Figure 1 – ReadMe

Introduction

Criterion.xlsx is a general-purpose linear modeling program. It was written for my students to 1) remedy some difficulties seen in typical commercial computer package output, e.g., SPSS, SAS, Stata, Minitab, etc., 2) provide theoretical narrative for consideration that is specific to each result, 3) provide relevant results in APA tables and 4) provide relief from the necessity of commercial packages. It is available for you to use as desired; all I ask is that citations to the work that led to this program be provided. **REDACTED**.

Input

The program is currently set up to accommodate up to 1,000 cases, and 20 predictor variables. On the "**Input**" tab, the number of predictors must be entered in cell **C2** and the number of cases in **C3**. In addition, the α used for inferential testing is entered in **C4.** As it is so common, I have ".05" there, but 1) feel free to change it, and realize that no decision is a decision for .05. Starting with the criterion (or "dependent" if you wish) variable, enter short variable names in row 8 starting in column **C** ("C" for **C**riterion). Then enter the scores for each of those variables for each case. You can also cut and paste all of these values from elsewhere if desired. In column **A**, starting in row **8** an "ok" will be rendered if the row meets the number of variables criterion in **C2**, and a green "OK" will be rendered in **A5** if **all** rows thusly check out. Only then will any analysis take place. So, until there is an OK, check your input. As well, unlike in aforementioned software packages, if you enter intractable data (e.g., a variable's variance is zero, or a set of predictors includes a perfect linear dependency), you will get a red message to that effect and no relevant results until you fix the issue. Rather than arbitrarily omitting variables due to one of these problems (as is common is commercial packages), I consider it **your** responsibility to consider these issues, diagnose, and make appropriate choices.

Output

Various sheets contain the results of the analyses, where appropriate in APA tabular or text format, and annotations to help clarify certain mistakes that are commonly made. Keep in mind that, except for analyses that require your further input, **all** relevant analyses are automatically calculated, thus relax, depending on the data set size, this may take some time.

The "**APA rs**" and "**APA rs Bon**" sheets include an intercorrelation matrix with embedded means and standard deviations, and the same using Bonferroni-type corrections.

A "**Scatterplot**" tab allows easy plotting of all bivariate relationships by selection of the variable number ("1", "2", etc.). A legend of numbers and names is provided for that selection. You can change these to get a scatterplot of any variable pair you like; it will be rendered automatically. As well, histograms and all potentially useful univariate statistics for each variable are calculated.

The "**APA Model**" tab includes both the typical table for overall regression model performance and a textual entry (in bold), as either may be preferred.

The “**APA Coefficients**” tab includes a typical table for consideration of regression coefficients. Because of the common misinterpretation of these inferential tests, the bivariate correlations of variables with the criterion are also shown as what is **not** being tested, but are, of course, not to be included in the table.

The “**APA Generalizability**” tab includes a variety of estimators that regard the generalizability of the regression model, with extensive discussion dedicated to the distinction between the goals of **Explanation** (**ρ^2**) and **Prediction** (**ρ_{cv}**^{2}).

The “**Residuals**” tab gives the usual regression residuals, but also the resampling PRESS residuals and diagnostics.

The “**ps**” tab gives all the p-values for the bivariate correlations. These will usually not be needed as the * type table note is used in the aforementioned APA correlation tables. This is included for rare situations in which it is needed.

The "**R2 Increment**" tab gives a flexible mechanism for investigating the unique contribution of any proper subset of predictors to the accuracy of the total model. This is done "real-time" with as many such tests as desired. (This should be **R^2 Increment**, but I have found no way to control fonts in tabs.)

The "**Moderation**" tab allows testing first order moderations including any two predictors. Again, this is done "real-time" so, as many such tests as desired can be accomplished. Non-dichotomous variables are automatically centered. In the moderation test of the product term, which variable is logically considered the "**M**oderator" is irrelevant, but, given a significant moderation test, it does matter for subsequent "split-sample" correlations, so you must specify which variable is the moderator (M). Then, if the test is significant, the correlation (and plots) between the other variable (X) and the criterion are automatically calculated for subjects at or below, and above, the median of the moderator . You can also specify an alternative split-point, other than the median. If the moderator is dichotomous, then the median, and any value you input are ignored, and the split is for subjects at each level of the dichotomy.

The "**Mediation**" tab allows testing of a simple mediation (X => M => Y) effect. Sobel's test (Aroian adjusted SE) is used. The indirect and direct effects are also apportioned as percentages.

The **"Ridge"** tab provides a ridge regression including plots of 1) the ridge trace, 2) reduction in R^2, 3) reduction in Beta variance, and 4) increase in correlation between the Betas and the original predictor- criterion correlations. Solutions for user input ks are also supported.

If the criterion variable is determined to be dichotomous (automatic) the **"MANDDA"** tab provides two-group contrasts for all multivariate (MANOVA) as well as univariate questions. All know effect sizes are also produced. As well, the mathematical linkage between the regression results, the multivariate tests, and the LDF are demonstrated. On the univariate side, all univariate contrasts are automatically presented with a Bonferroni adjustment for multiple hypothesis testing as suggested by Huberty and Morris (1989).

PDA Casewise, given a dichotomous criterion variable, renders a variety of classification models.

Regression classification accuracy assigns Ss to groups to which their regression predicted score is closest. This is the same as a linear PDA with prior probabilities estimated by group size and equal costs of misclassification.

A linear PDA is also produced which renders the probabilities of group membership, p(group|X), highlighted in red if the case was a "miss." As well, the "typicality" probability p(X|group) is produced. This is only of interest if small as a metric of the case being an outlier. Finally, for you to get a "feel" for how it works, exploration of unequal prior probabilities and costs of misclassification is available interactively, with resulting group probabilities and classification accuracies presented.

This analysis is with L-O-O [Huberty's (1994) label] cross-validated hits, using Lachenbruch's matrix shortcut. As far as I know, this joint ability is not available in commercial software packages, yet we statisticians often tell users to consider unequal costs of misclassification 😉

PDA Summary gives the inferential tests of hit-rate in respect to both the proportional and maximum chance criteria, as well as the ***I*** index of Huberty (1994). This is done for classification by 1) regression, 2) the PRESS cross-validated regression (equivalent to "L-O-O" or Lachenbruch U), 3) the linear PDA, 4) the linear PDA with input priors and cost of misclassification, and 5) the L-O-O cross-validated linear PDA with input priors and costs. All the casewise classifications for these are on the **PDA Casewise** sheet.

PDA Increment offers the same flexible format to investigate the contribution of any proper subset of variables to the classification accuracy of the full model as in the **R2 Increment** page for regression. The difference, as introduced in Morris & Huberty (1995), is that the joint distribution of hits and misses for the full and restricted model is needed, as the correct test statistic regards a correlated proportion McNemar's (1947) z, with more precise binomial mid-p (Fagerland, Lydersen & Laake, 2013) probabilities used. Moreover, for the two-group design treated herein, there are three such tests of increment to classification accuracy available -- that for each group as well as all subjects.

LR, given a dichotomous criterion, produces a logistic regression using the Newton-Raphson iteration method. If iteration fails because there is not a solution, such is noted, and **no results are presented**. This is unlike some commercial packages that have the misleading practice of producing the last iterated solution as a result with a small footnote that says that the maximum number of iterations allowed has been exceeded. All popular metrics of effect size are also included.

LR Firth presents Firth's (1993) penalized LR. This is done with the same tolerance as in **LR**, and regardless of LR's successful iteration. Iteration with Firth's penalized LR will converge regardless of complete or quasi-complete separation.

LR Increment is isomorphic to the PDA Increment tab, producing the same sort of results based on classification accuracy, but using a different mathematical model: maximum-likelihood for LR, rather than least-squares in PDA. The more traditional full vs. restricted contrast of the Log Likelihood is also produced, but I encourage you to also consider classification accuracy, which is, in my opinion given too little consideration in LR use.

PDA LR Comp renders a contrast between the classification performance of PDA and LR using the same technique as in the **PDA Increment** sheet, but this time the contrast is between mathematical algorithms rather than between variable subsets within algorithm.

Many other pages are only for calculations and are thus hidden. As well, the spreadsheet is locked so that we (I include myself) will not be able to destroy formulas.

