned computer program automatically provides the
teraction test so the author feels obligated to interpret the results.
at is equally disturbing is the last two bins, Pattern E. Here
teraction is not discussed until the interpretation stage — food for
hought.

'aper presented at the American Educational Research Association, Chicago,
pril, 1985

(from McNeil and Smith, 1985)


Interaction instances Y 44 N 351
In literature review Y 20 N 24
Hypothesis specified Y 10 N 10 Y 5 N 19
Hypothesis tested Y 8 N 2 Y 2 N 8 Y 5 N 0 Y 10 N 9
Correct inter-pretation Y 5 N 3 Y 1 N 1 Y 1 N 1 Y 0 N 8 Y 4 N 1 Y 0 N 0 Y 9 N 1 Y 6 N 3
Pattern A B G I H J x C F F x x D D E E

Note: Y=Yes
N=No

Review of Multiple Linear Regression Viewpoints for Applied Interaction Studies

Interaction hypotheses can easily be tested within the Multiple Linear Regression (MLR) approach, and there has been a history of MLR being taught alongside "complex behavioral science models incorporating interaction and non-linear variables (Kelly, Beggs, and McNeil, 1969; Fraser, 1979; Bottenberg and Ward, 1963). It was therefore predicted that a higher percentage of interaction hypotheses would appear in Multiple Linear Regression Viewpoints, (the journal of SIG/MLR) than in the two journals previously reviewed.

> When the predictors were used to predict the criterion "for the experimental and control groups separately, apparent differences were found in the two regression equations. It was these differences that led to the present consideration of the interaction of the predictors with experimental condition (Group) as a way of exploring the differences statistically."
> (Dinero, 1976)

So begins one of the few research studies which tests an interaction hypothesis in a meaningful way. All issues from 1975 through 1980 were reviewed. Only nine applied studies were found, with 49 of the 506 tests of significance involving interaction. Of the five studies which did consider an interaction hypothesis, two studies fit Pattern B (10 interaction instances), one Pattern F (28 interaction instances), and two Pattern D (11 interaction instances). In no case did the researcher include all four of what we consider to be crucial steps. Additionally, the percentage of interaction hypotheses is lower in Viewpoints than in the two applied education journals discussed earlier. This finding is particularly disconcerting because much has been written in Viewpoints about interaction and how easily one can test it within the MLR framework. The following (selected) review is intended to once again reinforce these interaction notions. Fraser (1979) provides a comprehensive approach to research with MLR. Researchers who haven't "interacted" within the last five years ought to reread the article.

Review of Multiple Linear Regression Viewpoints for Interaction Comments

Why so few researchers test interaction questions remains a puzzle. All canned ANOVA computer programs routinely provide a test for interaction. All stat texts discuss the concept, most in a negative light though. (The Kelly, Beggs, and McNeil (1969) text had the audacity to place curvilinear interaction on the text's cover.) Of most relevance to the members of SIG/MLR is the paucity of good applied interaction studies outside our journal. This is particularly disconcerting given the extensive discussion by numerous authors in Viewpoints. Upon rereading the early volumes of Viewpoints, we were astounded at the frequency and quality of interaction discussions. Desiring the work of these early "interactive pioneers" to not remain shelved, we will quote liberally.

Construction of interaction variables

An interaction variable is reflected in MLR as a product of two variables. If both variables ~~are dichotomous~~ then traditional ANOVA designs are being reflected. If one of the variables is dichotomous and the other continuous, then a difference between groups is being considered (evaluating "the question of homogeneous slopes [Jennings, 1972) or the difference between two correlation coefficients [Hoedt and Newman, 1984]). It has been shown that in the test for homogeneity of regression slopes, both methods of calculating analysis of covariance - traditional ANCOVA and MLR -- are exactly the same" (Newman and Fry 1972). (See also Jennings, 1972 and Williams, Naresh, and Peebles, 1972.)

