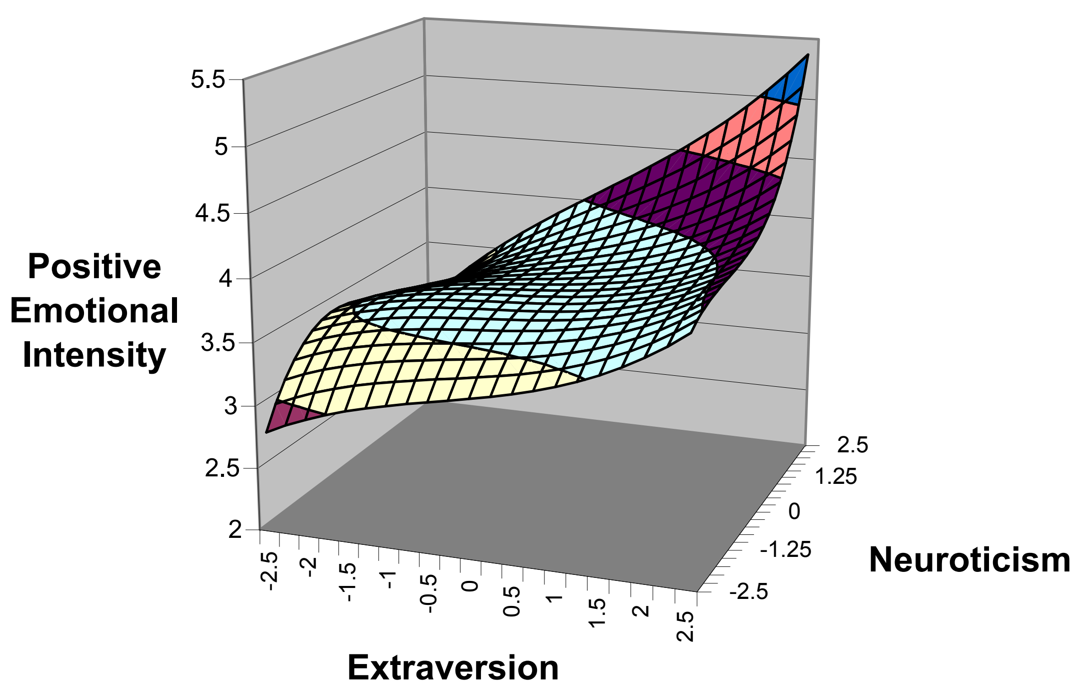
Psychology 513
Quantitative Models in Psychology

Class Notes

Spring 2014

Robert M. McFatter
University of Louisiana
Lafayette



SYLLABUS

Psychology 513 - Quantitative Models in Psychology
Instructor: Bob McFatter
Office: Girard Hall 222A; 482-6589
Internet page: *www.ucs.louisiana.edu/~rmm2440/*
E-mail: *mcfatter@louisiana.edu*

Texts: Kutner, M.H., Nachtsheim, C.J. & Neter, J. (2004). *Applied linear regression models, fourth edition.* New York: McGraw-Hill/Irwin **OR** Kutner, M.H., Neter, J., Nachtsheim, C.J. & Li, W. (2004). *Applied linear statistical models, 5th international edition.* New York: McGraw-Hill/Irwin.
JMP Statistical Discovery Software, by SAS Institute.
(available through UL site license at *http://helpdesk.louisiana.edu*)

Course Description

This course is designed to serve two main purposes. One objective is to provide an introduction to multiple correlation/regression and general linear models as basic analytical tools in psychological research. The other main objective is to familiarize students with a statistical package (JMP) available here and widely used throughout the country. Students should gain the practical skills necessary to enter, analyze, and interpret results for a variety of data sets using this package.

The first part of the course includes an introduction to the JMP statistical package mentioned above. Students use JMP to do familiar descriptive statistics, data screening, histograms, scatterplots, *t*-tests, etc.

A review of bivariate correlation and regression comes next if the backgrounds of the students require it. Topics covered include the relationship between correlation and regression, relevance of assumptions, effects of outliers, hypothesis testing, effects of measurement error and restricted variability, matrix formulation of regression analysis, the relation between dummy variable regression and *t*-test, and the interpretation of computer output including residual plots.

The remainder of the course is devoted to multiple regression and closely related topics. Consideration is given to the meaning and interpretation of regression weights, part and partial correlations, enhancer and suppressor effects, stepwise regression, multicollinearity problems, polynomial, interactive and nonlinear regression, logistic regression, analysis of covariance, and the relationship between regression and analysis of variance. Two exams, a midterm and a final, are given along with numerous homework assignments involving use of the computer to illustrate the theoretical aspects of the course.

Emergency Evacuation Procedures: A map of this floor is posted near the elevator marking the evacuation route and the ***Designated Rescue Areas.*** These are areas where emergency service personnel will go first to look for individuals who need assistance in exiting the building. Students who may need assistance should identify themselves to the teaching faculty.

READING ASSIGNMENTS

Sections to be covered in Kutner, M.H., Nachtsheim, C.J. & Neter, J. (2004). *Applied linear regression models, fourth edition.*

Appendix A - Basic Results
A.1, A.3-A.7
Chapter 1 - Linear Regression with One Independent Variable
All sections
Chapter 2 - Inferences in Regression Analysis
All sections
Chapter 3 - Diagnostics and Remedial Measures
3.1-3.4, 3.8
Chapter 4 - Effect of Measurement Errors
4.5
Chapter 5 - Matrix Approach to Regression Analysis
All sections
Chapter 6 - Multiple Regression - I
6.1-6.6, 6.9
Chapter 7 - Multiple Regression - II
All sections
Chapter 8 - Regression Models for Quantitative and Qualitative Predictors
All sections
Chapter 9 - Building the Regression Model I: Model Selection and Validation
9.4-9.5
Chapter 14 - Logistic Regression, Poisson Regression, and Generalized Linear Models
14.1-14.4

513 NOTES OUTLINE

A concise and useful review of basic statistical ideas is given in Kutner et al.'s Appendix A, sections A.1, A.3-A.7.

The model underlying general linear model (of which regression analysis and ANOVA are special cases) follows the general form: *Data = Fit + Residual.* One way this idea can be applied to the linear regression case is to describe each individual's score in the data set as being composed of 2 pieces: the predicted Y from some model, $\hat{Y}$, and a residual or error component, e.

$$Y = \hat{Y} + e \qquad [1]$$

It follows that $e = Y - \hat{Y}$ and reflects how far away the actual Y score is from the predicted Y score. The idea here is to hypothesize a model of the data and find estimates of the parameters of the model that make it fit the data as well as possible (i.e., minimize in some way the size of the residuals). The most common method is to find estimates of the parameters that make the sum of the squared residuals in the sample as small as possible. This is called 'least squares' estimation.

The model we focus on first is the linear regression model. The simplest case of linear regression analysis is the bivariate case: one predictor variable, X, and one criterion variable, Y. This involves fitting the sample data with a straight line of the form

$$\hat{Y} = b_0 + b_1 X \qquad [2]$$

where b_0 is the Y-intercept, and b_1 is the slope in the sample (these values being estimates of the population parameters β_0 and β_1).

The least squares estimators, b_0 and b_1, can be shown to be as follows:

$$b_1 = r\left(\frac{s_Y}{s_X}\right) = \frac{\sum XY - \frac{\left(\sum X\right)\left(\sum Y\right)}{n}}{\sum X^2 - \frac{\left(\sum X\right)^2}{n}} \qquad \text{and} \qquad [3]$$

$$b_0 = M_Y - b_1 M_X \qquad [4]$$

Here r is the correlation between X and Y, s_Y and s_X are the standard deviations of Y and X, respectively, and M_Y and M_X are the means of Y and X, respectively.

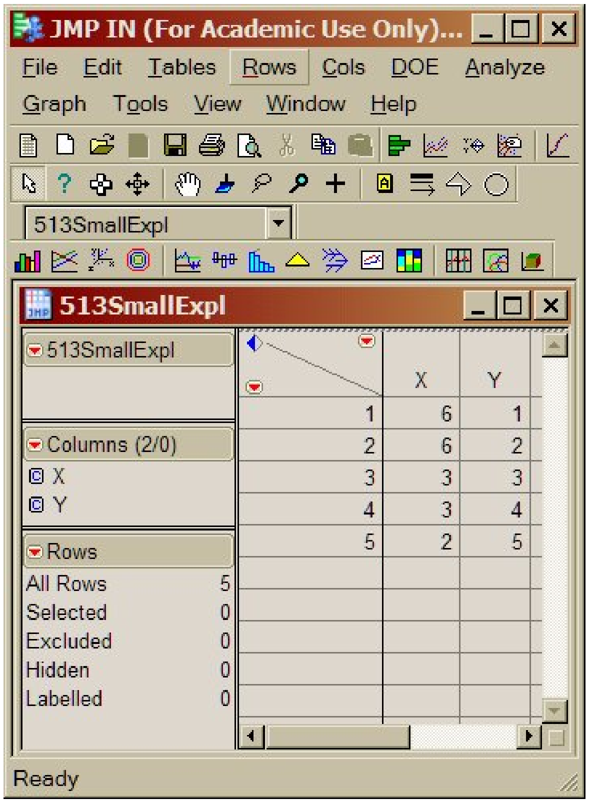
Example. Consider the following small data set.

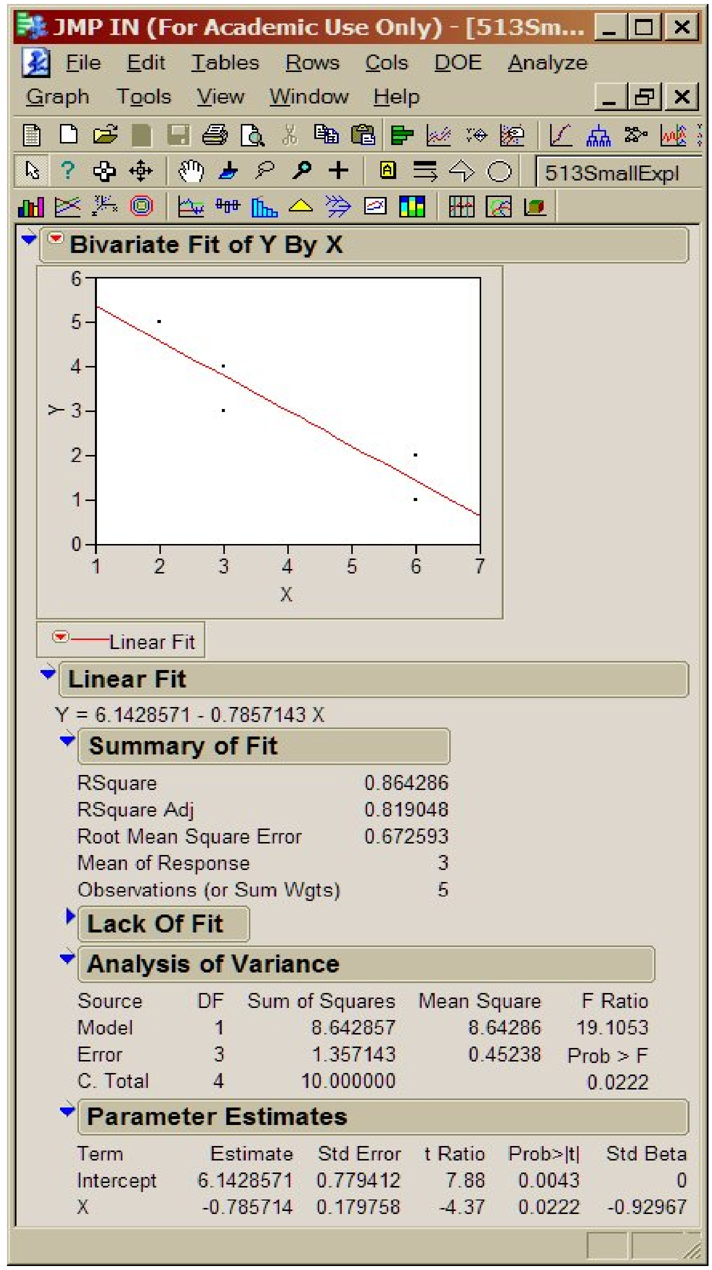
X	Y
6	1
6	2
3	3
3	4
2	5

The raw calculations may be done with $\sum X = 20, \sum Y = 15, \sum XY = 49, \sum X^2 = 94$ and

$M_X = 4$ and $M_Y = 3$, leading to $\hat{Y} = 6.1428571 - 0.7857143X$

One way to do the JMP analysis using the Fit Y By X platform is as follows:





Predicted *Y*s and residuals are obtained by plugging values of *X* into the regression equation. In JMP they may be obtained by clicking the red triangle by Linear Fit and selecting Save Predicteds or Save Residuals. For values of *X* not in the data set, predicted values and residuals may be obtained by simply including the desired *X* in the data set with a missing *Y* value and saving predicteds and residuals.

Meaning of b_0 and b_1

The parameter estimates, b_0 and b_1, are interpreted straightforwardly. The intercept b_0 is the predicted value of *Y* when $X = 0$. The slope b_1 is the predicted change in *Y* for a 1 unit increase in *X*.

Standardized and Deviation Score Regression Equations

The regression equation estimated above is sometimes called the *raw score* or *unstandardized* regression equation because raw scores on *X* plugged into the equation yield predicted raw scores on *Y*. If the *X* and *Y* scores are first transformed into deviation scores before computing the regression equation, the resulting equation is called the deviation score regression equation. To convert scores to deviation scores one simply subtracts the mean from each score (e.g., $y = Y - M_Y$, and $x = X - M_X$. I use lower case letters to indicate deviation scores). The deviation score regression equation will be of the form $\hat{y} = b_1 x$, where b_1 is exactly the same as the slope in the raw score equation. Notice that the intercept in the deviation score regression equation is zero.

It is actually more common to transform the *X* and *Y* scores to standardized (*Z*) scores than to deviation scores. When this is done, the resulting equation is called the *standardized regression equation*. The form of the standardized regression equation is $\hat{Z}_Y = rZ_X$. Notice that the intercept of the standardized regression equation is zero and the slope is *r*, the correlation coefficient between *X* and *Y*. This equation can be obtained in JMP by right-clicking in the Parameter Estimates table and selecting Columns|Std Beta. It is, unfortunately, common practice to refer to standardized regression coefficients (or weights or slopes; they all mean the same thing) as 'standardized betas' even though the coefficients are not population parameters.

One interpretation of the correlation coefficient *r*, then, is as the slope of the standardized regression equation. Thus, an *r* of .6 would mean that a 1 *standard deviation* increase in the value of *X* would lead us to predict a .6 *standard deviation* increase in the value of *Y*.

The predictions made are the same regardless of which form (raw score, deviation score, standard score) of the regression equation is used.

Example. Consider the following summary statistics for ACT scores and GPA.

	ACT	GPA
Mean	20	2.6
Std Dev	5	0.8
r	0.6	

What are the predicted GPAs for individuals with ACTs of 25, 10, 27, and 20? For the raw score regression equation, b_1 = .6(.8/5) = .096, and b_0 = 2.6 - .096 (20) = .68. Thus, $\hat{Y} = 0.68 + 0.096X$, and plugging *X* values into the equation gives

X	$\hat{Y}$
25	3.08
10	1.64
27	3.272
20	2.6

Note that the standardized regression equation ($\hat{Z}_Y = rZ_X$) yields the same predicted *Y* values:

For *X* = 25, Z_X = 1, $\hat{Z}_Y = .6$, $\hat{Y} = 2.6{+}.6(.8) = 3.08$

For *X* = 10, Z_X = -2, $\hat{Z}_Y = -1.2$, $\hat{Y} = 2.6 - 1.2(.8) = 1.64$

For *X* = 27, Z_X = 1.4, $\hat{Z}_Y = .84$, $\hat{Y} = 2.6{+}.84(.8) = 3.272$

For *X* = 20, Z_X = 0, $\hat{Z}_Y = 0$, $\hat{Y} = 2.6 + 0(.8) = 2.6$

One interesting and important phenomenon implied by the standardized regression equation is called *regression toward the mean*. Because it is always the case that $0 \le |r| \le 1$, the predicted value of Z_Y will be $\le$ in absolute value than Z_X, that is, generally closer to the mean of zero (in standard deviation units) than Z_X.

Assumptions

Certain assumptions are commonly made to aid in drawing inferences in regression analysis. Those assumptions may be stated as follows:

1) The means of all the conditional distributions of $Y|X$ lie on a straight line (linearity).
2) The variances of the conditional distributions of $Y|X$ are all equal (homoscedasticity).
3) The conditional distributions of $Y|X$ are all normal.
4) The points in the sample are a random sample from the population of points.

If homoscedasticity holds, then the population variance of any conditional distribution of $Y|X$ is equal to the variance of the population of residuals, and an estimate of that variance may be used as an estimate of the variance of *Y*s for any given *X*. The most commonly used estimate of the variance of the residuals is the mean square error (*MSE*) where $MSE = \Sigma\, e^2/(n - 2) = SSE/df_E$. The square root of this quantity (or Root *MSE*) is sometimes called the *standard error of estimate* for the regression analysis and would, of course, be an estimate of the standard deviation of each of the conditional distributions of $Y|X$. It can be thought of as how far off typically one would expect actual *Y* values to be from the predicted *Y* of the regression equation.

ANOVA Partitioning

It is common to construct an ANOVA partitioning of the variation (*SS*) of the *Y* scores in a regression analysis.

The total variation of the raw *Y* scores is measured by *SSY* or $SSTO = \sum (Y - M_Y)^2 = \sum y^2$.

The variation of the predicted *Y* scores is $SSR = \sum (\hat{Y} - M_Y)^2 = \sum \hat{y}^2$.

The variation of the residuals is $SSE = \sum (Y - \hat{Y})^2 = \sum e^2$

It is straightforward to show that $SSTO = SSR + SSE$. This is another example of the *Data = Fit + Residual* idea. The total variation of the *Y* scores can be broken down into two pieces: (1) predictable variation due to the model (*SSR*); and (2) unpredictable, noise, or error variation (*SSE*). It also turns out that the proportion of variation in *Y* that is predictable is equal to r^2:

$$r_{XY}^2 = r_{Y\hat{Y}}^2 = \frac{SSR}{SSTO} . \qquad [5]$$

Sums of squares (*SS*) are converted into sample variance estimates or mean squares (*MS*) by dividing by the appropriate *df*. For the ANOVA breakdown here, $df_{TO} = n - 1$, $df_R = 1$, and $df_E = n - 2$, so that $MSTO = SSTO/df_{TO}$, $MSR = SSR/df_R$, and $MSE = SSE/df_E$.

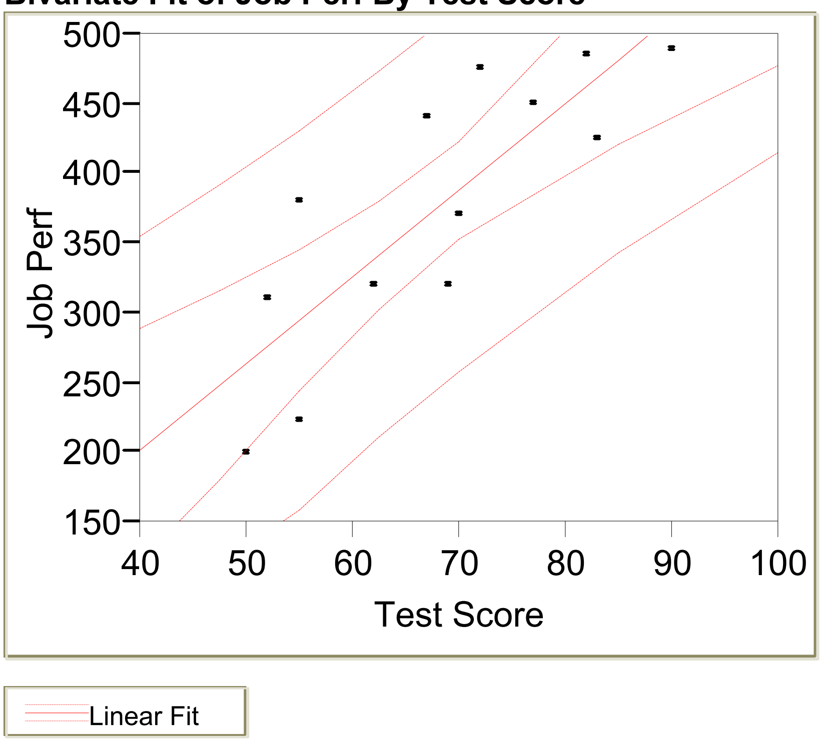
MSTO would thus be simply the sample variance of the *Y* scores, and *MSE* would be the sample variance of the residuals using $n - 2$ as the *df*.

The overall *F*-ratio from the ANOVA breakdown, $F(df_R, df_E) = MS_R/MS_E$, tests the null hypothesis that none of the variation in *Y* is linearly predictable from *X*. In the bivariate case this is exactly equivalent to the null hypothesis that the population slope, β_1, is equal to zero.

Example. The following hypothetical predictive validity regression analysis will illustrate the ideas in this and the following sections. Test score is the *X* variable and Job performance is the criterion.

Test Score	Job Perf
69	320
55	223
67	440
52	310
82	485
62	320
70	370
90	490
50	200
77	450
83	425
72	475
55	380

Bivariate Fit of Job Perf By Test Score



Linear Fit

Job Perf = -46.69521 + 6.216106 Test Score

Summary of Fit

RSquare	0.67861
RSquare Adj	0.649393
Root Mean Square Error	57.13172
Mean of Response	376
Observations (or Sum Wgts)	13

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	75811.63	75811.6	23.2264
Error	11	35904.37	3264.0	Prob > F
C. Total	12	111716.00		0.0005

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Lower 95%	Upper 95%	Std Beta
Intercept	-46.69521	89.12735	-0.52	0.6107	-242.8632	149.47277	0
Test Score	6.216106	1.289816	4.82	0.0005	3.3772398	9.0549723	0.823778

Test Score	Job Perf	Pred Formula Job Perf	Residual Job Perf	PredSE Job Perf	StdErr Indiv Job Perf	Studentized Resid Job Perf
69	320	382.216106	-62.216106	15.8978975	59.3024188	-1.133774
55	223	295.190622	-72.190622	23.0701597	61.6138459	-1.3811989
67	440	369.783894	70.216106	15.8978975	59.3024188	1.27955927
52	310	276.542304	33.4576962	26.0186032	62.777396	0.65779759
82	485	463.025484	21.9745158	24.0239498	61.9772853	0.42393072
62	320	338.703364	-18.703364	17.6343428	59.7913353	-0.3441782
70	370	388.432212	-18.432212	16.0540967	59.3444838	-0.3361718
90	490	512.754332	-22.754332	32.5003751	65.7290508	-0.4842707
50	200	264.110092	-64.110092	28.1086155	63.6720346	-1.2889375
77	450	431.944954	18.0550459	19.6426374	60.4141286	0.33654107
83	425	469.24159	-44.24159	25.007905	62.3652872	-0.8612729
72	475	400.864424	74.1355759	16.6642591	59.5124463	1.35661719
55	380	295.190622	84.8093782	23.0701597	61.6138459	1.6226293
60	.	326.271152	.	18.909034	60.1796087	.

Sampling Distributions of b_0 and b_1

Because b_0 and b_1 are sample statistics, their values will fluctuate from sample to sample from the same population of X and Y values. It is common to assume a model to describe the population of X and Y values, and consider b_0 and b_1 to be estimates of the parameters of the model. The most commonly assumed model is the *normal error regression model*:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad [6]$$

where:

- Y_i is the observed value of the criterion variable for the ith observation
- X_i is a known constant, the value of the predictor variable for the ith observation
- β_0 and β_1 are parameters
- the residuals, ε_i, are independent and $N(0, \sigma^2)$ [i.e., normally distributed with mean 0 and variance σ^2]
- $i = 1, ..., n$

Thus, b_0 and b_1 are estimates of β_0 and β_1, and will, like all statistics, have sampling distributions associated with them. Under the normal error regression model above, the sampling distributions of b_0 and b_1 will both be normal. The means of the sampling distributions will be $E\{b_0\} = \beta_0$ and $E\{b_1\} = \beta_1$, respectively. Thus, b_0 and b_1 are both *unbiased* estimators of their respective population parameters.

The variances of the two sampling distributions may be shown to be:

For b_0:

$$\sigma^2\{b_0\} = \sigma^2\left[\frac{1}{n} + \frac{M_X^2}{\sum (X_i - M_X)^2}\right]. \qquad [7]$$

For b_1:

$$\sigma^2\{b_1\} = \frac{\sigma^2}{\sum (X_i - M_X)^2}. \qquad [8]$$

Sample estimates of these variances may be obtained by replacing σ^2 with its sample estimate, the mean square error (*MSE)*. Thus,

$$s^2\{b_0\} = MSE\left[\frac{1}{n} + \frac{M_X^2}{\sum (X_i - M_X)^2}\right] \text{ and} \qquad [9]$$

$$s^2\{b_1\} = \frac{MSE}{\sum (X_i - M_X)^2}. \qquad [10]$$

The square roots of these quantities are called the *standard errors* of the parameter estimates and are given by most regression programs (including JMP) in a column next to the estimates themselves. The standard errors of b_0 and b_1 reflect how far off, typically, one would expect the sample values, b_0 and b_1, to be from β_0 and β_1, respectively. The standard errors of b_0 and b_1 can be used to construct confidence intervals around b_0 and b_1 and to conduct hypothesis tests about β_0 and β_1. To avoid the somewhat cumbersome curly brace notation, I will generally use s_{b0} and s_{b1} to refer to the standard errors of b_0 and b_1.

In the predictive validity example above, s_{b0} = 89.12735 and s_{b1} =1.289816.

Hypothesis Tests and Confidence Intervals for β_1, β_0, and ρ

It is common to test hypotheses and construct confidence intervals for β_1, β_0 and ρ where ρ is the population correlation between X and Y. Because

$$\beta_1 = \rho\left(\frac{\sigma_Y}{\sigma_X}\right), \qquad [11]$$

it is clear that $\beta_1 = 0$ when $\rho = 0$, and vice versa. Therefore, testing the null hypothesis that either of these is equal to zero is exactly equivalent to testing the hypothesis that the other is zero. These equivalent hypotheses are the most commonly tested null hypotheses in bivariate regression analyses because they test whether there is a relation between X and Y, often the hypothesis of most interest in the analysis.

When the sampling distribution of a statistic is normal, and the standard deviation of the sampling distribution of that statistic (i.e., the standard error of the statistic) can be estimated, a t-test may be constructed to test null hypotheses about the value of the population parameter the statistic estimates. The general form of the t-test is

$$t = \frac{Statistic - Hypothesized\ value}{Estimated\ Std\ Error\ of\ the\ Statistic}. \qquad [12]$$

Thus, to test H_0: $\beta_1 = 0$, the appropriate t-test is

$$t = \frac{b_1 - 0}{s_{b_1}} = \frac{b_1}{s_{b_1}}.$$

In the example above, $t(11) = 6.216106/1.289816 = 4.82$, $p = .0005$, and the H_0: $\beta_1 = 0$, would be rejected even at $\alpha = .001$.

The estimated standard error of r when $\rho = 0$ is $s_r = \sqrt{\dfrac{1-r^2}{n-2}}$, so testing the H_0: $\rho = 0$ can be carried out with $t = \dfrac{r}{s_r} = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$. The *df* associated with the *t*-tests for these two null hypotheses are $n - 2$.

Because the two *t*-tests are testing equivalent null hypotheses, the values of *t* obtained in the two tests will be exactly the same. It is also important to note that the overall *F*-ratio from the bivariate regression analysis described above, $F(df_R, df_E) = F(1, n - 2) = MS_R/MS_E$, is exactly equivalent to the square of the *t* value obtained testing the null hypotheses $\beta_1 = 0$ or $\rho = 0$. So, these three hypothesis tests all test the same basic question and will all have exactly the same *p*-level.

In the example above, $r = 0.823778$, and plugging this value into the equation above yields $t(11) = 4.82$, $p = .0005$. Also, $F(1, 11) = 23.23$, $p = .0005$. Note that $23.23 = 4.82^2$.

Similar logic yields a test of H_0: $\beta_0 = 0$ of the form $t = \dfrac{b_0}{s_{b_0}}$, but this hypothesis is usually not of as much interest as the ones described above. JMP provides values for all the crucial components of these hypothesis tests including *t* values, *F*-ratio, and *p*-levels.

In addition, confidence intervals around b_0 and b_1 may be obtained in the usual way. For example, the $1 - \alpha$ confidence limits for β_1 would be $b_1 \pm t(1 - \alpha/2, n - 2)\, s_{b1}$. In JMP the 95% confidence limits for β_0 and β_1 may be obtained by right-clicking on the Parameter Estimates table and selecting Columns|Lower 95% and Upper 95%.

