

Multiple Regression

Statistical Associates
Blue Book Series



G. David Garson
School of Public & International Affairs
North Carolina State University



www.statisticalassociates.com

@c 2014 by G. David Garson and Statistical Associates Publishing. All rights reserved worldwide in all media. No permission is granted to any user to copy or post this work in any format or any media.

ISBN-10: 1626380317

ISBN-13: 978-1-62638-031-8

The author and publisher of this eBook and accompanying materials make no representation or warranties with respect to the accuracy, applicability, fitness, or completeness of the contents of this eBook or accompanying materials. The author and publisher disclaim any warranties (express or implied), merchantability, or fitness for any particular purpose. The author and publisher shall in no event be held liable to any party for any direct, indirect, punitive, special, incidental or other consequential damages arising directly or indirectly from any use of this material, which is provided “as is”, and without warranties. Further, the author and publisher do not warrant the performance, effectiveness or applicability of any sites listed or linked to in this eBook or accompanying materials. All links are for information purposes only and are not warranted for content, accuracy or any other implied or explicit purpose. This eBook and accompanying materials is © copyrighted by G. David Garson and Statistical Associates Publishing. No part of this may be copied, or changed in any format, sold, or used in any way under any circumstances other than reading by the downloading individual.

Contact:

G. David Garson, President
Statistical Publishing Associates
274 Glenn Drive
Asheboro, NC 27205 USA

Email: [gdavidgarson@gmail.com](mailto:g davidgarson@gmail.com)
Web: www.statisticalassociates.com

Table of Contents

Page numbering works in the full edition. This is a preview edition of the first 30 pages.

Overview	13
Data examples in this volume.....	16
Key Terms and Concepts.....	17
OLS estimation	17
The regression equation	18
Dependent variable	20
Independent variables.....	21
Dummy variables.....	21
Interaction effects.....	22
Interactions.....	22
Centering	23
Significance of interaction effects.....	23
Interaction terms with categorical dummies	24
Plotting interactions through simple slope analysis.....	24
Separate regressions	27
Predicted values.....	28
SPSS.....	28
SAS	28
Stata.....	29
Adjusted predicted values.....	Error! Bookmark not defined.
Residuals.....	Error! Bookmark not defined.
Centering.....	Error! Bookmark not defined.
OLS regression in SPSS	Error! Bookmark not defined.
Example	Error! Bookmark not defined.
SPSS input.....	Error! Bookmark not defined.
SPSS Output.....	Error! Bookmark not defined.
The regression coefficient, b	Error! Bookmark not defined.
Interpreting b for dummy variables	Error! Bookmark not defined.
Confidence limits on b	Error! Bookmark not defined.
Beta weights	Error! Bookmark not defined.
Zero-order, partial, and part correlations	Error! Bookmark not defined.
R ² and the “Model Summary” table	Error! Bookmark not defined.
The Anova table.....	Error! Bookmark not defined.
Tolerance and VIF collinearity statistics	Error! Bookmark not defined.
SPSS plots	Error! Bookmark not defined.
SPSS “Plots” dialog.....	Error! Bookmark not defined.
Plot of standardized residuals against standardized predicted values.....	Error! Bookmark not defined.

Histogram of standardized residuals.....	Error! Bookmark not defined.
Normal probability (P-P) plot	Error! Bookmark not defined.
OLS regression in SAS.....	Error! Bookmark not defined.
Example	Error! Bookmark not defined.
SAS input	Error! Bookmark not defined.
SAS output.....	Error! Bookmark not defined.
The regression coefficient, b	Error! Bookmark not defined.
Interpreting b for dummy variables	Error! Bookmark not defined.
Confidence limits on b	Error! Bookmark not defined.
Beta weights	Error! Bookmark not defined.
Zero order, partial, and part correlation	Error! Bookmark not defined.
R-Squared and the Anova table.....	Error! Bookmark not defined.
Tolerance and VIF collinearity statistics	Error! Bookmark not defined.
SAS Plots.....	Error! Bookmark not defined.
SAS plotting options	Error! Bookmark not defined.
Plot of residuals against predicted values.....	Error! Bookmark not defined.
Histogram and kernel density plot of standardized residuals.....	Error! Bookmark not defined.
Normal probability (P-P) plot	Error! Bookmark not defined.
Normal quantile-quantile (Q-Q) plot.....	Error! Bookmark not defined.
Other SAS plots.....	Error! Bookmark not defined.
OLS regression in Stata	Error! Bookmark not defined.
Example	Error! Bookmark not defined.
Stata input	Error! Bookmark not defined.
Stata output	Error! Bookmark not defined.
The regression coefficient, b	Error! Bookmark not defined.
Interpreting b coefficients.....	Error! Bookmark not defined.
Confidence limits on b.....	Error! Bookmark not defined.
Beta weights	Error! Bookmark not defined.
R-Squared and the Anova table.....	Error! Bookmark not defined.
Zero order, partial, and part correlation	Error! Bookmark not defined.
Tolerance and VIF collinearity statistics	Error! Bookmark not defined.
Other Stata postestimation output	Error! Bookmark not defined.
Stata Plots.....	Error! Bookmark not defined.
Stata plotting options	Error! Bookmark not defined.
Plot of standardized residuals against standardized predicted values.....	Error! Bookmark not defined.
Histogram of standardized residuals.....	Error! Bookmark not defined.
Normal probability (P-P) plot	Error! Bookmark not defined.
Margin plots.....	Error! Bookmark not defined.
Robust regression	Error! Bookmark not defined.
Overview	Error! Bookmark not defined.
When to use robust regression	Error! Bookmark not defined.
Robust regression in SPSS	Error! Bookmark not defined.

Overview..... Error! Bookmark not defined.

SPSS input Error! Bookmark not defined.

SPSS output..... Error! Bookmark not defined.

Robust regression in SAS..... Error! Bookmark not defined.

 SAS input..... Error! Bookmark not defined.

 SAS output Error! Bookmark not defined.

Robust regression in Stata Error! Bookmark not defined.

 Stata input Error! Bookmark not defined.

 Stata output..... Error! Bookmark not defined.

Hierarchical multiple regression Error! Bookmark not defined.

 Overview Error! Bookmark not defined.

 Examples Error! Bookmark not defined.

Difference in differences regression..... Error! Bookmark not defined.

 Overview Error! Bookmark not defined.

 The parallel trend assumption Error! Bookmark not defined.

 Example data..... Error! Bookmark not defined.

 Data setup..... Error! Bookmark not defined.

 The model..... Error! Bookmark not defined.

Difference modeling in SPSS Error! Bookmark not defined.

 SPSS input Error! Bookmark not defined.

 Should the dependent variable be linear or logarithmic? Error! Bookmark not defined.

 SPSS output..... Error! Bookmark not defined.

Difference modeling in SAS..... Error! Bookmark not defined.

 SAS input..... Error! Bookmark not defined.

 Should the dependent variable be linear or logarithmic? Error! Bookmark not defined.

 SAS output Error! Bookmark not defined.