If both of the variables are continuous then "continuous interaction" (McNeil and McNeil, 1975) or moderator variables are being investigated. Moderator variables "lend somewhat limited support for the use of more complex models. Moderators improve preciction by acknowledging possible interactive effects of the moderator variable with other variables in the regression anlaysis." (Reed, Feldhusen, and Van Modfrans, 1971)

If the variables are actually the same variable, then a higher order effect (curvilinearity) is being implemented. This extension of interaction into curvilinearity was first brought to the senior author's atttention by Jack Byrne during his Doctoral prelims. Dinero (1977) later makes the connection: "Now that one has decided to use interaction terms in his prediction model, he has to decide which ones to include. The predictors raised to the first power, these variables squared or cubed or any of their cross-products may be used."

Dinero (1977) also reiterates the ease and value of conceptualizing research within the MLR approach. "Once a researcher understands how to generate interactions, more avenues of investigation are open. The regression model brings with its flexibility a set of decisions many researchers in the past have either ignored or been unaware of."

Interpretation of interaction

Many researchers avoid interactions because of interpretation problems. Here is what Viewpoints authors have to say about the interpretation issue.

> "A significant interaction hampers the interpretation of main effects, but the positive view is that a significant F test of interaction tells us how to appropriately limit our generalization" (Spaner, 1977).

> "A final word of warning is that second and higher order interactions must be interpreted with great care, if meaningless or erroneous conclusions are not to be drawn from research data." (Brebner, 1972)

> "In general, significant three-way interaction is seen to reflect different two-way interactions: if the ABC interaction is significantly different from zero, then either AB varies across C, AC varies across B, or BC varies across A. In any case, these differences would be manifest by significant cross-products of the standardized predictors." (Dinero, 1977)

> "Indeed the value of need for interaction tests has been grossly underemphasized in MLR studies. I suspect that this phenomenon arises out of a misunderstanding, perhaps even fear, of a significant interaction finding." (Spaner, 1977)

> "Indeed the value of need for interaction tests has been grossly underemphasized in MLR studies. I suspect that this phenomenon arises out of a misunderstanding, perhaps even fear, of a significant interaction finding." (Spaner, 1977)

McNeil and Beggs (1971) accepted the reality of interaction and challenged researchers to think about directional interactions — thus fully utilizing the power of their statistical test. No directional hypotheses have appeared in our review of Viewpoints.

Nonlinear predictors

> "Since many of the simplest functional relationships in the physical sciences have been found to be non-linear or interactive, we find it interesting that few non-linear relationships have been established in the behavioral sciences, especially since most behavioral scientists would maintain that human behavior is no less complicated than physical behavior." (McNeil, Evans, and McNeil, 1979)

> There are "two reasons for including non-linear terms - either the expected functional relationship is non-linear, or the way the construct has been originally measured needs to be modified." (McNeil, 1976)

> "A more important situation occurs when there is theoretical or empirical justification for the inclusion of such a variable." (McNeil and Spaner, 1971)

Interpretation problems with non-linear terms have been addressed.

> "When quadratic and interaction terms are significant, however, interpretation is made more difficult. Still, an attempt at interpretation seems somewhat better than ignoring the problem or assuming it does not exist." (Reed, Feldhusen, and Van Modfrans, 1971)

> "The range of manipulations available in order to test forms of curvilinearity is endless. However, contrived departure from linearity in regression models will not make trivial predictors into important ondes." (Jordan, 1971)

Nonlinear criterion

There are two instances that come to mind when a nonlinear criterion would be used. One instance is when the functional relationship is indeed nonlinear (McNeil, Evans, and McNeil, 1979). The Pythagorean Theorem is one such example. Any criterion that is a ratio of one variable to another is another example. A second instance when a nonlinear criterion would be used is when the measure of the construct does not map the construct, and some rescaling of the measure is necessary (McNeil, et al 1979).

Potential problems

When continuous variables are multiplied to reflect the interaction term several potential problems must be avoided. One potential problem is that the product is dependent on the means and variances of the original scores. Thus, researchers might want to standardize the variables before obtaining the product (Dinero, 1977). McNeil and McNeil (1975) also discussed the scaling effect on the resultant R^2. The product of two continuous predictor variables may not accurately reflect the interaction. The precictor variables must be rescaled such that the product term does match the expectations of the criterion.