In the example above, the 95% confidence interval for β_1 would be (from the output): (3.38, 9.05).

Confidence intervals around *r* may be obtained using the Fisher's Z_r transformation:
$Z_r = \tanh^{-1}(r) = \dfrac{1}{2}\ln\left(\dfrac{1+r}{1-r}\right)$. Fisher's Z_r values are approximately normally distributed with a standard deviation of $\sigma_{Z_r} = \dfrac{1}{\sqrt{n-3}}$. Ordinary normal distribution methods can thus be used to find confidence intervals around the Z_r values. The ends of the confidence interval obtained may then be transformed back to *r* values using $r = \tanh(Z_r)$. [See p. 85-86 in Kutner et al. or an introductory statistics text such as HyperStat Online (http://davidmlane.com/hyperstat/) for more details and examples.]

Two Kinds of Prediction (E{Y|X} and Y|X) [See p. 52 and 55 in Kutner et al.]

There are two kinds of prediction that one might consider following a regression analysis:
1) prediction of the mean value of *Y* for all observations with a given *X* score (i.e., $E\{Y|X\}$);
2) prediction of the value of *Y* for an individual observation with a given *X* score (i.e., $Y|X$).

It is important to recognize that the predicted value of *Y* will be the same for both cases, namely, $\hat{Y}$. However, the uncertainty associated with the two kinds of prediction will differ.

In the first case we are trying to predict the Y value of the population regression line for the given X, and the uncertainty associated with that prediction will be simply how far typically we would expect sample regression lines to 'wobble' in the Y direction around the population regression line.

In the second case we must consider not only this wobble of sample regression lines around the population regression line, but also the variability of individual Ys around the sample regression line. Thus, the variability in the second case will simply be the variability in the first case (i.e., wobble) plus an additional component reflecting the variability of the Ys around the sample regression line.

Kutner et al. show that a sample estimate of the (wobble) variance of the $\hat{Y}$ s around the population regression line is

$$s^2\left\{\hat{Y}_h\right\} = MSE\left[\frac{1}{n} + \frac{\left(X_h - M_X\right)^2}{\sum\left(X_i - M_X\right)^2}\right], \quad [13]$$

where X_h is the value of X for which we wish to estimate the mean Y value (i.e., $\hat{Y}_h$). Notice that the uncertainty depends on how far away X_h is away from the mean of X, M_X. The further X_h is from M_X, the greater the uncertainty. The square root of this quantity is called the 'Std Error of Predicted' by JMP and can be obtained for each observation from the Fit Model platform after a regression analysis has been performed by right-clicking on a section title and selecting Save Columns|Std Error of Predicted. In the above example, the predicted job performance for individuals with a test score of 60 is 326.27. One would expect this value to be off typically by 18.91 points from the *true mean performance of individuals* with a score of 60 on the test.

A sample estimate of the variance of the individual Ys around the sample regression line would then be the sum of the 'wobble' variance above and the MSE. Kutner et al. call this $s^2\{pred\}$:

$$s^2\left\{pred\right\} = MSE\left[1 + \frac{1}{n} + \frac{\left(X_h - M_X\right)^2}{\sum\left(X_i - M_X\right)^2}\right]. \quad [14]$$

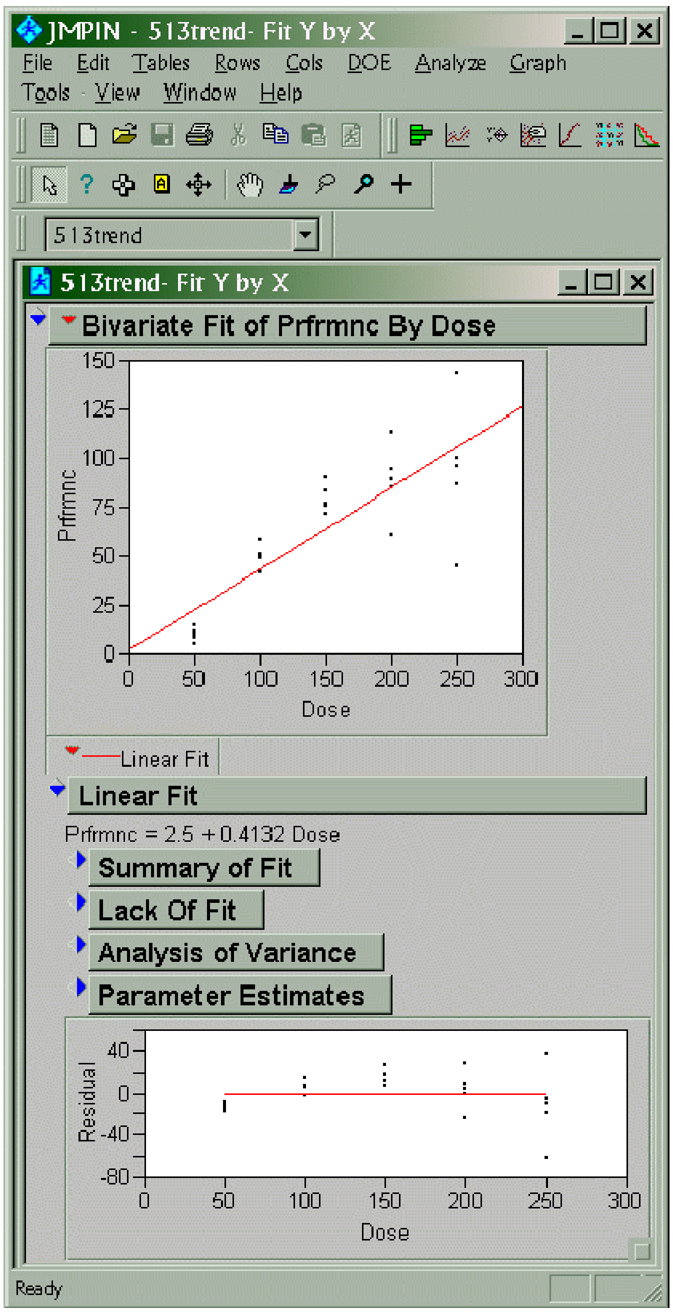
The square root of this quantity is called 'Std Error of Individual' by JMP and can be obtained for each observation by selecting Save Columns|Std Error of Individual. In the above example, the predicted job performance for an individual with a test score of 60 is 326.27. One would expect this value to be off typically by 60.18 points from the *actual performance of an individual* with a score of 60 on the test. It is instructive to compare this interpretation with the interpretation at the end of the last paragraph.

The standard errors of predicted and individual scores may be used with the t distribution in the usual way to construct confidence intervals for the two kinds of prediction. In the Fit Y by X platform of JMP, plots of the 95% confidence intervals for the two kinds of prediction may be obtained by clicking the Linear Fit red triangle and selecting 'Confid Curves Fit' and 'Confid Curves Individual,' respectively. (See the figure in the output for the example above.) Actual upper and lower bounds of the 95% confidence intervals for each observation may be obtained from the Fit Model platform regression analysis by right-clicking on a section title and selecting Save Columns|Mean Confidence Interval or Indiv Confidence Interval, respectively.

Residuals Analysis and Outliers

Examining the residuals, e_i, from a regression analysis is often helpful in detecting violations of the assumptions underlying the model. The least squares estimation procedure forces the residuals to sum to zero and to have a zero correlation with the predictor variable *X*. However, important diagnostic information can often be gained by analyzing the obtained residuals in various ways. It is common, for example, to plot the residuals as a function of *X* or $\hat{Y}$ (the two plots would look exactly the same except for the scale of the horizontal axis–make sure you see why). This kind of plot can easily reveal nonlinearity of the relation or violations of the homoscedasticity assumption. For example, the following scatterplot and residual plot illustrates violations of both the linearity and homoscedasticity assumptions.

Outliers are extreme observations whose values lie far from the values of the other observations in the data. These observations can strongly affect the fitting of a model to the data. It is common practice to examine data for outliers and perhaps remove extreme observations before fitting a model. Observations that lie more than 3 or 4 standard deviations away from the mean of the other observations are usually considered outliers.



In the bivariate regression case the residuals (or better, the 'studentized residuals,' i.e., the residuals divided by their standard errors) can be helpful in identifying outliers. Observations with studentized residuals >3 or 4 would be considered outliers. JMP provides these studentized residuals from the Fit Model platform under the Save Columns menu.

In multivariate situations the Multivariate platform in JMP provides an outlier analysis that uses a measure of distance (called Mahalanobis distance) of observations from the centroid of the data in multidimensional space. Again, points with distances of > 3 or 4 from the other points would be considered outliers.

The presence of outliers can, of course, strongly affect the size of the obtained correlation between two variables. Other factors that can affect the size of the correlation are
(1) restriction of variability in either *X* or *Y*;
(2) measurement error or unreliability in the measurement of *X* or *Y*;
(3) differences in the shapes of the marginal distributions of *X* and *Y*; and
(4) nonlinearity of the relation between *X* and *Y*.

All of these factors tend to make the correlation between *X* and *Y* smaller than would otherwise be the case. Instructive online simulation demonstrations of (1) and (2) can be found at the HyperStat Online site or directly at

http://davidmlane.com/hyperstat/Instructional_Demos.html.

Two-Predictor Models

When there are two predictors in the model the simplest fitted regression equation becomes

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 . \quad [15]$$

As we have seen, the simple bivariate regression equation (Equation [2]) represents the finding of the best fitting straight line to a two dimensional scatterplot of *X* and *Y*. When there are two predictors in the model, the equation in [15] represents the finding of the best fitting plane to a three dimensional cloud of points in 3-space. The 'Fit' part of *Data = Fit + Residual* is now a plane rather than a straight line. By adding predictors to, or changing the functional form of predictors in, equation [15] we may generate a huge variety of models of data, but the basic principles of model fitting, estimation of parameters, and hypothesis testing remain the same.

The sample statistics b_0, b_1, and b_2 in [15] are estimates of the population parameters β_0, β_1, and β_2. Formulas for the least squares estimates of the population parameters β_1 and β_2 may be found and involve the sample variances and covariances of the *X*s and *Y*. These formulas quickly become very complex with more than two predictors, and it is generally most satisfactory to solve the least squares estimation problem in matrix form. We will consider that later. It is impractical to compute these estimates with a calculator, so we generally use a statistics program. The intercept b_0, however, will always be $b_0 = M_Y - b_1M_1 - b_2M_2 \ldots$ regardless of the number of predictors, where M_Y is the mean of *Y*, M_1 the mean of X_1, etc.

Example.

The following example with hypothetical data from a study using IQ and Extraversion scores to predict Sale Success will illustrate the analysis. Here is the JMP output from the Fit Model platform.

Raw Data

IQ	Extr	Success
98	38	4
95	33	3
104	47	5
108	43	5
77	41	5
115	44	4
103	50	6
108	46	6
109	46	4
100	39	5

Variable	N	Mean	Std Dev	Min	Max
Extr	10	42.7	5.03432661	33	50
IQ	10	101.7	10.4780193	77	115
Success	10	4.7	0.9486833	3	6

Multivariate Correlations

	IQ	Extr	Success
IQ	1.0000	0.4510	0.0458
Extr	0.4510	1.0000	0.7003
Success	0.0458	0.7003	1.0000

MODEL 1 - IQ ALONE

Response Success
Summary of Fit

RSquare	0.0021
RSquare Adj	-0.12264
Root Mean Square Error	1.005173
Mean of Response	4.7
Observations (or Sum Wgts)	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.0170124	0.01701	0.0168
Error	8	8.0829876	1.01037	Prob > F
C. Total	9	8.1000000		0.9000

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta
Intercept	4.2780083	3.267579	1.31	0.2268	0
IQ	0.0041494	0.031977	0.13	0.9000	0.045829

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
IQ	1	1	0.01701245	0.0168	0.9000

Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
IQ	1	1	0.01701245	0.0168	0.9000

MODEL 2 - EXTR ALONE

Response Success
Summary of Fit

RSquare	0.490369
RSquare Adj	0.426665
Root Mean Square Error	0.718333
Mean of Response	4.7
Observations (or Sum Wgts)	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	3.9719860	3.97199	7.6976
Error	8	4.1280140	0.51600	Prob > F
C. Total	9	8.1000000		0.0241

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta
Intercept	-0.934678	2.043575	-0.46	0.6596	0
Extr	0.1319597	0.047562	2.77	0.0241	0.700263

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Extr	1	1	3.9719860	7.6976	0.0241

Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
Extr	1	1	3.9719860	7.6976	0.0241

Model 3 - IQ First

Response Success

Summary of Fit

RSquare	0.581862
RSquare Adj	0.462394
Root Mean Square Error	0.69559
Mean of Response	4.7
Observations (or Sum Wgts)	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	4.7130814	2.35654	4.8704
Error	7	3.3869186	0.48385	Prob > F
C. Total	9	8.1000000		0.0473

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta
Intercept	0.9560894	2.499998	0.38	0.7135	0
IQ	-0.030684	0.024793	-1.24	0.2558	-0.3389
Extr	0.1607603	0.051602	3.12	0.0170	0.853098

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
IQ	1	1	0.7410954	1.5317	0.2558
Extr	1	1	4.6960689	9.7057	0.0170

Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
IQ	1	1	0.0170124	0.0352	0.8566
Extr	1	1	4.6960689	9.7057	0.0170

Model 4 - Extr First

Response Success

Summary of Fit

RSquare	0.581862
RSquare Adj	0.462394
Root Mean Square Error	0.69559
Mean of Response	4.7
Observations (or Sum Wgts)	10

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	4.7130814	2.35654	4.8704
Error	7	3.3869186	0.48385	Prob > F
C. Total	9	8.1000000		0.0473

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	Std Beta
Intercept	0.9560894	2.499998	0.38	0.7135	0
Extr	0.1607603	0.051602	3.12	0.0170	0.853098
IQ	-0.030684	0.024793	-1.24	0.2558	-0.3389

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Extr	1	1	4.6960689	9.7057	0.0170
IQ	1	1	0.7410954	1.5317	0.2558

Sequential (Type 1) Tests

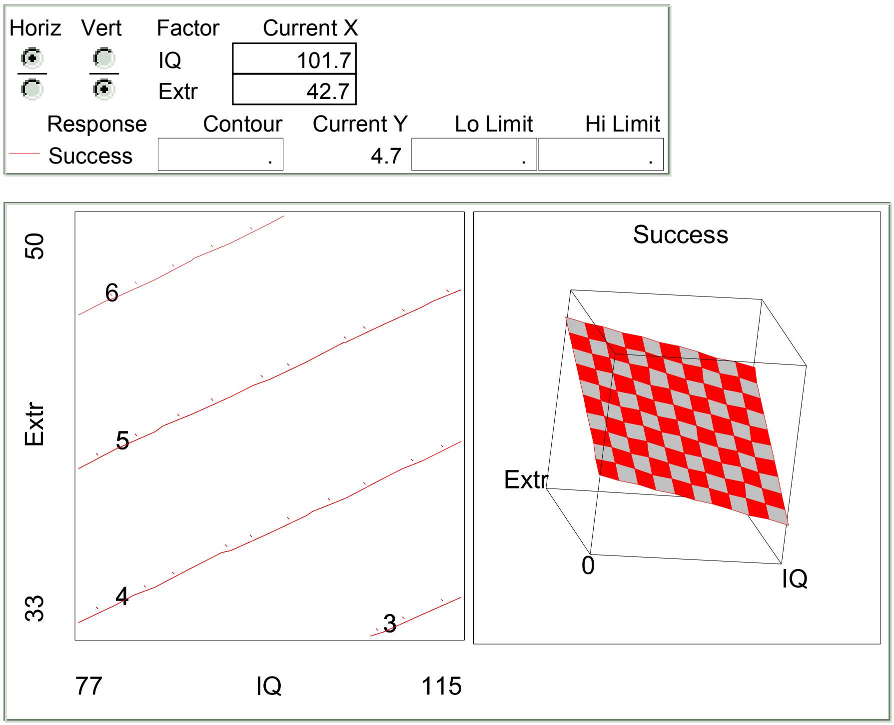
Source	Nparm	DF	Seq SS	F Ratio	Prob > F
Extr	1	1	3.9719860	8.2092	0.0242
IQ	1	1	0.7410954	1.5317	0.2558

It is very instructive to closely examine what stays the same and what changes from model to model in the four models above. Models 3 and 4 differ only in the order in which IQ and Extr are put into the model. The only differences in the output between these two models are in the Sequential (Type 1) Tests.

Interpretation of the Parameter Estimates

The following contour and 3-D plot from the Fit Model platform (select Factor Profiling|Contour Profiler) can help illustrate the interpretation of the regression coefficients b_0, b_1, and b_2.

Response Success
Contour Profiler



If we let X_1 = IQ and X_2 = Extr, the raw score (unstandardized) regression equation with two predictors is $\hat{Y} = .956 - .031X_1 + .161X_2$. The interpretation of the intercept, b_0 = 0.956 is similar to the interpretation for the bivariate case. The predicted Sales Success score for individuals who have scores of zero on both IQ and Extr is 0.956. This is not a particularly meaningful value, however, because no one had scores close to zero on either IQ or Extr.

In a multiple regression analysis like this, the interpretation of the regression coefficients is *always conditional.* This conditionality may be expressed in various ways, but it *must always be expressed* in the interpretation of the weights. A correct interpretation of b_1 = -0.031 in the above example would be '*holding individuals constant on extraversion*, a one point increase in IQ score would lead us to predict a .031 point decrease in sales success. That is, the interpretation of the coefficient of any variable in the equation is always conditional on holding all other variables in the equation constant. Here are some other ways this conditionality might be expressed in interpreting b_1:

(1) 'For individuals who are similar on extraversion, a one point increase in IQ . . .'
(2) 'Controlling for extraversion, a one point increase in IQ . . .'
(3) 'Partialing out the effect of extraversion, a one point increase in IQ . . .'

The third expression uses terminology that comes from the fact that the coefficients in a multiple regression equation are sometimes called *partial* regression coefficients. This usage perhaps comes from the fact that the coefficients, b_1 and b_2, in equation [15] are the partial derivatives of $\hat{Y}$ with respect to X_1 and X_2, respectively.

Graphically, the meaning of b_1 and b_2 may be seen by examining the 3-D surface plot in the JMP output above. The line on the front face of the cube, for example, shows the relation between predicted Sales Success and IQ when Extr is held constant at its lowest value. Notice that the slope is negative. In fact, it would be -0.031. Note also that when Extr is held constant at *any* value (e.g., at its highest value, putting us on the back face of the cube) the slope of the relation between predicted Sales Success and IQ is still the same, -0.031. That is, *on a plane* such as the one in the figure, the lines showing the relation between predicted Sales Success and IQ *for any given value of Extr* are all parallel.

The interpretation of $b_2 = 0.161$ is directly analogous: holding individuals constant on IQ, a one point increase in extraversion would lead us to predict a 0.161 point increase in Sales Success. Notice that lines showing this positive slope would be ones parallel to the lines on the two side faces of the cube in the figure from the output.

Standardized Regression Equation

The standardized form of the multiple regression equation results if the scores on X_1, X_2, and Y are standardized to Z-scores before carrying out the regression analysis. The standardized regression equation may be written as

$$\hat{Z}_Y = b_1^* Z_1 + b_2^* Z_2 + \ldots \quad [16]$$

for an arbitrary number of predictors, where Z_1 represents the Z-scores of X_1, etc., and b^*_1 and b^*_2 are the standardized regression weights in the equation. Notice that the intercept in this equation is zero. It may be shown that regardless of the number of predictors the standardized regression weight for the jth predictor is related to the unstandardized regression weight for the jth predictor by the following:

$$b_j = b_j^* \left(\frac{s_Y}{s_j} \right) \quad [17]$$

where s_j is the standard deviation of the jth predictor. JMP provides the standardized regression weights under Columns|Std Beta. In the Sales Success example the standardized regression equation from Model 3 would be

$$\hat{Z}_Y = -0.34 Z_1 + 0.85 Z_2 \quad [18]$$

and the interpretation would be exactly the same as in the unstandardized case, except that the units involved would be standard deviations of the variables rather than raw points. For example, one could interpret $b^*_1 = -0.34$ by saying that, for individuals who all had the same extraversion score, a one *standard deviation* increase in IQ would lead us to predict a 0.34 *standard deviation* decrease in Sales Success.

In the two predictor case, the standardized regression weights may be computed simply from the correlations among the three variables. For example, the value of b^*_2 would be

$$b^*_2 = (r_{Y2} - r_{Y1} r_{12})/(1 - r^2_{12}). \quad [19]$$

The formula for b^*_1 would be the same with the 1's and 2's exchanged. For more than two predictors, the formulas become much longer.

ANOVA Partitioning in Multiple Regression

The overall ANOVA partitioning of the variation (*SS*) of the *Y* scores in a multiple regression analysis is a straightforward generalization of the bivariate case. Just as in the bivariate case, $SSTO = SSR + SSE$, with the definitions of each component of this equation being the same as in the bivariate case. The only difference is that in the multiple predictor case the predictable variation in *Y*, i.e., *SSR*, comes from several sources. Also, because there are several *X*s it is not possible to talk about a single r^2_{XY}. However, as equation [5] indicates, r^2_{XY} in the bivariate case is equal to $r^2_{Y\hat{Y}} = \frac{SSR}{SSTO}$, and this latter quantity serves as an appropriate generalization in the multiple predictor case. The proportion of variation in Y that is predictable from multiple predictors is called the squared multiple correlation and is denoted by R^2:

$$R^2_{Y.1,2,\ldots k} = r^2_{Y\hat{Y}} = \frac{SSR}{SSTO}, \qquad [20]$$

where *k* is the number of predictors in the regression equation.

Hypothesis Tests as Comparisons of Full vs. Reduced (Restricted) Models
[see p. 72-73 in Kutner et al.]

The hypothesis tests carried out in the general linear model may be understood as comparisons of the adequacy of *full* vs. *reduced* or *restricted* models. For example, in the two predictor case the full model would be

[*Full Model*] $$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i. \qquad [21]$$

A test of $\beta_1 = 0$ may be understood as a test of whether a restricted model that requires β_1 to be zero does essentially as well a job at predicting *Y* as does a model that allows β_1 to be different from zero. The restricted model (obtained by setting β_1 to be zero in equation [21]) would be

[*Restricted Model*] $$Y_i = \beta_0 + \beta_2 X_{2i} + \varepsilon_i. \qquad [22]$$

The logic of the hypothesis test of $\beta_1 = 0$ is to ask if the *SSE* is significantly smaller for the model of equation [21] than for the model of equation [22]. The models are compared by looking at the decrease in *SSE* (or, equivalently, the increase in *SSR*) that occurs when going from the restricted model to the full model. In terms of *SSE* the *F*-ratio for carrying out a test comparing the full model with the restricted model is as follows (where R and F refer to restricted and full, respectively):

$$F(df_R - df_F, df_F) = \left[\frac{SSE(R) - SSE(F)}{df_R - df_F}\right] \div \left[\frac{SSE(F)}{df_F}\right] \qquad [23]$$

where df_F and df_R are error *df* for the full and restricted models, respectively. The above formula has the form of a *MSR*/*MSE*. The numerator can also be thought of as $[SSR(F) - SSR(R)]/[df_R - df_F]$. The error *df* for any model will always be $n - k - 1$, where *n* is the number of observations, and *k* is the number of predictors in the model

This formula may also be cast in terms of R^2s for the full and restricted models [see equation 7.19 on p. 266 of Kutner et al.]:

$$F(df_R - df_F, df_F) = \left[\frac{R^2_F - R^2_R}{df_R - df_F}\right] \div \left[\frac{1 - R^2_F}{df_F}\right]. \qquad [24]$$

Thus, another way of conceptualizing the hypothesis test of, say, $\beta_1 = 0$, in a model that has two predictors would be as a test of whether adding the predictor X_1 to an equation that already has X_2 in it significantly improves the prediction of Y beyond what can be predicted by X_2 alone. It is important to note that this interpretation is exactly equivalent to asking whether there is a relation between Y and X_1 when X_2 is held constant.

Type I (sequential) & Type II *SS*: Alternative Breakdowns of *SSR* [see p. 256-271]

There are several ways of decomposing the *SSR* in a regression equation with several predictors. Decomposing or partitioning the *SSTO* into pieces *that add up to SSTO* always uses what is commonly called Type I (or sequential) *SS*. JMP does not by default provide the Type I *SS*, but one can always obtain them from the Standard Least Squares analysis of the Fit Model platform: Simply right-click and select Estimates|Sequential Tests.

The default *SS* provided by JMP in the Effect Tests section are commonly called Type II *SS*. Type II *SS* provide the most commonly used hypothesis tests, but they do not add up (generally) to the *SSR*.

In the single predictor case the Type I and Type II *SS* will be identical and equal to *SSR*, but with more than one predictor there are several breakdowns possible for the Type I *SS*, depending on the order in which variables are entered into the regression equation. Type II *SS* do not depend on the order of variables entered as indicated below. For one predictor the breakdown is as follows:

$$SSTO = SSR(X_1) + SSE(X_1).$$

With two predictors the *SS* breakdown can be represented as follows:

$$SSTO = SSR(X_1) + SSR(X_2|X_1) + SSE(X_1, X_2) = SSR(X_1, X_2) + SSE(X_1, X_2). \quad [25]$$

$SSR(X_1, X_2) = SSR(X_1) + SSR(X_2|X_1)$ Type I *SS* with X_1 entered first
$SSR(X_1, X_2) = SSR(X_2) + SSR(X_1|X_2)$ Type I *SS* with X_2 entered first

With three predictors the breakdown would be:

$SSR(X_1, X_2, X_3) = SSR(X_1) + SSR(X_2|X_1) + SSR(X_3|X_1, X_2)$ Type I *SS* with X_1 entered first, X_2 second, X_3 last. The extension to more than 3 predictors is obvious.

Type II *SS* is the Type I *SS* for each variable when it is entered last:

$SSR(X_1|X_2, X_3)$
$SSR(X_2|X_1, X_3)$
$SSR(X_3|X_1, X_2)$

Type II *SS* is the usual *SS* output by JMP and used in the usual hypothesis tests.

The Type I *SS* breakdown of *SSR* leads to partial and semipartial (sometimes called part) correlations. The squared partial and semipartial correlations are as follows:

Squared partial correlation examples [see comments on p. 271]:

$$r^2_{Y2.1} = SSR(X_2|X_1)/SSE(X_1) = (r_{Y2} - r_{Y1}r_{12})^2/[(1 - r^2_{12})(1 - r^2_{Y1})] \quad [26]$$

$$r^2_{Y2.13} = SSR(X_2|X_1, X_3)/SSE(X_1, X_3)$$

Squared semipartial (part) correlation examples:

$$r^2_{Y(2.1)} = SSR(X_2|X_1)/SSTO = (r_{Y2} - r_{Y1}r_{12})^2/(1 - r^2_{12}) \quad [27]$$

$$r^2_{Y(2.13)} = SSR(X_2|X_1, X_3)/SSTO$$

Note that the standardized regression weight for b_2 has the same numerator as both the partial and semipartial correlations so that they all have the same sign and when one is zero the others are as well:

$$b^*_2 = (r_{Y2} - r_{Y1}r_{12})/(1 - r^2_{12}) \quad [28]$$

Thus, the significance test for the regression coefficient is also a significance test for the partial and semipartial correlations.

Note that the above Type I *SS* breakdown of *SSR* means that $R^2_{Y.123...k}$ may be partitioned as follows:

$$R^2_{Y.123...k} = r^2_{Y1} + r^2_{Y(2.1)} + r^2_{Y(3.12)} + ... \quad [29]$$

Note also that there are many possible breakdowns of $R^2_{Y.123...k}$ depending on the order that the *X*s are entered into the equation. Each squared semipartial correlation represents the additional increase in R^2 that one obtains by adding a new predictor to the equation. It is also important to realize that if the predictors (*X*s) are uncorrelated among themselves (orthogonal) the breakdown becomes

$$R^2_{Y.123...k} = r^2_{Y1} + r^2_{Y2} + r^2_{Y3} + ... \quad [30]$$

and it does not matter what order the predictors are entered into the equation. The Type I sequential *SS* are the same as the Type II *SS*.

Also, notice in the IQ, Extraversion and Sales Success example that the total $R^2_{Y.12}$ is greater than the sum $r^2_{Y1} + r^2_{Y2}$. Or another way of expressing the same phenomenon is to note that $SSR(X_2|X_1) > SSR(X_2)$ and $SSR(X_1|X_2) > SSR(X_1)$. This will not usually be the case, but it is certainly possible, as the example shows. The predictor variables seem to be enhancing one another in their prediction of *Y*. The inclusion of one predictor enhances the predictive effectiveness of the other predictor. This is one example of a phenomenon I call *enhancement* for obvious reasons. Sometimes in the literature predictors exhibiting this phenomenon are called *suppressor* variables for reasons that are not so obvious. See McFatter (1979) [The use of structural equation models in interpreting regression equations including suppressor and enhancer variables, *Applied Psychological Measurement, 3,*123-135], for additional examples and discussion.