Figure 2 – Input (Regression)

	A	B	C	D	E	F	G
2		*# Predictors =*	4				
3		*N (cases) =*	30				
4		*α =*	.050				
5	*OK*						
6		Enter Short Variable names in ***Row 8*** ; Criterion in column C, Predictors in D and on →					
7	***Check:***		***Criterion***	***Predictors →***			
8	***ok***	***Case # ↓***	**GPA**	**GRE-Q**	**GRE-V**	**MAT**	**AR**
9	***ok***	***1***	3.2	625	540	65	2.7
10	***ok***	***2***	4.1	575	680	75	4.5
11	***ok***	***3***	3	520	480	65	2.5
12	***ok***	***4***	2.6	545	520	55	3.1
13	***ok***	***5***	3.7	520	490	75	3.6
14	***ok***	***6***	4	655	535	65	4.3
15	***ok***	***7***	4.3	630	720	75	4.6
16	***ok***	***8***	2.7	500	500	75	3
17	***ok***	***9***	3.6	605	575	65	4.7
18	***ok***	***10***	4.1	555	690	75	3.4
19	***ok***	***11***	2.7	505	545	55	3.7
20	***ok***	***12***	2.9	540	515	55	2.6
21	***ok***	***13***	2.5	520	520	55	3.1
22	***ok***	***14***	3	585	710	65	2.7
23	***ok***	***15***	3.3	600	610	85	5
24	***ok***	***16***	3.2	626	540	65	2.7
25	***ok***	***17***	4.1	575	680	75	4.5
26	***ok***	***18***	3	520	480	65	2.5
27	***ok***	***19***	2.6	545	520	55	3.1
28	***ok***	***20***	3.7	520	490	75	3.6
29	***ok***	***21***	4	655	535	65	4.3
30	***ok***	***22***	4.3	630	720	75	4.6
31	***ok***	***23***	2.7	500	500	75	3
32	***ok***	***24***	3.6	605	575	65	4.7
33	***ok***	***25***	4.1	555	690	75	3.4
34	***ok***	***26***	2.7	505	545	55	3.7
35	***ok***	***27***	2.9	540	515	55	2.6
36	***ok***	***28***	2.5	520	520	55	3.1
37	***ok***	***29***	3	585	710	65	2.7
38	***ok***	***30***	3.3	600	610	85	5

Read Me | **Input** | APA rs | APA rs Bon | Scatterplot | APA Model | APA Coefficients

Figure 3 – APA R

This is the standard APA format for presentation of an intercorrelation matrix. This would be true even if interest is only in the correlations among variables, with no regression. If that is the case, then you may simply enter your variables ignoring which is designated “Criterion” in the Input tab. Then ignore all other tabs except "APA rs Bon." You may find need to adjust column widths for your particular data in any table, but the basic structure and "lines" will remain the same.

Table 1

Title such as: Correlation of Criterion with all Predictor Variables

Variable	M	SD	1	2	3	4	5
1. GPA	3.31	.60	—				
2. GREV	565.37	48.66	.61*	—			
3. GREQ	575.33	83.03	.58*	.47*	—		
4. MAT	67.00	9.25	.60*	.27	.43*	—	
5. AFR	3.57	.84	.62*	.51*	.41*	.52*	—

*p < .05

Figure 4 – APA R Bon (Bonferroni adjustment and alternative suggestions)

This is the intercorrelation matrix using Bonferonni corrected ps for all correlations.
With p variables, there are p (p - 1) / 2 correlations, thus the corrected α= .00500
Possibilities in addition to Bonferroni (Holm, Hochberg, Sidak, FDR, etc.) are available and should also be considered as contenders.

Table 1

Title such as: Correlation of Criterion with all Predictor Variables

Variable	M	SD	1	2	3	4	5
1. GPA	3.31	.60	—				
2. GREV	565.37	48.66	.61*	—			
3. GREQ	575.33	83.03	.58*	.47	—		
4. MAT	67.00	9.25	.60*	.27	.43	—	
5. AFR	3.57	.84	.62*	.51*	.41	.52*	—

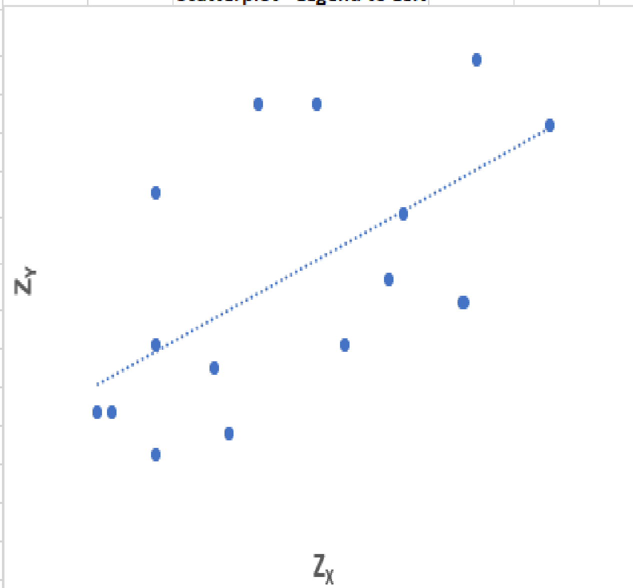
*p < .00500

Figure 5 – Scatterplot (of any pair of variables)

	A	B	C
8	Variable	2	1
9	1. GPA	X=ZGRE-Q	Y=ZGPA
10	2. GRE-Q	1.226	-.189
11	3. GRE-V	.198	1.311
12	4. MAT	-.932	-.522
13	5. AR	-.419	-1.189
14		-.932	.645
15		1.842	1.145
16		1.328	1.645
17		-1.343	-1.022
18		.814	.478
19		-.213	1.311
20		-1.241	-1.022
21		-.521	-.689
22		-.932	-1.356
23		.403	-.522
24		.712	-.022
25		1.246	-.189
26		.198	1.311
27		-.932	-.522
28		-.419	-1.189
29		-.932	.645
30		1.842	1.145
31		1.328	1.645
32		-1.343	-1.022
33		.814	.478
34		-.213	1.311
35		-1.241	-1.022
36		-.521	-.689
37		-.932	-1.356
38		.403	-.522
39		.712	-.022

←*Enter the numbers of two variables in B8 and C8 from the legend for which you wish a scatterplot.*

Scatterplot - Legend to Left



Univariate Descriptives

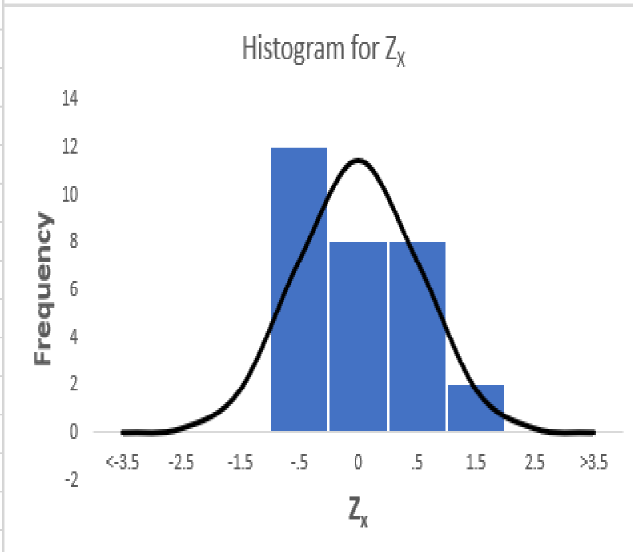
	GRE-Q	GPA
Mean	565.3667	3.3133
Median	555.0000	3.2000
Mode	520.0000	2.7000
SD	48.6603	.5998
Variance	2367.8264	.3598
SE_{mean}	8.8841	.1095
Skewness	.3300	.2887
Kurtosis	-1.1585	-1.3759
Min	500.0000	2.5000
Max	655.0000	4.3000
Semi-Quartile Range	42.5000	.6500

As SE_{skew} and $SE_{kurtosis}$ estimates depend only on N, their values are the same for all variables.

SE_{skew} = .4269

$SE_{kurtosis}$ = .8327

r = .6105



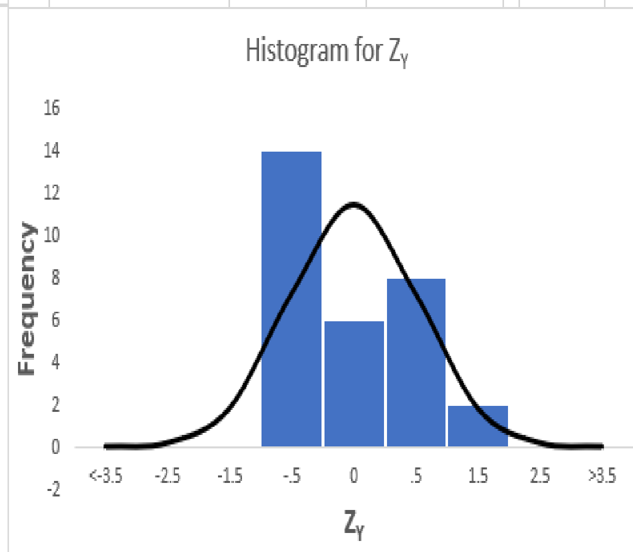


Figure 6 – APA Model

Presentation of information regarding accuracy of the total regression model is often included in text, rather than in a table. If such a table is deemed necessary, it is provided below. In respect to SPSS's version, the R^2 is added, as this is the effect that is being tested. Cohen's f^2, a usual index of effect size, is also included. Note that the "Adjusted R Square" of SPSS is NOT to be included, at least with that name. The explanation and all "generalization" information, more correctly documented, is in a separate table in the "Generalization" tab.

The presentation in text is usually something like:

"The model predicted a significant percentage of criterion variance, $R^2 = .640$, $F(4,25) = 11.129$, $p < .001$, $f^2 = 1.781$."

Table 1

Title such as: R^2, Sum of and Mean Squares, dfs, F, p and Cohen's f^2

Source	Sum of Squares	df	Mean Square	R^2	F	p	f^2
Regression	6.682	4	1.671	.640	11.129	<.001	1.781
Residual	3.753	25	.150				
Total	10.435	29					

Figure 7 – APA Coefficients

This table includes information about individual coefficients and is usually included in regression publications. Take special care in discussing the tests of the coefficients. These tests are for ***partial*** slopes (the Bs); stated alternately, for each predictor, it is a test of whether that predictor **adds** significant predictive accuracy to the model, in addition to that afforded by the remaining predictors. For this reason, an additional statistic, $\mathbf{\Delta R^2}$, is added to this table in respect to that which SPSS includes. This represents the increment to R^2 afforded by the addition of the respective variable to the model having all the remaining predictors in it. This, after all, is that which is being tested. This is **not** a test of the predictive accuracy afforded by only that row variable. If you simply wish to attend to whether a variable, by itself, is a significant predictor, the bivariate correlation should be used. Those are in the intercorrelation matrix (APA rs or APA rs Bon tabs). This interpretive mistake occurs so often in the literature that those r_{xy}s are included here for contrast, but they should not be in the table as they are not tested there. The βs are hidden in column D if you wish them, however the CI is in respect to B, not β. The βs are not in the APA (v 7) regression example table.