Miscellaneous techniques

The search for interaction in the hypothesis generating mode has been well stated by Dinero (1977).

> "Given the problem of shrinkage, any regression anlaysis should be run in two phases, the first to estimate and the second to corroborate. This being the case, it may be just as wise to explore with the data of the first phase, to the extent of plotting the scatter diagrams, and use this information to select the interaction term to be used in the second phase. This type of exploration would seem to be almost a necessity in educational and phychological studies where there is little such comparative data available, where interaction has been something more to be avoided than awaited, and where complex aptitude-treatment interactions could bring exciting new interpretations to old data."

A computer program has been written to assist in finding the interactions which account for the most variance.

> "The primary value of AID-4 to the task scientist is its ability to identify the maximum amount of variance in the criterion which can be accounted for by the predictors available; it relieves the task scientist of the trial-and-error task of attempting to identify the various relevant combinations of linear and non-linear interaction terms presently required by the multiple linear regression technique. The splitting process of AID-4, being based upon maximizing the between sums-of-squares and minimizing the within sums-of-squares, automatically takes all present interaction into account, indiciting the maximum variance predictable in the cirterion from the predictors."
> (Koplyay, 1972)

Finally, the dectection of interaction is one of the major advantages of the "regression model" in evaluating compensatory education programs (McNeil and Findlay, 1980).

Discussion

The purpose for providing all the quotes in the previous sections was to document the interaction efforts made by authors in Viewpoints. The fact that the majority of these references are over 10 years old reflects more our concern for being aware of, and implementing existing methodology, rather than our lack of concern for improving existing methodology.

Given that this methodology exists for studying interaction questions, why don't more researchers look at interaction? We don't have the answer, but we have some thoughts, and we will present them grouped by the four major hypothesis testing steps.

With respect to literature review, most authors do not review interaction results, and when they do, they review them poorly. Furthermore, part of ~~the~~ publish or perish mentality is to invent new predictor variables, rather than try to increase the amount of variance accounted for. Finally, most researchers do not understand that different results from two studies implies an underlying interaction variable

In this world of posthoc orthogonal contrast coding and alpha protection levels few researchers realize that an interaction hypothesis can be specified all by itself, if no other question is of interest. But most of the statistics texts insist on a step-by-step procedure, looking at interaction in particular ways. What ever happened to the notion of the research question guiding the statistical tool?

With respect to the actual testing of the hypothesis, we have three major concerns. First, canned ANOVA programs generally don't allow for testing specific interaction questions. Second, canned MLR programs encourage the inclusion of linear terms first. (Stepwise linear programs, though of value for some purposes, totally ignore the testing of a specific hypothesis.) Third, most statistics texts still present the interaction question as being valuable only for meeting assumptions -- to reject so that main effects can be tested.

The fourth step in hypothesis testing, interpretation, also causes some problems for those considering interaction questions. Unfortunately most of our quoted Viewpoints authors acknowledge that interpreting an interaction result can be difficult. But if interaction is significant, then that is reflecting reality -- and shouldn't it be more valuable to make a "difficult" interpretation of reality as it is, than to make some "easier" statement about some constrained aspect of reality. Perhaps researchers need to become more familiar with significant interaction.

Summary

Fortunately, for us, the summary of our paper was published in Viewpoints over 12 years ago.

> "Perhaps one of the most overused assumptions within multivariate studies in educational research is that only simple linear relationships exist among the variables. Although interactive effects have been acknowledged within analysis of variance studies, the logical extension to regression analysis has rarely been actualized (Reed, Feldhusen, and Van Modfrans, 1971).
>
> "Too often, even plausible interactions are ignored and all subjects are lumped together and, hence, treated as similar. Our conceptual theories have long ago turned to distinct groupings, and it is about time that our statistical procedures reflect this empirical possibility." (Newman, Lewis, and McNeil, 1973).