Matrix Algebra and Matrix Formulation of Regression Analysis

Matrix algebra is a very powerful tool for mathematical and statistical analysis. It is very commonly applied to problems in statistics and, in particular, to regression analysis. Here are some basic definitions and operations used in matrix algebra. Bold face letters will be used to indicate matrices. Lower case bold letters are usually used to indicate vectors (one-dimensional matrices). Matrices are often written with subscripts to indicate the *order* (i.e., number of rows and columns) of the matrix. Matrix operations like the ones described below may be very easily carried out in Excel or Quattro Pro. *In Excel, it is important to remember that to carry out a matrix operation it is necessary to select the entire shape of the final matrix before entering the formula, enter an = sign to begin the formula, and end the formula with a Cntl-Shift-Enter simultaneous key press.*

Definitions:

SQUARE MATRIX. A square matrix is a matrix with the same number of rows as columns.

MATRIX TRANSPOSE. The transpose of matrix **A**, denoted by **A'**, is obtained by interchanging the rows and columns of **A**. The TRANSPOSE function in Excel will carry out this function.

$$\mathbf{A}_{3\times 2} = \begin{bmatrix} 1 & 5 \\ 9 & 4 \\ 0 & 3 \end{bmatrix} \qquad \mathbf{A'}_{2\times 3} = \begin{bmatrix} 1 & 9 & 0 \\ 5 & 4 & 3 \end{bmatrix}$$

SYMMETRIC MATRIX. A symmetric matrix is a square matrix whose rows are the same as its columns. That is, **A** = **A'**. For example, matrix **A** below is symmetric.

$$\mathbf{A}_{3\times 3} = \begin{bmatrix} 1 & 5 & 9 \\ 5 & 2 & 4 \\ 9 & 4 & 3 \end{bmatrix} \qquad \mathbf{A'}_{3\times 3} = \begin{bmatrix} 1 & 5 & 9 \\ 5 & 2 & 4 \\ 9 & 4 & 3 \end{bmatrix}$$

A common example of a symmetric matrix is a correlation matrix.

DIAGONAL MATRIX. A diagonal matrix is a square matrix whose off-diagonal elements are all zero.

$$\mathbf{A}_{3\times 3} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

IDENTITY MATRIX. An identity matrix, $\mathbf{I}_{n\times n}$ is a diagonal matrix with all diagonal elements equal to one.

$$\mathbf{I}_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Matrix Addition Example: $\mathbf{A}_{r\times c} + \mathbf{B}_{r\times c} = \mathbf{C}_{r\times c}$ [31]

$$\begin{bmatrix} 1 & 9 \\ 5 & 6 \\ 8 & 6 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 4 & 5 \\ 4 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 9 \\ 9 & 11 \\ 12 & 7 \end{bmatrix}$$

Matrix subtraction works analogously (i.e., element by element addition or subtraction). Note that only matrices with identical orders may be added or subtracted, and that the resultant matrix has the same order as the component matrices.

Matrix Multiplication Example: $\mathbf{A}_{p \times q}\,\mathbf{B}_{r \times s} = \mathbf{C}_{p \times s}$ [32]

$$\begin{bmatrix} 1 & 5 & 6 \\ 9 & 7 & 6 \end{bmatrix} \begin{bmatrix} 4 & 6 & 8 & 0 \\ 1 & 2 & 1 & 5 \\ 3 & 7 & 1 & 4 \end{bmatrix} = \begin{bmatrix} 27 & 58 & 19 & 49 \\ 61 & 110 & 85 & 59 \end{bmatrix}$$

In matrix multiplication, each row of the first matrix is multiplied with each column of the second matrix and the result summed. For example, in matrix **C** above, 27 = 1(4) + 5(1) + 6(3), and 110 = 9(6) + 7(2) + 6(7). Note that the order of the resultant matrix, **C**, is the number of columns of **A** times the number of rows of **B**. The function in Excel that performs matrix multiplication is MMULT.

Note that in order for two matrices to be 'conformable' for multiplication the column order of the first matrix must be equal to the row order of the second matrix, i.e., $q = r$. Moreover, in general, in matrix algebra $\mathbf{AB} \neq \mathbf{BA}$. In fact, both multiplications may not even be possible. This contrasts with ordinary algebra where, necessarily, $ab = ba$.

Note also that $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$, where **I** is the appropriate identity matrix. Thus, in matrix algebra the identity matrix functions like the number 1 in ordinary algebra.

Matrix Inverse Example: $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ [33]

Division, as such, is not defined in matrix algebra. However, the operation that functions analogously to division in ordinary algebra is multiplication by the inverse matrix. In ordinary algebra, division is equivalent to multiplication by the reciprocal (or inverse) of a number—dividing by a is equivalent to multiplying by $1/a = a^{-1}$. Thus $a\,(a^{-1}) = 1$. In matrix algebra, the inverse of a *square* matrix **A** is the matrix $\mathbf{A}^{-1}$ which when multiplied by **A** yields the identity matrix **I**.

$$\mathbf{A} = \begin{bmatrix} 6 & 4 \\ 1 & 2 \end{bmatrix} \qquad \mathbf{A}^{-1} = \begin{bmatrix} \frac{2}{8} & -\frac{4}{8} \\ -\frac{1}{8} & \frac{6}{8} \end{bmatrix} \qquad \mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad [34]$$

Finding the inverse of a square matrix is computationally tedious for anything more than a 2 × 2 matrix. The function that does this in Excel is the MINVERSE function.

The computation of the inverse of a square matrix involves the computation of a quantity called the determinant of the matrix. The determinant of **A** is sometimes denoted by $|\mathbf{A}|$. To find the inverse of **A**, a matrix called the adjoint of **A** (**Adj A**) is multiplied by the reciprocal of the determinant of **A**, i.e.,

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\,\mathbf{Adj\ A}\,. \quad [35]$$

In the 2 x 2 case, the determinant of matrix $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is $|\mathbf{A}| = ad - bc$. And the adjoint of A is

$\mathbf{Adj\ A} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. Thus, in the numerical example above, the determinant is 8 and the adjoint of **A**

is $\begin{bmatrix} 2 & -4 \\ -1 & 6 \end{bmatrix}$.

Because computation of the inverse of a matrix involves division of numbers by the determinant of the matrix, the operation becomes undefined when the determinant of the matrix is zero. When a matrix has a determinant that is zero, the matrix is said to be *singular*, and its inverse does not exist.

Linear Dependence and the Rank of a Matrix

A matrix will be singular when its rows or columns are *linearly dependent*. For example, if one of the columns of a matrix can be expressed as an exact linear combination of the other columns of the matrix, then the columns are linearly dependent. The following matrix is singular (its determinant is zero) because its columns are linearly dependent. The third column is equal to 3 times the first column minus 2 times the second column.

$$\mathbf{A} = \begin{bmatrix} 6 & 4 & 10 \\ 7 & 3 & 15 \\ 4 & 5 & 2 \end{bmatrix}. \qquad \text{No inverse exists for this matrix.}$$

When the rows and columns of a matrix are not linearly dependent the matrix is said to be of *full rank*. If a matrix is not of full rank, then it is singular. The rank of a matrix is defined to be the maximum number of linearly independent columns (or rows) of the matrix. The matrix above is of rank 2 because only two of the columns are linearly independent. It can be shown that the rank of an r x c matrix cannot exceed the minimum of r and c. For example, the rank of a 6 x 10 matrix can at most be 6. Also, when two matrices are multiplied, the rank of the resulting matrix can be at most the rank of the matrix with the smallest rank. Linear dependence and rank become important in regression analysis because the computation of regression coefficients involves finding the inverse of a matrix. If that matrix is singular, no solution exists.

Basic Theorems from Matrix Algebra

Below are some basic theorems for manipulation of matrices in matrix algebra. I have reproduced these from Kutner et al., p. 193:

$$\begin{aligned}
\mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A} \\
(\mathbf{A} + \mathbf{B}) + \mathbf{C} &= \mathbf{A} + (\mathbf{B} + \mathbf{C}) \\
(\mathbf{AB})\mathbf{C} &= \mathbf{A}(\mathbf{BC}) \\
\mathbf{C}(\mathbf{A} + \mathbf{B}) &= \mathbf{CA} + \mathbf{CB} \\
\lambda(\mathbf{A} + \mathbf{B}) &= \lambda\mathbf{A} + \lambda\mathbf{B} \qquad (\lambda \text{ is any scalar}) \\
(\mathbf{A}')' &= \mathbf{A} \\
(\mathbf{A} + \mathbf{B})' &= \mathbf{A}' + \mathbf{B}' \\
(\mathbf{AB})' &= \mathbf{B}'\mathbf{A}' \\
(\mathbf{ABC})' &= \mathbf{C}'\mathbf{B}'\mathbf{A}' \\
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} \\
(\mathbf{ABC})^{-1} &= \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \\
(\mathbf{A}^{-1})^{-1} &= \mathbf{A} \\
(\mathbf{A}')^{-1} &= (\mathbf{A}^{-1})'
\end{aligned} \qquad [36]$$

Notice that a major difference between matrix algebra and ordinary algebra is that sequential order of multiplication is crucial in matrix algebra whereas it is not in ordinary algebra.

Matrix Formulation of Regression Analysis

The general multiple regression model and the solution of its equations may be formulated very compactly using matrix notation:

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times (k+1)} \boldsymbol{\beta}_{(k+1) \times 1} + \boldsymbol{\varepsilon}_{n \times 1} \quad [37]$$

where k is the number of predictors in the regression model.

In the simple bivariate case using equation [2] and the simple data set from p. 3, the matrix formulation would be

$$\mathbf{Y}_{5 \times 1} = \mathbf{X}_{5 \times 2}\, \mathbf{b}_{2 \times 1} + \mathbf{e}_{5 \times 1} \quad \text{where}$$

$$\mathbf{Y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 6 \\ 1 & 6 \\ 1 & 3 \\ 1 & 3 \\ 1 & 2 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{bmatrix}$$

The ‘normal equations’ which reflect the solution to the least squares estimation of the model’s parameters can be shown to be (p. 199-200, Kutner et al.)

$$\mathbf{X'X\, b} = \mathbf{X'Y}. \quad [38]$$

This is a set of equations with the parameter estimates in **b** being the unknowns. The solution to this matrix equation may be found using the matrix algebra operations described above. We wish to solve for **b**. Because **X'X** is a square matrix its inverse may be found as long as it is nonsingular. Premultiplying both sides of equation [38] by $\mathbf{(X'X)}^{-1}$ leads to

$$\mathbf{(X'X)}^{-1}\mathbf{X'X\, b} = \mathbf{(X'X)}^{-1}\, \mathbf{X'Y}, \quad [39]$$

but $\mathbf{(X'X)}^{-1}\mathbf{X'X} = \mathbf{I}$, and $\mathbf{Ib} = \mathbf{b}$, so we have

$$\mathbf{b} = \mathbf{(X'X)}^{-1}\, \mathbf{X'Y}. \quad [40]$$

Equation [40] is the general matrix solution for the regression coefficients of any multiple regression problem regardless of the number of predictors or their nature.

The following Excel spreadsheet shows the computation of **b** as well as many of the other commonly calculated quantities in the regression analysis for the small bivariate example above.

X	
1	6
1	6
1	3
1	3
1	2

Y
1
2
3
4
5

X'X	
5	20
20	94

(X'X)inv	
1.34286	-0.2857
-0.2857	0.0714

X'Y
15
49

b=(X'X)inv*X'Y
6.142857
-0.78571

Yhat
1.429
1.429
3.786
3.786
4.571

e
-0.4
0.57
-0.8
0.21
0.43

e'e	
1.3571	=SSE

Var{b}	
0.60748	-0.1293
-0.1293	0.0323

MSE
0.4524

S{b}	
0.77941	
	0.1798

Hat=X(X'X)invX'				
0.486	0.486	0.06	0.06	-0.09
0.486	0.486	0.06	0.06	-0.09
0.057	0.057	0.27	0.27	0.343
0.057	0.057	0.27	0.27	0.343
-0.09	-0.09	0.34	0.34	0.486

Following are some of the important quantities we have looked at regression analysis expressed in matrix terms:

$$\hat{\mathbf{Y}} = \mathbf{Xb} \quad [41]$$

$$\mathbf{e} = \mathbf{Y} - \hat{\mathbf{Y}} \quad [42]$$

$$SSE = \mathbf{e'e} = \mathbf{Y'Y} - \mathbf{b'X'Y} \quad [43]$$

$$MSE = \mathbf{e'e}/(n\text{-}k\text{-}1) \quad [44]$$

$$\mathbf{s}^2\{\mathbf{b}\} = MSE\,(\mathbf{X'X})^{-1} \quad [45]$$

When the X and Y variables have been standardized ($M = 0$, $s = 1$), the normal equations [38] become

$$\mathbf{R}_{XX}\,\mathbf{b}^* = \mathbf{r}_{YX} \quad [46]$$

where $\mathbf{R}_{XX}$ is the correlation matrix of the X variables, $\mathbf{r}_{YX}$ is the column vector of correlations between Y and the X variables, and $\mathbf{b}^*$ is the column vector of standardized regression coefficients. Solving this matrix equation for $\mathbf{b}^*$ leads to

$$\mathbf{b}^* = \mathbf{R}^{-1}_{XX}\,\mathbf{r}_{YX}. \quad [47]$$

Thus, one could easily obtain the standardized regression coefficients from correlation matrix of all the variables in the analysis.

Multicollinearity

When the predictor variables in a regression analysis are correlated with one another they are said to be multicollinear. The degree of relationship (or multicollinearity) among the predictor variables is related to a number of problems in interpreting the results of a regression analysis. Although moderate levels of multicollinearity do not ordinarily produce major difficulties of interpretation, the multicollinearity does affect the interpretation of the equation, and extreme multicollinearity can make it impossible, or next to impossible, to even estimate the equation, let alone interpret it. It should be noted that some authors restrict use of the term ‘multicollinearity’ to these situations of extreme multicollinearity.

If the predictors are all uncorrelated with one another (i.e., no multicollinearity) then interpretation of the analysis is quite straightforward. The relative contribution that each variable makes to the prediction of the criterion variable is simply the square of the correlation of that predictor with the criterion, and the total R^2 may be partitioned into components that unambiguously reflect the contribution of each predictor as equation [30] shows.

Testing Blocks of Variables

The following is the output from a SAS analysis testing the effect of a block of variables in an omnibus test:

```
1               Example for Testing Blocks of variables     1

 Model: MODEL1
 Dependent Variable: DEP

                          Analysis of Variance

                          Sum of         Mean
  Source          DF      Squares       Square      F Value       Prob>F

  Model            9   2159.38844    239.93205        5.868       0.0001
  Error          294  12021.86156     40.89069
  C Total        303  14181.25000

      Root MSE       6.39458     R-square       0.1523
      Dep Mean       9.62500     Adj R-sq       0.1263
      C.V.          66.43722

                          Parameter Estimates

                   Parameter      Standard    T for H0:
   Variable  DF     Estimate         Error   Parameter=0    Prob > |T|

   INTERCEP   1    17.258054    3.84824995         4.485        0.0001
   IMP        1     0.234499    0.21341461         1.099        0.2728
   SOC        1    -0.772616    0.14636501        -5.279        0.0001
   AB1        1     0.010397    0.80341100         0.013        0.9897
   AB2        1    -0.594332    0.67680400        -0.878        0.3806
   AB3        1    -0.170676    0.53212308        -0.321        0.7486
   AB4        1     2.024047    1.16603605         1.736        0.0836
   AB5        1    -2.240639    0.63629947        -3.521        0.0005
   AB6        1     1.566799    1.00810390         1.554        0.1212
   AB7        1    -2.094335    0.95305308        -2.198        0.0288


   Variable  DF     Type I SS    Type II SS

   INTERCEP   1         28163    822.396981
   IMP        1      6.897263     49.369243
   SOC        1   1137.985210   1139.401802
   AB1        1     14.906676      0.006848
   AB2        1    180.175464     31.532414
   AB3        1     18.265787      4.206740
   AB4        1     21.949314    123.208791
   AB5        1    552.665954    507.043403
   AB6        1     29.081151     98.773434
   AB7        1    197.461618    197.461618
```

Example for Testing Blocks of variables 2

Model: MODEL2
Dependent Variable: DEP

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	9	2159.38844	239.93205	5.868	0.0001
Error	294	12021.86156	40.89069		
C Total	303	14181.25000			

Root MSE	6.39458	R-square	0.1523
Dep Mean	9.62500	Adj R-sq	0.1263
C.V.	66.43722		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	17.258054	3.84824995	4.485	0.0001
AB1	1	0.010397	0.80341100	0.013	0.9897
AB2	1	-0.594332	0.67680400	-0.878	0.3806
AB3	1	-0.170676	0.53212308	-0.321	0.7486
AB4	1	2.024047	1.16603605	1.736	0.0836
AB5	1	-2.240639	0.63629947	-3.521	0.0005
AB6	1	1.566799	1.00810390	1.554	0.1212
AB7	1	-2.094335	0.95305308	-2.198	0.0288
IMP	1	0.234499	0.21341461	1.099	0.2728
SOC	1	-0.772616	0.14636501	-5.279	0.0001

Variable	DF	Type I SS	Type II SS
INTERCEP	1	28163	822.396981
AB1	1	32.823095	0.006848
AB2	1	176.975460	31.532414
AB3	1	6.427272	4.206740
AB4	1	17.201841	123.208791
AB5	1	553.739673	507.043403
AB6	1	54.363970	98.773434
AB7	1	162.753435	197.461618
IMP	1	15.701890	49.369243
SOC	1	1139.401802	1139.401802

```
Dependent Variable: DEP
Test: ABUSE Numerator:    144.9294  DF:    7   F value:   3.5443
            Denominator:  40.89069  DF:  294   Prob>F:    0.0011

Dependent Variable: DEP
Test: EXTR  Numerator:    577.5518  DF:    2   F value:  14.1243
            Denominator:  40.89069  DF:  294   Prob>F:    0.0001
```

Stepwise Selection of Predictors

The following is the output from a JMP analysis using the forward selection and backwards elimination techniques to select predictors in a regression analysis predicting subjective well-being (W) from demographic and social relationship variables:

Stepwise Fit

Response:
w

Stepwise Regression Control

Prob to Enter	0.030
Prob to Leave	0.030

Direction:

Current Estimates

SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
1622.202	139	11.67052	0.4162	0.3994	10.72091	358.7286

Lock

Entered

Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
Intercept	8.05864558	1	0	0.000	1.0000
sex	.	1	41.72756	3.643	0.0584
race	.	1	6.257591	0.534	0.4660
empl	.	1	0.686317	0.058	0.8094
spouse	.	1	7.123479	0.609	0.4366
mom	2.31127751	1	121.2652	10.391	0.0016
dad	.	1	39.5474	3.448	0.0654
sibs	2.32482843	1	133.2412	11.417	0.0009
rel	.	1	0.542101	0.046	0.8303
numso	.	1	6.752823	0.577	0.4488
timeso	.	1	34.81695	3.027	0.0841
e	0.35800083	1	259.1405	22.205	0.0000
n	-0.3067543	1	344.9606	29.558	0.0000
l	.	1	2.138198	0.182	0.6702
acq	.	1	21.91027	1.889	0.1715
frnd	.	1	45.22244	3.957	0.0486

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
1	e	Entered	0.0000	479.4523	0.1725	65.118	2
2	n	Entered	0.0000	348.7243	0.2980	36.008	3
3	sibs	Entered	0.0001	207.0159	0.3726	19.539	4
4	mom	Entered	0.0016	121.2652	0.4162	10.721	5

Stepwise Fit

Response:
w

Stepwise Regression Control

Prob to Enter	0.030
Prob to Leave	0.030

Current Estimates

SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
1434.7743	128	11.20917	0.4836	0.4231	16	363.0487

Lock

Entered

Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
Intercept	6.60702504	1	0	0.000	1.0000
sex	-1.2972898	1	50.85201	4.537	0.0351
race	0.50082994	1	7.165031	0.639	0.4255
empl	-0.0623098	1	0.12838	0.011	0.9149
spouse	0.44541743	1	5.984767	0.534	0.4663
mom	1.46533811	1	37.36558	3.333	0.0702
dad	1.20316572	1	35.57343	3.174	0.0772
sibs	2.16149578	1	104.601	9.332	0.0027
rel	-0.7012345	1	11.16948	0.996	0.3201
numso	-0.0573344	1	0.44578	0.040	0.8422
timeso	0.34307195	1	22.82063	2.036	0.1561
e	0.27976401	1	131.9864	11.775	0.0008
n	-0.3231491	1	336.0496	29.980	0.0000
l	-0.0953823	1	3.246401	0.290	0.5914
acq	0.03976322	1	4.032358	0.360	0.5497
frnd	0.14580884	1	29.51324	2.633	0.1071

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
1	e	Entered	0.0000	479.4523	0.1725	65.118	2
2	n	Entered	0.0000	348.7243	0.2980	36.008	3
3	sibs	Entered	0.0001	207.0159	0.3726	19.539	4
4	mom	Entered	0.0016	121.2652	0.4162	10.721	5

Stepwise Fit

Response:
w

Stepwise Regression Control

Prob to Enter	0.030
Prob to Leave	0.030

Direction:

Current Estimates

SSE	DFE	MSE	RSquare	RSquare Adj	Cp	AIC
1521.3101	137	11.10445	0.4525	0.4285	5.720085	353.482

Entered

Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
Intercept	8.93337147	1	0	0.000	1.0000
sex	-1.2945208	1	55.66942	5.013	0.0268
race	.	1	7.209445	0.648	0.4224
empl	.	1	0.096685	0.009	0.9261
spouse	.	1	3.273621	0.293	0.5890
mom	1.94450411	1	82.84942	7.461	0.0071
dad	.	1	36.98623	3.389	0.0678
sibs	2.2709447	1	126.754	11.415	0.0009
rel	.	1	3.567762	0.320	0.5727
numso	.	1	0.000593	0.000	0.9942
timeso	.	1	25.4668	2.315	0.1304
e	0.2798742	1	137.3623	12.370	0.0006
n	-0.3241244	1	380.7799	34.291	0.0000
l	.	1	2.254577	0.202	0.6539
acq	.	1	1.229298	0.110	0.7407
frnd	0.18183481	1	59.16431	5.328	0.0225

Step History

Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p
1	e	Entered	0.0000	479.4523	0.1725	65.118	2
2	n	Entered	0.0000	348.7243	0.2980	36.008	3
3	sibs	Entered	0.0001	207.0159	0.3726	19.539	4
4	mom	Entered	0.0016	121.2652	0.4162	10.721	5
5	empl	Removed	0.9149	0.12838	0.4836	14.011	15
6	numso	Removed	0.8448	0.427913	0.4834	12.05	14
7	l	Removed	0.5918	3.189925	0.4823	10.334	13
8	acq	Removed	0.6265	2.612655	0.4814	8.5673	12
9	race	Removed	0.5129	4.698999	0.4797	6.9865	11
10	spouse	Removed	0.5187	4.551578	0.4780	5.3926	10
11	rel	Removed	0.2944	11.9927	0.4737	4.4625	9
12	timeso	Removed	0.1569	21.9474	0.4658	4.4204	8
13	dad	Removed	0.0678	36.98623	0.4525	5.7201	7

'Intrinsically Linear' Nonlinear Functions

It is common to desire to fit a function more complex than a linear one to a set of data. Of course, there are a huge number of possible nonlinear functions that one could fit. Iterative methods are commonly used to estimate many nonlinear functions, but for a class of nonlinear functions called 'intrinsically linear,' the usual methods of standard least squares multiple regression analysis may be used to estimate the model. Intrinsically linear functions are ones that, through some combination of transformations of the variables involved, may be brought into the form of the general linear regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \varepsilon_i. \qquad [48]$$

Some examples of intrinsically linear functions are the simple 'power function'

$$Y = \alpha X^{\beta} \varepsilon, \qquad [49]$$

which may be transformed into a linear function by taking the logarithm of both sides:

$$\ln(Y) = \ln(\alpha) + \beta \ln(X) + \ln(\varepsilon) \qquad [50]$$

and estimated by simply regressing the logarithm of Y on the logarithm of X using usual least squares methods,

or $$Y = \alpha e^{\beta X} \varepsilon \qquad [51]$$

which may also be transformed into a linear function by taking the logarithm of both sides:

$$\ln(Y) = \ln(\alpha) + \beta X + \ln(\varepsilon) \qquad [52]$$

and estimated by simply regressing the logarithm of Y on untransformed X using usual least squares methods.

Polynomial Regression

Another family of intrinsically linear functions is the class of polynomial functions:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 \ldots + \varepsilon. \qquad [53]$$

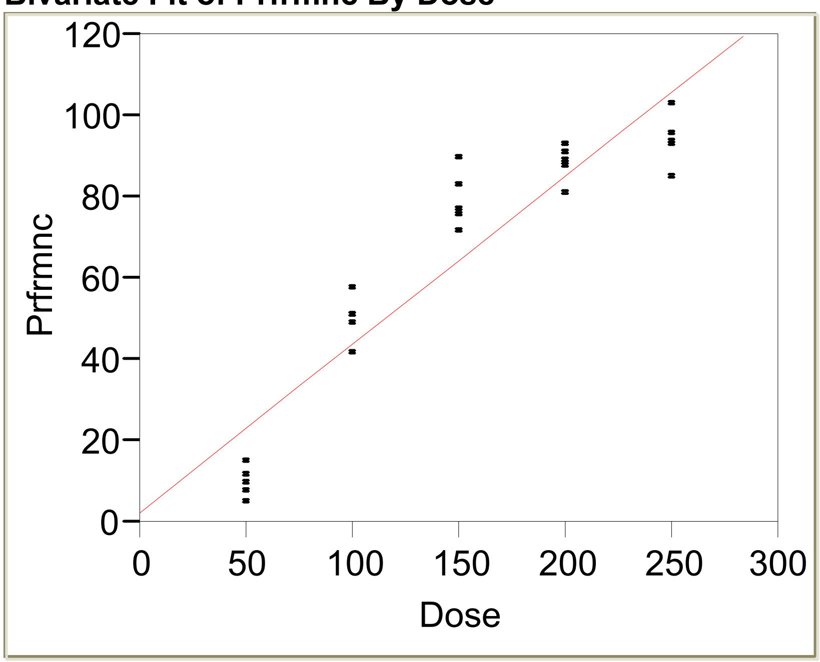
These functions may be estimated simply by regressing Y on the untransformed, squared, cubed, etc. values of the Xs.

Polynomial Regression Example.

Consider the following hypothetical set of data from a study examining performance as a function of drug dosage with subjects randomly assigned to receive one of five dosages (50mg, 100mg, 150mg, 200mg, 250mg) of a certain drug. There are five subjects in each condition.