Difference modeling in Stata..... Error! Bookmark not defined.

 Stata input Error! Bookmark not defined.

 Should the dependent variable be linear or logarithmic? Error! Bookmark not defined.

 Stata output..... Error! Bookmark not defined.

Panel data regression..... Error! Bookmark not defined.

 Overview Error! Bookmark not defined.

 Types of panel data regression Error! Bookmark not defined.

 Software for panel data regression..... Error! Bookmark not defined.

Stepwise Multiple Regression..... Error! Bookmark not defined.

 Overview Error! Bookmark not defined.

 Forward, backward, and stepwise regression..... Error! Bookmark not defined.

 Warning Error! Bookmark not defined.

 Other problems of stepwise regression Error! Bookmark not defined.

 Dummy variables in stepwise regression Error! Bookmark not defined.

 Example Error! Bookmark not defined.

 Stepwise regression in SPSS..... Error! Bookmark not defined.

 Overview..... Error! Bookmark not defined.

SPSS input	Error! Bookmark not defined.
SPSS Output	Error! Bookmark not defined.
Stepwise regression in SAS.....	Error! Bookmark not defined.
Overview.....	Error! Bookmark not defined.
Example	Error! Bookmark not defined.
SAS input.....	Error! Bookmark not defined.
SAS output	Error! Bookmark not defined.
Stepwise regression in Stata	Error! Bookmark not defined.
Overview.....	Error! Bookmark not defined.
Example	Error! Bookmark not defined.
Stata input	Error! Bookmark not defined.
Stata output.....	Error! Bookmark not defined.
Model selection regression in SPSS	Error! Bookmark not defined.
Overview	Error! Bookmark not defined.
Example	Error! Bookmark not defined.
The “Fields” tab	Error! Bookmark not defined.
The “Build Options” tab.....	Error! Bookmark not defined.
The “Model Options tab	Error! Bookmark not defined.
Model Viewer	Error! Bookmark not defined.
Default output	Error! Bookmark not defined.
Model Viewer interface.....	Error! Bookmark not defined.
Model Viewer: Automatic Data Preparation Table	Error! Bookmark not defined.
Model Viewer: Model Building Summary window.....	Error! Bookmark not defined.
Model Viewer: Coefficients window	Error! Bookmark not defined.
Coefficient importance.....	Error! Bookmark not defined.
Model Viewer: Effects window.....	Error! Bookmark not defined.
Model Viewer: Predicted by Observed window.....	Error! Bookmark not defined.
Model Viewer: Estimated Means window	Error! Bookmark not defined.
Model Viewer: Residuals window	Error! Bookmark not defined.
Model Viewer: Outliers window.....	Error! Bookmark not defined.
Example 2: A boosted ensemble model.....	Error! Bookmark not defined.
Model Viewer: Model Summary window.....	Error! Bookmark not defined.
Model Viewer: Predictor Frequency window.....	Error! Bookmark not defined.
Model Viewer: Saving and printing output	Error! Bookmark not defined.
Model selection regression in SAS.....	Error! Bookmark not defined.
Overview	Error! Bookmark not defined.
Selection criteria.....	Error! Bookmark not defined.
Example	Error! Bookmark not defined.
PROC REG/SELECTION method	Error! Bookmark not defined.
SAS input.....	Error! Bookmark not defined.
SAS output	Error! Bookmark not defined.
Summary.....	Error! Bookmark not defined.
PROC GLMSELECT method	Error! Bookmark not defined.

Overview.....	Error! Bookmark not defined.
SAS input.....	Error! Bookmark not defined.
SAS output.....	Error! Bookmark not defined.
Quantile Regression.....	Error! Bookmark not defined.
Overview.....	Error! Bookmark not defined.
Introduction.....	Error! Bookmark not defined.
Pseudo-R ²	Error! Bookmark not defined.
Standard errors and coefficient significance.....	Error! Bookmark not defined.
Interpreting quantile regression coefficients.....	Error! Bookmark not defined.
Conditional vs. unconditional quantile regression.....	Error! Bookmark not defined.
Generalized quantile regression (GQR and IVGQR).....	Error! Bookmark not defined.
Quantile Regression for Panel Data (QRPD).....	Error! Bookmark not defined.
Example.....	Error! Bookmark not defined.
Quantile regression in SPSS.....	Error! Bookmark not defined.
Overview.....	Error! Bookmark not defined.
Heteroscedasticity test.....	Error! Bookmark not defined.
SPSS input.....	Error! Bookmark not defined.
SPSS output.....	Error! Bookmark not defined.
Quantile regression in SAS.....	Error! Bookmark not defined.
Overview.....	Error! Bookmark not defined.
Heteroscedasticity test.....	Error! Bookmark not defined.
SAS input.....	Error! Bookmark not defined.
SAS output.....	Error! Bookmark not defined.
Model selection quantile regression in SAS.....	Error! Bookmark not defined.
Overview.....	Error! Bookmark not defined.
Example.....	Error! Bookmark not defined.
SAS input.....	Error! Bookmark not defined.
SAS output.....	Error! Bookmark not defined.
Quantile regression in Stata.....	Error! Bookmark not defined.
Overview.....	Error! Bookmark not defined.
Heteroscedasticity test.....	Error! Bookmark not defined.
Stata input.....	Error! Bookmark not defined.
Stata output.....	Error! Bookmark not defined.
More about significance tests in regression models.....	Error! Bookmark not defined.
Overview.....	Error! Bookmark not defined.
Complex samples.....	Error! Bookmark not defined.
F test.....	Error! Bookmark not defined.
Overall test of the model and R ²	Error! Bookmark not defined.
Small samples.....	Error! Bookmark not defined.
The partial F test.....	Error! Bookmark not defined.
t-tests.....	Error! Bookmark not defined.
One- vs. two-tailed t tests.....	Error! Bookmark not defined.
t-tests for dummy variables.....	Error! Bookmark not defined.