Table 1

Title such as: *Regression Weights, R^2 Increments, Tests, and VIFs.*

Variable	B	SE	ΔR^2	95% CI		t(25)	p	VIF	r_{xy}
				LL	UL				
Constant	-1.734	.950		-3.691	.222	-1.826	.080		
GREV	.004	.002	.068	.000	.008	2.182	.039	1.5	.61*
GREQ	.002	.001	.030	-.001	.004	1.453	.159	1.5	.58*
MAT	.021	.010	.069	.001	.041	2.187	.038	1.5	.60*
AFR	.145	.113	.024	-.088	.377	1.281	.212	1.7	.62*

Note. CI = confidence interval; LL = lower limit; UL = upperlimit.

Figure 8 – APA Generalizability

R^2 is the proportion of variance in the criterion predicted by the multiple regression model in the sample at hand. Because we are almost always interested in applying the model outside of our sample, the question quite naturally arises as to how well the model will **generalize**. However, we must be precise in our meaning of "generalize." One question regards how accurately our sample based R^2 estimates the population "R^2" (herein labeled $\boldsymbol{\rho^2}$). That is, what R^2 would we obtain if the entire population were available? Another question regards how accurately our sample weights will be able to predict the criterion if applied to the population (labeled $\boldsymbol{\rho_{cv}}^2$ – "cv" for cross validated). These are often regarded, as goals of **explanation** and **prediction**, respectively. These generalizability questions apparently appear similar, as they are often confused in the literature (see documentation in Huberty & Mourad, 1980), but they are indeed different. My experience has been that researchers are most often interested in prediction, but frequently present an estimator of $\boldsymbol{\rho^2}$ rather than $\boldsymbol{\rho_{cv}}^2$ due to this confusion; perhaps because a $\boldsymbol{\rho^2}$ estimator is readily available. There are several estimators for $\boldsymbol{\rho^2}$ and $\boldsymbol{\rho_{cv}}^2$. This page provides essentially all these estimators. I strongly advise against naively using SPSS's "Adjusted R Square" simply because it is available. It is an estimator of $\boldsymbol{\rho^2}$ due to Ezekiel (1930); if explanation is your goal, then it is fine. In my view, part of the problem arises from a name (Adjusted R Square) that communicates nothing about the rationale of the index. Other contender estimators of $\boldsymbol{\rho^2}$ included herein are those of Olkin and Pratt (1958) and Wherry (1931).

However, if your goal is to estimate how accurately your regression weights will **predict** the criterion upon application to new samples, then use one of the formula estimators of $\boldsymbol{\rho_{cv}}^2$ (Browne, 1975; Darlington, 1968; Lord, 1950; Nicholson, 1948; Rozeboom, 1978; Stein, 1960), or alternatively, and in my opinion preferably, you can use the distribution free resampling estimate of cross-validated prediction accuracy, "PRESS," by Allen (1971). You could also offer all the relevant estimators. The most important lesson here is to know what you are estimating, use the appropriate estimator(s), and document what they are. If you elect to use this table in a publication, you will need to set this Word page to landscape. As PRESS is from raw data, an MSE is also available; it is = ***.1881***

Table 1

Title such as: Generalizability in respect to ρ^2 and ρ_{cv}^2 for model

Estimators							
ρ^2_{cv} (Prediction)					ρ^2 (Explanation)		
ρ^2_{NL}	ρ^2_{SD}	ρ^2_{B}	ρ^2_{R}	R^2_{P}	$\rho^2_{E(SPSS\ uses)}$	ρ^2_{W}	ρ^2_{OP}
Nicholson(1948) Lord (1950)	Stein (1960) Darlington (1968)	Browne(1975) using ρ^2_{OP}	Rozeboom(1978)	Allen(1971)	Ezekiel(1930)	Wherry(1931)	Olkin & Pratt(1958)
.5133	.4971	.5564	.5297	.4765	.5828	.5989	.6013

Figure 9 – R^2 Increment

Consideration of the contribution of a subset of variables to the predictive accuracy of a model that contains other variables is one of the most important theory building tools we have. The question is, for a model containing k_1 variables, what is the contribution of a subset of those variables containing k_2 variables? The k_2 variables need to constitute a "proper" subset; therefore, all the k_2 variables are among the k_1 variables, but there are also variables among the k_1 that are not among the k_2, therefore $k_1>k_2$. This theory testing method is valid for consideration of any model (logically with at least two variables), but in the case of multiple regression an F statistic is available: $F(k_1-k_2),(N-k_1-1) = (R_{k1}^2 – R_{k2}^2)/((1-R_{k1}^2)/(N-k_1-1))$.

This sheet allows you to test the contribution of any proper subset of variables to the full model introduced in the **Input** sheet. The variables' names for the model are automatically listed in row 15. Simply enter an "X" under each variable (row 16) to be included in the subset to be tested, and an "Enter" after each. This is "real-time"; as you add or omit Xs, the ΔR^2 (the difference in R^2s between the "k_1" and "k_2" models, an appropriate Effect Size), F and p change accordingly. Therefore, you may consider any subsets of interest.

Place an X below variables for which Increment test is sought:

GRE-Q	GRE-V	MAT	AR
X	X		

ΔR^2 (ES)	F(2,25)	p
.1481	5.146	***.013***

Figure 10 – Moderation

In its simplest form, a Moderation question regards whether the relationship between two variables, X and Y, is consistent across the range of a third variable, M. **Interaction** is a synonym that tends to be used more in ANOVA; however, ANOVA can only be used with nominal variables, whereas the more general least-squares linear model employed in regression can be used with predictors manifesting any scale of measurement. This **Moderation** sheet is for exploring possible moderating effects of predictors on other predictors. The mathematics of testing a moderation is through a product term, thus, to consider whether M moderates the relationship between X and Y, we would create a three-predictor regression model predicting Y from X, M and X*M. It is the test of the **product term** (highlighted in **green**) that is a test of moderation. To control multicollinearity incurred in using variables and functions of variables in the same model, we often "center" the predictors; this simply means subtracting the mean from the variable. This can be done with a transformation in statistical packages but is done automatically here. (The necessity of centering is not without argument, for instance see Kromrey & Foster-Johnson, 1998; I typically fall on the side of centering.) If a moderation term is significant, we usually explore more deeply to consider the exact magnitude of the relationship between X and Y at "high" and "low" levels of M. These levels are often determined with a median split on M, but the split can be done at any interesting level you wish if group sizes do not become too small. You can also consider the relationship at multiple split levels.

To obtain a Moderation analysis all you need to do is to put ***one*** "X" and and ***one*** "M" (denoting the variable's role) under (Row 16) for any two of your predictors (listed here). You may explore as many moderation analyses with different variables and split points as you wish; the calculations are automatically updated.

Place an X and an M (the Moderator) below two variables for which a moderation test is sought:

GRE-Q	GRE-V	MAT	AR	
		x	m	
			AR Median:	3.40

If desired, alternative to median for split on AR (E22):
(Ignored if M dichotomous.)

Table 1

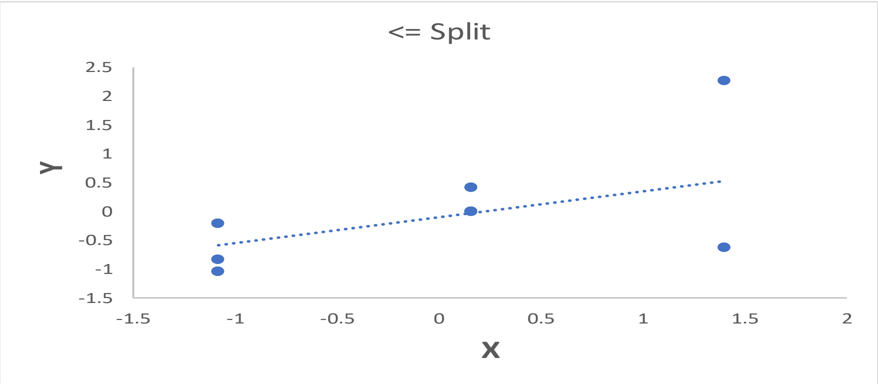
Title such as: *Moderation Test for predictors "X" and "Z," Regression Weights, R^2 Increments, Tests, and VIFs.*

Variable	B	SE	β	ΔR^2	95% CI LL	95% CI UL	t(22)	p	VIF
Constant	3.404	.088			3.224	3.583	38.534	.000	
MAT	.028	.010	.430	.131	.008	.048	2.791	.010	1.409
AR	.350	.112	.489	.165	.122	.577	3.130	.004	1.449
Product	**-.023**	**.011**	**-.284**	**.070**	**-.046**	**.000**	**-2.044**	**.051**	**1.146**

Note . CI = confidence interval; LL = lower limit; UL = upper limit.