Obs	Prfrmnc	Dose	DoseDev	Dose2	Dose3	Dose4	Predicted Prfrmnc
1	15	50	-100	2500	125000	6250000	10
2	5	50	-100	2500	125000	6250000	10
3	8	50	-100	2500	125000	6250000	10
4	12	50	-100	2500	125000	6250000	10
5	10	50	-100	2500	125000	6250000	10
6	58	100	-50	10000	1000000	100000000	50.2
7	51	100	-50	10000	1000000	100000000	50.2
8	49	100	-50	10000	1000000	100000000	50.2
9	42	100	-50	10000	1000000	100000000	50.2
10	51	100	-50	10000	1000000	100000000	50.2
11	90	150	0	22500	3375000	506250000	79.6
12	83	150	0	22500	3375000	506250000	79.6
13	77	150	0	22500	3375000	506250000	79.6
14	76	150	0	22500	3375000	506250000	79.6
15	72	150	0	22500	3375000	506250000	79.6
16	88	200	50	40000	8000000	1600000000	88.4
17	91	200	50	40000	8000000	1600000000	88.4
18	81	200	50	40000	8000000	1600000000	88.4
19	89	200	50	40000	8000000	1600000000	88.4
20	93	200	50	40000	8000000	1600000000	88.4
21	103	250	100	62500	15625000	3906250000	94.2
22	85	250	100	62500	15625000	3906250000	94.2
23	96	250	100	62500	15625000	3906250000	94.2
24	93	250	100	62500	15625000	3906250000	94.2
25	94	250	100	62500	15625000	3906250000	94.2

Bivariate Fit of Prfrmnc By Dose



Linear Fit

Prfrmnc = 2.5 + 0.4132 Dose

Summary of Fit

RSquare	0.856541
RSquare Adj	0.850304
Root Mean Square Error	12.46641
Mean of Response	64.48
Observations (or Sum Wgts)	25

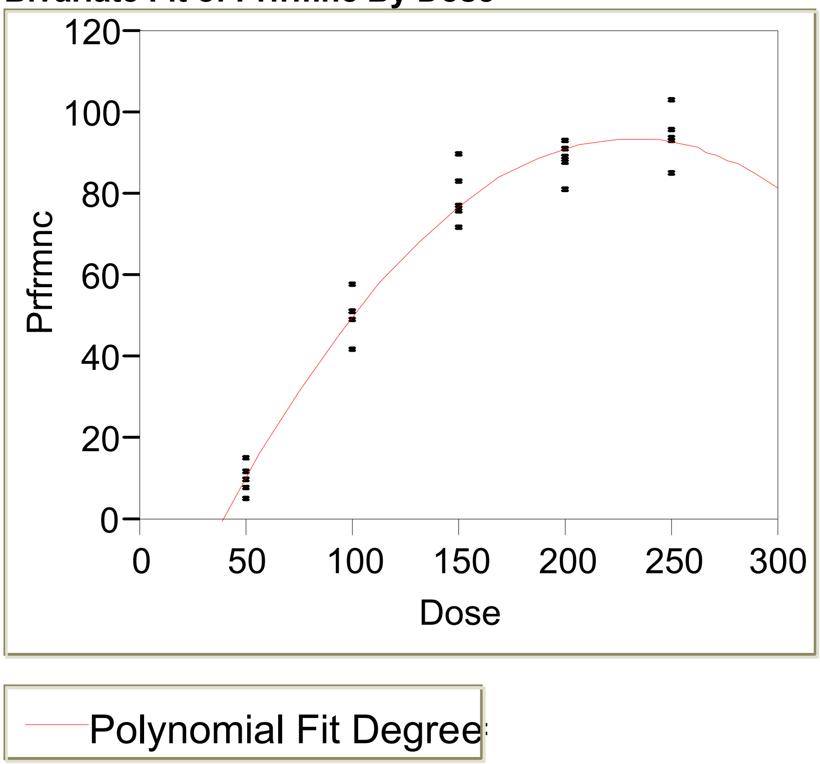
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	21341.780	21341.8	137.3245
Error	23	3574.460	155.4	Prob > F
C. Total	24	24916.240		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.5	5.847263	0.43	0.6730
Dose	0.4132	0.03526	11.72	<.0001

Bivariate Fit of Prfrmnc By Dose



Polynomial Fit Degree=2

Prfrmnc = 15.271429 + 0.4132 Dose - 0.0025543 (Dose-150)^2

Summary of Fit

RSquare	0.971101
RSquare Adj	0.968474
Root Mean Square Error	5.720957
Mean of Response	64.48
Observations (or Sum Wgts)	25

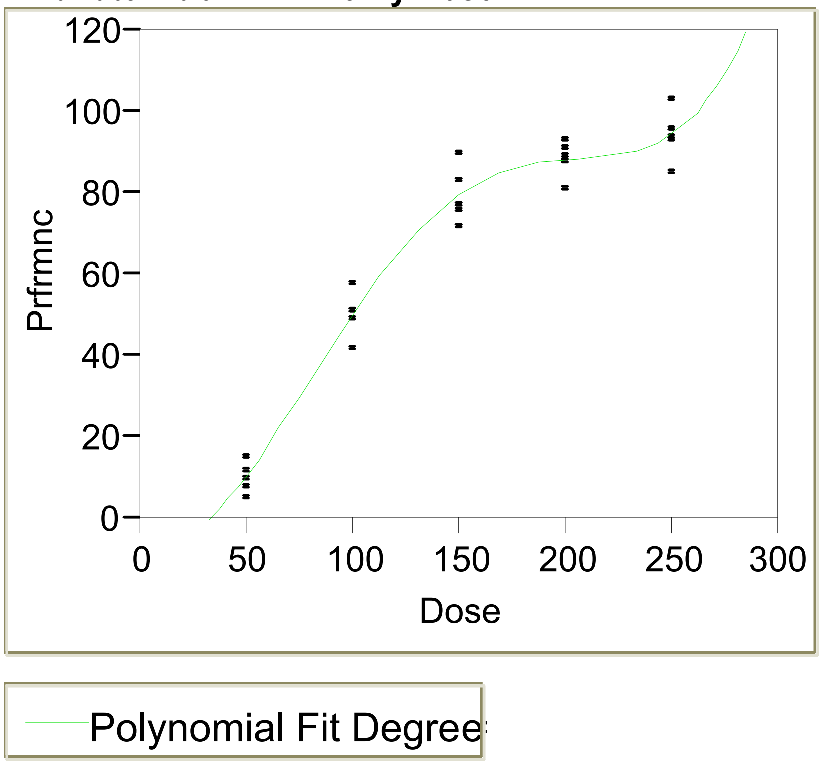
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	24196.194	12098.1	369.6406
Error	22	720.046	32.7	Prob > F
C. Total	24	24916.240		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	15.271429	3.011761	5.07	<.0001
Dose	0.4132	0.016181	25.54	<.0001
(Dose-150)^2	-0.002554	0.000274	-9.34	<.0001

Bivariate Fit of Prfrmnc By Dose



Polynomial Fit Degree=4

Prfrmnc = 24.25 + 0.369 Dose - 0.0045767 (Dose-150)^2 + 0.0000052 (Dose-150)^3 + 1.8267e-7 (Dose-150)^4

Summary of Fit

RSquare	0.974474
RSquare Adj	0.969369
Root Mean Square Error	5.639149
Mean of Response	64.48
Observations (or Sum Wgts)	25

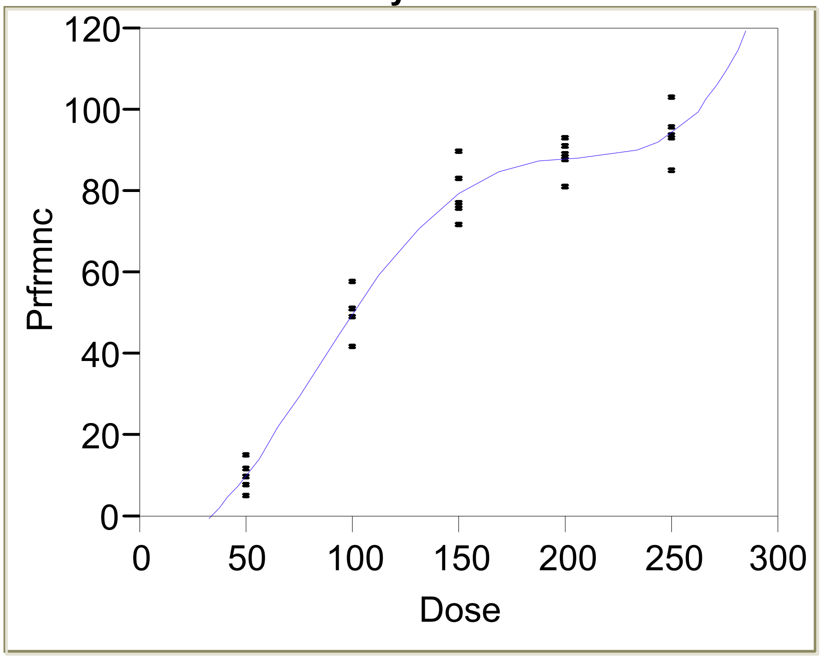
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	24280.240	6070.06	190.8824
Error	20	636.000	31.80	Prob > F
C. Total	24	24916.240		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	24.25	7.618071	3.18	0.0047
Dose	0.369	0.047924	7.70	<.0001
(Dose-150)^2	-0.004577	0.001581	-2.90	0.0089
(Dose-150)^3	0.0000052	0.000005	0.98	0.3397
(Dose-150)^4	1.8267e-7	1.407e-7	1.30	0.2089

Bivariate Fit of Prfrmnc By Dose



Polynomial Fit Degree

Polynomial Fit Degree=4

Prfrmnc = -3.8 - 0.373 Dose + 0.0177433 Dose^2 - 0.0001044 Dose^3 + 1.8267e-7 Dose^4

Summary of Fit

RSquare	0.974474
RSquare Adj	0.969369
Root Mean Square Error	5.639149
Mean of Response	64.48
Observations (or Sum Wgts)	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	24280.240	6070.06	190.8824
Error	20	636.000	31.80	Prob > F
C. Total	24	24916.240		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-3.8	39.95447	-0.10	0.9252
Dose	-0.373	1.468045	-0.25	0.8020
Dose^2	0.0177433	0.017598	1.01	0.3254
Dose^3	-0.000104	0.000085	-1.23	0.2313
Dose^4	1.8267e-7	1.407e-7	1.30	0.2089

Response Prfrmnc
Summary of Fit

RSquare	0.974474
RSquare Adj	0.969369
Root Mean Square Error	5.639149
Mean of Response	64.48
Observations (or Sum Wgts)	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	24280.240	6070.06	190.8824
Error	20	636.000	31.80	Prob > F
C. Total	24	24916.240		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-3.8	39.95447	-0.10	0.9252
Dose	-0.373	1.468045	-0.25	0.8020
Dose*Dose	0.0177433	0.017598	1.01	0.3254
Dose*Dose*Dose	-0.000104	0.000085	-1.23	0.2313
Dose*Dose*Dose*Dose	1.8267e-7	1.407e-7	1.30	0.2089

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Dose	1	1	2.052891	0.0646	0.8020
Dose*Dose	1	1	32.327780	1.0166	0.3254
Dose*Dose*Dose	1	1	48.465534	1.5241	0.2313
Dose*Dose*Dose*Dose	1	1	53.625714	1.6863	0.2089

Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
Dose	1	1	21341.780	671.1252	<.0001
Dose*Dose	1	1	2854.414	89.7615	<.0001
Dose*Dose*Dose	1	1	30.420	0.9566	0.3397
Dose*Dose*Dose*Dose	1	1	53.626	1.6863	0.2089

Response Prfrmnc
Summary of Fit

RSquare	0.974474
RSquare Adj	0.969369
Root Mean Square Error	5.639149
Mean of Response	64.48
Observations (or Sum Wgts)	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	24280.240	6070.06	190.8824
Error	20	636.000	31.80	Prob > F
C. Total	24	24916.240		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	79.6	2.521904	31.56	<.0001
DoseDev	0.369	0.047924	7.70	<.0001
DoseDev*DoseDev	-0.004577	0.001581	-2.90	0.0089
DoseDev*DoseDev*DoseDev	0.0000052	0.000005	0.98	0.3397
DoseDev*DoseDev*DoseDev*DoseDev	1.8267e-7	1.407e-7	1.30	0.2089

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
DoseDev	1	1	1885.3062	59.2864	<.0001
DoseDev*DoseDev	1	1	266.6378	8.3848	0.0089
DoseDev*DoseDev*DoseDev	1	1	30.4200	0.9566	0.3397
DoseDev*DoseDev*DoseDev*DoseDev	1	1	53.6257	1.6863	0.2089

Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
DoseDev	1	1	21341.780	671.1252	<.0001
DoseDev*DoseDev	1	1	2854.414	89.7615	<.0001
DoseDev*DoseDev*DoseDev	1	1	30.420	0.9566	0.3397
DoseDev*DoseDev*DoseDev*DoseDev	1	1	53.626	1.6863	0.2089

Response Prfrmnc
Summary of Fit

RSquare	0.974474
RSquare Adj	0.969369
Root Mean Square Error	5.639149
Mean of Response	64.48
Observations (or Sum Wgts)	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	24280.240	6070.06	190.8824
Error	20	636.000	31.80	Prob > F
C. Total	24	24916.240		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	64.48	1.12783	57.17	<.0001
Dose[50]	-54.48	2.25566	-24.15	<.0001
Dose[100]	-14.28	2.25566	-6.33	<.0001
Dose[150]	15.12	2.25566	6.70	<.0001
Dose[200]	23.92	2.25566	10.60	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Dose	4	4	24280.240	190.8824	<.0001

Effect Details
Dose
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
50	10.000000	2.5219040	10.0000
100	50.200000	2.5219040	50.2000
150	79.600000	2.5219040	79.6000
200	88.400000	2.5219040	88.4000
250	94.200000	2.5219040	94.2000

Lack of Fit

Because the cubic and quartic trends in the above analyses are both nonsignificant (from the sequential Type 1 tests) the best model appears to be one that has both linear and quadratic components only. The model on the following page shows the analysis with only those two components. Note that this model and the previous ones have really only one predictor, Dose, and it is simply the form of the model (i.e., how complex the polynomial should be) that is in question. Note also that we have only five levels of the predictor, Dose, and multiple observations at each level of Dose. There is a limit, therefore, to the degree of fit that any model that includes only Dose could have to this set of observations. Any model that perfectly fits the means of the five groups will have obtained the maximum fit possible. Additional variation of Performance scores around those means represents *pure* (irreducible) *error*.

It is common to evaluate how well a model fits a set of data by testing whether the error variation left over after the model fit is significantly greater than pure error. If this *lack of fit* error is not significant, then we conclude that the model is a good fit to the data. In the polynomial trend analysis example above, the pure error estimate would be the *MSE* for the quartic model that perfectly fits the five means. No model with only Dose as a predictor will do better than perfectly fitting the five means. Is the quadratic model (with only linear and quadratic components) a good fit to the data? The Lack of Fit section of the JMP output tests this by testing whether the *block* of two additional predictors that would lead to a perfect fit (cubic and quartic) significantly improves prediction over the linear and quadratic. Because it does not, $F(2, 20) = 1.32$, $p = .2890$, we conclude that the quadratic model is a good fit.

Response Prfrmnc
Summary of Fit

RSquare	0.971101
RSquare Adj	0.968474
Root Mean Square Error	5.720957
Mean of Response	64.48
Observations (or Sum Wgts)	25

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	24196.194	12098.1	369.6406
Error	22	720.046	32.7	Prob > F
C. Total	24	24916.240		<.0001

Lack Of Fit

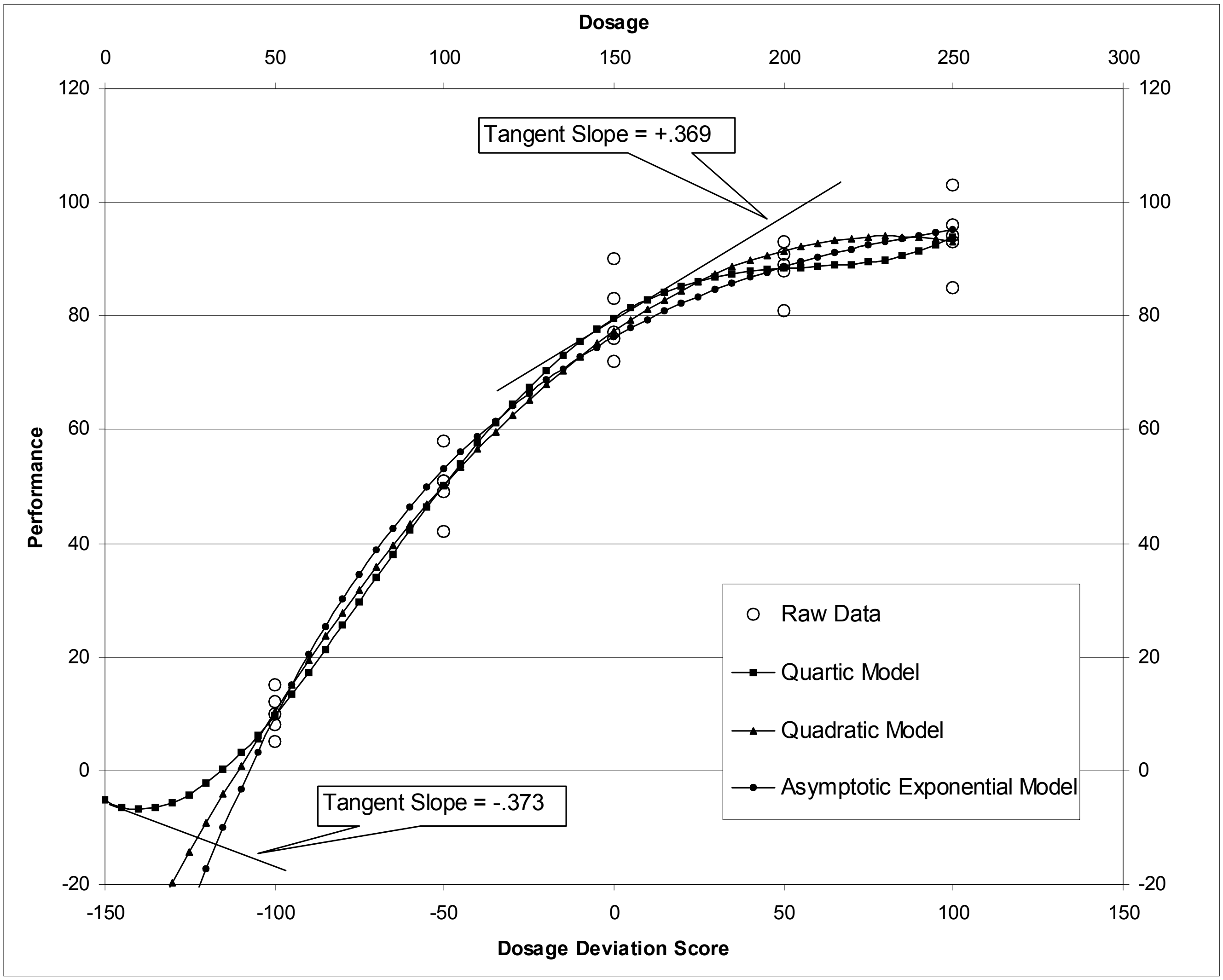
Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	2	84.04571	42.0229	1.3215
Pure Error	20	636.00000	31.8000	Prob > F
Total Error	22	720.04571		0.2890
				Max RSq
				0.9745

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	77.251429	1.783094	43.32	<.0001
DoseD	0.4132	0.016181	25.54	<.0001
DoseD*DoseD	-0.002554	0.000274	-9.34	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
DoseD	1	1	21341.780	652.0685	<.0001
DoseD*DoseD	1	1	2854.414	87.2127	<.0001


Dosage
0
50
100
150
200
250
300
120
100
80
60
40
20
0
-20
Performance
Tangent Slope = +.369
Tangent Slope = -.373
Raw Data
Quartic Model
Quadratic Model
Asymptotic Exponential Model
-150
-100
-50
0
50
100
150
Dosage Deviation Score

Models with Cross-Product (Interaction) Terms

In the simple multiple regression model with two predictors,

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 , \qquad [54]$$

it is important to notice that the value of b_1, which is the conditional or partial effect of X_1 on Y holding X_2 constant, does not depend on the value at which one holds X_2 constant. That is, the effect of X_1 on Y is the same, namely, b_1, whether X_2 is held constant at a very low level or at a high level. This is clearly reflected in the parallel lines in the 3-D surface plot of the best-fitting plane (see the plot in the IQ, extraversion, sales success example above).

If one includes a cross-product term in the equation, however, the response surface is no longer a plane, but rather a warped surface with nonparallel lines in the 3-D plot. Such a model would be

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_1 X_2 . \qquad [55]$$

Notice that because X_1 appears in the equation more than once, the conditional effect of X_1 on Y, holding X_2 constant, is not simply b_1. It is easy to see what the conditional effect of X_1 is, however, by simply algebraically rearranging the equation:

$$\hat{Y} = b_0 + \left(b_1 + b_3 X_2\right) X_1 + b_2 X_2 . \qquad [56]$$

Notice that the coefficient of X_1 in this equation is the quantity ($b_1 + b_3 X_2$). This quantity is the conditional effect of X_1 on Y, holding X_2 constant, and its value depends on the value of X_2. If X_2 is zero, then the effect of X_1 is b_1. However, if $X_2 = 1$, for example, then the effect of X_1 is $b_1 + b_3$. When the effect of one independent or predictor variable on the criterion depends on the level of a second predictor, the two predictors are said to *interact*. Thus, a model that contains a cross-product term is an *interaction* model, as opposed to a purely additive model.

The interaction effect works the same way for X_2. A rearrangement of equation [55] isolating X_2 leads to the conditional effect of X_2 being ($b_2 + b_3 X_1$).

Notice, in particular, what the implications of these considerations are for the interpretations of the three regression coefficients, b_1, b_2, and b_3, and their significance tests. It is a common error for researchers to interpret b_1 and b_2 as 'main effects' similar to the main effects in an ANOVA. It is important to see exactly why in general that is not correct. As equation [56] shows, b_1 is the conditional effect of X_1 on Y, holding X_2 constant *at zero*. Therefore, b_1 only has a meaningful interpretation for individuals who have a score of zero on X_2. If zero is not a possible score for X_2, or a score of interest, then b_1 and its significance test do not have much interpretive value. Certainly one could not interpret it as the 'main effect' or average or simple effect of X_1. The same goes for b_2.

The interpretation of b_3 is much less problematic. It is how much the conditional effect (slope) of X_1 on Y, holding X_2 constant, changes as X_2 increases by one unit. Notice that it is also how much the conditional effect (slope) of X_2 on Y, holding X_1 constant, changes as X_1 increases by one unit.

The problems in interpreting b_1 and b_2 occur because of the presence of the higher order term, X_1X_2, in the equation as well as the fact that zero may not be an interesting or meaningful value for X_1 and X_2. The coefficients b_1 and b_2 become much more interpretable if one rescales X_1 and X_2 before the analysis so that zero becomes an interesting value for the two variables. The most common way to do this is to standardize X_1 and X_2 to have means of zero and standard deviations of one before computing the cross-product. When that is done, b_1 becomes the effect of X_1 on Y for individuals who are at the mean on X_2. The significance test of b_1 tests whether that conditional effect is different from zero.

In models that have higher order terms (e.g., cross-products, squared terms) it is crucial in interpreting the lower order coefficients to pay careful attention to what a zero on the variables means. Without meaningful zeros on the variables the lower order coefficients and their significance tests become uninterpretable. Consequently, I recommend that unless a variable has a naturally interesting and interpretable zero, it should be standardized (or at least centered) prior to making cross-products or higher order terms in a regression model.

It is important to recognize that this procedure is not the same as simply doing a raw score regression analysis with higher order terms and then looking at the standardized regression weights from the analysis. The standardized regression weights from such an analysis are uninterpretable. This abominable practice is, unfortunately, all too common in the literature.

It is also important for interpretability of the coefficients that if a higher order term is in the model (e.g., X_1X_2), then all the components of the higher order term must also be in the model as lower order terms.

Extended Example. Predicting Subjective Well-being (W) from Extraversion (E) and Neuroticism (N).

Response W

Summary of Fit

RSquare	0.353161
RSquare Adj	0.349765
Root Mean Square Error	3.860245
Mean of Response	16.54427
Observations (or Sum Wgts)	384

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	3099.7785	1549.89	104.0090
Error	381	5677.4689	14.90	Prob > F
C. Total	383	8777.2474		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	209	3012.5404	14.4141	0.9303
Pure Error	172	2664.9286	15.4938	Prob > F
Total Error	381	5677.4689		0.6917
				Max RSq
				0.6964

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	16.544271	0.196992	83.98	<.0001
Std E	1.9318696	0.198589	9.73	<.0001
Std N	-1.87633	0.198589	-9.45	<.0001

Response W

Summary of Fit

RSquare	0.366178
RSquare Adj	0.359488
Root Mean Square Error	3.831274
Mean of Response	16.54427
Observations (or Sum Wgts)	384

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	3214.0339	803.508	54.7399
Error	379	5563.2135	14.679	Prob > F
C. Total	383	8777.2474		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	207	2898.2849	14.0014	0.9037
Pure Error	172	2664.9286	15.4938	Prob > F
Total Error	379	5563.2135		0.7576
				Max RSq
				0.6964

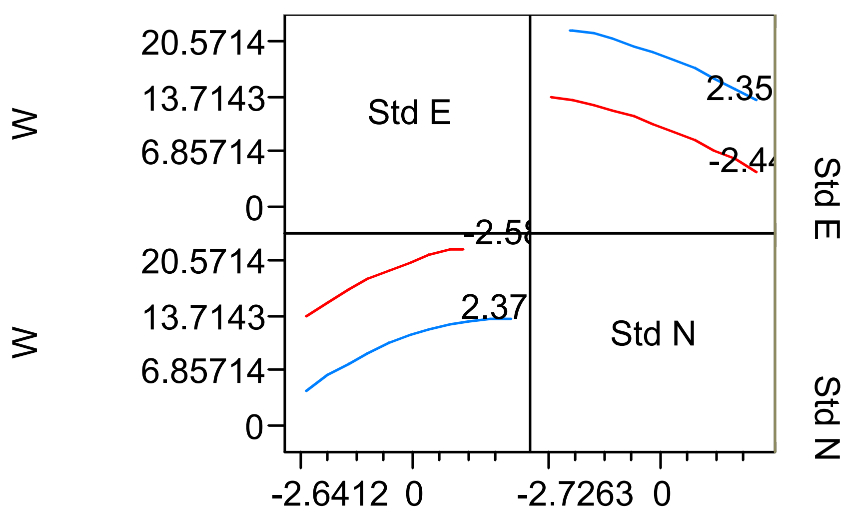
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	17.144342	0.296088	57.90	<.0001
Std E	1.8246413	0.201551	9.05	<.0001
Std E*Std E	-0.398624	0.16596	-2.40	0.0168
Std N	-1.892636	0.197274	-9.59	<.0001
Std N*Std N	-0.203014	0.158653	-1.28	0.2015

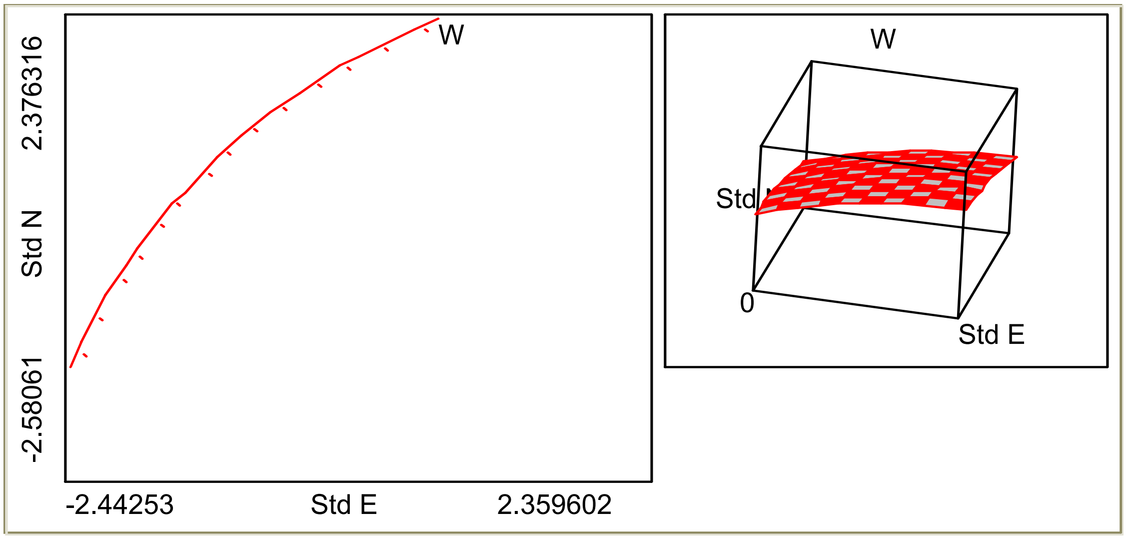
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Std E	1	1	1203.0119	81.9565	<.0001
Std E*Std E	1	1	84.6845	5.7692	0.0168
Std N	1	1	1351.0858	92.0442	<.0001
Std N*Std N	1	1	24.0350	1.6374	0.2015

Interaction Profiles



Contour Profiler



Response W
Summary of Fit

RSquare	0.373968
RSquare Adj	0.369025
Root Mean Square Error	3.802645
Mean of Response	16.54427
Observations (or Sum Wgts)	384

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	3282.4055	1094.14	75.6658
Error	380	5494.8419	14.46	Prob > F
C. Total	383	8777.2474		<.0001