Confidence limits and standard errors..... **Error! Bookmark not defined.**
 Confidence intervals and prediction intervals..... **Error! Bookmark not defined.**
 The confidence interval of the regression coefficient..... **Error! Bookmark not defined.**
 The confidence interval of y, the dependent variable **Error! Bookmark not defined.**
 The prediction interval of y, the dependent variable **Error! Bookmark not defined.**
 Standard error of estimate (SEE) / root mean square error (MSE)**Error! Bookmark not defined.**
 Standard error and mean standard error of predicted values [SE(Pred) and MSEP]**Error! Bookmark not defined.**
 More about effect size measures in multiple regression **Error! Bookmark not defined.**
 Beta weights..... **Error! Bookmark not defined.**
 Overview **Error! Bookmark not defined.**
 Using beta weights in model comparisons..... **Error! Bookmark not defined.**
 Significance of beta..... **Error! Bookmark not defined.**
 Unique v. joint effects **Error! Bookmark not defined.**
 Standardization and comparability of variables..... **Error! Bookmark not defined.**
 Beta weights over 1.0 **Error! Bookmark not defined.**
 Labeling: b and beta **Error! Bookmark not defined.**
 Correlation..... **Error! Bookmark not defined.**
 Zero-order correlation, r..... **Error! Bookmark not defined.**
 Semipartial (part) correlation **Error! Bookmark not defined.**
 Partial correlation squared **Error! Bookmark not defined.**
 R-squared **Error! Bookmark not defined.**
 Overview..... **Error! Bookmark not defined.**
 Warning regarding R-square differences between samples.. **Error! Bookmark not defined.**
 Adding variables **Error! Bookmark not defined.**
 Adjusted R-square **Error! Bookmark not defined.**
 R-squared difference tests **Error! Bookmark not defined.**
 Level-importance **Error! Bookmark not defined.**
 The intercept **Error! Bookmark not defined.**
 Residual analysis and diagnostics **Error! Bookmark not defined.**
 Outliers, influence, leverage, and distance..... **Error! Bookmark not defined.**
 Overview..... **Error! Bookmark not defined.**
 Outliers and residuals **Error! Bookmark not defined.**
 Leverage and distance **Error! Bookmark not defined.**
 Influence **Error! Bookmark not defined.**
 Example **Error! Bookmark not defined.**
 Types of residuals and plots..... **Error! Bookmark not defined.**
 Unstandardized residuals **Error! Bookmark not defined.**
 Standardized residuals..... **Error! Bookmark not defined.**
 Standardized deleted residuals **Error! Bookmark not defined.**
 Studentized residuals **Error! Bookmark not defined.**
 Studentized deleted residuals **Error! Bookmark not defined.**

Which type of residual to use?	Error! Bookmark not defined.
Types of residual plots	Error! Bookmark not defined.
Residual plots.....	Error! Bookmark not defined.
Partial regression plots	Error! Bookmark not defined.
Partial residual plots	Error! Bookmark not defined.
Error histograms	Error! Bookmark not defined.
Normal probability-probability (P-P) plots	Error! Bookmark not defined.
Normal quantile-quantile (Q-Q) plots	Error! Bookmark not defined.
Coefficients flagging unusual observations	Error! Bookmark not defined.
Different outlier status definitions	Error! Bookmark not defined.
Leverage values	Error! Bookmark not defined.
Cook's distance	Error! Bookmark not defined.
Mahalanobis distance.....	Error! Bookmark not defined.
DFFITS	Error! Bookmark not defined.
Standardized DFFITS	Error! Bookmark not defined.
DFBETA.....	Error! Bookmark not defined.
Standardized DFBETA	Error! Bookmark not defined.
Covariance ratio (COVRATIO)	Error! Bookmark not defined.
What to do about outliers.....	Error! Bookmark not defined.
Residual analysis in SPSS	Error! Bookmark not defined.
Obtaining residuals-related statistics	Error! Bookmark not defined.
Saving residuals to file	Error! Bookmark not defined.
Listing cases with the largest residuals	Error! Bookmark not defined.
Checking for serial independence	Error! Bookmark not defined.
Checking for homoscedastic error.....	Error! Bookmark not defined.
Checking for nonlinearity	Error! Bookmark not defined.
Checking for normally distributed error.....	Error! Bookmark not defined.
Checking for outliers.....	Error! Bookmark not defined.
Residual analysis in SAS.....	Error! Bookmark not defined.
Obtaining residuals-related statistics	Error! Bookmark not defined.
Saving residuals to file	Error! Bookmark not defined.
Scatterplots of bivariate relationships	Error! Bookmark not defined.
Listing cases with the largest residuals	Error! Bookmark not defined.
Checking for serial independence	Error! Bookmark not defined.
Checking for homoscedastic error.....	Error! Bookmark not defined.
Checking for nonlinearity	Error! Bookmark not defined.
Checking for normally distributed error.....	Error! Bookmark not defined.
Checking for outliers.....	Error! Bookmark not defined.
Other ODS plot options	Error! Bookmark not defined.
Residual analysis in Stata	Error! Bookmark not defined.
Obtaining residuals-related statistics	Error! Bookmark not defined.
Saving residuals to file	Error! Bookmark not defined.
Listing cases with the largest residuals	Error! Bookmark not defined.

Checking for serial independence	Error! Bookmark not defined.
Checking for homoscedastic error.....	Error! Bookmark not defined.
Checking for nonlinearity	Error! Bookmark not defined.
Checking for normally distributed error.....	Error! Bookmark not defined.
Checking for outliers.....	Error! Bookmark not defined.
Additional plotting options in Stata.....	Error! Bookmark not defined.
Multicollinearity.....	Error! Bookmark not defined.
Types of multicollinearity.....	Error! Bookmark not defined.
The correlation matrix.....	Error! Bookmark not defined.
Tolerance.....	Error! Bookmark not defined.
VIF (variance-inflation factor)	Error! Bookmark not defined.
Condition index values and variance proportions	Error! Bookmark not defined.
Checking multicollinearity in SPSS	Error! Bookmark not defined.
VIF and tolerance.....	Error! Bookmark not defined.
Condition index and variance proportions.....	Error! Bookmark not defined.
Checking multicollinearity in SAS.....	Error! Bookmark not defined.
SAS syntax.....	Error! Bookmark not defined.
VIF and tolerance.....	Error! Bookmark not defined.
Condition index and variance proportions.....	Error! Bookmark not defined.
Checking multicollinearity in Stata.....	Error! Bookmark not defined.
VIF and tolerance.....	Error! Bookmark not defined.
Condition index and variance proportions.....	Error! Bookmark not defined.
Assumptions.....	Error! Bookmark not defined.
Proper specification of the model.....	Error! Bookmark not defined.
Spuriousness.....	Error! Bookmark not defined.
Suppression	Error! Bookmark not defined.
Ramsey’s RESET test for misspecification	Error! Bookmark not defined.
Proper specification of the research question.....	Error! Bookmark not defined.
Appropriate modeling of control variables.....	Error! Bookmark not defined.
Population error is assumed to be uncorrelated with each of the independent variables	Error!
Bookmark not defined.	
Non-recursivity.....	Error! Bookmark not defined.
No overfitting	Error! Bookmark not defined.
Absence of perfect multicollinearity.....	Error! Bookmark not defined.
Absence of high partial multicollinearity	Error! Bookmark not defined.
Linearity.....	Error! Bookmark not defined.
Nonlinear transformations.....	Error! Bookmark not defined.
Nonlinear link functions	Error! Bookmark not defined.
Data level.....	Error! Bookmark not defined.
Multivariate normality	Error! Bookmark not defined.
Normally distributed residuals (error)	Error! Bookmark not defined.
Homoscedasticity	Error! Bookmark not defined.
Robust standard errors.....	Error! Bookmark not defined.