As moderation term is significant, correlations at or below, and above, Moderator Split (input or median):

Split OK	<= Split (N=16)	< Split (N=14)
r(MAT,GPA)	**.6145**	**.4128**



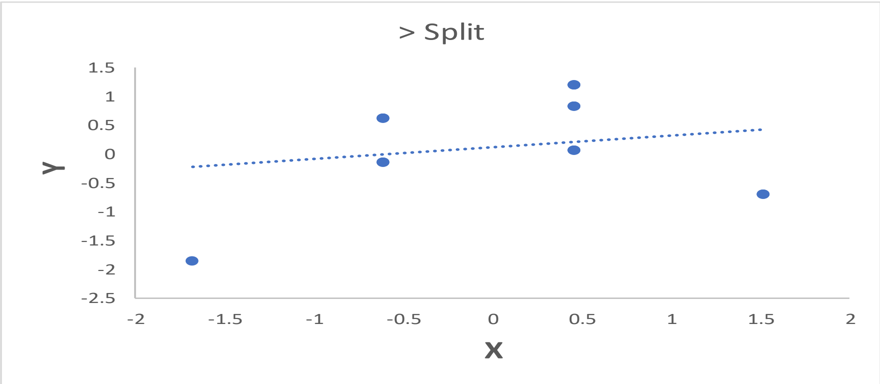


Figure 11 – Mediation

Like moderation, a **Mediation** question always involves at least three variables. An important distinction is that mediation always posits a **causal** proposition, whereas moderation does not depend on such. In its simplest form, this entails the premise that a variable, **X**, causes variation in **Y**, but that the variation arises, at least partially, **"through"** a mediating variable, **M.** That is, the model proposes that at least part of the influence of X on Y is indirect; it occurs because X affects M, and then, due to X's influence on M, M affects Y. Referring to the usual mediation graphic (available in the Mediation pdf), whether there is such an indirect effect (**a*b**) or not, X may also influence Y directly (**c'**). Parsing the total influence of X on Y (**c**) into the indirect and direct "paths" is the point.

There are several methods used to test mediation. Barron and Kenny's (1986) approach involves multiple steps. Sobel (1982) produced a single direct test of mediation, and variants (in respect to the SE) of that test have been proffered. Preacher and Hayes (2008) proposed use of the bootstrap. None of these tests are natively available in statistical packages (e.g., SPSS, SAS), although Hayes (2013) offers a macro for SPSS and SAS that will accomplish both the Sobel test and the bootstrapped estimation. This sheet allows you to select any two variables to be X and M. A mediation test (p in green -- Sobel, with a SE consistent with Aroian, 1944/1947), estimates and tests all parameters, and parsing the proportion of effect that is indirect and direct (except in the case of "Inconsistent" mediation, MacKinnon, Fairchild & Fritz, 2007), will then be accomplished. This is done "real-time", so you may serially select as many X and M combinations as you like, receiving the result for each.

Place an X and an M (the Mediator) below two variables for which a mediation test is sought:

GRE-Q	GRE-V	MAT	AR
		X	M
	Effect	t	p
a	.0476	3.2622	.0029
b	.2998	2.6013	.0149
c'	.0249	2.3858	.0243
c	.0392	4.0126	.0004

Test of Mediation (Indirect) Effect: ***MAT => AR => GPA***

Effect	SE	z	p
.0143	.0072	1.9778	.0479

Direct % =	63.61%
Indirect % =	36.39%

Figure 12 – Ridge Regression (part 1: Contextual discussion and calculations)

Ridge regression stabilizes the βs by reducing their variance, which led Darlington (1978) to helpfully dub the technique (among others) as "Reduced Variance Regression." The cost of that bias (in respect to **O**rdinary **L**east **S**quares) is a commensurate reduction in $R^2_{y.\hat{y}(ridge)}$. In "generalized ridge," this is done by adding a small constant "k" to the predictor intercorrelation matrix before inversion as $\boldsymbol{\beta^*} = (\mathbf{R_{xx}} + k\mathbf{I})^{-1}$, where **β*** are the ridge estimated Betas. Of course, if k=0, **β=β*** you are back to OLS. Articles and presentations, most frequently regarding the estimation of k, are far too numerous to include. Hoerl and Kennard (1970), the originators of ridge regression, used a graphical method, the "Ridge Trace," to consider where stabilization occurs. This is simply a plot of the β*s as a function of k. Of interest is that, as k increases and the variance of the β*s decreases, at some point (the trend is not necessarily monotonic) the β*s become more and more closely proportional to the simple correlations, r_{xy}, between predictors and the criterion (Darlington, 1978). Philosophically, one can see this as movement away from least-squares to naïve weighting by simple predictor-criterion "validities" wherein multicollinearity has less and less effect. Here, as is traditional, the range of k plotted is [0,1]. In addition, row (#37) illustrates an inordinately large k=100. Row (#38) uses $\mathbf{r_{xy}}$ as weights and a "k" as "↑∞" to signify "as k increases to infinity" as these are the **β*** to which ridge is proportionately asymptotic. Row (#39) uses an easy to calculate k (1/F) proposed by Lawless and Wang (1976). The final row (40) is for you to explore. Put in any positive k you wish -- see what happens:-) Note that in addition to ridge regression, the Lasso and Elastic Net techniques are also used for taming unruly least-squares **β**s.

					*Predictor β*s→*			
	$r(\beta, r_{xy})$	Var(β)	$R^2_{y.\hat{y}(ridge)}$	**k**	**GRE-Q**	**GRE-V**	**MAT**	**AR**
	.166	.003372	.640373	.00	.32348	.21124	.32195	.20226
	.183	.002506	.640303	.05	.30961	.21146	.30810	.20615
	.199	.001922	.640150	.10	.29781	.21070	.29631	.20812
	.215	.001511	.639965	.15	.28752	.20929	.28602	.20880
	.231	.001212	.639771	.20	.27838	.20745	.27689	.20857
	.246	.000990	.639579	.25	.27015	.20531	.26865	.20772
	.261	.000820	.639396	.30	.26264	.20297	.26115	.20641
	.276	.000688	.639224	.35	.25574	.20051	.25425	.20477
	.290	.000583	.639062	.40	.24935	.19797	.24786	.20290
	.304	.000500	.638912	.45	.24339	.19538	.24191	.20087
	.318	.000432	.638772	.50	.23781	.19277	.23633	.19871
	.331	.000376	.638643	.55	.23256	.19016	.23108	.19648
	.344	.000330	.638522	.60	.22760	.18756	.22613	.19419
	.357	.000291	.638410	.65	.22290	.18499	.22144	.19188
	.370	.000258	.638306	.70	.21844	.18245	.21698	.18955
	.383	.000230	.638209	.75	.21419	.17994	.21274	.18721
	.395	.000206	.638118	.80	.21013	.17748	.20869	.18489
	.407	.000186	.638033	.85	.20625	.17506	.20482	.18258
	.419	.000168	.637953	.90	.20254	.17268	.20111	.18030
	.430	.000152	.637879	.95	.19898	.17035	.19756	.17804
	.442	.000139	.637808	1.00	.19555	.16807	.19415	.17582
Very large k (100):	.999	<.000001	.635650	**100**	.00597	.00568	.00591	.00606
r_{xy} as coefficients:	1.000	inapt	.635616	↑∞	.61048	.58145	.60423	.62072
Lawless & Wang k:	.196	.002024	.640185	**.0899**	.30006	.21092	.29856	.20784
Explore Enter your k:	.176	.002810	.640345	.03	.31487	.21152	.31335	.20487

Figure 13 – Ridge Regression (part 2: Plots)

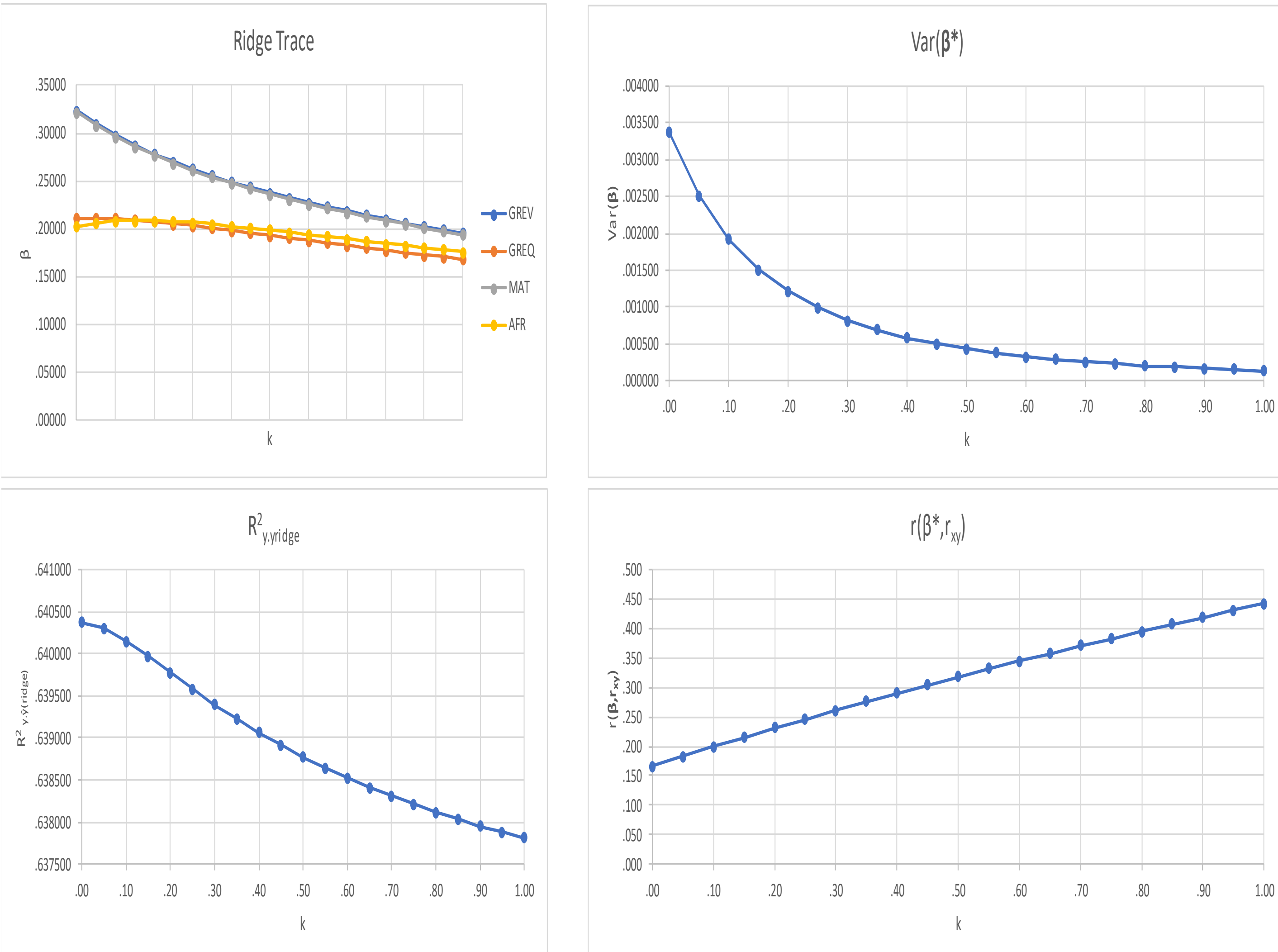


Figure 14 – Ridge Regression (Plots for Longley Data)