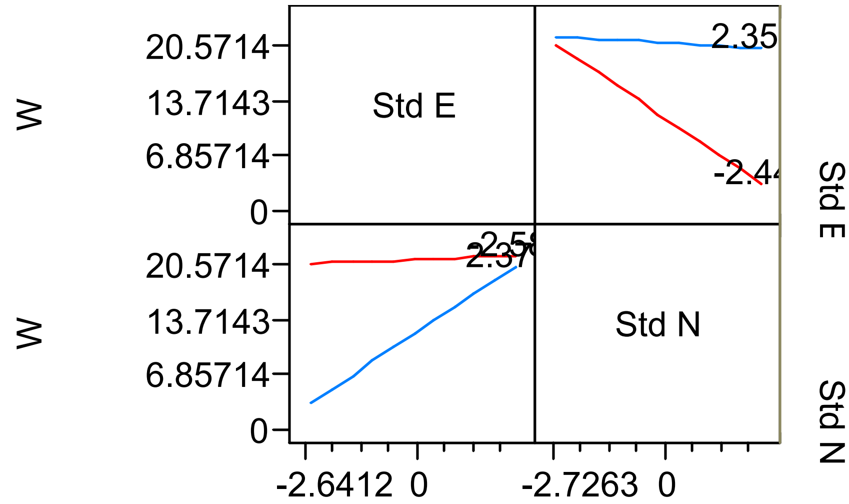
Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	208	2829.9133	13.6054	0.8781
Pure Error	172	2664.9286	15.4938	Prob > F
Total Error	380	5494.8419		0.8151
				Max RSq
				0.6964

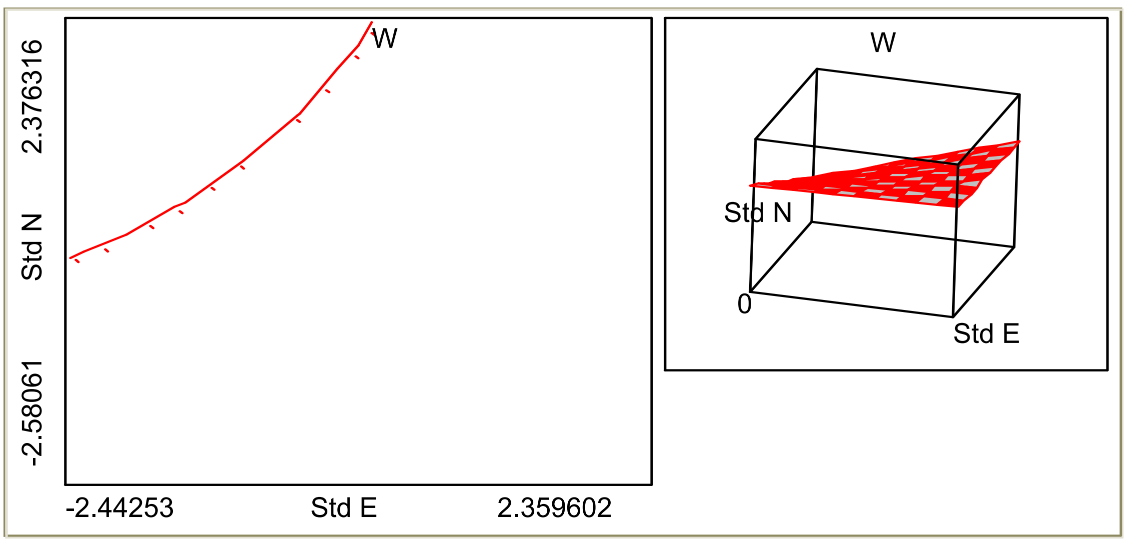
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	16.621353	0.195261	85.12	<.0001
Std E	1.9239807	0.195638	9.83	<.0001
Std N	-1.861427	0.195671	-9.51	<.0001
Std E*Std N	0.6664521	0.187531	3.55	0.0004

Interaction Profiles



Contour Profiler



Response W

Summary of Fit

RSquare	0.383923
RSquare Adj	0.375773
Root Mean Square Error	3.782256
Mean of Response	16.54427
Observations (or Sum Wgts)	384

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	3369.7836	673.957	47.1119
Error	378	5407.4638	14.305	Prob > F
C. Total	383	8777.2474		<.0001

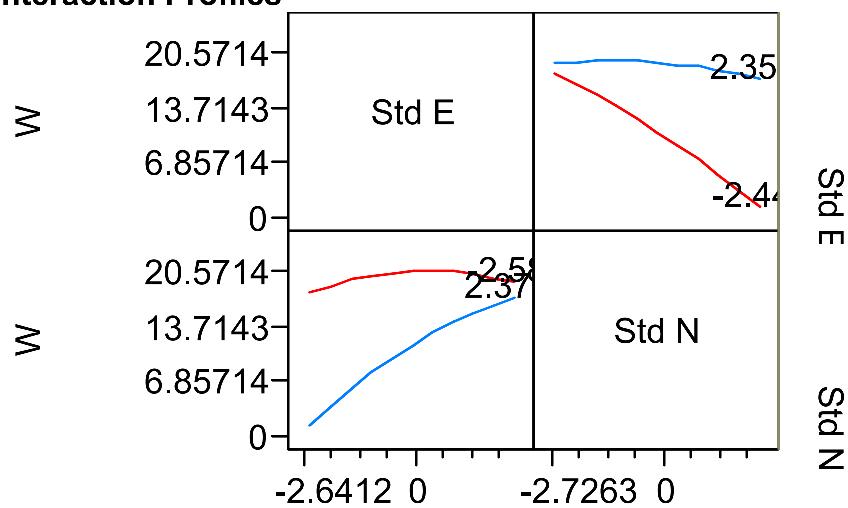
Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	206	2742.5352	13.3133	0.8593
Pure Error	172	2664.9286	15.4938	Prob > F
Total Error	378	5407.4638		0.8516
				Max RSq
				0.6964

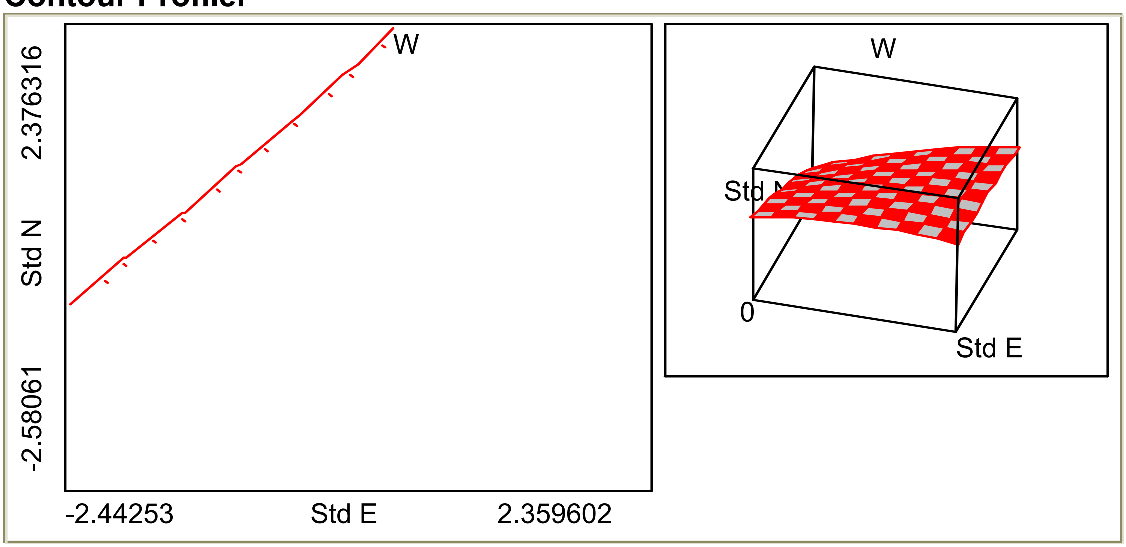
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	17.144204	0.2923	58.65	<.0001
Std E	1.830522	0.198981	9.20	<.0001
Std E*Std E	-0.349243	0.164519	-2.12	0.0344
Std N	-1.876914	0.194808	-9.63	<.0001
Std N*Std N	-0.180504	0.156771	-1.15	0.2503
Std E*Std N	0.6187539	0.187523	3.30	0.0011

Interaction Profiles



Contour Profiler



Response W

Summary of Fit

RSquare	0.390899
RSquare Adj	0.377905
Root Mean Square Error	3.775793
Mean of Response	16.54427
Observations (or Sum Wgts)	384

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	8	3431.0180	428.877	30.0827
Error	375	5346.2294	14.257	Prob > F
C. Total	383	8777.2474		<.0001

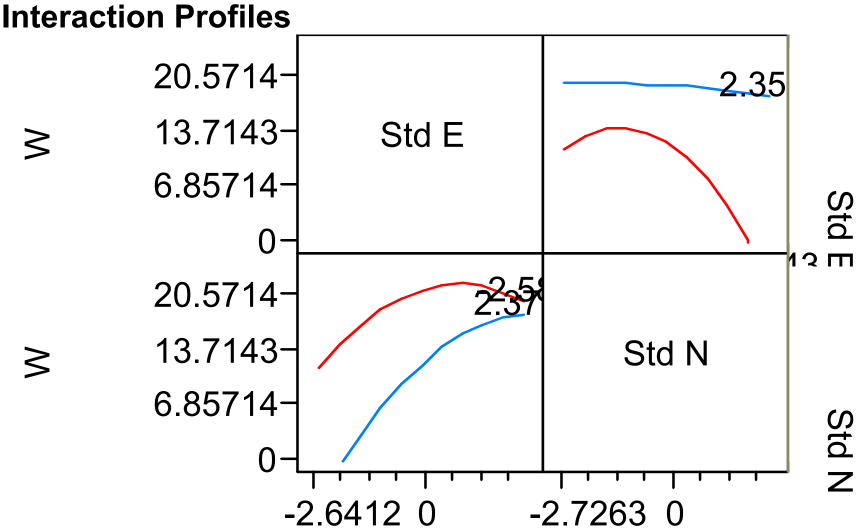
Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	203	2681.3008	13.2084	0.8525
Pure Error	172	2664.9286	15.4938	Prob > F
Total Error	375	5346.2294		0.8629
				Max RSq
				0.6964

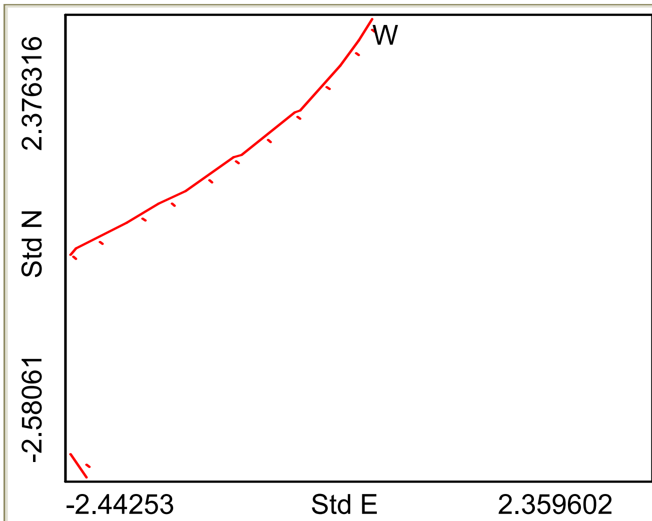
Parameter Estimates

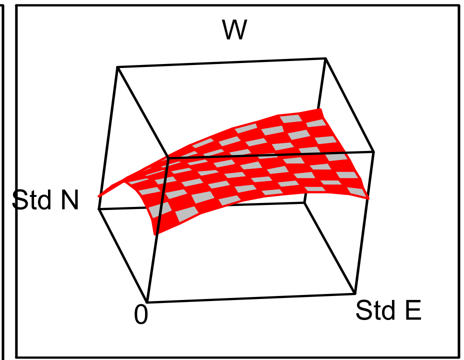
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	17.037417	0.32351	52.66	<.0001
Std E	1.5332287	0.258934	5.92	<.0001
Std E*Std E	-0.210954	0.217855	-0.97	0.3335
Std N	-1.871318	0.25667	-7.29	<.0001
Std N*Std N	-0.071266	0.208559	-0.34	0.7328
Std E*Std N	0.6716616	0.193684	3.47	0.0006
Std E*Std N*Std E	-0.006914	0.157244	-0.04	0.9649
Std E*Std N*Std N	0.2725322	0.153589	1.77	0.0768
Std E*Std N*Std E*Std N	-0.11746	0.128814	-0.91	0.3624

Interaction Profiles



Contour Profiler





Response W
Summary of Fit

RSquare	0.389549
RSquare Adj	0.378184
Root Mean Square Error	3.774947
Mean of Response	16.54427
Observations (or Sum Wgts)	384

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	7	3419.1639	488.452	34.2768
Error	376	5358.0835	14.250	Prob > F
C. Total	383	8777.2474		<.0001

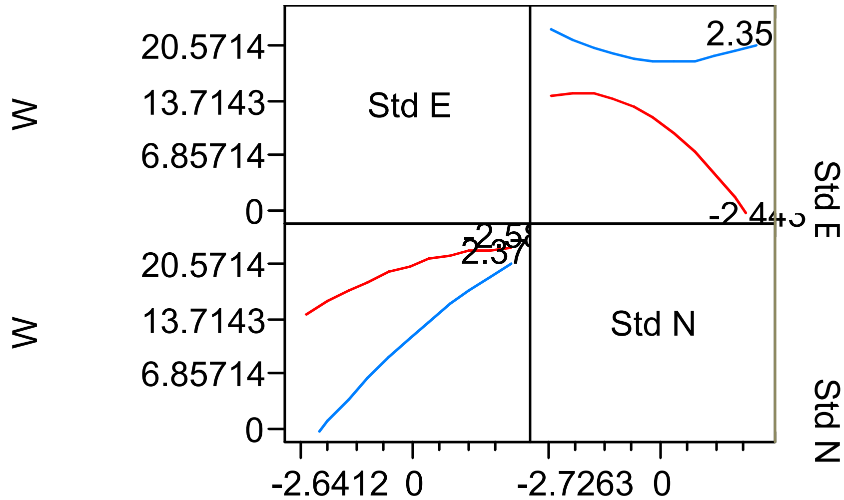
Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	204	2693.1549	13.2017	0.8521
Pure Error	172	2664.9286	15.4938	Prob > F
Total Error	376	5358.0835		0.8639
				Max RSq
				0.6964

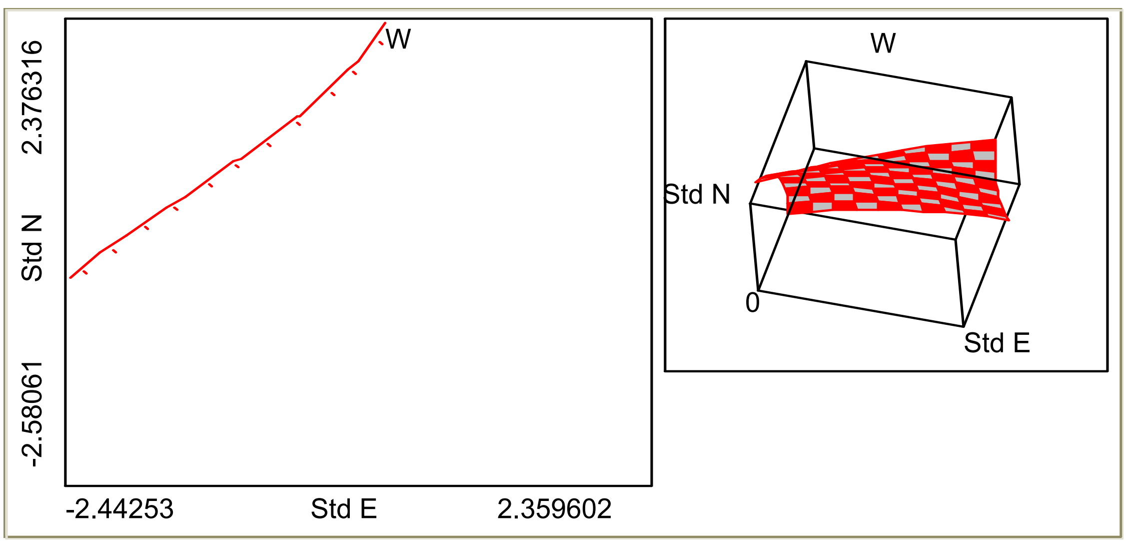
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	17.164222	0.292031	58.78	<.0001
Std E	1.5236879	0.258665	5.89	<.0001
Std E*Std E	-0.341289	0.164374	-2.08	0.0385
Std N	-1.863745	0.256478	-7.27	<.0001
Std N*Std N	-0.195864	0.157527	-1.24	0.2145
Std E*Std N	0.6699262	0.193631	3.46	0.0006
Std E*Std N*Std E	0.0018896	0.156912	0.01	0.9904
Std E*Std N*Std N	0.2847394	0.15297	1.86	0.0635

Interaction Profiles



Contour Profiler



Response W

Summary of Fit

RSquare	0.373968
RSquare Adj	0.369025
Root Mean Square Error	3.802645
Mean of Response	16.54427
Observations (or Sum Wgts)	384

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	3282.4055	1094.14	75.6658
Error	380	5494.8419	14.46	Prob > F
C. Total	383	8777.2474		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	16.621353	0.195261	85.12	<.0001
Std E	1.9239807	0.195638	9.83	<.0001
Std N	-1.861427	0.195671	-9.51	<.0001
Std E*Std N	0.6664521	0.187531	3.55	0.0004

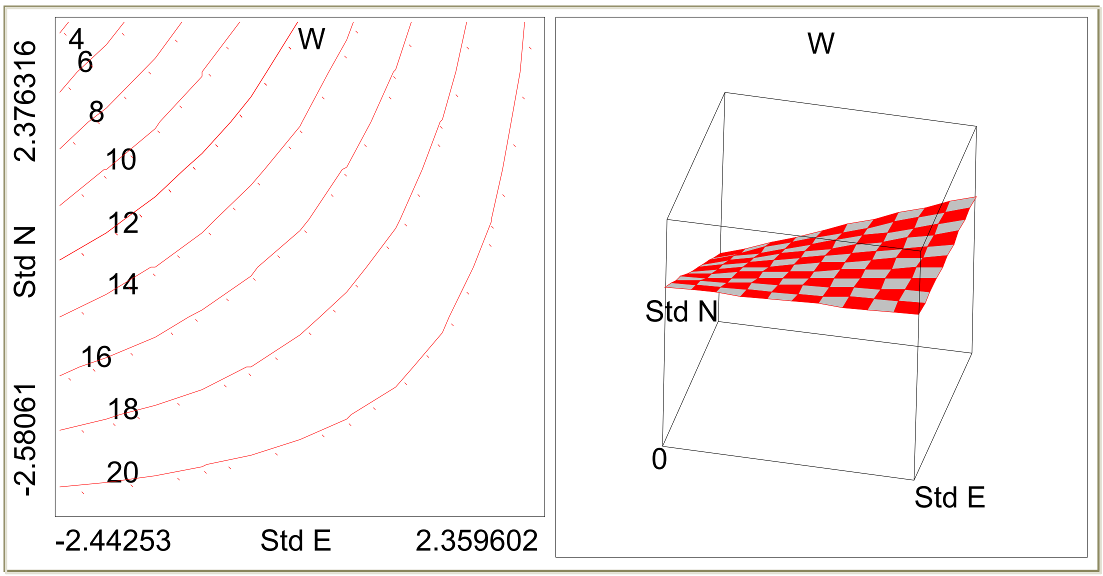
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Std E	1	1	1398.5069	96.7148	<.0001
Std N	1	1	1308.6137	90.4982	<.0001
Std E*Std N	1	1	182.6270	12.6297	0.0004

Contour Profiler

Horiz	Vert	Factor	Current X
◉	○	Std E	3.701e-17
○	◉	Std N	-5.67e-17

Response	Contour	Current Y	Lo Limit	Hi Limit
W	12	16.621353	.	.



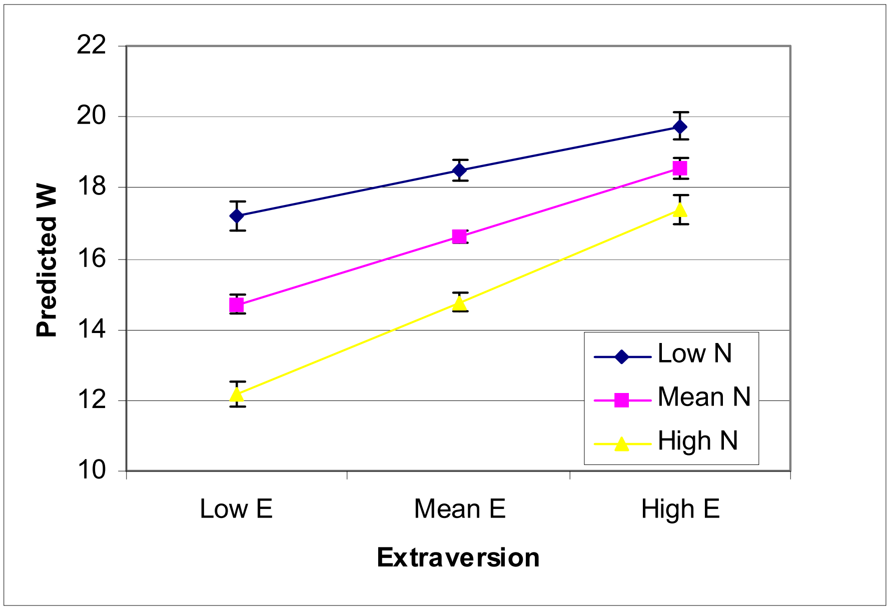
Excel Plot for Mood Model Example - E×N Interactive Model

513 Mood Example Predicting W from E and N

Model with E, N, and E×N

Predicted W	Low E	Mean E	High E
Low N	17.22525	18.48278	19.74031
Mean N	14.69737	16.62135	18.54533
High N	12.16949	14.75993	17.35036

Std Err	Low E	Mean E	High E
Low N	0.407458	0.276102	0.367412
Mean N	0.276582	0.195261	0.276234
High N	0.361576	0.27676	0.40948



INTERACTION SLOPE TESTS WITH MOOD DATA EXAMPLE

Key Idea: *To test the slope for E at different levels of N (e.g., N = 1), use JMP's Custom Test… option, which tests linear combinations of the model parameters. The key thing to understand is that the slope for E at N = 0 (Mean N) is simply the regression weight for E. The slope for E at N = 1 (High N) is the regression weight for E plus the regression weight for the interaction. Similarly, the slope for E at N = -1 (Low N) is the regression weight for E minus the regression weight for the interaction. Thus, by using 1 or -1 at the appropriate places in the Custom Test… dialog the desired test may be obtained.*

Response W
Summary of Fit

RSquare	0.373968
RSquare Adj	0.369025
Root Mean Square Error	3.802645
Mean of Response	16.54427
Observations (or Sum Wgts)	384

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Ratio**
Model	3	3282.4055	1094.14	75.6658
Error	380	5494.8419	14.46	**Prob > F**
C. Total	383	8777.2474		<.0001*

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	208	2829.9133	13.6054	0.8781
Pure Error	172	2664.9286	15.4938	**Prob > F**
Total Error	380	5494.8419		0.8151
				Max RSq
				0.6964

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	16.621353	0.195261	85.12	<.0001*
Std E	1.9239807	0.195638	9.83	<.0001*
Std N	-1.861427	0.195671	-9.51	<.0001*
Std E*Std N	0.6664521	0.187531	3.55	0.0004*

Custom Test

Slope for E at Low N (-1)

Parameter	
Intercept	0
Std E	1
Std N	0
Std E*Std N	-1
=	0

Value	1.2575285725
Std Error	0.2725338848
t Ratio	4.6142099847
Prob>|t|	5.4032577e-6
SS	307.86925112

Sum of Squares	307.86925112
Numerator DF	1
F Ratio	21.290933783
Prob > F	5.4032577e-6

Custom Test

Slope for E at High N (+1)

Parameter	
Intercept	0
Std E	1
Std N	0
Std E*Std N	1
=	0

Value	2.5904328197
Std Error	0.2694616313
t Ratio	9.6133642742
Prob>|t|	9.967099e-20
SS	1336.3567274

Sum of Squares	1336.3567274
Numerator DF	1
F Ratio	92.416772669
Prob > F	9.967099e-20

Interpretation of the Lower-Order Terms in the Output on pp. 50-51

The intercept (16.62) in the regression equation reflects the predicted well-being (W) score for individuals who are at the mean on both E and N because it is the predicted *Y* when all predictors are zero. The regression weight for E (1.92) is the slope of the relation between E and W for individuals who are at the mean on neuroticism (N = 0). That is, for individuals at the mean on neuroticism, a one *SD* increase in E would lead us to predict a 1.92 *point* increase in W (note that the difference in scaling between E and W is reflected in the interpretation). An analogous interpretation applies to the weight for N. That is, the regression weight for N (-1.86) is the slope of the relation between N and W for individuals who are at the mean on extraversion (E = 0). For individuals at the mean on extraversion, a one *SD* increase in N would lead us to predict a 1.86 point *decrease* in W.

Interpretation of the Two-Way Interaction in the Output on pp. 50-51

The 2-way E × N weight in this output is 0.67. The weight, 0.67, reflects the increase in the slope of the relation between E and W that results from a one unit increase in N. Since E and N are both standardized, one unit is one standard deviation. Thus, an increase of one standard deviation in N would lead to an increase of 0.67 in the *slope* of the relation between E and W. The slope of the relation between E and W at the mean on N (N = 0) is 1.92 (from the output). Thus, for individuals at the mean on N a one standard deviation increase in E would lead us to predict a 1.92 point increase in positive intensity. However, for neurotic individuals at one standard deviation above the mean on N (N = 1), the slope of the relation between E and W would be 1.92 + (1)*0.67 = 2.59 (see the highlighted value in the second Custom Test in the output above). Thus, for neurotic individuals, a one standard deviation increase in E would lead us to predict a 2.59 point increase in well-being. Thus, while the relation between extraversion and well-being is positive for both those at the mean on neuroticism (b = 1.92) and those at one standard deviation above the mean on neuroticism (b = 2.59), the relation between E and W is significantly stronger for neurotics than for those at the mean on N. The significant $t(380) = 3.55$ tests the significance of the *difference* (0.67) between the two slopes.

Note also that for *stable* individuals, at one standard deviation *below* the mean on N (N = -1), the slope of the relation between E and W is still significantly positive: 1.92 + (-1)*.67 = 1.26 (see the highlighted value in the first Custom Test in the output above), though significantly weaker than that for individuals at the mean on N (b = 1.92). Note, as well, that the 3 slopes discussed above (1.92, 2.59, and 1.26) are *not* predicted W scores. They are predicted *changes* in W scores for a one *SD* increase in E.