Robust regression **Error! Bookmark not defined.**

Outliers **Error! Bookmark not defined.**

Reliability **Error! Bookmark not defined.**

Additivity **Error! Bookmark not defined.**

Independent observations (absence of autocorrelation) **Error! Bookmark not defined.**

 Overview **Error! Bookmark not defined.**

 Graphical test of serial independence **Error! Bookmark not defined.**

 The Durbin-Watson coefficient **Error! Bookmark not defined.**

Mean population error of zero **Error! Bookmark not defined.**

Random sampling **Error! Bookmark not defined.**

Validity **Error! Bookmark not defined.**

Other data requirements **Error! Bookmark not defined.**

Frequently Asked Questions **Error! Bookmark not defined.**

 How do I report regression results? **Error! Bookmark not defined.**

 What is the logic behind the calculation of regression coefficients in multiple regression?
 **Error! Bookmark not defined.**

 How large a sample size do I need to do multiple regression? .. **Error! Bookmark not defined.**

 Can R-squared be interpreted as the percent of the cases explained?**Error! Bookmark not defined.**

 When may ordinal data be used in regression? **Error! Bookmark not defined.**

 When testing for interactions, is there a strategy alternative to adding multiplicative interaction terms to the equation and testing for R² increments?**Error! Bookmark not defined.**

 How do margin plots reveal interaction effects?..... **Error! Bookmark not defined.**

 The regress command **Error! Bookmark not defined.**

 The margins command **Error! Bookmark not defined.**

 The marginsplot command..... **Error! Bookmark not defined.**

 How do I code dummy variables in regression? **Error! Bookmark not defined.**

 What is "attenuation" in the context of regression? **Error! Bookmark not defined.**

 Is multicollinearity only relevant if there are significant findings?**Error! Bookmark not defined.**

 What can be done to handle multicollinearity? **Error! Bookmark not defined.**

 What can be done to handle autocorrelation?..... **Error! Bookmark not defined.**

 How does stepwise multiple regression relate to multicollinearity?**Error! Bookmark not defined.**

 What are forward inclusion and backward elimination in stepwise regression?.....**Error! Bookmark not defined.**

 Should I keep dropping non-significant independent variables one at a time until only significant ones remain in my model? **Error! Bookmark not defined.**

 What are different types of sums of squares used in F tests?.... **Error! Bookmark not defined.**

 Can regression be used in place of Anova for analysis of categorical independents affecting an interval dependent? **Error! Bookmark not defined.**

Does regression analysis require uncorrelated independent variables?**Error! Bookmark not defined.**

How can you test the significance of the difference between two R-squareds?**Error! Bookmark not defined.**

How do I compare b coefficients after I compute a model with the same variables for two subgroups of my sample? **Error! Bookmark not defined.**

How do I compare regression results obtained for one group of subjects to results obtained in another group, assuming the same variables were used in each regression model?....**Error! Bookmark not defined.**

What do I do if I have censored, truncated, or sample-selected data?**Error! Bookmark not defined.**

What do I do if I am measuring the same independent variable at both the individual and group level?..... **Error! Bookmark not defined.**

What is a “relative effects” regression model? **Error! Bookmark not defined.**

How do I test to see what effect a quadratic or other nonlinear term makes in my regression model?..... **Error! Bookmark not defined.**

What is "smoothing" in regression and how does it relate to dealing with nonlinearities in OLS regression?..... **Error! Bookmark not defined.**

What is nonparametric regression for nonlinear relationships? **Error! Bookmark not defined.**

What is Poisson regression?..... **Error! Bookmark not defined.**

SPSS questions **Error! Bookmark not defined.**

 What is the command syntax for linear regression in SPSS?..... **Error! Bookmark not defined.**

 How do I standardize variables in SPSS?..... **Error! Bookmark not defined.**

 How do I create interaction variables in SPSS?..... **Error! Bookmark not defined.**

 All I want is a simple scatterplot with a regression line. Why won't SPSS give it to me?...**Error! Bookmark not defined.**

 What is categorical regression in SPSS?..... **Error! Bookmark not defined.**

Acknowledgments..... **Error! Bookmark not defined.**

Bibliography **Error! Bookmark not defined.**

Multiple Regression

Overview

Multiple regression, a time-honored statistical technique going back to Karl Pearson's use of it in 1908, is employed to account for the variance in a continuous dependent variable, based on linear combinations of interval, dichotomous, or dummy independent variables. Often called OLS regression because of its reliance on ordinary least squares estimation, multiple regression can establish whether a set of independent variables explains a proportion of the variance in a dependent variable (through R^2 , which is percent of variance explained) at a significant level (through a significance test of R^2), and can establish the relative predictive importance of the independent variables (by comparing beta weights, which are standardized regression coefficients).

In multiple regression...

- Power terms can be added as independent variables to explore curvilinear effects.
- Cross-product terms can be added as independent variables to explore interaction effects.
- The researcher can test the significance of the difference between two models, with and without a given predictor variable, to determine if adding an independent variable to the model would improve the model significantly.
- Using hierarchical regression (a.k.a. “block regression”, not to be confused with hierarchical linear regression, which is multilevel regression), the researcher can see how much variance in the dependent variable can be explained by one or a set of new independent variables, over and above that explained by an earlier set.
- Finally, the parameter estimates (the b coefficients and the constant) can be used to construct a prediction equation and generate predicted scores of the dependent variable for further analysis.

A multiple regression equation takes the form $y = b_1x_1 + b_2x_2 + \dots + b_nx_n + c$. The b parameter estimates are the regression coefficients, representing the amount the dependent variable (y) changes when the corresponding independent variable

changes 1 unit, controlling for other variables in the model. The c is the constant, indicating where the regression line intercepts the y axis, representing the average magnitude of the dependent variable when all the independent variables are held to 0. If the data have been centered, this will be controlling all other variables at their means. The standardized version of the b coefficients are the beta weights, and the ratio of the beta weights is often interpreted as the ratio of the relative predictive power of the independent variables. Associated with multiple regression is R^2 (a.k.a multiple correlation), which is the percent of variance in the dependent variable explained collectively by all of the independent variables in the model.

Multiple regression shares all the assumptions of correlation: linearity of relationships, the same level of relationship throughout the range of the dependent variable (homoscedasticity), interval or near-interval measurement level, absence of outliers, and data whose range is not truncated. In addition, it is important that the model being tested is correctly specified. The exclusion of important causal variables or the inclusion of extraneous variables can change markedly the b coefficients and beta weights and hence the interpretation of the importance of the independent variables.

There are many of alternatives to ordinary least squares (OLS) regression. These are treated in separate volumes of the Statistical Associates "Blue Book" series. Some of these topics are listed alphabetically below.

- Cox regression may be used to analyze time-to-event as well as proximity, and preference data. See the separate Statistical Associates "Blue Book" volume on "Cox Regression", listed at the end of this volume.
- Curve estimation lets the researcher explore how linear regression compares to nonlinear models, useful for exploring which statistical procedures and models may be appropriate for relationships in one's data. See the separate Statistical Associates "Blue Book" volume on "Curve Estimation & Nonlinear Regression".
- Discriminant function analysis. is used when the dependent variable is a dichotomy but other assumptions of multiple regression can be met, making it more powerful than the logistic alternatives. See the separate Statistical Associates "Blue Book" volume on "Discriminant Function Analysis".