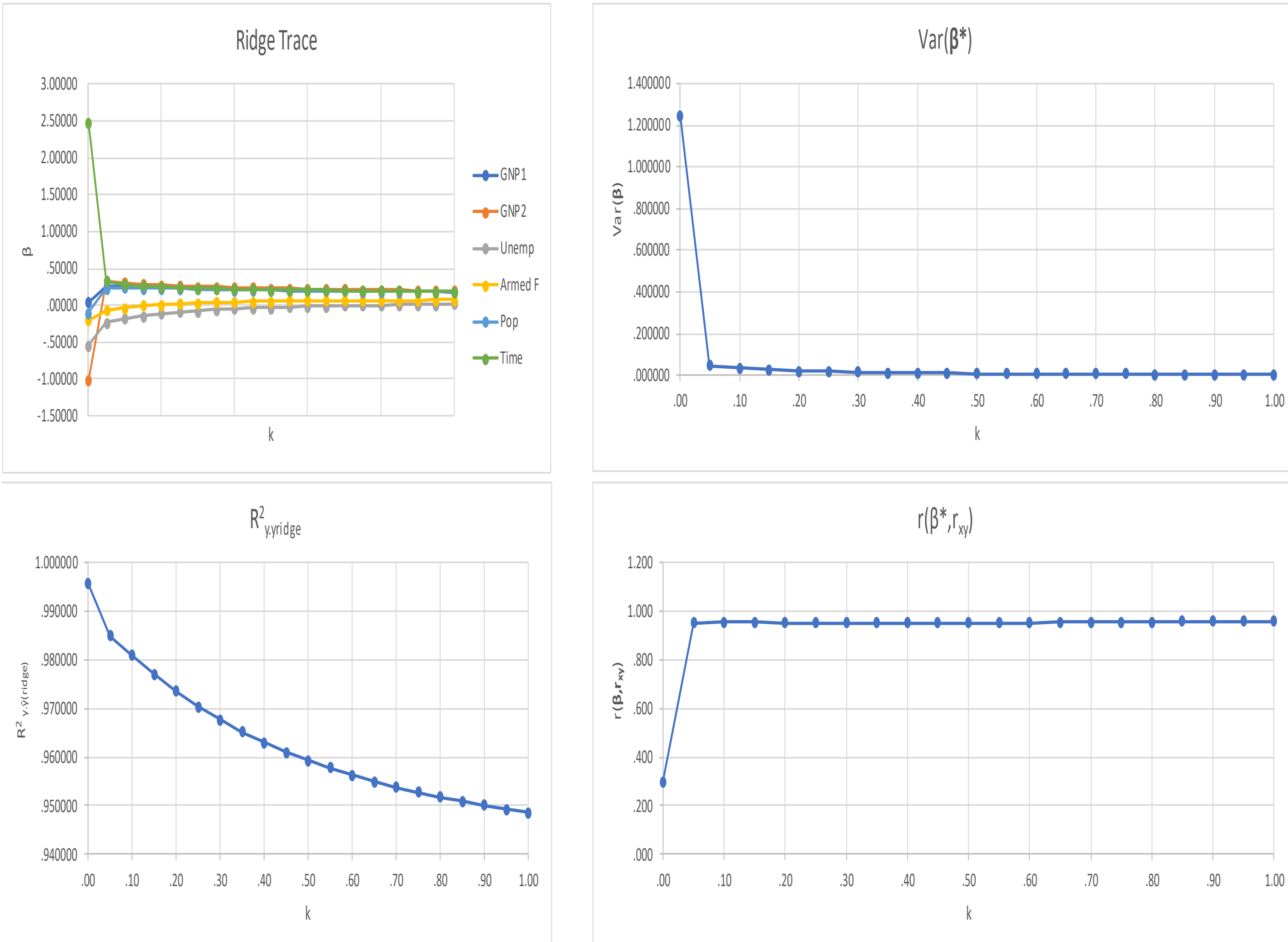


Figure 15 – Input (Dropout with 8 predictors)

	A	B	C	D	E	F	G	H	I	J	K
2		*# Predictors =*	8								
3		*N (cases) =*	162								
4		*α =*	.050								
5	*OK*										
6		Enter Short Variable names in ***Row 8*** ; Criterion in column C, Predictors in D and on →									
7	***Check:***		***Criterion***	***Predictors →***							
8	*ok*	***Case # ↓***	**dostat**	**schools8**	**repeats8**	**reading8**	**math8**	**lang8**	**science8**	**socst8**	**dfsf8**
9	*ok*	***1***	2	4	1	3	3	3	3	3	4
10	*ok*	***2***	2	3	1	3	3	3	3	2	6
11	*ok*	***3***	2	3	2	3	3	3	3	3	5
12	*ok*	***4***	2	5	1	3	3	3	3	3	7
13	*ok*	***5***	2	3	1	2	2	2	2	2	3
14	*ok*	***6***	2	4	1	2	3	2	2	3	4
15	*ok*	***7***	2	6	1	2	3	3	2	2	1
16	*ok*	***8***	2	4	0	2	2	2	3	3	0
17	*ok*	***9***	2	5	0	2	2	3	2	2	0
18	*ok*	***10***	2	4	1	2	2	2	2	2	3
19	*ok*	***11***	2	3	0	2	3	2	2	2	5
20	*ok*	***12***	2	3	0	2	3	2	2	2	4
21	*ok*	***13***	2	4	0	4	4	3	5	3	1
22	*ok*	***14***	2	4	1	3	2	2	2	3	7
23	*ok*	***15***	2	3	1	1	1	1	1	1	7
24	*ok*	***16***	2	7	1	3	3	3	3	3	1
25	*ok*	***17***	2	7	1	3	3	3	2	2	2
26	*ok*	***18***	2	3	2	2	3	3	3	3	3
27	*ok*	***19***	2	5	2	3	2	2	2	2	4
28	*ok*	***20***	2	5	0	3	3	3	3	3	3
29	*ok*	***21***	2	3	1	2	2	2	3	3	2
30	*ok*	***22***	2	4	0	2	3	3	2	2	1
31	*ok*	***23***	2	5	0	3	2	3	3	3	4

Figure 16 – MANDDA

If the criterion is dichotomous, that fact will automatically be detected, as well as the indices that determine the two groups established. If dichotomous, this sheet calculates 1) all multivariate tests (including a test of homogeneous covariance matrices) and effect sizes for a MANOVA between the two groups, 2) all univariate tests (selection denoted by a test of HOV) and effect sizes, and 3) the DDA LDF. The parallel between the multivariate tests and regression on this dichotomous criterion (imported from other sheets, as such is automatically calculated) is illustrated; the test of that R^2 is necessarily the same as all multivariate tests. As well, all known multivariate effect size estimates (except $\omega^{2,}$ which is based on unbiased variance estimators, and is thus, also different in the univariate case) are shown to be the same as R^2 in the two-group case. In addition, that the DDA LDF is proportional to the regression weights is also demonstrated. Results for the appropriate t-test (Gosset as "Student", 1908; or Welch, 1947) is highlighted in green. Although multivariate and univariate contrasts are both provided, in no case should one consider the multivariate "MANOVA" test as omnibus in respect to the univariate tests. As well, if multiple univariate tests are the interest, the accumulated error-rate in respect to α must be controlled (Huberty & Morris, 1989). As mentioned in the "**APA rs Bon**" sheet, there are several alternatives (Bonferroni, Holm, Hochberg, Sidak, FDR, etc.). All ps that are < the Bonferroni adjusted α entered on the "**Input**" sheet are ***bolded and italicized*** . You may use alternative corrections; the use of Bonferroni is because it is simple, and to guide you to do **something**. Multivariate tests require each group N to be greater than the number of predictors, thus, if this is not the case, multivariate results will not be provided. Univariate tests require each group N > 2, else they will not be provided. Obviously, larger Ns are required for stable inference.