Response W

Summary of Fit

RSquare	0.383923
RSquare Adj	0.375773
Root Mean Square Error	3.782256
Mean of Response	16.54427
Observations (or Sum Wgts)	384

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	5	3369.7836	673.957	47.1119
Error	378	5407.4638	14.305	Prob > F
C. Total	383	8777.2474		<.0001

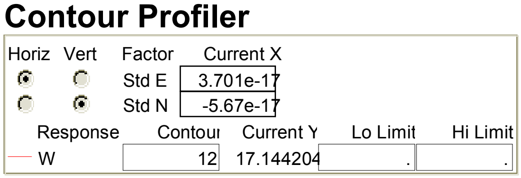
Parameter Estimates

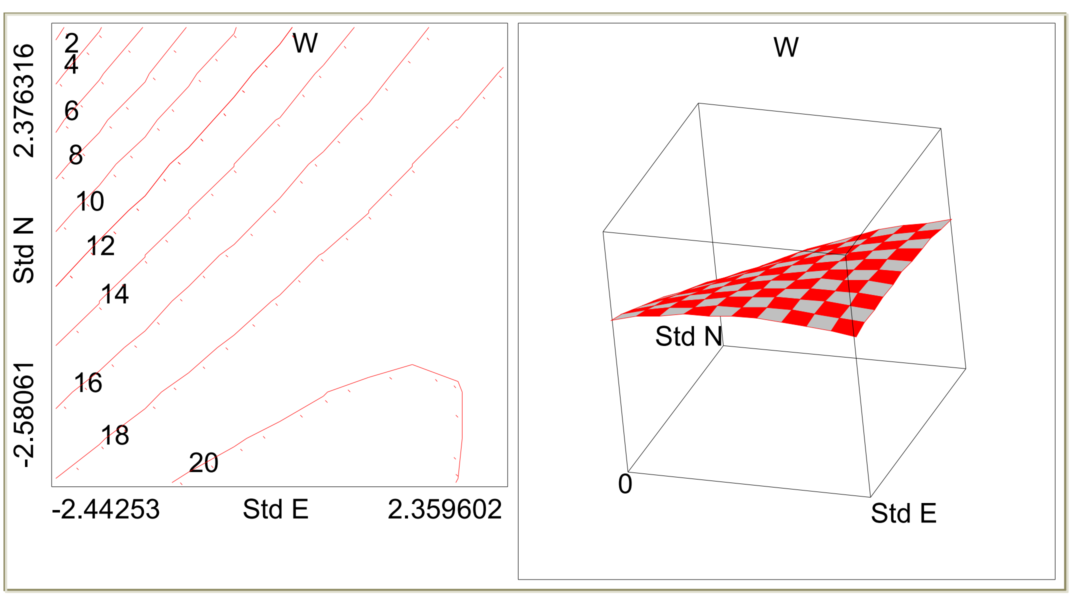
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	17.144204	0.2923	58.65	<.0001
Std E	1.830522	0.198981	9.20	<.0001
Std E*Std E	-0.349243	0.164519	-2.12	0.0344
Std N	-1.876914	0.194808	-9.63	<.0001
Std N*Std N	-0.180504	0.156771	-1.15	0.2503
Std E*Std N	0.6187539	0.187523	3.30	0.0011

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Std E	1	1	1210.6817	84.6307	<.0001
Std E*Std E	1	1	64.4651	4.5063	0.0344
Std N	1	1	1327.9371	92.8273	<.0001
Std N*Std N	1	1	18.9646	1.3257	0.2503
Std E*Std N	1	1	155.7497	10.8874	0.0011

Contour Profiler





Response W
Summary of Fit

RSquare	0.381762
RSquare Adj	0.375237
Root Mean Square Error	3.783881
Mean of Response	16.54427
Observations (or Sum Wgts)	384

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	3350.8191	837.705	58.5081
Error	379	5426.4283	14.318	Prob > F
C. Total	383	8777.2474		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	16.975262	0.252912	67.12	<.0001
Std E	1.833212	0.199052	9.21	<.0001
Std E*Std E	-0.359275	0.164359	-2.19	0.0294
Std N	-1.867677	0.194726	-9.59	<.0001
Std E*Std N	0.6281495	0.187426	3.35	0.0009

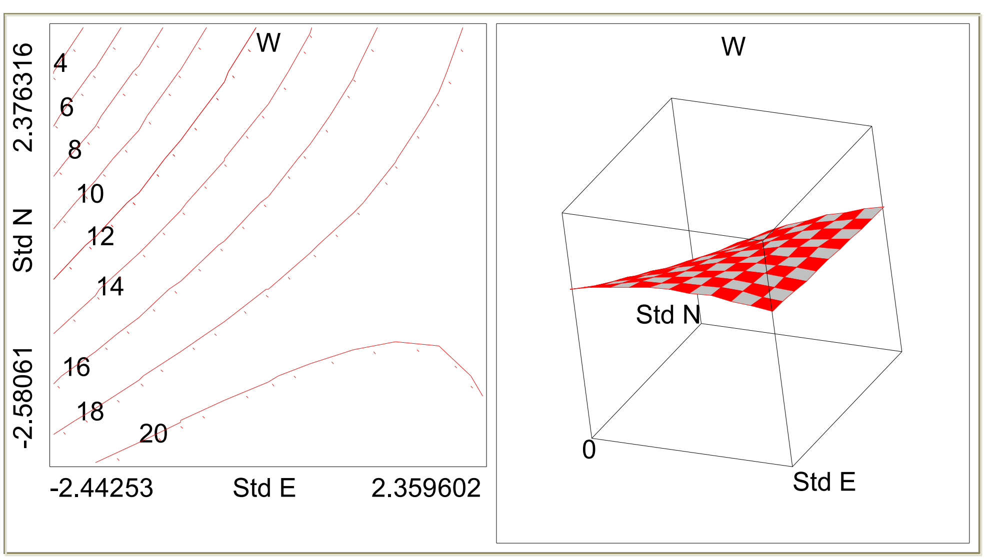
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Std E	1	1	1214.4101	84.8185	<.0001
Std E*Std E	1	1	68.4136	4.7782	0.0294
Std N	1	1	1317.1326	91.9930	<.0001
Std E*Std N	1	1	160.8202	11.2322	0.0009

Contour Profiler

Horiz	Vert	Factor	Current X
⦿	○	Std E	3.701e-17
○	⦿	Std N	-5.67e-17

Response	Contour	Current Y	Lo Limit	Hi Limit
W	12	16.975262	.	.



Qualitative, Indicator or Dummy Variable Models and ANOVA

The general linear regression model which is given generally by the equation

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \varepsilon_i. \quad [57]$$

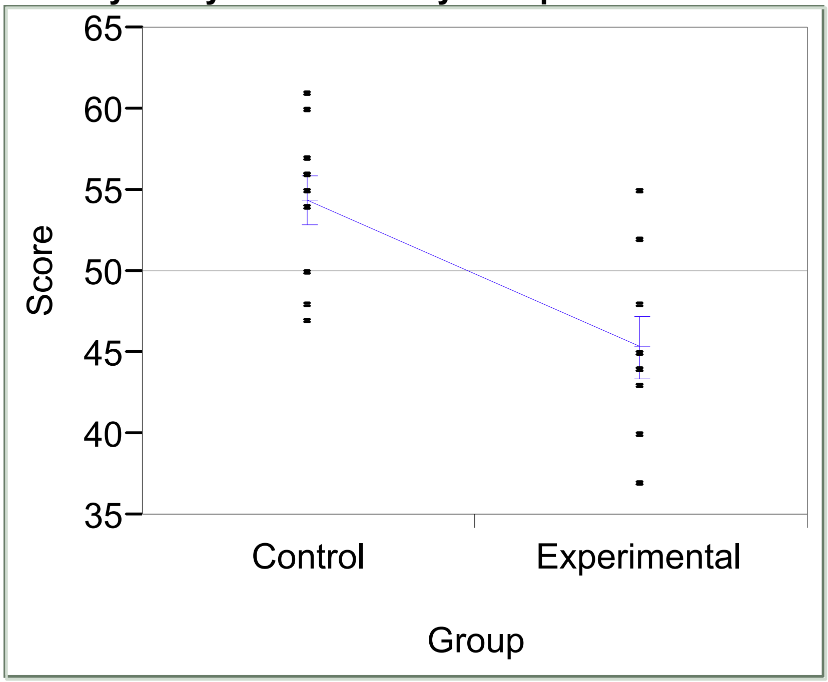
may be used to test hypotheses about the means and differences between means of different groups identified by way of variables called qualitative, indicator or dummy variables. These variables code group membership in a numerical code so that the variable may be included in a regression analysis. The most common coding is to use 0 and 1 to encode group membership. Variables that use 0 and 1 in this way are called dummy variables.

In the simplest case, two groups of subjects (e.g., males and females) are coded with one group receiving a 1, the other a 0. If this group dummy variable is correlated with another (criterion) variable, the correlation is often called a point biserial correlation, and a test of H_0: $\rho = 0$ is equivalent to a test of whether the means of the two groups on the criterion variable are equal.

Point-biserial Correlation Example

Subject	Score	Group Dummy	Group
1	45	1	Experimental
2	52	1	Experimental
3	40	1	Experimental
4	55	1	Experimental
5	43	1	Experimental
6	44	1	Experimental
7	48	1	Experimental
8	37	1	Experimental
9	44	1	Experimental
10	56	0	Control
11	54	0	Control
12	61	0	Control
13	48	0	Control
14	57	0	Control
15	50	0	Control
16	56	0	Control
17	55	0	Control
18	47	0	Control
19	60	0	Control

Oneway Analysis of Score By Group



Oneway Anova

Summary of Fit

Rsquare	0.461477
Adj Rsquare	0.429799
Root Mean Square Error	5.170049
Mean of Response	50.10526
Observations (or Sum Wgts)	19

t-Test

	Difference	t-Test	DF	Prob > |t|
Estimate	9.0667	3.817	17	0.0014
Std Error	2.3755			
Lower 95%	4.0549			
Upper 95%	14.0785			

Assuming equal variances

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Group	1	389.38947	389.389	14.5678	0.0014
Error	17	454.40000	26.729		
C. Total	18	843.78947			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
Control	10	54.4000	1.6349	50.951	57.849
Experimental	9	45.3333	1.7233	41.697	48.969

Std Error uses a pooled estimate of error variance

Means and Std Deviations

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
Control	10	54.4000	4.74225	1.4996	51.008	57.792
Experimental	9	45.3333	5.61249	1.8708	41.019	49.647

Multivariate

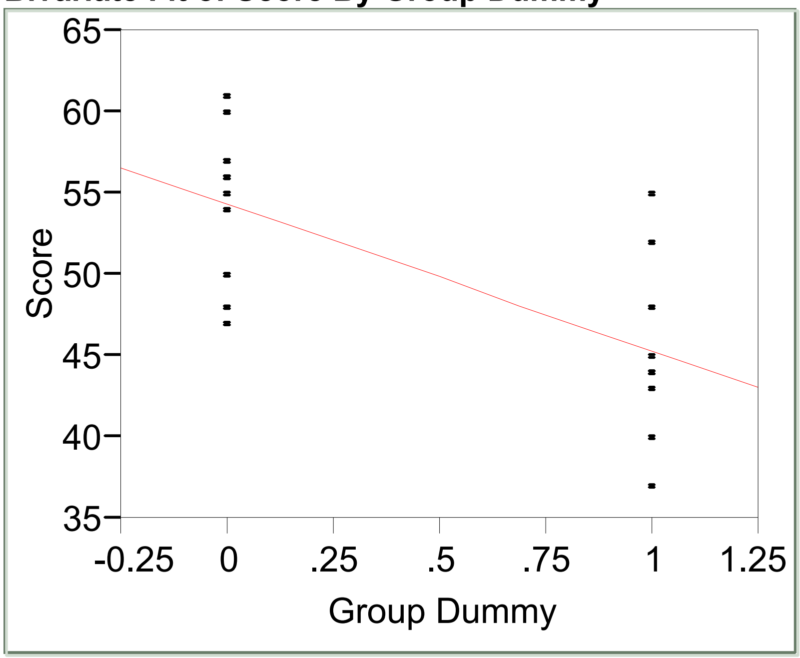
Correlations

	Score	Group Dummy
Score	1.0000	-0.6793
Group Dummy	-0.6793	1.0000

Pairwise Correlations

Variable	by Variable	Correlation	Count	Signif Prob	
Group Dummy	Score	-0.6793	19	0.0014	---------------------------

Bivariate Fit of Score By Group Dummy



Linear Fit

Score = 54.4 - 9.0666667 Group Dummy

Summary of Fit

RSquare	0.461477
RSquare Adj	0.429799
Root Mean Square Error	5.170049
Mean of Response	50.10526
Observations (or Sum Wgts)	19

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	389.38947	389.389	14.5678
Error	17	454.40000	26.729	Prob > F
C. Total	18	843.78947		0.0014

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	54.4	1.634913	33.27	<.0001
Group Dummy	-9.066667	2.375474	-3.82	0.0014

Example for different coding methods

Assume we have scores on 3 subjects in each of 4 groups:

a_1	a_2	a_3	a_4
2	9	5	9
4	6	5	8
3	3	2	4

We may code group membership in several ways as shown in the table below. The first set of three predictors is 'dummy coding,' the second is 'effects coding,' the third 'Helmert coding,' the fourth 'effects coding' for a two way factorial ANOVA, and the last columns just use index numbers to indicate group membership.

Perf	D1	D2	D3	E1	E2	E3	H1	H2	H3	A	B	A×B	Betw	FacA	FacB
2	1	0	0	1	0	0	3	0	0	1	1	1	1	1	1
4	1	0	0	1	0	0	3	0	0	1	1	1	1	1	1
3	1	0	0	1	0	0	3	0	0	1	1	1	1	1	1
9	0	1	0	0	1	0	-1	2	0	1	-1	-1	2	1	2
6	0	1	0	0	1	0	-1	2	0	1	-1	-1	2	1	2
3	0	1	0	0	1	0	-1	2	0	1	-1	-1	2	1	2
5	0	0	1	0	0	1	-1	-1	1	-1	1	-1	3	2	1
5	0	0	1	0	0	1	-1	-1	1	-1	1	-1	3	2	1
2	0	0	1	0	0	1	-1	-1	1	-1	1	-1	3	2	1
9	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	1	4	2	2
8	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	1	4	2	2
4	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	1	4	2	2

```
1                      Example for various coding methods                 1
                     General Linear Models Procedure
                 Number of observations in data set = 12
```

Dummy Coding

```
                                                                  2
                     General Linear Models Procedure

 Dependent Variable: PERF
                                    Sum of          Mean
 Source                  DF        Squares        Square  F Value     Pr > F

 Model                    3      30.000000     10.000000     2.00     0.1927
 Error                    8      40.000000      5.000000
 Corrected Total         11      70.000000

                  R-Square           C.V.      Root MSE            PERF Mean
                  0.428571       44.72136        2.2361               5.0000

 Source                  DF      Type I SS   Mean Square  F Value     Pr > F
 D1                       1      16.000000     16.000000     3.20     0.1114
 D2                       1       0.500000      0.500000     0.10     0.7599
 D3                       1      13.500000     13.500000     2.70     0.1390

 Source                  DF    Type III SS   Mean Square  F Value     Pr > F
 D1                       1      24.000000     24.000000     4.80     0.0598
 D2                       1       1.500000      1.500000     0.30     0.5988
 D3                       1      13.500000     13.500000     2.70     0.1390

                                    T for H0:     Pr > |T|    Std Error of
 Parameter            Estimate    Parameter=0                   Estimate
 INTERCEPT         7.000000000           5.42       0.0006      1.29099445
 D1               -4.000000000          -2.19       0.0598      1.82574186
 D2               -1.000000000          -0.55       0.5988      1.82574186
 D3               -3.000000000          -1.64       0.1390      1.82574186
```

Effects Coding

```
                                                                3
                     General Linear Models Procedure
                 Number of observations in data set = 12
                          This is effects coding                          4
                     General Linear Models Procedure

 Dependent Variable: PERF
                                    Sum of          Mean
 Source                  DF        Squares        Square  F Value     Pr > F
 Model                    3      30.000000     10.000000     2.00     0.1927
 Error                    8      40.000000      5.000000
 Corrected Total         11      70.000000

                  R-Square           C.V.      Root MSE            PERF Mean
                  0.428571       44.72136        2.2361               5.0000

 Source                  DF      Type I SS   Mean Square  F Value     Pr > F
 E1                       1      24.000000     24.000000     4.80     0.0598
 E2                       1       2.000000      2.000000     0.40     0.5447
 E3                       1       4.000000      4.000000     0.80     0.3972

 Source                  DF    Type III SS   Mean Square  F Value     Pr > F
 E1                       1      16.000000     16.000000     3.20     0.1114
 E2                       1       4.000000      4.000000     0.80     0.3972
 E3                       1       4.000000      4.000000     0.80     0.3972

                                    T for H0:     Pr > |T|    Std Error of
 Parameter            Estimate    Parameter=0                   Estimate
 INTERCEPT         5.000000000           7.75       0.0001      0.64549722
 E1               -2.000000000          -1.79       0.1114      1.11803399
 E2                1.000000000           0.89       0.3972      1.11803399
 E3               -1.000000000          -0.89       0.3972      1.11803399
```

Helmert Contrast Coding

5

```
            General Linear Models Procedure
        Number of observations in data set = 12
            This is helmert contrast coding                  6
            General Linear Models Procedure

Dependent Variable: PERF
                                   Sum of          Mean
Source                   DF       Squares        Square  F Value     Pr > F
Model                     3     30.000000     10.000000     2.00     0.1927
Error                     8     40.000000      5.000000

Corrected Total          11     70.000000

                 R-Square          C.V.      Root MSE           PERF Mean
                 0.428571      44.72136        2.2361              5.0000

Source                   DF     Type I SS   Mean Square  F Value     Pr > F
H1                        1     16.000000     16.000000     3.20     0.1114
H2                        1      0.500000      0.500000     0.10     0.7599
H3                        1     13.500000     13.500000     2.70     0.1390

Source                   DF   Type III SS   Mean Square  F Value     Pr > F
H1                        1     16.000000     16.000000     3.20     0.1114
H2                        1      0.500000      0.500000     0.10     0.7599
H3                        1     13.500000     13.500000     2.70     0.1390

                                   T for H0:     Pr > |T|    Std Error of
Parameter             Estimate    Parameter=0                  Estimate
INTERCEPT          5.000000000           7.75      0.0001      0.64549722
H1                -0.666666667          -1.79      0.1114      0.37267800
H2                 0.166666667           0.32      0.7599      0.52704628
H3                -1.500000000          -1.64      0.1390      0.91287093
```

2×2 Factorial Design Effects Coding

7

```
            General Linear Models Procedure
        Number of observations in data set = 12
         This is a 2×2 factorial design coding               8
            General Linear Models Procedure

Dependent Variable: PERF
                                   Sum of          Mean
Source                   DF       Squares        Square  F Value     Pr > F
Model                     3     30.000000     10.000000     2.00     0.1927
Error                     8     40.000000      5.000000
Corrected Total          11     70.000000

                 R-Square          C.V.      Root MSE           PERF Mean
                 0.428571      44.72136        2.2361              5.0000

Source                   DF     Type I SS   Mean Square  F Value     Pr > F
A                         1      3.000000      3.000000     0.60     0.4609
B                         1     27.000000     27.000000     5.40     0.0486
AXB                       1      0.000000      0.000000     0.00     1.0000

Source                   DF   Type III SS   Mean Square  F Value     Pr > F
A                         1      3.000000      3.000000     0.60     0.4609
B                         1     27.000000     27.000000     5.40     0.0486
AXB                       1      0.000000      0.000000     0.00     1.0000

                                   T for H0:     Pr > |T|    Std Error of
Parameter             Estimate    Parameter=0                  Estimate
INTERCEPT          5.000000000           7.75      0.0001      0.64549722
A                 -0.500000000          -0.77      0.4609      0.64549722
B                 -1.500000000          -2.32      0.0486      0.64549722
AXB                0.000000000           0.00      1.0000      0.64549722
```

Usual One-Way ANOVA

The JMP output for the usual one-way ANOVA (using the nominal factor Between as the between subjects factor) is

Response Perf

Summary of Fit

RSquare	0.428571
RSquare Adj	0.214286
Root Mean Square Error	2.236068
Mean of Response	5
Observations (or Sum Wgts)	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	30.000000	10.0000	2.0000
Error	8	40.000000	5.0000	Prob > F
C. Total	11	70.000000		0.1927

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	5	0.645497	7.75	<.0001
Between[1]	-2	1.118034	-1.79	0.1114
Between[2]	1	1.118034	0.89	0.3972
Between[3]	-1	1.118034	-0.89	0.3972

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Between	3	3	30.000000	2.0000	0.1927

Effect Details

Between

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	3.0000000	1.2909944	3.00000
2	6.0000000	1.2909944	6.00000
3	4.0000000	1.2909944	4.00000
4	7.0000000	1.2909944	7.00000

Note that the parameter estimates are the same as for the effects coded regression analysis. Thus, JMP by default uses effects coding in carrying out ANOVAs.

Usual Two-Way ANOVA

The JMP output for the usual two-way ANOVA (using the nominal factors FacA and FacB and their interaction as the between subjects factors) is

Response Perf

Summary of Fit

RSquare	0.428571
RSquare Adj	0.214286
Root Mean Square Error	2.236068
Mean of Response	5
Observations (or Sum Wgts)	12

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	30.000000	10.0000	2.0000
Error	8	40.000000	5.0000	Prob > F
C. Total	11	70.000000		0.1927

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	5	0.645497	7.75	<.0001
FacA[1]	-0.5	0.645497	-0.77	0.4609
FacB[1]	-1.5	0.645497	-2.32	0.0486
FacA[1]*FacB[1]	0	0.645497	0.00	1.0000

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
FacA	1	1	3.000000	0.6000	0.4609
FacB	1	1	27.000000	5.4000	0.0486
FacA*FacB	1	1	0.000000	0.0000	1.0000

Effect Details

FacA

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	4.5000000	0.91287093	4.50000
2	5.5000000	0.91287093	5.50000

FacB

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	3.5000000	0.91287093	3.50000
2	6.5000000	0.91287093	6.50000

FacA*FacB

Least Squares Means Table

Level	Least Sq Mean	Std Error
1,1	3.0000000	1.2909944
1,2	6.0000000	1.2909944
2,1	4.0000000	1.2909944
2,2	7.0000000	1.2909944

Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
FacA	1	1	3.000000	0.6000	0.4609
FacB	1	1	27.000000	5.4000	0.0486
FacA*FacB	1	1	0.000000	0.0000	1.0000

Meaning of Regression Coefficients for Coded Predictors

Although the omnibus ANOVA will be identical for all the coding methods, the values of the estimated regression coefficients and their associated hypothesis tests will vary depending on the coding scheme used. For the three coding schemes used in the example above (dummy coding, effects coding, orthogonal contrast coding, and factorial effects coding) the values of the estimated regression weights and their associated hypothesis tests are as follows:

Dummy Coding

$b_0 = \overline{Y_4}$ (the mean of the group always coded zero) H$_0$: $\mu_4 = 0$

$b_1 = \overline{Y_1} - \overline{Y_4}$ H$_0$: $\mu_1 - \mu_4 = 0$

$b_2 = \overline{Y_2} - \overline{Y_4}$ H$_0$: $\mu_2 - \mu_4 = 0$

$b_3 = \overline{Y_3} - \overline{Y_4}$ H$_0$: $\mu_3 - \mu_4 = 0$

Effects Coding

$b_0 = \overline{Y} = \hat{\mu}$ H$_0$: $\mu = 0$

$b_1 = \overline{Y_1} - \overline{Y} = \hat{\alpha}_1$ H$_0$: $\alpha_1 = \mu_1 - \mu = 0$

$b_2 = \overline{Y_2} - \overline{Y} = \hat{\alpha}_2$ H$_0$: $\alpha_2 = \mu_2 - \mu = 0$

$b_3 = \overline{Y_3} - \overline{Y} = \hat{\alpha}_3$ H$_0$: $\alpha_3 = \mu_3 - \mu = 0$

Orthogonal Contrast Coding

In general, for orthogonal contrast coding the regression coefficients will be equal to $\frac{\hat{\psi}}{\sum c_i^2}$, where $\hat{\psi} = \sum c_i \overline{Y_i}$ is the contrast among the group means. Thus, we have

$b_0 = \overline{Y} = \hat{\mu}$ H$_0$: $\mu = 0$

$b_1 = \frac{\hat{\psi}_1}{\sum c_{1i}^2}$ H$_0$: $\psi_1 = \sum c_{1i}\mu_i = 0$,

and similarly for b_2 and b_3.

Factorial Effects Coding

$b_0 = \overline{Y} = \hat{\mu}$ H$_0$: $\mu = 0$

$b_1 = \overline{A_1} - \overline{Y} = \hat{\alpha}_1$ H$_0$: $\alpha_1 = \mu_{A1} - \mu = 0$

$b_2 = \overline{B_1} - \overline{Y} = \hat{\beta}_1$ H$_0$: $\beta_1 = \mu_{B1} - \mu = 0$

$b_3 = \overline{AB_{11}} - \overline{A_1} - \overline{B_1} + \overline{Y} = \left(\hat{\alpha}\hat{\beta}\right)_{11}$ H$_0$: $(\alpha\beta)_{11} = 0$

Correlations Among Coded Predictor Variables

It is instructive to consider the correlations among the coded predictor variables. For the example above, summary statistics for the variables are as follows:

Variable	N	Mean	Std Dev	Min	Max
Perf	12	5	2.522625	2	9
D1	12	0.25	0.452267	0	1
D2	12	0.25	0.452267	0	1
D3	12	0.25	0.452267	0	1
E1	12	0	0.738549	-1	1
E2	12	0	0.738549	-1	1
E3	12	0	0.738549	-1	1
H1	12	0	1.809068	-1	3
H2	12	0	1.279204	-1	2
H3	12	0	0.738549	-1	1
A	12	0	1.044466	-1	1
B	12	0	1.044466	-1	1
A×B	12	0	1.044466	-1	1

Note that all the coding schemes except dummy coding are deviation scores ($M = 0$); thus the intercept in all these schemes will be the grand mean. The correlations among the variables are as follows:

	Perf	D1	D2	D3	E1	E2	E3	H1	H2	H3	A	B	A×B
Perf	1	-0.48	0.239	-0.24	-0.59	-0.15	-0.44	-0.48	0.085	-0.44	-0.21	-0.62	0
D1	-0.48	**1**	**-0.33**	**-0.33**	0.817	0	0	1	0	0	0.577	0.577	0.577
D2	0.239	**-0.33**	**1**	**-0.33**	0	0.817	0	-0.33	0.943	0	0.577	-0.58	-0.58
D3	-0.24	**-0.33**	**-0.33**	**1**	0	0	0.817	-0.33	-0.47	0.817	-0.58	0.577	-0.58
E1	-0.59	0.817	0	0	**1**	**0.5**	**0.5**	0.817	0.289	0.5	0.707	0.707	0
E2	-0.15	0	0.817	0	**0.5**	**1**	**0.5**	0	0.866	0.5	0.707	0	-0.71
E3	-0.44	0	0	0.817	**0.5**	**0.5**	**1**	0	0	1	0	0.707	-0.71
H1	-0.48	1	-0.33	-0.33	0.817	0	0	**1**	**0**	**0**	0.577	0.577	0.577
H2	0.085	0	0.943	-0.47	0.289	0.866	0	**0**	**1**	**0**	0.817	-0.41	-0.41
H3	-0.44	0	0	0.817	0.5	0.5	1	**0**	**0**	**1**	0	0.707	-0.71
A	-0.21	0.577	0.577	-0.58	0.707	0.707	0	0.577	0.817	0	**1**	**0**	**0**
B	-0.62	0.577	-0.58	0.577	0.707	0	0.707	0.577	-0.41	0.707	**0**	**1**	**0**
A×B	0	0.577	-0.58	-0.58	0	-0.71	-0.71	0.577	-0.41	-0.71	**0**	**0**	**1**

Note that only the orthogonal Helmert contrasts and the orthogonal factorial effects coding have uncorrelated predictors. Thus, it is only in these two cases that the Type 1 sequential *SS* and the Type 2 (or 3) *SS* will be the same.

Nonorthogonal (Unbalanced) Analysis of Variance

When cell sizes are equal in a factorial ANOVA the design is said to be *balanced*. When this is the case, as in the example above, the A, B, and A×B effects are independent (i.e., orthogonal or uncorrelated) as the above correlation matrix shows. If cell sizes are unequal, however, this orthogonality of the predictors doesn't hold. The design is then said to be unbalanced or nonorthogonal. This nonorthogonality has implications for hypothesis testing because the Type 1 sequential *SS* now depend on the order in which factors are entered into the model. Furthermore, there will be two different types of marginal means, *weighted* and *unweighted* (or in SAS and JMP terminology, *means* and *least squares means* or *LS means*).

By removing one of the fives from the a_2b_1 group in the data set above the design may be made nonorthogonal. The correlations between A, B, and A×B are no longer zero but now

	A	B	A×B
A	1.0000	0.1000	-0.1000
B	0.1000	1.0000	0.1000
A×B	-0.1000	0.1000	1.0000

Below are several JMP analyses of the new data with different orders of the A, B, and A×B effects. Note the differences in the Type 1 sequential *SS* and the distinction between means and LS means in the first analysis (the means in the subsequent analyses would be identical to the ones in the first analysis, so I have left them out).