- The general linear modeling (univariate GLM) is primarily used for analysis of variance designs but also may be used to implement multiple regression. See the separate Statistical Associates “Blue Book” volume on “GLM Univariate, ANOVA, and ANCOVA”.
- The general linear model (multivariate GLM). Multivariate GLM can implement regression models with multiple dependent variables. See the separate Statistical Associates “Blue Book” volume on “GLM Multivariate, MANOVA, and MANCOVA”.
- Generalized linear models (GZLM) is the generalization of linear modeling to a form covering almost any dependent distribution with almost any link function, thus supporting linear regression, Poisson regression, gamma regression, and many others. See the separate Statistical Associates “Blue Book” volume on “Generalized Linear Models: Generalized Estimating Equations”.
- Generalized estimating equations (GEE) is similar to GZLM but handles repeated measures and other non-independent data. See above.
- Linear mixed models (LMM), also called hierarchical linear models, implement regression in the context of multilevel data and linear effects of higher levels on lower levels. See Garson (2013).
- Generalized linear mixed models (GLMM) implement regression for multilevel data, supporting a variety of nonlinear link functions.
- Logistic regression is used for dichotomous and multinomial dependent variables, implemented with stand-alone logistic procedures or through generalized linear models. See the separate Statistical Associates “Blue Book” volume on “Logistic Regression: Binomial & Multinomial”.
- Logit regression is an equivalent to logistic regression, using log-linear techniques to predict one or more categorical dependent variables. See the separate Statistical Associates “Blue Book” volume on “Log-linear Models”.
- Multinomial regression handles research where the dependent variable is categorical. See the separate Statistical Associates “Blue Book” volume on “Logistic Regression: Binomial & Multinomial”.
- Nonlinear regression is used when the model is inherently nonlinear (nonlinearities cannot be dealt with using link functions in generalized linear models or by power and other transformations in general linear models, including regression). See the separate Statistical Associates “Blue Book” volume on “Curve Estimation & Nonlinear Regression”.

- Ordinal regression has more power than multinomial regression when the levels of a categorical dependent variable are ordered. See the separate Statistical Associates “Blue Book” volume on “Ordinal Regression”.
- Partial least squares regression, which merges regression and factor analysis techniques, may be used even with small datasets to predict a set of response variables from a set of independent variables. See the separate Statistical Associates “Blue Book” volume on “Partial Least Squares Regression and Structural Equation Models”.
- Partial proportional odds (PPO) regression is an alternative to ordinal regression when the parallel lines test fails. See the separate Statistical Associates “Blue Book” volume on “Ordinal Regression”.
- Poisson regression is used for count data in survival (event history) analysis and other techniques, implemented with stand-alone general loglinear procedures and through generalized linear models. See above.
- Two-stage least squares regression (2SLS) may be used when one or more independent variables are correlated with the error term, when omitted variables may be treated using instrumental variables, and when the model involves interdependence among predictor variables and thus is recursive. See the separate Statistical Associates “Blue Book” volume on “Two-Stage Least Squares Regression”.
- Weighted least squares (WLS) regression may be used when the OLS regression assumption of homoscedasticity has been violated. See the separate Statistical Associates “Blue Book” volume on “Weighted Least Squares Regression”.

Data examples in this volume

The example datasets used in this volume are listed below in order of use, with versions for SPSS (.sav), SAS (.sas7bdat), and Stata (.dta).

Except where otherwise noted, examples in this volume are from a modified version of the GSS93subset.sav General Social Survey data file originally supplied in the SPSS Samples directory.

- Click [here](#) to download GSS93subset.sav for SPSS.
- Click [here](#) to download GSS93subset.sas7bdat for SAS.
- Click [here](#) to download GSS93subset.dta for Stata.

A section [below](#) on interaction effects uses the “cars” dataset in a version modified from that found in the SPSS Samples directory. Click [here](#) to download cars.sav for SPSS.

- Click [here](#) to download cars.sav for SPSS.
- Click [here](#) to download cars.sas7bdat for SAS.
- Click [here](#) to download cars.dta for Stata.

A section [below](#) on quantile regression uses the “auto” dataset in a version based on that provided by Stata.

- Click [here](#) to download auto.sav for SPSS.
- Click [here](#) to download auto.sas7bdat for SAS.
- Click [here](#) to download auto.dta for Stata.

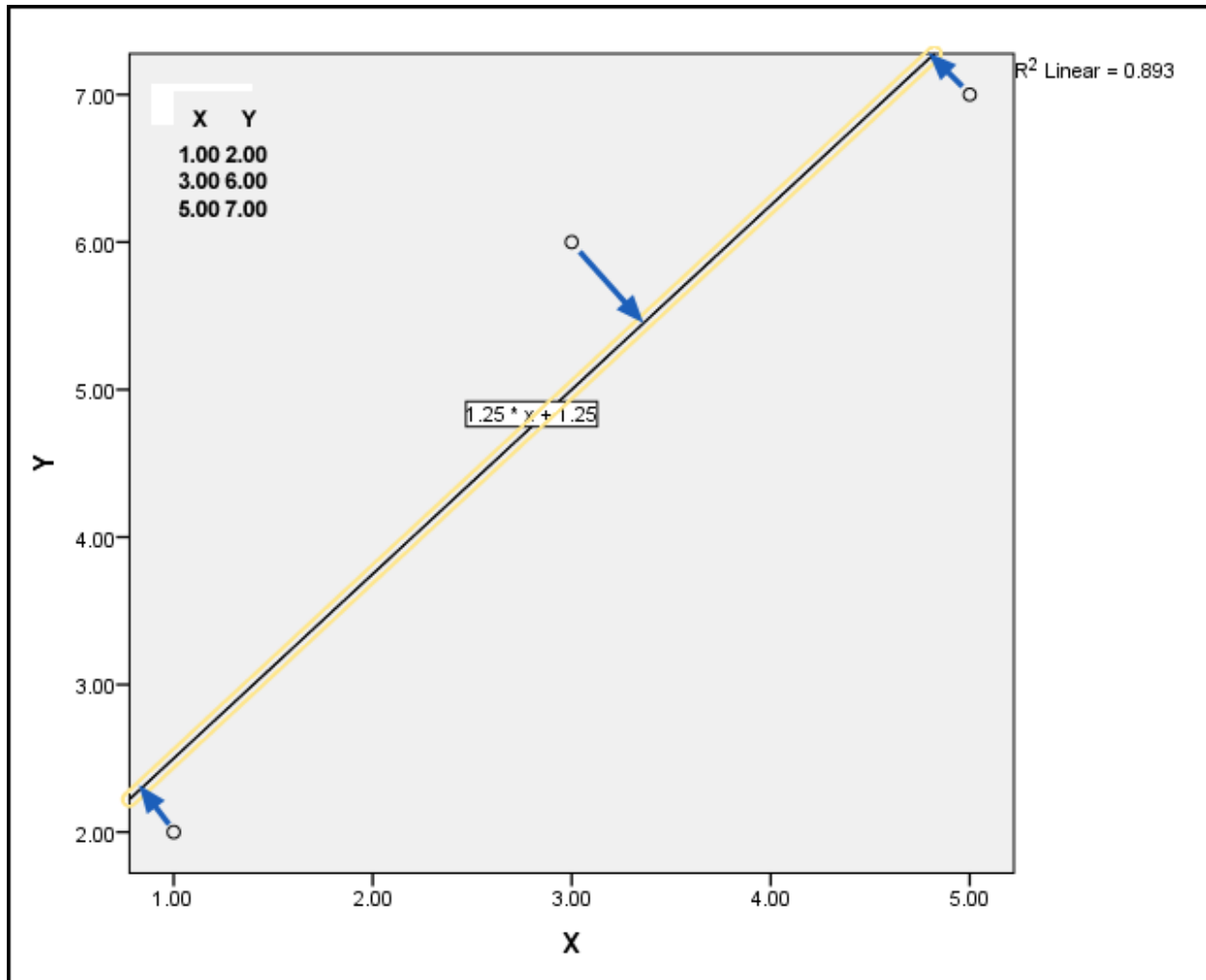
A section [below](#) on difference-in-difference regression uses the “kielmclainsubset” dataset in a version based on that provided by Stata.