Box M:	64.541485	df1	36	Wilks' F:	7.163209
Box "c":	.063870	df2	34679.3	B-Pillai F:	7.163209
c':	.005175	b	38.49889	Roy F:	7.163209
c'-c^2:	.001096	Hot F	7.163209		

dostat	N
1	51
2	111
Total	162

Univariate Tests

	dostat=1		dostat=2		HOV?		Gosset (=σ²)		Welch (≠σ²)			Univariate ES		DDA α Regression	
	Mean	SD	Mean	SD	Bartlett χ2(1)	p	t(160)	p	t	df	p	d (pooled)	rxy²	LDF	B/LDF
schools8	2.31	1.05	3.40	1.42	5.653	***.017***	4.877	***<.001***	5.439	128.03	***<.001***	.8250	.129	-.6227	-.1589
repeats8	.24	.89	.68	.80	.723	.395	3.147	***.002***	3.030	88.84	***.003***	.3456	.058	-.4067	-.1589
reading8	3.29	.76	2.67	.99	4.723	***.030***	4.005	***<.001***	4.425	125.07	***<.001***	.5028	.091	.5545	-.1589
math8	3.35	.96	2.79	.90	.290	.590	3.620	***<.001***	3.534	91.72	***<.001***	.4343	.076	-.0320	-.1589
lang8	3.49	.92	2.77	1.01	.477	.490	4.309	***<.001***	4.447	105.08	***<.001***	.5577	.104	.2987	-.1589
science8	3.22	.99	2.65	.97	.021	.885	3.440	***<.001***	3.418	95.65	***<.001***	.4372	.069	.0270	-.1589
socst8	3.25	1.07	2.70	1.06	.015	.902	3.070	***.003***	3.053	95.88	***.003***	.4187	.056	-.3096	-.1589
dsfs8	1.31	1.56	2.92	2.26	8.559	***.003***	4.592	***<.001***	5.250	135.95	***<.001***	1.0987	.116	-.2950	-.1589

Regression (Imported)

R² (Canonical r²):	.2725	
	Statistic	**p**
F(8,153):	7.1632	***<.001***

Multivariate Test

= Covariance Matrix Test		
Box M	64.541	—
χ2(36):	60.419	***.007***
F(36,34679):	1.676	***.007***

Tests of equal mean vector		
Hotelling T²:	59.927	***<.001***
Wilks' Λ:	.728	***<.001***
Bartlett-Pillai:	.272	***<.001***
Roy:	.375	***<.001***

Multivariate Effect Sizes		
η² (Eta²):	.2725	—
τ² (Tau²):	.2725	—
ζ² (Zeta²):	.2725	—
ξ² (Xi²):	.2725	—
ω² (Omega²):	.2667	—

Multivariate Distance between group		
Mahalanobis D²:	1.715	—
Mean LDF Grp 1:	1.175	—
Mean LDF Grp 2:	-.540	—

Figure 17 – PDA Casewise

PDA Casewise, given a dichotomous criterion variable, renders regression classification accuracy – assigning Ss to groups to which their regression predicted score is closest. This is the same as a linear PDA with prior probabilities estimated by group size and equal costs of misclassification. A linear PDA is also produced which renders the probabilities of group membership, p(group|X), highlighted in **red** if the case was a "miss." As well, the "typicality" probability p(X|group) is produced. This is only of interest if small as a metric of the case being an outlier. Finally, exploration of unequal prior probabilities and costs of misclassification is available interactively, with resulting group probabilities and classification accuracies presented. This is done with L-O-O cross-validated classification.

		From Regression Priors α to group size		Linear PDA (not CV) - Equal Priors, Covariance Matrices, & Costs						
		Predicted	Predicted	dostat =						
	Actual	(Not CV)	PRESS CV	1		2			'Typicality'	L-O-O CV
S #	dostat	dostat	dostat	Distance	p(group|X)	Distance	p(group|X)	LDF Score	p(X|group)	Predicted
1	2	2	2	4.1544	.1793	1.1119	.8207	-1.2037	.6122	2
2	2	2	2	8.9459	.2352	6.5879	.7648	-0.8614	.8060	2
3	2	2	2	7.6496	.1679	4.4490	.8321	-1.2827	.5706	2
4	2	2	2	13.8054	.0461	7.7474	.9539	-2.7114	.0973	2
5	2	2	2	3.5360	.2419	1.2516	.7581	-0.8246	.8279	2
6	2	2	2	10.3020	.0831	5.4990	.9169	-2.0839	.2384	2
7	2	2	2	11.9492	.1040	7.6420	.8960	-1.8360	.3223	2
8	2	2	2	9.7352	.3196	8.2238	.6804	-0.4381	.9380	2
9	2	2	2	10.1959	.3106	8.6017	.6894	-0.4795	.9632	2
10	2	2	2	4.8036	.1462	1.2739	.8538	-1.4473	.4884	2
11	2	2	2	7.5544	.2047	4.8394	.7953	-1.0399	.7026	2
12	2	2	2	6.0671	.2568	3.9422	.7432	-0.7449	.8756	2
13	2	2	2	13.7562	***.5859***	14.4503	***.4141***	0.6646	.3577	***1***
14	2	2	2	12.5546	.0630	7.1546	.9370	-2.3824	.1594	2
15	2	2	2	11.8796	.0541	6.1578	.9459	-2.5433	.1261	2
16	2	2	2	13.5217	.0756	8.5130	.9244	-2.1868	.2085	2

	dostat	
	1	2
Prior Probabilities	.5000	.5000
Cost of Misclassification	1.0000	1.0000

L-O-O CV Classification - Input Priors and Costs

			Predicted		
		dostat	1	2	Total
Actual	Hit #	1	***38***	13	51
		2	39	***72***	111
		Total Hit #=	***110***		162
	Hit %	1	***74.51%***	25.49%	31.48%
		2	35.14%	***64.86%***	68.52%
		Total Hit %=	***67.90%***		100.00%

Figure 18 – PDA Summary

If the criterion is dichotomous, **PDA Summary** gives the inferential tests of hit-rate in respect to both the proportional and maximum chance criteria, as well as the ***I*** index of Huberty (1994). This is done for the regression classification (equivalent to a PDA with priors established by group size and equal costs of misclassification), and the PRESS cross-validated regression classification (equivalent to "L-O-O" or Lachenbruch U), and the linear PDA with and without cross-validation. All the case-wise classifications are on the **PDA Casewise** sheet.

Regression -- priors from group sizes and costs of missclassification assumed.

				For Each Group				For Total Sample				
Group	N	# Hits	Hit Rate	Chance Exp	z	p	I		Chance Exp	z	p	I
1	51	28	54.90%	16.06	3.60	***<.001***	34.18%	Proportional	92.11	5.53	***<.001***	49.92%
2	111	99	89.19%	76.06	4.69	***<.001***	65.66%	Maximum	111.00	2.71	***.003***	31.37%
Total	162	127	78.40%									

PRESS CV regression -- priors from group sizes and costs of missclassification assumed.

				For Each Group				For Total Sample				
Group	N	# Hits	Hit Rate	Chance Exp	z	p	I		Chance Exp	z	p	I
1	51	23	45.10%	16.06	2.09	***.018***	19.87%	Proportional	92.11	4.27	***<.001***	38.47%
2	111	96	86.49%	76.06	4.08	***<.001***	57.07%	Maximum	111.00	1.35	.088	15.69%
Total	162	119	73.46%									

Linear PDA (Not CV)-- equal priors and costs of missclassification assumed.

				For Each Group				For Total Sample				
Group	N	# Hits	Hit Rate	Chance Exp	z	p	I		Chance Exp	z	p	I
1	51	42	82.35%	16.06	7.82	***<.001***	74.24%	Proportional	92.11	3.63	***<.001***	32.75%
2	111	73	65.77%	76.06	.00	.500	-8.74%	Maximum	111.00	.68	.249	7.84%
Total	162	115	70.99%									

Linear PDA (Not CV) -- priors and costs of missclassification input.

				For Each Group				For Total Sample				
Group	N	# Hits	Hit Rate	Chance Exp	z	p	I		Chance Exp	z	p	I
1	51	42	82.35%	16.06	7.82	***<.001***	74.24%	Proportional	92.11	3.63	***<.001***	32.75%
2	111	73	65.77%	76.06	.00	.500	-8.74%	Maximum	111.00	.68	.249	7.84%
Total	162	115	70.99%									

Linear PDA L-O-O CV -- priors and costs of missclassification input.

				For Each Group				For Total Sample				
Group	N	# Hits	Hit Rate	Chance Exp	z	p	I		Chance Exp	z	p	I
1	51	38	74.51%	16.06	6.62	***<.001***	62.80%	Proportional	92.11	2.84	***.002***	25.60%
2	111	72	64.86%	76.06	.00	.500	-11.61%	Maximum	111.00	.00	.500	-1.96%
Total	162	110	67.90%									

Figure 19 – PDA Increment

Consideration of the contribution of a subset of variables to the predictive accuracy of a classification model containing other variables is just as useful in theory building as in the case of a criterion variable that is continuous. The question is isomorphic to the multiple regression case – simply with a different criterion, thus objective. For a model containing k_1 variables, what is the contribution of a subset of those variables containing k_2 variables? Indeed, the ΔR^2 with a dichotomous criterion can be legitimately tested, but, in general, it is not the gain in percentage of prediction accuracy that is of interest, rather it is the increment to classification accuracy afforded by the hypothesized variables that one wishes to know. That which makes the computation more difficult, and not available from simple classification analyses for the full and restricted models from commercial software, is that, as the same subjects are involved, the effects are correlated. Therein, a four-fold table containing the conditional hits and misses from both the full and restricted model is necessary. Further, to allow consideration of the effect of the increment to classification accuracy in each group and the total sample, such a table is needed for each. This method and associated software were introduced by Morris and Huberty (1995). Working just like the R^2 Increment sheet, this sheet allows you to test the contribution of any proper subset of variables to the full model introduced in the Input sheet. The variables' names for the model are automatically listed in row 15. Simply enter an "X" under each variable (row 16) to be included in the subset to be tested. This is real-time; as you add or omit Xs, the tabled hit-rates and the McNemar's z change accordingly. Therefore, you may consider any subsets of interest.