Order of Effects: A, B, A×B

Response Perf
Summary of Fit

RSquare	0.45
RSquare Adj	0.214286
Root Mean Square Error	2.345208
Mean of Response	5
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	31.500000	10.5000	1.9091
Error	7	38.500000	5.5000	Prob > F
C. Total	10	70.000000		0.2166

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.875	0.71807	6.79	0.0003
FacA[1]	-0.375	0.71807	-0.52	0.6176
FacB[1]	-1.625	0.71807	-2.26	0.0581
FacA[1]*FacB[1]	0.125	0.71807	0.17	0.8667

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
FacA	1	1	1.500000	0.2727	0.6176
FacB	1	1	28.166667	5.1212	0.0581
FacA*FacB	1	1	0.166667	0.0303	0.8667

Effect Details
FacA
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	4.5000000	0.9574271	4.50000
2	5.2500000	1.0704360	5.60000

FacB
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	3.2500000	1.0704360	3.20000
2	6.5000000	0.9574271	6.50000

FacA*FacB

Least Squares Means Table

Level	Least Sq Mean	Std Error
1,1	3.0000000	1.3540064
1,2	6.0000000	1.3540064
2,1	3.5000000	1.6583124
2,2	7.0000000	1.3540064

Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
FacA	1	1	3.300000	0.6000	0.4639
FacB	1	1	28.033333	5.0970	0.0585
FacA*FacB	1	1	0.166667	0.0303	0.8667

Order of Effects: B, A, A×B

Response Perf

Summary of Fit

RSquare	0.45
RSquare Adj	0.214286
Root Mean Square Error	2.345208
Mean of Response	5
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	31.500000	10.5000	1.9091
Error	7	38.500000	5.5000	Prob > F
C. Total	10	70.000000		0.2166

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.875	0.71807	6.79	0.0003
FacB[1]	-1.625	0.71807	-2.26	0.0581
FacA[1]	-0.375	0.71807	-0.52	0.6176
FacA[1]*FacB[1]	0.125	0.71807	0.17	0.8667

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
FacB	1	1	28.166667	5.1212	0.0581
FacA	1	1	1.500000	0.2727	0.6176
FacA*FacB	1	1	0.166667	0.0303	0.8667

Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
FacB	1	1	29.700000	5.4000	0.0531
FacA	1	1	1.633333	0.2970	0.6027
FacA*FacB	1	1	0.166667	0.0303	0.8667

Order of Effects: A×B, A, B

Response Perf

Summary of Fit

RSquare	0.45
RSquare Adj	0.214286
Root Mean Square Error	2.345208
Mean of Response	5
Observations (or Sum Wgts)	11

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	31.500000	10.5000	1.9091
Error	7	38.500000	5.5000	Prob > F
C. Total	10	70.000000		0.2166

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	4.875	0.71807	6.79	0.0003
FacA[1]*FacB[1]	0.125	0.71807	0.17	0.8667
FacA[1]	-0.375	0.71807	-0.52	0.6176
FacB[1]	-1.625	0.71807	-2.26	0.0581

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
FacA*FacB	1	1	0.166667	0.0303	0.8667
FacA	1	1	1.500000	0.2727	0.6176
FacB	1	1	28.166667	5.1212	0.0581

Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
FacA*FacB	1	1	0.000000	0.0000	1.0000
FacA	1	1	3.333333	0.6061	0.4618
FacB	1	1	28.166667	5.1212	0.0581

Note that the Type 1 sequential *SS F*-tests test different hypotheses depending on the order of entry of factors into the model. These hypothesis tests depend on the number of observations in the cells. The usual Type 2 *SS F*-tests test hypotheses about the least squares (unweighted) marginal means, which are obtained by averaging the *cell means* rather than the individual scores in each row or column. These are the hypothesis tests most commonly recommended (for good reason) in the nonorthogonal ANOVA, but it is important to bear in mind always that they are testing the differences between the least squares means rather than the weighted means.

Analysis of Covariance

When indicator variables (such as dummy variables) are combined in an additive (no interaction) regression model with continuous (or many-valued) predictor variables, the model is said to be an analysis of covariance (ANCOVA) model. The continuous predictor variable is called the covariate and is routinely centered (converted to deviation scores) first. The model for a simple one-way ANCOVA with a single covariate is

$$Y_{ij} = \mu + \alpha_i + \beta x_j + \varepsilon_{ij} \qquad [58]$$

where α_i is the treatment effect for the *i*th group, x_j is the score for the *j*th subject on the *centered* covariate (i.e., deviation scores: $x = X - M_X$), and β is the slope of the common or pooled regression line relating the covariate to Y in the several groups.

The values of the α_is could be estimated by using effects coded predictor variables along with the covariate in a multiple regression analysis. Note that there is only one slope in the model, so the model assumes that the slope of the regression line relating Y to X is the same in each group. This assumption of parallel regression lines is called the *homogeneity of regression slope* assumption.

Just as with any ANOVA model, different coding schemes for group membership are possible. For example, with two groups one could use a simple dummy coded variable, D, to indicate group membership and estimate the simple ANCOVA model by obtaining the regression equation

$$\hat{Y} = b_0 + b_1 D + b_2 x \ . \qquad [59]$$

Note that here b_0 would be the *adjusted* mean of Y for the group coded 0 for subjects who are at the mean (of zero) on X. The coefficient b_1 would be the difference between the adjusted means of the two groups for individuals at the mean on X. The coefficient b_2 would be the common slope of the regression line for the two groups. Below is an example analysis using JMP.

Example.

I have included two possible covariates (Cov and Cov2), along with their deviation scores in this data set to illustrate some issues. The criterion variable is DV.

DV	Trtmt	Cov	TrtEff	Cov2	D	CovDev	Cov2Dev
13	1	13	1	43	1	-7.0625	-14.5625
22	1	23	1	53	1	2.9375	-4.5625
23	1	25	1	55	1	4.9375	-2.5625
20	1	22	1	52	1	1.9375	-5.5625
24	1	26	1	56	1	5.9375	-1.5625
17	1	19	1	49	1	-1.0625	-8.5625
5	1	6	1	36	1	-14.0625	-21.5625
16	1	18	1	48	1	-2.0625	-9.5625
11	2	11	-1	56	0	-9.0625	-1.5625
23	2	22	-1	67	0	1.9375	9.4375
24	2	22	-1	67	0	1.9375	9.4375
16	2	14	-1	59	0	-6.0625	1.4375
25	2	25	-1	70	0	4.9375	12.4375
26	2	27	-1	72	0	6.9375	14.4375
28	2	27	-1	72	0	6.9375	14.4375
20	2	21	-1	66	0	0.9375	8.4375

The regression analysis predicting DV from CovDev2 and D using the Fit Model platform is

Response DV
Summary of Fit

RSquare	0.97885
RSquare Adj	0.975596
Root Mean Square Error	0.964621
Mean of Response	19.5625
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	559.84108	279.921	300.8301
Error	13	12.09642	0.930	Prob > F
C. Total	15	571.93750		<.0001

Parameter Estimates

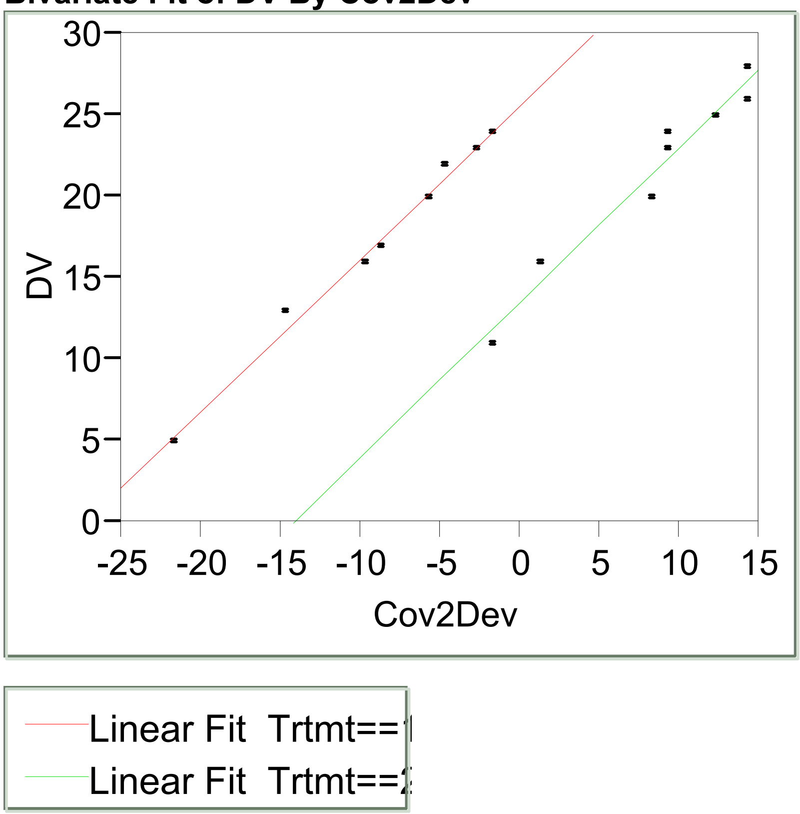
Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	13.564021	0.489141	27.73	<.0001
Cov2Dev	0.9414282	0.04095	22.99	<.0001
D	11.996959	0.851125	14.10	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Cov2Dev	1	1	491.77858	528.5136	<.0001
D	1	1	184.87128	198.6809	<.0001

Separate analyses for the two groups using the Fit Y by X platform are as follows:

Bivariate Fit of DV By Cov2Dev



Linear Fit Trtmt==1

DV = 25.493473 + 0.9335443 Cov2Dev

Summary of Fit

RSquare	0.990632
RSquare Adj	0.98907
Root Mean Square Error	0.658841
Mean of Response	17.5
Observations (or Sum Wgts)	8

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	275.39557	275.396	634.4471
Error	6	2.60443	0.434	Prob > F
C. Total	7	278.00000		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	25.493473	0.393662	64.76	<.0001
Cov2Dev	0.9335443	0.037063	25.19	<.0001

Linear Fit Trtmt==2

DV = 13.474719 + 0.9518577 Cov2Dev

Summary of Fit

RSquare	0.958179
RSquare Adj	0.951209
Root Mean Square Error	1.254749
Mean of Response	21.625
Observations (or Sum Wgts)	8

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	216.42864	216.429	137.4679
Error	6	9.44636	1.574	Prob > F
C. Total	7	225.87500		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	13.474719	0.824632	16.34	<.0001
Cov2Dev	0.9518577	0.081184	11.72	<.0001

Note that the slopes of the two regression lines, though similar, are slightly different. The ANCOVA, by contrast, uses a pooled common slope for the two regression lines so that they are forced to be exactly parallel.

A single equation that combines the separate analyses for the two groups and allows separate slopes can be carried out by estimating an equation that includes the *Dx* interaction term:

$$\hat{Y} = b_0 + b_1 D + b_2 x + b_3 Dx \ . \qquad [60]$$

The JMP output for this analysis is

Response DV
Summary of Fit

RSquare	0.97893
RSquare Adj	0.973662
Root Mean Square Error	1.002114
Mean of Response	19.5625
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	559.88671	186.629	185.8423
Error	12	12.05079	1.004	Prob > F
C. Total	15	571.93750		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	13.474719	0.658598	20.46	<.0001
D	12.018754	0.890099	13.50	<.0001
Cov2Dev	0.9518577	0.064838	14.68	<.0001
Cov2Dev*D	-0.018313	0.085918	-0.21	0.8348

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
D	1	1	183.09489	182.3232	<.0001
Cov2Dev	1	1	216.42864	215.5164	<.0001
Cov2Dev*D	1	1	0.04562	0.0454	0.8348

It is worth comparing the above analyses to analyses in JMP that use Trtmt as a nominal variable along with the additional CovDev2 as the covariate:

Response DV
Summary of Fit

RSquare	0.97885
RSquare Adj	0.975596
Root Mean Square Error	0.964621
Mean of Response	19.5625
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	559.84108	279.921	300.8301
Error	13	12.09642	0.930	Prob > F
C. Total	15	571.93750		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	19.5625	0.241155	81.12	<.0001
Trtmt[1]	5.9984794	0.425562	14.10	<.0001
Cov2Dev	0.9414282	0.04095	22.99	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Trtmt	1	1	184.87128	198.6809	<.0001
Cov2Dev	1	1	491.77858	528.5136	<.0001

Effect Details
Trtmt
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	25.560979	0.48914123	17.5000
2	13.564021	0.48914123	21.6250

Cov2Dev

Response DV
Summary of Fit

RSquare	0.97893
RSquare Adj	0.973662
Root Mean Square Error	1.002114
Mean of Response	19.5625
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	559.88671	186.629	185.8423
Error	12	12.05079	1.004	Prob > F
C. Total	15	571.93750		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	19.484096	0.44505	43.78	<.0001
Trtmt[1]	6.0093772	0.44505	13.50	<.0001
Cov2Dev	0.942701	0.042959	21.94	<.0001
Cov2Dev*Trtmt[1]	-0.009157	0.042959	-0.21	0.8348

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Trtmt	1	1	183.09489	182.3232	<.0001
Cov2Dev	1	1	483.58246	481.5442	<.0001
Cov2Dev*Trtmt	1	1	0.04562	0.0454	0.8348

Effect Details
Trtmt
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	25.493473	0.59876966	17.5000
2	13.474719	0.65859849	21.6250

Cov2Dev
Cov2Dev*Trtmt

Response DV
Summary of Fit

RSquare	0.119003
RSquare Adj	0.056075
Root Mean Square Error	5.999256
Mean of Response	19.5625
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	68.06250	68.0625	1.8911
Error	14	503.87500	35.9911	Prob > F
C. Total	15	571.93750		0.1907

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	19.5625	1.499814	13.04	<.0001
Trtmt[1]	-2.0625	1.499814	-1.38	0.1907

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Trtmt	1	1	68.062500	1.8911	0.1907

Effect Details
Trtmt
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	17.500000	2.1210573	17.5000
2	21.625000	2.1210573	21.6250

A test of the significance of the *Dx* interaction term equation [60] is a test of the homogeneity of regression slopes assumption of the ANCOVA. This kind of test should be routinely carried out as a preliminary to an ANCOVA.

ANOVA vs. ANCOVA

It is instructive to compare the results of an ordinary ANOVA with two groups with the ANCOVA with those same two groups. Here is the JMP output for these two analyses.

ANOVA output

Response DV
Summary of Fit

RSquare	0.119003
RSquare Adj	0.056075
Root Mean Square Error	5.999256
Mean of Response	19.5625
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	68.06250	68.0625	1.8911
Error	14	503.87500	35.9911	Prob > F
C. Total	15	571.93750		0.1907

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	19.5625	1.499814	13.04	<.0001
Trtmt[1]	-2.0625	1.499814	-1.38	0.1907

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Trtmt	1	1	68.062500	1.8911	0.1907

Effect Details
Trtmt
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	17.500000	2.1210573	17.5000
2	21.625000	2.1210573	21.6250

The highlighted quantities are worth particular attention and comparison with the ANCOVA below.

ANCOVA output

Response DV
Summary of Fit

RSquare	0.97885
RSquare Adj	0.975596
Root Mean Square Error	0.964621
Mean of Response	19.5625
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	559.84108	279.921	300.8301
Error	13	12.09642	0.930	Prob > F
C. Total	15	571.93750		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	19.5625	0.241155	81.12	<.0001
Trtmt[1]	5.9984794	0.425562	14.10	<.0001
Cov2Dev	0.9414282	0.04095	22.99	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Trtmt	1	1	184.87128	198.6809	<.0001
Cov2Dev	1	1	491.77858	528.5136	<.0001

Effect Details
Trtmt
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
1	25.560979	0.48914123	17.5000
2	13.564021	0.48914123	21.6250

Cov2Dev
Sequential (Type 1) Tests

Source	Nparm	DF	Seq SS	F Ratio	Prob > F
Trtmt	1	1	68.06250	73.1467	<.0001
Cov2Dev	1	1	491.77858	528.5136	<.0001

Note the reduction in *MSE* that results from including the covariate in the model. This is the major advantage of ANCOVA.

Lord's Paradox

Bock (1975, *Multivariate Statistical Methods in Behavioral Research*, p. 490ff) describes what has come to be known as Lord's Paradox:

> Suppose a large university obtains measurements, at the beginning and end of the school year, of the weight of each student who takes his meals in the university dining halls. . . Suppose two statisticians analyze these data for differences in weight gain of men versus women. The first statistician analyzes simple gain scores [spring weight minus fall weight] and concludes that 'as far as these data are concerned, there is no evidence of any interesting effect of the school diet (or of anything else) on student weight; in particular, there is no evidence of any differential effect on the two sexes, since neither group shows any systematic change.
>
> The second statistician, on the other hand, decides to do an analysis of covariance. 'After some necessary preliminaries, he determines that the slope of the regression line of final weight on initial weight is essentially the same for the two sexes. This is fortunate since it makes possible a fruitful comparison of the intercepts of the regression lines . . . He finds that the difference between the intercepts is statistically highly significant. The second statistician concludes . . . that the [men] showed significantly more gain in weight than the [women] when proper allowance is made for differences in initial weight between the two sexes.' (from Lord, 1967)
>
> As they are stated, the conclusions of the two statisticians are contradictory, and some form of paradox seems implied. On closer inspection, however, it is seen that these alternative methods of analyzing the data are actually directed toward different inferential problems. Moreover, each method provides the correct solution of the problem to which it is relevant.
>
> These inferential problems may be described briefly as *unconditional* and *conditional*, respectively. The first statistician correctly analyzes gain scores to answer the unconditional question, "Is there a difference in the average gain in weight of the populations?" . . . The answer to this question, "No, there is no difference in average gain represented by the two sexes. At the same time, the second statistician correctly employs analysis of covariance to answer the conditional question, "Is a man expected to show a greater weight gain than a woman, given that they are initially of the same weight?" . . . The answer to this question is, "Yes, the man will be expected to gain more, for if he is initially of the same weight as the woman, he is either underweight and will be expected to gain, or the woman is overweight and will be expected to lose." Because the regression lines are parallel, this expectation is independent of the given initial weight.

JMP example illustrating Lord's paradox.

Sex	Fall	Spring	D=Sp-Fa
Male	213	214	1
Male	148	155	7
Male	135	149	14
Male	154	168	14
Male	129	142	13
Male	174	188	14
Male	258	265	7
Male	180	194	14
Female	185	184	-1
Female	108	112	4
Female	101	112	11
Female	153	164	11
Female	95	93	-2
Female	88	92	4
Female	75	89	14
Female	127	119	-8

Response D=Sp-Fa
Summary of Fit

RSquare	0.221042
RSquare Adj	0.165402
Root Mean Square Error	6.396846
Mean of Response	7.3125
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	162.56250	162.563	3.9727
Error	14	572.87500	40.920	Prob > F
C. Total	15	735.43750		0.0661

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	7.3125	1.599212	4.57	0.0004
Sex[Female]	-3.1875	1.599212	-1.99	0.0661

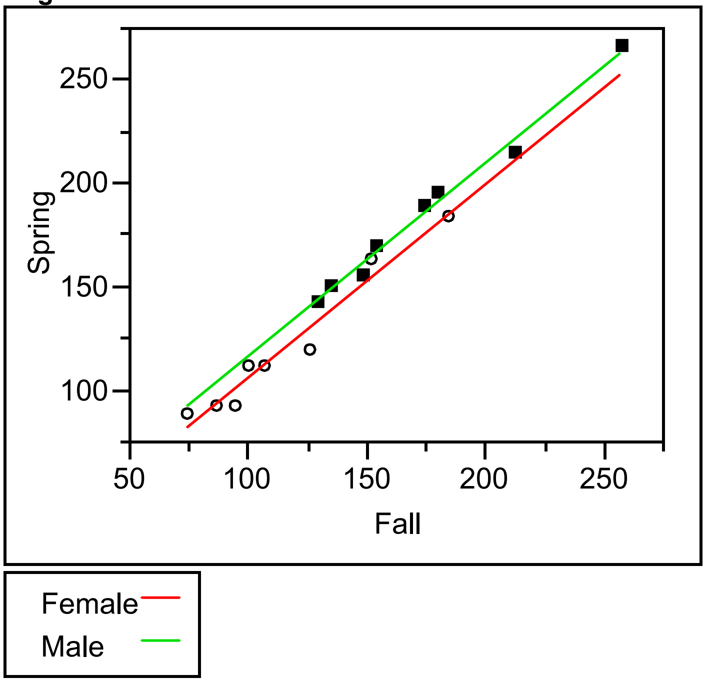
Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Sex	1	1	162.56250	3.9727	0.0661

Effect Details
Sex
Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
Female	4.125000	2.2616267	4.1250
Male	10.500000	2.2616267	10.5000

Response Spring
Regression Plot



Summary of Fit

RSquare	0.986965
RSquare Adj	0.98496
Root Mean Square Error	6.053748
Mean of Response	152.5
Observations (or Sum Wgts)	16

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	36073.578	18036.8	492.1648
Error	13	476.422	36.6	Prob > F
C. Total	15	36550.000		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	16.780885	6.029403	2.78	0.0155
Fall	0.9347851	0.040199	23.25	<.0001
Sex[Female]	-5.058352	1.902727	-2.66	0.0197

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
Fall	1	1	19817.328	540.7499	<.0001
Sex	1	1	259.008	7.0675	0.0197

Effect Details

Fall

Sex

Least Squares Means Table

Level	Least Sq Mean	Std Error	Mean
Female	147.44165	2.4312266	120.625
Male	157.55835	2.4312266	184.375

Example with Both Interactions and Indicator Variables

Below is the SAS output for a regression analysis predicting positive emotional intensity (PI) from three predictors, extraversion (E), neuroticism (N), and sex (coded in various ways). E and N were standardized ($M = 0$, $s = 1$) prior to creating cross-products.

EPI F94 & S95 2

SEXD: 0=Males; 1=Females

NI2 is non-anger negative intensity
SEXE: -1=Males; 1=Females

General Linear Models Procedure

Dependent Variable: PI

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	32.010227	4.572890	26.87	0.0001
Error	1497	254.792180	0.170202		
Corrected Total	1504	286.802407			

R-Square	C.V.	Root MSE	PI Mean
0.111611	10.89072	0.4126	3.7881

Source	DF	Type I SS	Mean Square	F Value	Pr > F
E	1	11.299174	11.299174	66.39	0.0001
N	1	10.903866	10.903866	64.06	0.0001
SEXD	1	8.203511	8.203511	48.20	0.0001
E*N	1	0.008766	0.008766	0.05	0.8205
E*SEXD	1	0.224650	0.224650	1.32	0.2508
N*SEXD	1	0.052366	0.052366	0.31	0.5792
E*N*SEXD	1	1.317892	1.317892	7.74	0.0055

Source	DF	Type III SS	Mean Square	F Value	Pr > F
E	1	8.1300794	8.1300794	47.77	0.0001
N	1	2.3810579	2.3810579	13.99	0.0002
SEXD	1	7.4903956	7.4903956	44.01	0.0001
E*N	1	1.0346205	1.0346205	6.08	0.0138
E*SEXD	1	0.2074294	0.2074294	1.22	0.2698
N*SEXD	1	0.0320519	0.0320519	0.19	0.6644
E*N*SEXD	1	1.3178923	1.3178923	7.74	0.0055

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
INTERCEPT	3.694665501	206.40	0.0001	0.01790010
E	0.122413262	6.91	0.0001	0.01771182
N	0.064418417	3.74	0.0002	0.01722294
SEXD	0.150048878	6.63	0.0001	0.02261843
E*N	0.041980318	2.47	0.0138	0.01702697
E*SEXD	-0.025116632	-1.10	0.2698	0.02275141
N*SEXD	0.009693299	0.43	0.6644	0.02233709
E*N*SEXD	-0.061604284	-2.78	0.0055	0.02213875

EPI F94 & S95 2

SEXD: 0=Females; 1=Males

NI2 is non-anger negative intensity
SEXE: -1=Males; 1=Females

General Linear Models Procedure

Dependent Variable: PI

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	32.010227	4.572890	26.87	0.0001
Error	1497	254.792180	0.170202		
Corrected Total	1504	286.802407			

R-Square	C.V.	Root MSE	PI Mean
0.111611	10.89072	0.4126	3.7881

Source	DF	Type I SS	Mean Square	F Value	Pr > F
E	1	11.299174	11.299174	66.39	0.0001
N	1	10.903866	10.903866	64.06	0.0001
SEXD	1	8.203511	8.203511	48.20	0.0001
E*N	1	0.008766	0.008766	0.05	0.8205
E*SEXD	1	0.224650	0.224650	1.32	0.2508
N*SEXD	1	0.052366	0.052366	0.31	0.5792
E*N*SEXD	1	1.317892	1.317892	7.74	0.0055

Source	DF	Type III SS	Mean Square	F Value	Pr > F
E	1	7.9013987	7.9013987	46.42	0.0001
N	1	4.6206966	4.6206966	27.15	0.0001
SEXD	1	7.4903956	7.4903956	44.01	0.0001
E*N	1	0.3273855	0.3273855	1.92	0.1657
E*SEXD	1	0.2074294	0.2074294	1.22	0.2698
N*SEXD	1	0.0320519	0.0320519	0.19	0.6644
E*N*SEXD	1	1.3178923	1.3178923	7.74	0.0055

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
INTERCEPT	3.844714378	278.06	0.0001	0.01382679
E	0.097296629	6.81	0.0001	0.01427999
N	0.074111716	5.21	0.0001	0.01422379
SEXD	-0.150048878	-6.63	0.0001	0.02261843
E*N	-0.019623966	-1.39	0.1657	0.01414944
E*SEXD	0.025116632	1.10	0.2698	0.02275141
N*SEXD	-0.009693299	-0.43	0.6644	0.02233709
E*N*SEXD	0.061604284	2.78	0.0055	0.02213875

EPI F94 & S95 2
SEXD: 0=Males; 1=Females
NI2 is non-anger negative intensity
SEXE: -1=Males; 1=Females

General Linear Models Procedure

Dependent Variable: PI

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	32.010227	4.572890	26.87	0.0001
Error	1497	254.792180	0.170202		
Corrected Total	1504	286.802407			

R-Square	C.V.	Root MSE	PI Mean
0.111611	10.89072	0.4126	3.7881

Source	DF	Type I SS	Mean Square	F Value	Pr > F
E	1	11.299174	11.299174	66.39	0.0001
N	1	10.903866	10.903866	64.06	0.0001
SEXE	1	8.203511	8.203511	48.20	0.0001
E*N	1	0.008766	0.008766	0.05	0.8205
E*SEXE	1	0.224650	0.224650	1.32	0.2508
N*SEXE	1	0.052366	0.052366	0.31	0.5792
E*N*SEXE	1	1.317892	1.317892	7.74	0.0055

Source	DF	Type III SS	Mean Square	F Value	Pr > F
E	1	15.872553	15.872553	93.26	0.0001
N	1	6.546354	6.546354	38.46	0.0001
SEXE	1	7.490396	7.490396	44.01	0.0001
E*N	1	0.173564	0.173564	1.02	0.3127
E*SEXE	1	0.207429	0.207429	1.22	0.2698
N*SEXE	1	0.032052	0.032052	0.19	0.6644
E*N*SEXE	1	1.317892	1.317892	7.74	0.0055

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
INTERCEPT	3.769689940	333.33	0.0001	0.01130922
E	0.109854945	9.66	0.0001	0.01137571
N	0.069265066	6.20	0.0001	0.01116855
SEXE	0.075024439	6.63	0.0001	0.01130922
E*N	0.011178176	1.01	0.3127	0.01106938
E*SEXE	-0.012558316	-1.10	0.2698	0.01137571
N*SEXE	0.004846650	0.43	0.6644	0.01116855
E*N*SEXE	-0.030802142	-2.78	0.0055	0.01106938

EPI F94 & S95 2
SEXD: 0=Males; 1=Females
NI2 is non-anger negative intensity
SEXE: -.5=Males; .5=Females

General Linear Models Procedure

Dependent Variable: PI

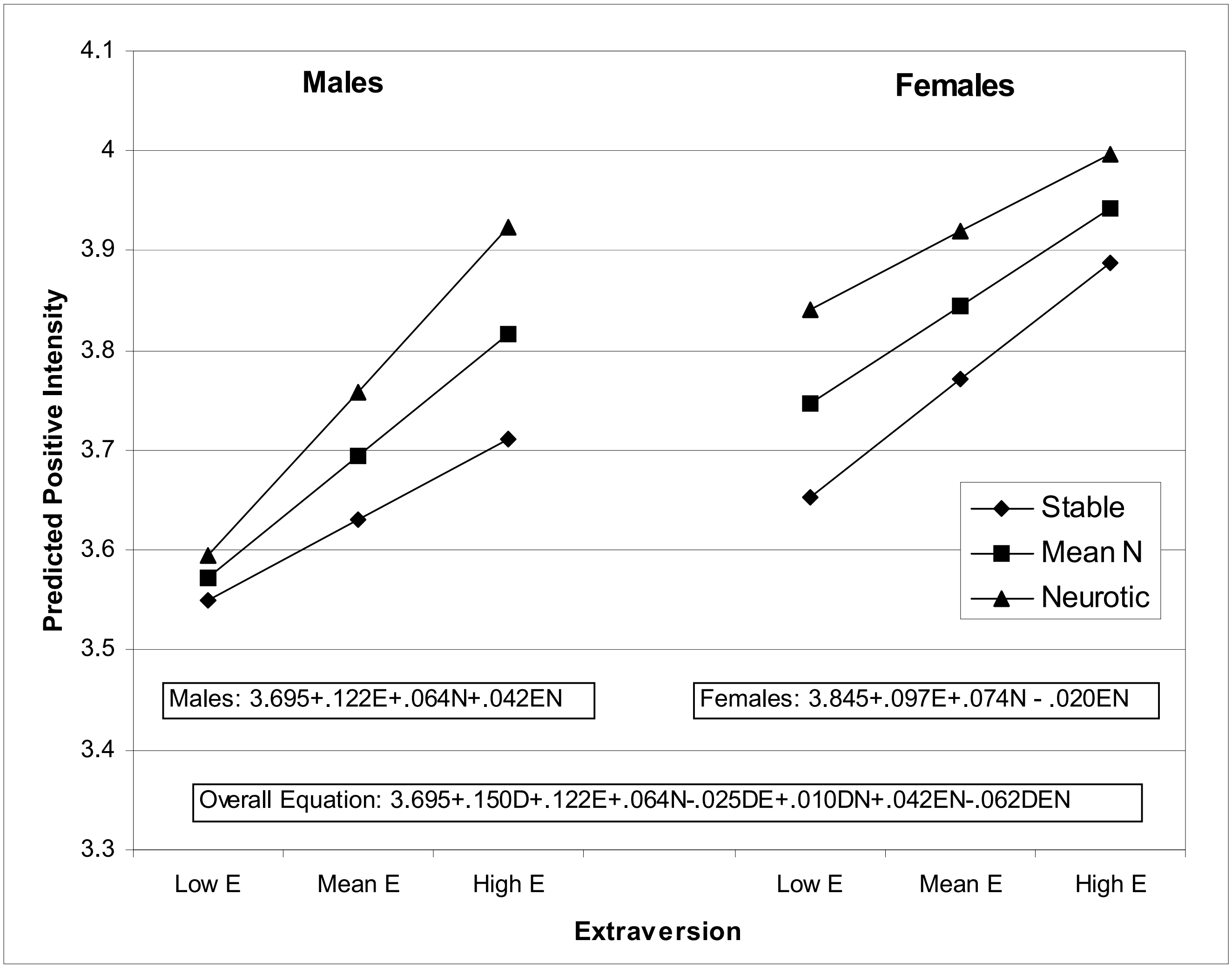
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	32.010227	4.572890	26.87	0.0001
Error	1497	254.792180	0.170202		
Corrected Total	1504	286.802407			

R-Square	C.V.	Root MSE	PI Mean
0.111611	10.89072	0.4126	3.7881

Source	DF	Type I SS	Mean Square	F Value	Pr > F
E	1	11.299174	11.299174	66.39	0.0001
N	1	10.903866	10.903866	64.06	0.0001
SEXE	1	8.203511	8.203511	48.20	0.0001
E*N	1	0.008766	0.008766	0.05	0.8205
E*SEXE	1	0.224650	0.224650	1.32	0.2508
N*SEXE	1	0.052366	0.052366	0.31	0.5792
E*N*SEXE	1	1.317892	1.317892	7.74	0.0055

Source	DF	Type III SS	Mean Square	F Value	Pr > F
E	1	15.872553	15.872553	93.26	0.0001
N	1	6.546354	6.546354	38.46	0.0001
SEXE	1	7.490396	7.490396	44.01	0.0001
E*N	1	0.173564	0.173564	1.02	0.3127
E*SEXE	1	0.207429	0.207429	1.22	0.2698
N*SEXE	1	0.032052	0.032052	0.19	0.6644
E*N*SEXE	1	1.317892	1.317892	7.74	0.0055

Parameter	Estimate	T for H0: Parameter=0	Pr > |T|	Std Error of Estimate
INTERCEPT	3.769689940	333.33	0.0001	0.01130922
E	0.109854945	9.66	0.0001	0.01137571
N	0.069265066	6.20	0.0001	0.01116855
SEXE	0.150048878	6.63	0.0001	0.02261843
E*N	0.011178176	1.01	0.3127	0.01106938
E*SEXE	-0.025116632	-1.10	0.2698	0.02275141
N*SEXE	0.009693299	0.43	0.6644	0.02233709
E*N*SEXE	-0.061604284	-2.78	0.0055	0.02213875



Excel plot illustrating the previous four regression analyses. Note how the equations in the figure are obtained from the first regression output (with Sex dummy coding: Males = 0; Females = 1) and how the equations are related to the other outputs.