- Click [here](#) to download kielmclainsubset.sav for SPSS.
- Click [here](#) to download kielmclainsubset.sas7bdat for SAS.
- Click [here](#) to download kielmclainsubset.dta for Stata.

Key Terms and Concepts

OLS estimation

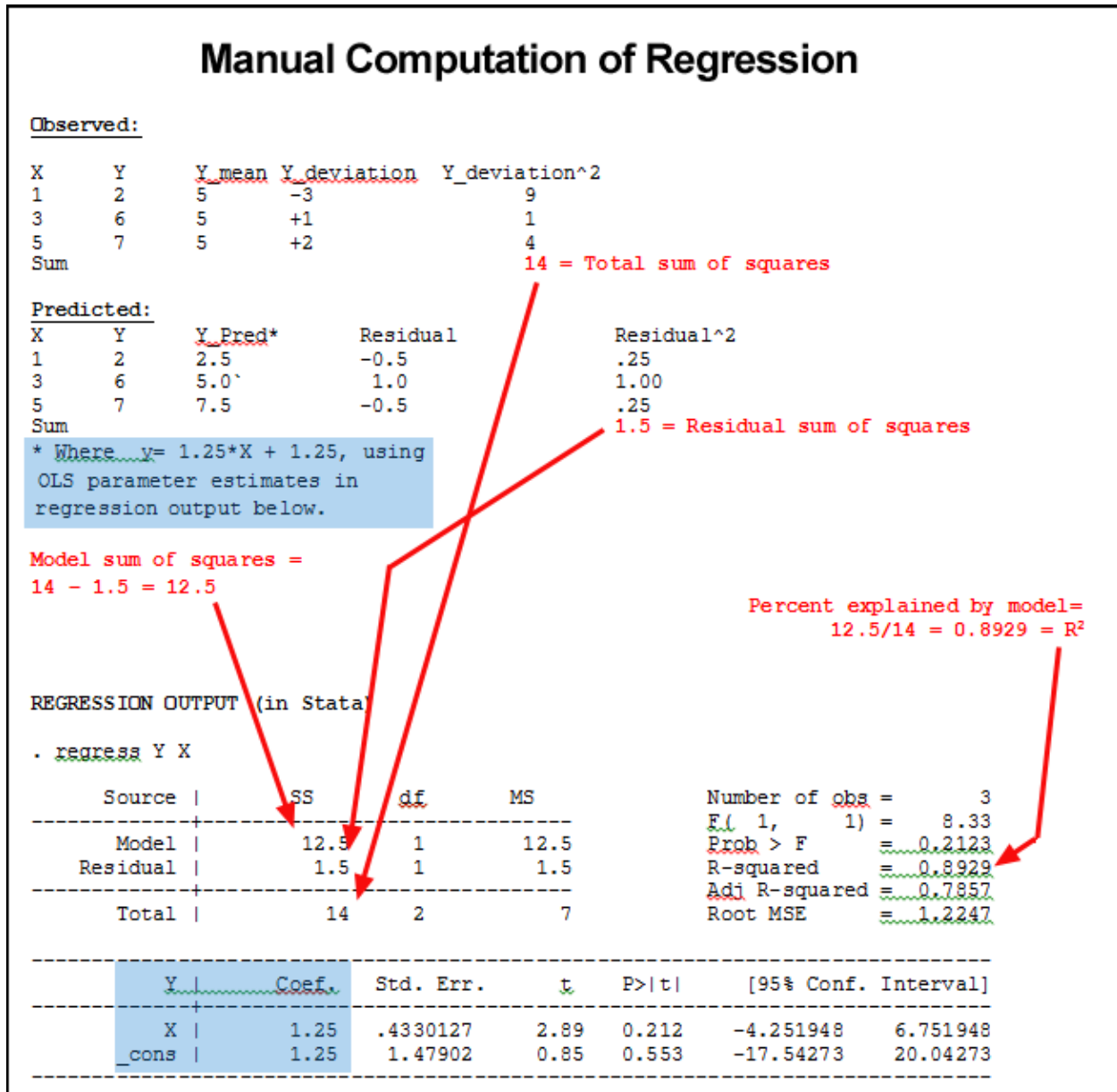
OLS stands for ordinary least squares. The name is derived from the criterion used to draw the best fit regression line: a line such that the sum of the squared deviations of the distances of all observed points to the line is minimized. In the illustration below, Y is regressed on X for three data points. The OLS algorithm creates a prediction equation (here, $Y = 1.25 * X + 1.25$) such that the squared distances to the prediction line are minimized.



The regression equation

The OLS regression equation takes the form $Y = b_1 * x_1 + b_2 * x_2 + c + e$, where Y is the raw dependent, the b's are the regression coefficients for the corresponding x (independent) terms, where c is the constant or intercept, and e is the error term reflected in the residuals. Sometimes this is expressed more simply as $y = b_1 * x_1 + b_2 * x_2 + c$, where y is the estimated dependent variable and c is the constant (which includes the error term). Equations such this, with no interaction effects (see below), are called *main effects models*.

In the figure below, key regression statistics are computed based on the three data points in the figure above and on the OLS-derived equation, $Y = 1.25 * X + 1.25$.



The simple example above illustrates some of the key components of regression:

1. *OLS parameter estimates*: Shown in blue in the figure above, the b coefficient for X is 1.25 and by coincidence the constant is also 1.25. The regression (prediction) equation is therefore $Y = 1.25 * X + 1.25$.
2. *Total sum of squares*: This is the sum of the squared deviations of the Y (dependent) variable from its mean.
3. *Residual sum of squares*: This is the sum of the squared residuals, where residuals are the differences between the observed Y and the predicted Y.