Place an X below variables for which Increment test is sought:

schools8	repeats8	reading8	math8	lang8	science8	socst8	dsfs8
X	X						X

Resubstitution Contrasts (Not Cross-validated) -- Input Priors and Costs of Misclassification

Hit-rate (as %)for both full and restricted models

	Groups		
	1	2	Total
Full	82.35%	65.77%	70.99%
Restricted	60.78%	68.47%	66.05%

Hit/Miss Matrix for Group 1

		Full 8 variable model	
		Misses	Hits
Restricted 5 variable model (less	Hits	1	30
the 3 variables you selected).	Misses	8	12
	McNemar p =	***.002***	

... for Group 2

	Full 8 variable model	
	Misses	Hits
Hits	14	62
Misses	24	11
McNemar p =	.557	

... for Total

	Full 8 variable model	
	Misses	Hits
Hits	15	92
Misses	32	23
McNemar p =	.200	

L-O-O Cross-Validated Contrasts -- Input Priors and Costs of Misclassification

Hit-rate (as %)for both full and restricted models

	Groups		
	1	2	Total
Full	74.51%	64.86%	67.90%
Restricted	54.90%	68.47%	64.20%

Hit/Miss Matrix for Group 1

		Full 8 variable model	
		Misses	Hits
Restricted 5 variable model (less	Hits	2	26
the 3 variables you selected).	Misses	11	12
	McNemar p =	***.007***	

... for Group 2

	Full 8 variable model	
	Misses	Hits
Hits	15	61
Misses	24	11
McNemar p =	.442	

... for Total

	Full 8 variable model	
	Misses	Hits
Hits	17	87
Misses	35	23
McNemar p =	.349	

Figure 20 – LR Logistic Regression

Logistic regression and discriminant analysis treat the same classification problem, but with different mathematical models; least-squares for discriminant analysis and maximum likelihood for logistic regression. Logistic regression directly produces probabilities of group membership, although as you can see (**PDA Casewise**), these are available as well in discriminant analysis. Discriminant analysis is the older technique being originally invented, in the two-group case, by Fisher (1936). The maximum likelihood solution cannot be presented as a formula; iterated must be employed to derive a solution. There are two ways to do this; in **LR** the Newton-Raphson method is used. The iteration usually works just fine, but in some circumstances, it may fail (with “complete” or “quasi-complete”) group separation. In such a case, it doesn’t matter how many iterations you pursue, a solution is not mathematically possible. One problem in commercial software is that a solution is often given in any case. SPSS is flagrant in this regard; weights are given as if they **are** the logistic weights, and only a tiny footnote about iterations being exceeded is a clue to the failure. Most don’t see this and believe that they have a logistic solution, when indeed, it does not exist. This page calculates an LR solution if possible, else reports its failure if not.

Input Iteration tolerance: .00100
Iterations = 6

Table 1

Title such as: Logistic Regression Weights, Tests, and VIFs.

Variable	B	SE	95% CI for EXP(B)		Exp(B)	Wald z	p	VIF
			LL	UL				(OLS)
Constant	-.975	1.241			.377	-.786	.432	
schools8	.796	.204	1.493	3.294	2.217	3.911	<.001	1.0
repeats8	.479	.306	.890	2.925	1.614	1.563	.118	1.1
reading8	-.579	.322	.300	1.048	.561	-1.798	.072	2.5
math8	-.047	.351	.482	1.888	.954	-.135	.893	2.7
lang8	-.265	.337	.399	1.476	.767	-.787	.431	2.9
science8	.047	.338	.544	2.020	1.048	.138	.890	3.0
socst8	.386	.329	.776	2.785	1.470	1.173	.241	3.3
dsfs8	.383	.144	1.108	1.940	1.466	2.656	.008	1.4

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

Logistic Regression Classification Summary.

				For Each Group					For Total Sample			
Group	N	# Hits	Hit Rate	Chance Exp	z	p	I		Chance Exp	z	p	I
1	51	31	60.78%	16.06	4.51	***<.001***	42.77%	Proportional	92.11	5.69	***<.001***	51.35%
2	111	97	87.39%	76.06	4.28	<.001	59.94%	Maximum	111.00	2.88	.002	33.33%
Total	162	128	79.01%									

Model Fit:

Null model -2*Log Likelihood =	-201.819	
Full model -2*Log Likelihood =	-146.013	p
χ2(8) =	55.807	***.001***
Wald (Full with intercept) χ2(9) =	36.364	***.001***

"Psuedo" R^2s -- Careful, estimates different notions of fit; can't be considered in the metric of an OLS R^2

McFadden =	.2765	Efron =	.3219
McFadden Adjusted =	.1873	Tjur =	.3164
Cox &Snell =	.2914	Count =	.7901 (Same as cell M20/N)
Cragg-Uhler/Naglekerke =	.4091	Adj Count =	.3333
McKelvey & Zavoina =	.5011		

Information Criteria

AIC =	164.0125
AIC/N =	1.0124
BIC =	191.8009

Figure 21 – LR Firth

Maximum Likelihood Estimation (MLE) is an asymptotic estimator; it, thus LR, requires large N. How large in a generic sense is as bit difficult to posit; there are several hotly contested rules of thumb for LR. In any case, keep in mind that it isn't just the N that is of concern, but the smaller n among the two groups. Maximum likelihood has specific problems modeling a dichotomous grouping in which one of the ns is small (see literature under "rare events."). So, a typical rule of thumb is from 10 to 20 subjects (in the smaller group)/predictor, but there are many opinions, and the case is that the answer isn't simple.

Firth's (1993) method was to overcome the small sample bias of the MLE solution with a "penalty" for small sample size, with that penalty disappearing for larger samples. An artifact of the method is that even with data that manifest "complete" or "quasi-complete" separation, iteration converges yielding weights. For that reason, many see (e.g., Heinze & Schemper, 2002 – see title) this as a "second level" analysis if LR iteration fails. I would prefer to see this as Firth originally intended – a generic method to overcome the effect of small samples on MLE (in this case, LR). Therefore, this sheet produces the Firth results, regardless of whether convergence obtains on the LR sheet. Note: Heinze & Schemper (2002) suggest caution in use of the Wald z for variables that present complete separation.

Iterations = 6

Table 1

Title such as: Firth Penalized Logistic Regression Weights, Tests, and VIFs.

Variable	B	SE	95% CI for EXP(B) LL	95% CI for EXP(B) UL	Exp(B)	Wald z	p	VIF (OLS)
Constant	-.875	1.206	.039	4.432	.417	-.725	.468	
schools8	.733	.194	1.421	3.046	2.081	3.768	<.001	1.0
repeats8	.386	.279	.852	2.543	1.471	1.384	.166	1.1
reading8	-.545	.314	.314	1.072	.580	-1.737	.082	2.5
math8	-.041	.344	.489	1.884	.960	-.120	.905	2.7
lang8	-.244	.329	.411	1.494	.783	-.742	.458	2.9
science8	.040	.332	.542	1.996	1.040	.119	.905	3.0
socst8	.364	.323	.764	2.709	1.439	1.128	.259	3.3
dsfs8	.358	.138	1.092	1.874	1.430	2.600	.009	1.4

Note. CI = confidence interval; LL = lower limit; UL = upper limit.

Firth Penalized Logistic Regression Classification Summary.

Group	N	# Hits	Hit Rate	For Each Group: Chance Exp	z	p	I	For Total Sample	Chance Exp	z	p	I
1	51	31	60.78%	16.06	4.51	***<.001***	42.77%	Proportional	92.11	5.69	***<.001***	51.35%
2	111	97	87.39%	76.06	4.28	<.001	59.94%	Maximum	111.00	2.88	.002	33.33%
Total	162	128	79.01%									

Firth Model Fit:

		p
Null model -2*Log Likelihood =	-201.819	
Full model -2*Log Likelihood =	-146.320	
$\chi 2(8)$ =	55.499	***<.001***
Wald (Full with intercept) $\chi 2(9)$ =	35.869	***<.001***

Firth "Pseudo" R2s -- Careful, estimates different notions of fit; can't be considered in the metric of an OLS R2

McFadden =	.2750	Efron =	.3192
McFadden Adjusted =	.1858	Tjur =	.3030
Cox &Snell =	.2901	Count =	.7901 (Same as cell M20/N)
Cragg-Uhler/Naglekerke =	.4072	Adj Count =	.3333
McKelvey & Zavoina =	.4569		

Firth Information Criteria

AIC =	164.3205
AIC/N =	1.0143
BIC =	192.1088

Figure 22 – LR Increment

©John D. Morris
mailto:jdmorris@fau.edu

LR ---Consideration of the contribution of a subset of variables to the predictive accuracy of a classification model containing other variables is just as useful in theory building as in the case of a criterion variable that is continuous. The question is isomorphic to the multiple regression case – simply with a different criterion, thus objective. For a model containing k_1 variables, what is the contribution of a subset of those variables containing k_2 variables? Indeed, the ΔR^2 with a dichotomous criterion can be legitimately tested, but, in general, it is not the gain in percentage of prediction accuracy that is of interest, rather it is the increment to classification accuracy afforded by the hypothesized variables that one wishes to know. That which makes the computation more difficult, and not available from simple classification analyses for the full and restricted models from commercial software, is that, as the same subjects are involved, the effects are correlated. Therein, a four-fold table containing the conditional hits and misses from both the full and restricted model is necessary. Further, to allow consideration of the effect of the increment to classification accuracy in each group and the total sample, such a table is needed for each. This method and associated software were introduced by Morris and Huberty (1995). Working just like the R^2 Increment sheet, this sheet allows you to test the contribution of any proper subset of variables to the full model introduced in the Input sheet. The variables' names for the model are automatically listed in row 15. Simply enter an "X" under each variable (row 16) to be included in the subset to be tested. This is real-time; as you add or omit Xs, the tabled hit-rates and the McNemar's z change accordingly. Therefore, you may consider any subsets of interest.