Interpretation of the Lower-Order Terms in the Output on p. 78

The intercept (3.695) in the regression equation reflects the predicted positive intensity (PI) score for males who are at the mean on both E and N because it is the predicted *Y* when all predictors are zero. The regression weight for E (0.1224) is the slope of the relation between E and PI for males (Sex = 0) who are at the mean on neuroticism (N = 0). An analogous interpretation applies to the weight for N. The weight for Sex (0.150) is the difference in predicted PI between males and females who are at the mean on both E and N, with the female predicted PI being 0.15 points higher than that for males. The interpretations of the 2-way interaction weights are given below in the discussion of the interpretation of the 3-way interaction.

Interpretation of the Two and Three-Way Interactions in the Output on p. 78

The 3-way interaction weight may be interpreted as the change in a 2-way interaction weight one would get for a one unit increase in the 3rd variable. That is, for example, the Sex × E × N weight reflects how the E × N 2-way interaction weight changes for a one unit increase in Sex. Since Sex is dummy coded here (M = 0, F = 1), a one unit increase in Sex means the change in going from male to female. The 2-way E × N weight for males in this output is 0.04198. This is the coefficient for males only because when Sex = 0 (males) the 3-way term goes away because it's multiplied by zero, and the only E × N term left is 0.04198. When Sex = 1 (females) there are 2 E × N terms in the equation: 0.04198 and -0.0616. The latter term becomes a simple E × N term because it's multiplied by Sex, which is 1. Therefore, the 2-way E × N term for females would thus be 0.04198 + (-0.0616) = -0.01962. These two E × N weights in fact show up in the separate equations shown in the figure on p. 82. The difference between the two 2-way E × N interaction weights is thus the 3-way interaction weight, and in this example it's highly significant. That means that the two E × N interactions shown in the figure are significantly different from one another. The one for males is pretty pronounced with a positive fan shape (as reflected in the 0.04198). The one for females is less pronounced and has a negative fan shape (as reflected in the -0.01962).

Because the 3-way interaction is significant here, the interpretation of the E × N interactions must be carried out separately for males and females. Interpreting the E × N interaction for males (0.04198) would go something like this (it needs to be explicitly stated that all of the following applies only to males): The weight, 0.04198, reflects the increase in the slope of the relation between E and PI (positive intensity) that results from a one unit increase in N. Since E and N are standardized, one unit is one standard deviation. Thus, an increase of one standard deviation in N would lead to an increase of 0.04198 in the *slope* of the relation between E and PI. The slope of the relation between E and PI for males at the mean on N (N = 0) is 0.1224 (from the output). Thus, for males at the mean on N a one standard deviation increase in E would lead us to predict a 0.1224 point increase in positive intensity. However, for neurotic males at one standard deviation above the mean on N (N = 1), the slope of the relation between E and PI would be 0.1224 + 0.04198 = 0.16438. Thus, for neurotic males, a one standard deviation increase in E would lead us to predict a 0.16438 point increase in positive intensity. Thus, while the relation between extraversion and positive intensity is positive for both those at the mean on neuroticism and those at one standard deviation above the mean on neuroticism, that relation between E and PI is much stronger (and significantly so) for neurotics than for those at the mean on N.

Logistic Regression

If the criterion variable in a regression analysis is dichotomous (0, 1), then special considerations arise in the analysis. The predicted value of *Y* for any given *X* may be considered to be the probability that *Y* takes on the value 1 for observations with the given *X*. As Kutner et al. point out, if one uses the ordinary simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad [61]$$

to analyze the data, at least three problems arise:

1) Nonnormal error terms
2) Violation of the homoscedasticity assumption
3) Predicted *Y*s that fall outside the 0 to 1 range.

The most common solution is to use a *logistic* response function rather than the simple linear model of equation [61] to model the data. The logistic response function (generalized to the case of several predictors) may be written as

$$\hat{Y} = \frac{\exp(b_0 + b_1 X_1 + b_2 X_2 + \ldots)}{1 + \exp(b_0 + b_1 X_1 + b_2 X_2 + \ldots)} = \frac{1}{1 + \exp(-b_0 - b_1 X_1 - b_2 X_2 - \ldots)}. \qquad [62]$$

This function is a monotonic, flattened S-shaped curve often called a *sigmoidal* curve which asymptotes at 0 and 1.

A nice property of the logistic response function is that it may be linearized by using the *logit* or *log odds* transform. If the predicted *Y* is considered to be the probability, *p*, of obtaining a 1 for the given value of *X*, and the model of equation [62] is assumed, then it can be shown that the natural log (i.e., logarithm to the base *e*) of the odds associated with *p* is a linear function of the *X*s:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X_1 + b_2 X_2 + \ldots \qquad [63]$$

The ratio $p/(1 - p)$ is called the odds associated with the probability, *p*. Values for b_0, b_1, etc. may be obtained, usually by iterative maximum likelihood estimation, by many programs for logistic regression analysis. JMP will do logistic regression with one predictor in the Fit Y by X platform and many predictors in the Fit Model platform.

Odds and Probability

Because odds play such a prominent role in logistic regression and have properties that in some respects are nicer than those of probabilities, it is worth considering some of the characteristics of odds and their relation to probability. Odds are usually represented as a ratio of the chances of an event occurring to the chances of the event not occurring. For example, odds of 3:2 mean 3 chances out of 5 that the event occurs, 2 chances out of 5 the event doesn’t occur. One could represent this as a fraction, 3/2 = 1.5 or odds of 1.5:1, and the meaning would be the same. Notice that the probability of the event occurring would be 3/5 = .6, and the probability of the event not occurring would be 2/5 = .4.

Note that the scale on which odds are measured goes from 0 to ∞, whereas the scale on which probabilities are measured goes from 0 to 1. The algebraic relations between odds (o) and probability (p) are as follows:

$$o = \frac{p}{1-p} \qquad [64]$$

and

$$p = \frac{o}{1+o}\,. \qquad [65]$$

Because the logistic regression equation [63] predicts the log of the odds, odds play an important role in the interpretation of the regression coefficients obtained from the logistic regression analysis. The first step in interpreting the coefficients is to take the antilog of each, i.e., exponentiate each coefficient. In the case of just one predictor, X_1, the interpretations of b_0 and b_1 would proceed as follows:

$e^{b0} = \exp(b_0)$ = the odds that Y takes on the value 1 when $X_1 = 0$. This value may be converted to a probability using equation [65] if one wishes.

$e^{b1} = \exp(b_1)$ = the ratio of the odds that Y takes on the value 1 when $X_1 = a$ to the odds that Y takes on the value 1 when $X_1 = a + 1$. This is sometimes called the *odds ratio*, and it reflects how the odds of Y taking on the value 1 change as X_1 increases by one point.

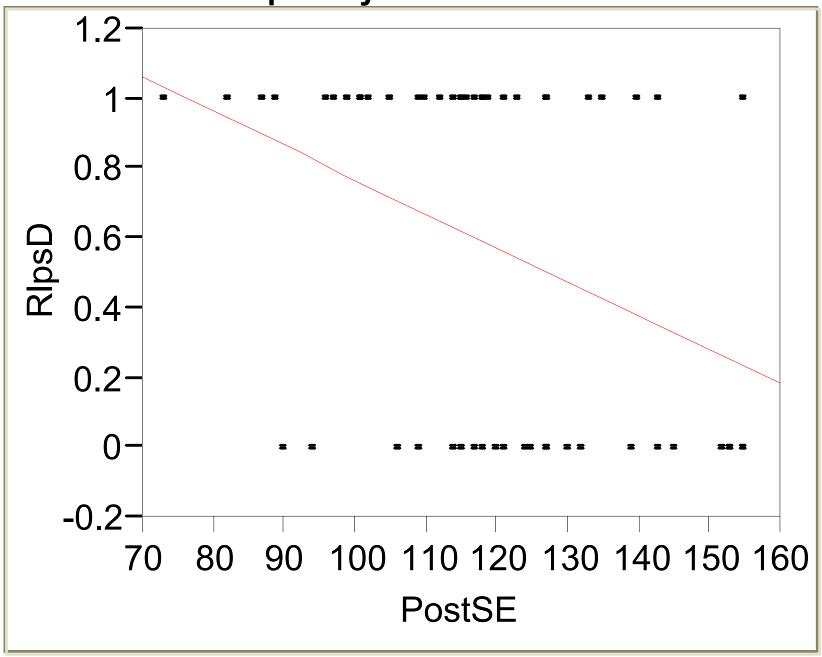
For example, if $\exp(b_1) = 1$, then a one point increase in X_1 predicts no change in the odds of Y taking on the value 1. If $\exp(b_1) = 1.10$, then a one point increase in X_1 predicts a 10% increase in the odds of Y taking on the value 1. If $\exp(b_1) = .80$, then a one point increase in X_1 predicts that the odds of Y taking on the value 1 would be only 80% as great as before the one point increase (in other words, the odds decrease as X_1 increases).

If one desires the odds ratio comparing points on X_1 that are, say, farther apart than one point, one simply multiplies b_1 by the difference one is interested in before exponentiating b_1. For example, if one were interested in the odds ratio for a 5 point difference in X_1, one would find $\exp(5b_1)$. By default, JMP will provide the odds ratio for the difference between the lowest and highest values of X_1 in the data set.

Example.

Below is the JMP analysis of some clinical smoking cessation data obtained by a local therapist who was interested in predicting whether or not a client would relapse as a function of post-therapy self-efficacy (for quitting smoking) score. The data file is available on my web site. First, the ordinary linear regression fit:

Bivariate Fit of RlpsD By PostSE



Linear Fit

RlpsD = 1.7430458 - 0.0097354 PostSE

Summary of Fit

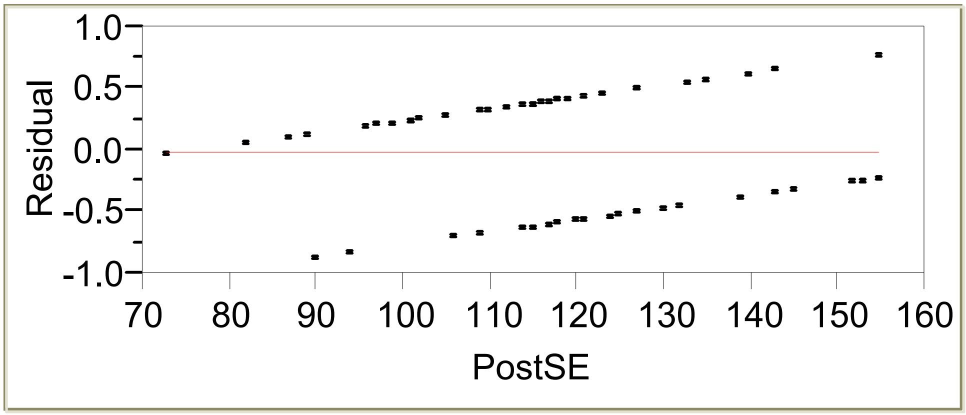
RSquare	0.146795
RSquare Adj	0.131559
Root Mean Square Error	0.462982
Mean of Response	0.586207
Observations (or Sum Wgts)	58

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2.065256	2.06526	9.6349
Error	56	12.003709	0.21435	Prob > F
C. Total	57	14.068966		0.0030

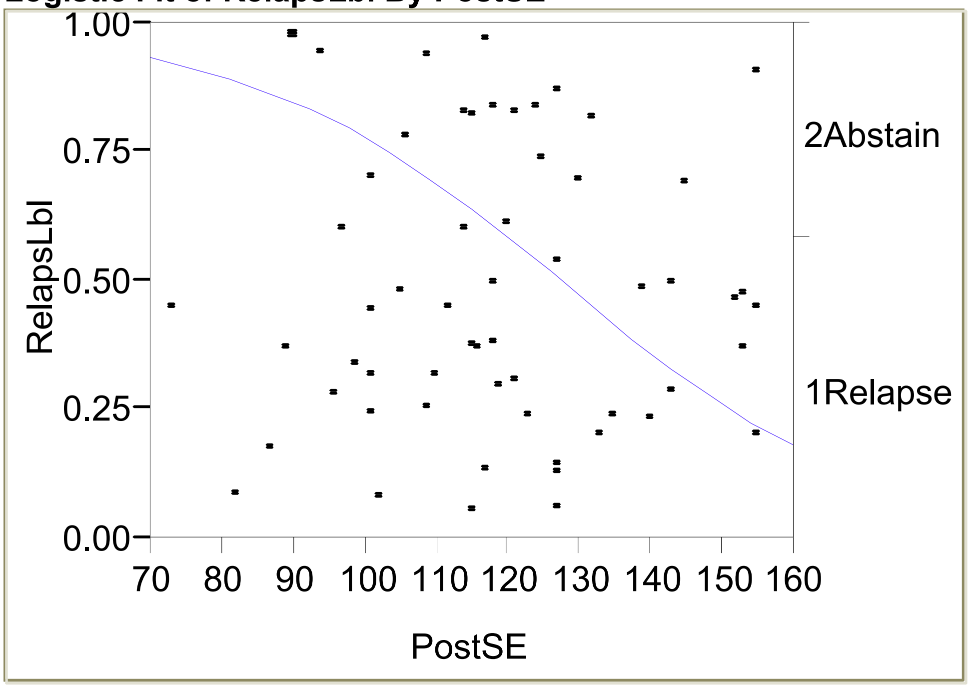
Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	1.7430458	0.377617	4.62	<.0001
PostSE	-0.009735	0.003136	-3.10	0.0030



Now the logistic regression model analysis:

Logistic Fit of RelapsLbl By PostSE



Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	4.546697	1	9.093394	0.0026
Full	34.789448			
Reduced	39.336145			

RSquare (U)	0.1156
Observations (or Sum Wgts)	58

Converged by Gradient

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	5.87454534	2.0638865	8.10	0.0044
PostSE	-0.0460273	0.0168984	7.42	0.0065

For log odds of 1Relapse/2Abstain

Nominal Logistic Fit for RelapsLbl
Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	4.546697	1	9.093394	0.0026
Full	34.789448			
Reduced	39.336145			

RSquare (U)	0.1156
Observations (or Sum Wgts)	58

Converged by Gradient

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	37	18.764290	37.52858
Saturated	38	16.025158	Prob>ChiSq
Fitted	1	34.789448	0.4449

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq	Odds Ratio
Intercept	5.87454534	2.0638865	8.10	0.0044	.
PostSE	-0.0460273	0.0168984	7.42	0.0065	0.02295465

For log odds of 1Relapse/2Abstain

Effect Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
PostSE	1	1	7.41885645	0.0065

Now a model that includes post-therapy self-efficacy along with number of prior attempts to quit as predictors along with the interaction:

Nominal Logistic Fit for RelapsLbl
Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	6.592714	3	13.18543	0.0043
Full	32.743431			
Reduced	39.336145			

RSquare (U)	0.1676
Observations (or Sum Wgts)	58

Converged by Gradient

Lack Of Fit

Source	DF	-LogLikelihood	ChiSquare
Lack Of Fit	50	28.061300	56.1226
Saturated	53	4.682131	Prob>ChiSq
Fitted	3	32.743431	0.2562

Parameter Estimates

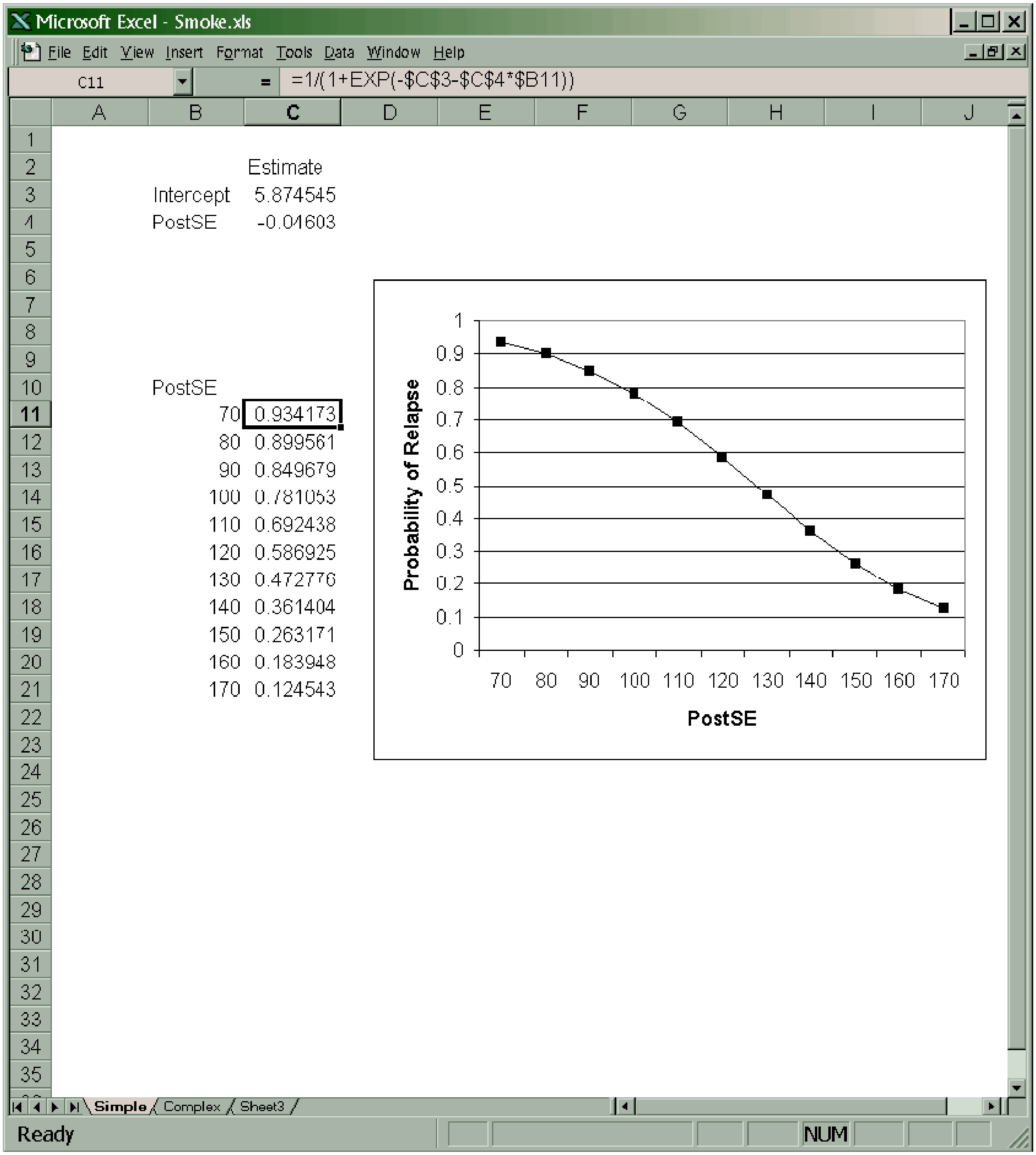
Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	8.77016896	3.2256597	7.39	0.0066
PostSE	-0.0734596	0.0267743	7.53	0.0061
PriorAt	-0.7084395	0.6222344	1.30	0.2549
PostSE*PriorAt	0.00712071	0.0051451	1.92	0.1664

For log odds of 1Relapse/2Abstain

Effect Wald Tests

Source	Nparm	DF	Wald ChiSquare	Prob>ChiSq
PostSE	1	1	7.52766897	0.0061
PriorAt	1	1	1.29627631	0.2549
PostSE*PriorAt	1	1	1.91536066	0.1664

Excel may be used to get nice probability plots for logistic regression models as the following two Excel pages illustrate. Note the formula in the formula window for the selected cell in each worksheet.



Microsoft Excel - Smoke.xls

C11 = =1/(1+EXP(-C3-C4*$B11))

	B	C
2		Estimate
3	Intercept	5.874545
4	PostSE	-0.04603
10	PostSE	
11	70	0.934173
12	80	0.899561
13	90	0.849679
14	100	0.781053
15	110	0.692438
16	120	0.586925
17	130	0.472776
18	140	0.361404
19	150	0.263171
20	160	0.183948
21	170	0.124543

Microsoft Excel - Smoke.xls

File Edit View Insert Format Tools Data Window Help

C11 = =1/(1+EXP(-C3-C4*$B11-$C$5*C$10-C6*$B11*C$10))

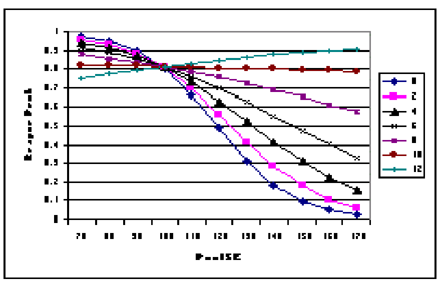
Unstandardized Term	Estimate	Standardized Term	Estimate	Standardized Term	Estimate
Intercept	8.77017	Intercept	0.4115	Intercept	0.44856
PartSE	-0.0735	Std PartSE	-1.0618	Std PartSE	-0.9745
PriorAt	-0.7084	Std PriorAt	0.37485	Std PriorAt	0.5048
PartSE*PriorAt	0.00712	Std PartSE*Std Pr	0.37901		

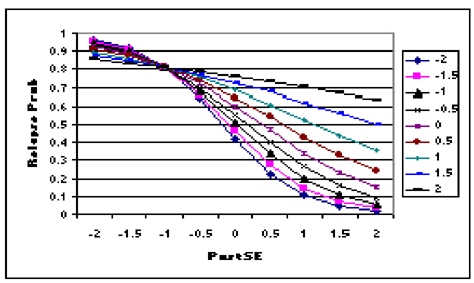
Prior Attempts

		0	2	4	6	8	10	12
	70	0.97417	0.96113	0.94203	0.91436	0.87524	0.82173	0.75178
	80	0.94752	0.93187	0.912	0.88702	0.85608	0.8184	0.77347
	90	0.89649	0.88326	0.86858	0.85238	0.83455	0.81503	0.79379
	100	0.806	0.80713	0.80826	0.80938	0.8105	0.81161	0.81272
	110	0.66588	0.69832	0.72888	0.75743	0.78386	0.80814	0.83029
PartSE	120	0.48876	0.56146	0.63163	0.69663	0.75462	0.80463	0.84652
	130	0.31441	0.41458	0.52234	0.62807	0.72281	0.80106	0.86146
	140	0.18032	0.28146	0.41088	0.55394	0.68858	0.79745	0.87516
	150	0.09546	0.17808	0.30787	0.47732	0.65216	0.79379	0.88768
	160	0.04818	0.10701	0.221	0.40176	0.61387	0.79007	0.89909
	170	0.02371	0.06216	0.15321	0.3306	0.57412	0.78631	0.90946

Prior Att

		-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
	-2	0.96448	0.9573	0.94874	0.93859	0.92657	0.91242	0.89585	0.87658	0.85431
	-1.5	0.91618	0.90844	0.90006	0.89101	0.88124	0.87073	0.85943	0.84732	0.83437
	-1	0.81482	0.81451	0.81419	0.81388	0.81356	0.81325	0.81293	0.81261	0.8123
	-0.5	0.63916	0.66025	0.68072	0.70051	0.71959	0.7379	0.75543	0.77214	0.78803
PartSE	0	0.41625	0.46238	0.50916	0.55579	0.60145	0.64541	0.68705	0.72587	0.76155
	0.5	0.22303	0.27569	0.33542	0.40093	0.47018	0.5406	0.60944	0.67417	0.73288
	1	0.10358	0.14417	0.19715	0.26362	0.34292	0.43207	0.52586	0.61786	0.70212
	1.5	0.04445	0.06938	0.10673	0.16072	0.23483	0.3297	0.44081	0.55819	0.6694
	2	0.01838	0.03194	0.05494	0.09291	0.15289	0.24127	0.3591	0.49679	0.63497

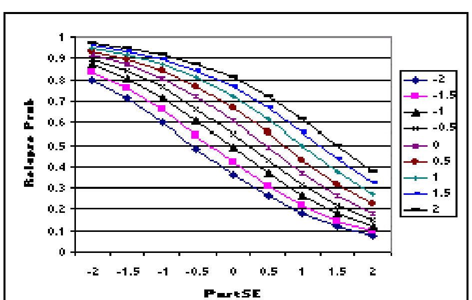




No interaction model

Prior Att

		-2	-1.5	-1	-0.5	0	0.5	1	1.5	2
	-2	0.80027	0.83759	0.86907	0.89522	0.91664	0.93401	0.94797	0.9591	0.96793
	-1.5	0.7111	0.76008	0.80306	0.83996	0.87106	0.89685	0.91797	0.93508	0.94882
	-1	0.60192	0.66058	0.71469	0.76327	0.80582	0.84231	0.87302	0.89847	0.91929
	-0.5	0.48156	0.54454	0.60612	0.6645	0.71826	0.76642	0.80855	0.84462	0.87495
PartSE	0	0.36331	0.42345	0.48594	0.54888	0.6103	0.6684	0.72179	0.76955	0.81125
	0.5	0.25955	0.3109	0.36737	0.42774	0.49033	0.55322	0.61446	0.67228	0.7253
	1	0.17718	0.21701	0.26294	0.31467	0.37146	0.43203	0.49471	0.55756	0.61861
	1.5	0.11683	0.14549	0.17975	0.22001	0.26635	0.31847	0.37557	0.43634	0.4991
	2	0.07515	0.09469	0.11865	0.14769	0.18235	0.22303	0.26979	0.32229	0.37969



Simple Complex Sheet3

Ready NUM