Unfortuanately these comments seem to still be appropriate today. Hopefully tomorrow they will not be appropriate.

Epilogue

An examination of why interaction studies are not conducted in one specific area may shed some light on possible solutions. The two authors have been involved with educational program evaluations for several years. As such, we function as the program evaluator, providing evaluation information to the program manager.

In order to study an interaction question, the evaluator first needs to understand interaction concepts and be able to calculate interaction effects. Second, the evaluator must be able to translate these concepts into terms that the program manager can understand. Third, the interaction question must become of interest to the program manager, a person who often wants to use only the simplest of statements.

Collection of interaction information

Program managers usually want all students to be provided the best possible educatinal opportunity. This notion is usually envisioned in the same treatment for all. Denying treatments or parts of treatments is often not desired, and obtaining additional information from students is sometimes difficult if not impossible.

Verbal outcome

The program manager has a vested outcome in the program. Often the program has been devised by the manager and therefore the manager "knows" that the best program has been devised. Providing the same program to all students probably costs less, is easier administratively, and is usually more defensible to outside interests. The program manager is hard put to take the neutral stance towards the program that evaluators easily take.

Implications if interaction is significant

First, the program evaluator must clearly communicate to the program manager the implications of a significant interaction. Then the program manager must incorporate this finding into next year's program, a task which requires additional administrative attention.

When programs are constructed around significant interactions much additional administrative work is required. Program descriptions and guidelines must clearly reflect such interactions. Alternative programs must be delineated and procedures must be identified to get the right students (and probably the right teachers) into those programs. Different teaching materials may be required for the various programs, as well as different staff development. Classroom monitoring and program evaluations will continually need to incorporate those interactive variables. Consequently, additional administrative effort and commitment is required. Significant interactions imply that the KISS (Keep It Simple Stupid) principle is no longer applicable.

Roadblocks to replacing significant interactions

Everyone, including program managers, knows that results need to be replicated. The extent to which replicated results can be generalized to different settings and different students is usually an interesting question. But in the educational arena programs are often changed due to factors unrelated to evaluation results: a) new local, state, or Federal mandates, b) change in program manager, c) availability of personnel to plan and implement the program, and d) availability of funds.

Some possible next steps for SIG/MLR members

Now that we've a) established that adequate methodology exists to investigate interactive questions, b) documented that few interactive questions are being investigated, and c) specified some of the roadblocks to studying interactions in our field, we would like to propose some remediation.

First, we should all strive in our own daily endeavors to consider interaction hypotheses. We understand the methodology and can provide exemplary behavior to other researchers.

Second, we could infuse other SIGs and the various AERA Divisions. We challenge each of you to become involved in another SIG, to spread the interaction hypothesis.

Third, many of you participate in other national or regional educational meetings where more program managers are in attendance. These program people need to know that interaction questions can be tested -- for behind every good program manager is an interaction hypothesis.

References

Bottenberg, R. A. and Ward, J. H. Applied Multiple Linear Regression. Lackland Air Force Base, Texas: Aerospace Medical Division, AD 413128, 1963.

Brebner, M. A. Conditions for no second-order interaction in multiple linear regression models for three factor anlysis of variance. Viewpoints, 1972, 3(1), 46-57.

Dinero, T. E. An empirical example of the use of interaction terms in the multiple regression model. Multiple Linear Regression Viewpoints, 1977, 7(2), 75-100.

Fraser, B. J. A multiple regression model for research on teacher effects. Multiple Linear Regression Viewpoints, 1979, 9(3), 37-52.

Hoedt, K. and Newman, I., Testing the hypothesis of a difference between P_1, and P_2 using independent and dependent samples. Paper presented at the meeting of the American Educational Research Association, New Orleans, March, 1984.

Jennings, E. Linear models underlying the anlysis of covariance, residual gain scores and raw gain scores. Viewpoints, 1972, 3(1), 17-24.

Jordan, T. E. Curvilinearity within early developmental variables. Viewpoints, 1971, 1(1), 53-77.