4. *Model sum of squares*: The total sum of squares minus the residual sum of squares is the model sum of squares.
5. R^2 : R-square is the model sum of squares as a percentage of the total sum of squares and is interpreted as the percent of variance in Y explained by the regression model. Here, 89.29% of the variance in Y is explained by X.

Second example. In the figure below, using SPSS output, the highest year of school completed (educ) is predicted from total family income (income91) and age when first married (agewed). Using the parameter estimates shown in blue, the regression equation is:

$$\text{educ} = .29 * \text{income91} + .10 * \text{agewed} + 6.251.$$

If 0 is not within the 95% confidence limits (shown in red below), then the regression coefficient is significant at at least the .05 significance level, as both are for this example.

Model		Coefficients ^a						
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	6.251	.405		15.423	.000	5.456	7.046
	Total Family Income	.290	.015	.491	19.588	.000	.261	.319
	Age When First Married	.100	.016	.161	6.429	.000	.069	.130

a. Dependent Variable: Highest Year of School Completed

Dependent variable

The dependent variable is the predicted variable in the regression equation (educ in the example above). Also called response, outcome, or criterion variables, dependent variables are assumed to be continuous, interval variables, though it is common to see binary or ordinal independent variables in linear regression. Use of binary variables as dependent variables is no longer acceptable since such variables cannot meet regression's normal distribution assumption - logistic regression is commonly used instead. Likewise, use of ordinal variables as dependent variables is now derogated in favor of ordinal regression and proportional odds models (treated in the "Ordinal Regression" title of the Statistical Associates "Blue Book" series).

Independent variables

Independent variables are the predictor variables in the regression equation (income91 and agewed in the example above). Predictor variables usually are continuous variables. It is, however, common to see the use of ordinal predictor variables in linear regression even though this violates the assumptions of OLS regression. Use of ordinal predictor variables with fewer than five levels is particularly derogated. It is acceptable to use binary predictor variables and to transform nominal and ordinal categorical variables into sets of dummy variables (coded 0, 1, and leaving one level out as the reference category to avoid perfect multicollinearity). See the discussion in the "Assumptions" section [below](#).

Dummy variables

Dummy variables are a way of adding the values of a nominal or ordinal variable to a regression equation. The standard approach to modeling categorical variables is to include the categorical variables in the regression equation by converting each level of each categorical variable into a variable of its own, usually coded 0 or 1. For instance, the categorical variable "region" may be converted into dummy variables such as "East," "West," "North," or "South." Typically "1" means the attribute of interest is present (ex., South = 1 means the case is from the region South).

Once a set of dummy variables is created, if we know an observation's value on all the levels of a categorical variable except one, that last one is determined. We have to leave one of the levels out of the regression model to avoid perfect multicollinearity (a.k.a. singularity or redundancy), which will prevent a solution. For example, we may leave out "West" to avoid singularity. The reference level should be the level of greatest interest or at least a level with known characteristics. Letting a residual category (e.g., "Other") be the reference category undermines the meaningfulness of reference statements.

The omitted category is the reference category because b coefficients must be interpreted with reference to it. In the example output below, three regions (North, South, Midwest, where West is the left-out reference category) are added as dummy predictors. Since West has the highest mean score on the dependent (highest year of school completed), all the b coefficients are negative, meaning that if a subject is in any other region than West, a lower value is predicted. Only

being in the South is significantly lower than being from the West (the reference category), however. Being in the South predicts 0.577 fewer years of education than being in the West, controlling for other variables in the model. See further discussion [below](#).

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	6.434	.412		15.616	.000	5.625	7.242
	Total Family Income	.288	.015	.488	19.374	.000	.259	.318
	Age When First Married	.100	.016	.161	6.391	.000	.069	.130
	North	-.324	.285	-.029	-1.136	.256	-.884	.236
	South	-.577	.213	-.070	-2.708	.007	-.995	-.159
	Midwest	-.135	.228	-.015	-.593	.553	-.583	.313

a. Dependent Variable: Highest Year of School Completed

Note that dummy variables are not used as dependent variables in a normal regression model. See further discussion in the FAQ section [below](#), including alternative methods of coding categorical dummy variables.

Interaction effects

Interactions

Interaction effects are sometimes called *moderator effects* because the interacting third variable which changes the relation between two original variables moderates the original relationship. However, an interaction effect is a special type of moderator effect which has an impact going beyond simple main effects of individual variables. For instance, the impact of being African American or of being a woman may be significant, but there may be an additional impact of being an African American woman over and above the main effects of race and gender.

Interaction terms may be added to the model to incorporate the joint effect of two variables (ex., race and gender) on a dependent variable (ex., income) over and above their separate effects. The researcher adds interaction terms to the model as crossproducts of the standardized independent variables and/or dummy independent variables, typically placing them after the simple "main effects" independent variables. In the example below, the interaction term "agedwed_by_income91" is added to the previously discussed regression model. It

is found to be significant at the .04 level, meaning that there is a combined effect of age when first wed with family income over and beyond that of either aged or income individually.

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	6.361	.413		15.404	.000	5.551	7.171
	Total Family Income	.289	.015	.490	19.443	.000	.260	.318
	Age When First Married	.102	.016	.164	6.523	.000	.071	.132
	North	-.346	.285	-.031	-1.215	.225	-.905	.213
	South	-.594	.213	-.072	-2.791	.005	-1.012	-.176
	Midwest	-.120	.228	-.014	-.526	.599	-.568	.327
	agedwed_by_income91	.148	.072	.051	2.054	.040	.007	.289

a. Dependent Variable: Highest Year of School Completed

Margin plots, discussed [below](#) in the FAQ section, can reveal interaction effects graphically, where they appear in the form on non-parallel lines which may cross.

For further reading, a leading reference for interaction effects in regression is Aiken & West (1991).

Centering

Crossproduct interaction terms may be highly correlated (multicollinear) with the corresponding simple independent variables in the regression equation, creating problems with assessing the relative importance of main effects and interaction effects. Centering the predictor variables (or standardizing, which incorporates centering) tends to reduce multicollinearity and is therefore recommended. Note also that there are alternatives to the crossproduct approach to analyzing interactions (see discussion [below](#)). See also general discussion of centering [below](#).

Significance of interaction effects

The significance of an interaction effect is the same as for any other variable, except in the case of a set of dummy variables representing a single categorical variable. When an ordinal variable has been entered as a set of dummy variables, the interaction of another variable with the ordinal variable will involve multiple interaction terms. In this case the [F-test](#) of the significance of the interaction of the two variables is the significance of the change of R-square of the equation

with the interaction terms and the equation without the set of terms associated with the categorical variable.