Place an X below variables for which Increment test is sought:

schools8	repeats8	reading8	math8	lang8	science8	socst8	dsfs8
X	X						X

	Full Model	Restricted	Full/Restricted Contrast $\chi^2(3)$	p
-2*Log Likelihood =	-146.013	-180.051	34.038	*<.001*
McFadden =	.2765	.1079		
McFadden Adjusted =	.1873	.0484		
Cox &Snell =	.2914	.1257		
Cragg-Uhler/Naglekerke =	.4091	.1765		
McKelvey & Zavoina =	.5011	.1974		
Efron =	.3219	.1160		
Tjur =	.3164	.1223		
Count =	.7901	.6852		
Adj Count =	.3333	.0000		
AIC =	164.0125	192.0506		
AIC/N =	1.0124	1.1855		
BIC =	191.8009	210.5761		

Resubstitution Classification Contrasts (Not CV)

Hit-rate (as %)for both Full and Restricted models

	Groups 1	2	Total
Full	60.78%	87.39%	79.01%
Restricted	25.49%	88.29%	68.52%

Hit/Miss Matrix for Group 1

		Full 8 variable model Misses	Hits
Restricted 5 variable model	Hits	1	12
	Misses	19	19
	McNemar p =	*<.001*	

... for Group 2

	Full 8 variable model Misses	Hits
Hits	7	91
Misses	7	6
McNemar p =	.791	

... for Total

	Full 8 variable model Misses	Hits
Hits	8	103
Misses	26	25
McNemar p =	*.003*	

Figure 23 – PDA LR Comp

More interest in logistic regression vs. Discriminant Analysis is apparent currently. This appears to be due to the novelty of logistic regression and due to its theoretical dependance on fewer assumptions. It is certainly a viable technique to use, but there is clear evidence that it should not wholly replace discriminant analysis. First, see the work of Efron (1975) showing discriminant analysis at an advantage. As well, in addition to the aforementioned potential iteration failures on the **LR** page, there are occasions in which logistic regression renders a solution that has very disproportionate accuracies for the two groups, sometimes rendering it useless. For these reasons, I suggest an empirical comparison between the classification accuracies of PDA and LR. This page performs that comparison. This is done by the same procedure as developed for Full vs. Restricted model testing (Morris & Huberty, 1995), so it will look familiar.

Hit-rate (as %)for both full and restricted models

	Groups		
	1	2	Total
PDA	82.35%	65.77%	70.99%
LR	60.78%	87.39%	79.01%

Hit/Miss Matrix for Group 1

		PDA	
		Misses	Hits
Logistic	Hits	0	31
Regression	Misses	9	11
	McNemar p =	***<.001***	

... for Group 2

	PDA	
	Misses	Hits
Hits	24	73
Misses	14	0
McNemar p =	***<.001***	

... for Total

	PDA	
	Misses	Hits
Hits	24	104
Misses	23	11
McNemar p =	***.029***	

Figure 24

References

Allen, D. M. (1971). *The prediction sum of squares as a criterion for selecting predictor variables*. Technical Report 23, Dept. of Statistics, University of Kentucky.

Aroian, L. A. (1944/1947). The probability function of the product of two normally distributed variables. *Annals of Mathematical Statistics*, *18*, 265-271.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society*, *4*, 268-282.

Bonferroni, C. E. (1936). Teoria statistica delle classi e calcola delle probabilita, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze.

Browne, M. W. (1975). Predictive validity of a linear regression equation. *British Journal of Mathematical and Statistical Psychology*, *28*, 70-87.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 133-159.

Darlington, R.B. (1978). Reduced variance regression. *Psychological Bulletin*, *85*, 1238-1255.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, *69*, 161-82.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association*, *70*, 892-898.

Ezekiel, M. (1930). *Methods of correlational analysis*. New York: Wiley.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*, 27-38.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179-188.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York: Guilford Press.

Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, *21*, 2409-2419.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*, 55-67.

Huberty, C. J, (1994). *Applied discriminant analysis*. New York: Wiley.

Huberty, C. J, & Morris, J. D. (1989). Multivariate analysis versus multiple univariate analysis. *Psychological Bulletin*, *105*, 302-308. [Reprinted in A. E. Kazdin (Ed.), *Methodological Issues and Strategies in Clinical Research* (1992, pp. 351-365). Washington, DC: APA.]

Huberty, C. J, & Mourad, S. A. (1980). Estimation in multiple correlation/prediction. *Educational and Psychological Measurement*, *40*, 101-112.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems, Proc. R. Soc. A, 186, 453-461.

Lachenbruch, P. A. (1967). An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. *Biometrics*, *23*, 639-645.

Lawless, J. F., & Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics – Theory and Methods*, *5*, 307-323.

Longley, J. W. (1967). An appraisal of least-squares programs from the point of view of the user. *Journal of the American Statistical Association, 62*, 819–841.

Lord, F. M. (1950). *Efficiency of prediction when a regression equation from one sample is used in a new sample*. (Research Bulletin No. 50-40) Princeton, N. J.: Educational Testing Service.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614.

McNemar, Q. (1947).Note on the sampling error of the differences between correlated proportions or percentages. *Psychometrika*, *12*, 153-157.

Morris, J. D., & Lieberman, M. G. (2021, April). *Pandemic Package for the Social Sciences – An Ongoing Effort Toward Access and Accuracy*. Paper to be presented at the annual meeting of the American Educational Research Association, Virtual Conference.

Morris, J. D., & Lieberman, M. G. (2018). Multicollinearity's effect on regression prediction accuracy with real data structures. *General Linear Model Journal*, *44*, 29-34.

Morris, J. D., & Lieberman, M. G. (2015). Prediction, explanation, multicollinearity, and validity concentration in multiple regression. *General Linear Model Journal*, *41*, 29-35.

Morris, J. D., & Huberty, C. J (1995). Full versus restricted model testing in predictive discriminant analysis. *Journal of Experimental Education*, *63*, 161-165.

Morris, J. D. (1986). Selecting a predictor weighting method by PRESS. *Educational and Psychological Measurement*, *46*, 853-870.

Morris, J. D. (1981). Updating the criterion for regression predictor variable selection. *Educational and Psychological Measurement*, *41*, 777-780.

Nicholson, G. E. (1948). *The application of a regression equation to a new sample*. Unpublished doctoral dissertation, University of North Carolina. [Condensed in G. E. Nicholson, Prediction in future samples. In I. Olkin et al, (Eds.), Contributions to probability and statistics. Stanford: Stanford University Press, 1960.]

Olkin, I., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics*, *29*, 201-

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879-891.

Rozeboom, W. W. (1978). Estimation of cross-validated multiple correlation: A clarification. *Psychological Bulletin*, *85*,1348-1351.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological Methodology 1982* (pp.290-312). San Francisco: Jossey-Bass.

Stein, C. (1960). Multiple regression. In I. Olkin et al. (Eds.), *Contributions to probability and statistics*. Stanford: Stanford

University
Press.

Student. (1908). The Probable Error of a Mean. *Biometrika*, *6*, 1–25.

Welch, B. L. (1947). The generalization of ‘student’s’ problem when several different population variances are involved. *Biometrika***,** *34***,** 28–35.

Wherry, R. J. (1975). Underprediction from overfitting: 45 years of shrinkage. *Personnel Psychology*, 28,1-18.

Wherry, R. J. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *Annals of Mathematical Statistics*, *2*, 440-451.

Table 1

Student Comments on Criterion.

Please provide any additional comments or suggestions related to your experience with Criterion. You may use as much space as you wish.	
#1	Using it through one drive was fantastic and worked like a charm. Loved using it.
#2	Criterion is user friendly, easy to use and a better system than SPSS. I did not have a great experience working with SPSS in my first stats class (I was required to use without another option). My classmates experienced the same
#3	Criterion saved this course and should replace spss.
#4	The only issue I had with criterion was the fact that I was unable to import/export data from the web version
#5	So much easier to use than SPSS. I spend a lot of time connecting and reconnecting to FAU virtual apps (timeouts, etc.) to use SPSS and then trying to remember what boxes to check. Thanksyou for this tools.
#6	I enjoyed working in Criterion during the semester and will continue using this program in the future. The program is very intuitive to use and makes all necessary tests and even more in seconds. The tables are prepared in the APA style. A very useful statistical program to use for all majors!
#7	I don't know of any other program that formats your output into an APA style table. I am sure this will save people incredible amounts of time!
#8	Not having the appropriate Excel was an issue for me.
#9	Honestly, I am truly grateful to you for sharing Criterion with us. It really is such a comprehensive and easy to use package. The fact that you have gone as far as to have the tables already in APA format is incredibly helpful. Thank you!
#10	Great program, made the concepts in class easier to understand.
#11	CrCriterion is much more user friendly than SPSS. By running statistical analysis through criterion I was able to get all of the information much faster than SPSS.
#12	Was much simpler to use and understand.