Kelly, F. J., Beggs, D. L., McNeil, K. A., Eichelberger, T. and Lyon, J. Research Design in the Behavioral Sciences: Multiple Regression Approach. Carbondale: Southern Illinois University Press, 1969.

Koplyay, J. B. Automatic interaction detector AID-4. Viewpoints, 1972, 3(1), 25-38.

McNeil, K. Position statement on the roles and relationships between stepwise regression and hypothesis testing regression. Multiple Linear Regression Viewpoints, 1976, 6(4), 46-49.

McNeil, K. A, & Beggs, D. L. Directional hypotheses with the multiple linear regression approach. Viewpoints, 1971, 1(1), 89-102.

McNeil, K., Evans, J., & McNeil, J. Nonlinear transformation of the criterion. Multiple Linear Regression Viewpoints, 1979, 9(5), 1-9.

McNeil, K., & Findlay, E. Evaluating Title I early childhood programs: Problems, the applicability of Model C, and several evaluation plans. Multiple Linear Regression Viewpoints, 1980, 10(4), 41-50.

McNeil, K., & McNeil, J. Some thoughts on continuous interaction. Viewpoints, 1975, 5(3), 41-46.

McNeil, K. A. and Smith, G. Educationally significant interaction. Paper presented at the meeting of the Southwest Educational Research Association, Austin, Texas, January, 1985.

McNeil, K. A. and Spaner, S. D. Brief report: Highly correlated predictor variables in multiple regression models. Multivariate Behavioral Research, 1971, 6, 117-125.

Newman, I., Lewis, E. L., & McNeil, K. A. Multiple linear regression models which more closely reflect Bayesian concerns. Viewpoints, 1972, 3(1), 71-77.

Newman, I., & Fry, J. Proof that the degrees of freedom for the traditional method of calculating analysis of covariance and the multiple regression method are exactly the same. Viewpoints, 1972, 3(1), 42-45.

Reed, C. L., Feldhusen, J. F., & Van Mondfrans, A. P. Regression models in educational research. Viewpoints, 1971, 1(1), 78-88.

Spaner, S. D. What inferences are allowable with a significant F in regression analysis? Multiple Linear Regression Viewpoints, 1977, 7(2), 62-74.

Williams, J. D., Maresh, R. T., & Peebles, J. D. A comparison of raw gain scores, residual gain scores, and analysis of covariance with two modes of teaching reading. Viewpoints, 1972, 3(1), 2-16.

If you are submitting a research article other than notes or comments, I would like to suggest that you use the following format if possible:

Title
Author and affiliation
Indented abstract (entire manuscript should be single spaced)
Introduction (purpose—short review of literature, etc.)
Method
Results
Discussion (conclusion)
References

All manuscripts should be sent to the editor at the above address. (All manuscripts should be camera-ready.)

It is the policy of the M.L.R. SIG-multiple linear regression and of *Viewpoints* to consider articles for publication which deal with the theory and the application of multiple linear regression. Manuscripts should be submitted to the editor as original, double-spaced, *camera-ready copy.* Citations, tables, figures and references should conform to the guidelines published in the most recent edition of the *APA Publication Manual* with the exception that figures and tables should be put into the body of the paper. A cost of $1 per page should be sent with the submitted paper. Reprints are available to the authors from the editor. Reprints should be ordered at the time the paper is submitted, and 20 reprints will cost $.50 per page of manuscript. Prices may be adjusted as necessary in the future.

A publication of the Multiple Linear Regression Special Interest Group of the American Educational Research Association, *Viewpoints* is published primarily to facilitate communication, authorship, creativity and exchange of ideas among the members of the group and others in the field. As such, it is not sponsored by the American Educational Research Association nor necessarily bound by the association's regulations.

"Membership in the Multiple Linear Regression Special Interest Group is renewed yearly at the time of the American Educational Research Association convention. Membership dues pay for a subscription to the *Viewpoints* and are either individual at a rate of $5, or institutional (libraries and other agencies) at a rate of $18. Membership dues and subscription requests should be sent to the executive secretary of the M.L.R. SIG."