Interaction terms with categorical dummies

To create an interaction term between a categorical variable and a continuous variable, first the categorical variable is dummy-coded, creating $(k - 1)$ new variables, one for each level of the categorical variable except the omitted reference category. The continuous variable is multiplied by each of the $(k - 1)$ dummy variables. The terms entered into the regression include the continuous variable, the $(k - 1)$ dummy variables, and the $(k - 1)$ cross-product interaction terms. Also, a regression is run without the interaction terms. The R-squared difference measures the effect of the interaction. The beta weights for the interaction terms in the regression which includes the interaction terms measure the relative predictive power of the effects of the continuous variable given specific levels of the categorical variable. It is also possible to include both interaction terms and power terms in the model by multiplying the dummies by the square of the continuous variable, but this can lead to an excessive number of terms in the model. One approach to dealing with this is to use stepwise regression's F test as a criterion to stop adding terms to the model.

Plotting interactions through simple slope analysis

Simple slope analysis serves as an exploration tool for understanding interaction effects more deeply. This subsection uses the “cars” dataset (see [above](#)), containing attributes of U.S. and foreign cars. Consider the table below, in which miles per gallon (MPG) is predicted from engine displacement in cubic inches (engine), from horsepower, and the interaction of the two (enginebyhorse). Both main effects and the interaction effect are all significant at better than the .001 level.

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	53.245	1.589		33.515	.000
	Engine Displacement (cu. inches)	-.095	.007	-1.272	-13.699	.000
	Horsepower	-.244	.020	-1.195	-11.972	.000
	enginebyhorse	.001	.000	1.626	10.879	.000

a. Dependent Variable: Miles per Gallon

We use the following notation:

- DV miles per gallon (MPG), the dependent variable
- X₁ engine
- B₁ regression coefficient for engine = -.095
- X₂ horsepower
- B₂ regression coefficient for horsepower = -.244
- X₁X₂ enginebyhorse, which is the engine*horsepower interaction
- B₃ the regression coefficient for the interaction effect = .001
- C the constant = 53.245
- Y' the estimated value of MPG

The regression equation is:

$$(1) Y' = C + B_1X_1 + B_2X_2 + B_3(X_1X_2)$$

When there is an interaction effect, such as one affecting the relationship of engine size (X1) and MPG (the DV), then the engine size slope (b1) will be different for different values of the interacting variable, here horsepower (X2). We can test this by seeing whether B1 at high levels of X2 differs from B1 at low levels of X2.

While any pair of values of X2 might be spotlighted (hence simple slope analysis may be called a form of “spotlight analysis), Cohen et al. (2003) state that plus or minus one standard deviation of the mean may be used as high and low values, absent any other theoretical basis for selecting particular high and low values. For X2 = horsepower, the mean is 104.93; the standard deviation is 38.522; plus 1 sd is 143.452; and -1 sd is 66.408.

To better obtain the regression formula for Y' at high and low values of X_2 , the regression formula above may be transformed by simple algebra as follows:

$$(2) Y' = C + B_2X_2 + B_1X_1 + B_3(X_1X_2)$$

$$(3) Y' = (C + B_2X_2) + (B_1 + B_3X_2) * X_1$$

Using formula (3) above, we may compute the regression formulas for Y' at high and low values of X_2 .

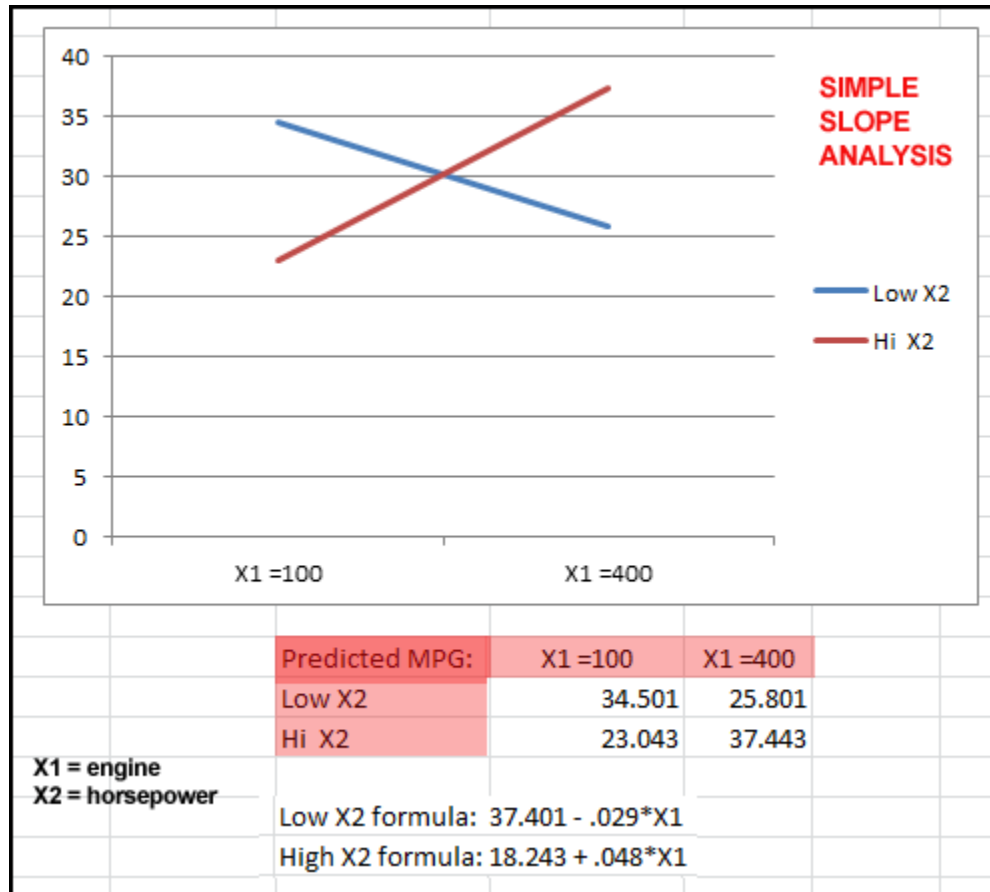
$$\begin{aligned} \text{High: } Y' &= (53.245 + (-.244)(143.452)) + (-.095 + .001(143.452)) * X_1 \\ &= (53.245 - 35.002) + (-.095 + .143) * X_1 \end{aligned}$$

$$(4) \quad = 18.243 + .048 * X_1$$

$$\begin{aligned} \text{Low: } Y' &= (53.245 + (-.244)(66.408)) + (-.095 + .001(66.408)) * X_1 \\ &= (53.245 - 16.204) + (-.095 + .066) * X_1 \end{aligned}$$

$$(5) \quad = 37.401 - .029 * X_1$$

We immediately see that when X_2 (horsepower) is high, the slope of X_1 (engine displacement) is $+.048$, but when X_2 is low, the slope of X_1 is $-.029$. That the slope of engine displacement is different at different levels of horsepower shows an interaction effect involving horsepower. This is shown graphically if equations (4) and (5) are plotted. S below (using Microsoft Excel), where the crossed slopes indicate an interaction effect. This is confirmed in the regression table [above](#), where the enginebyhorse interaction is significant.



For more on simple slope analysis, see Aiken & West (1991) and Tabachnick & Fidell (2013:158-161).

Separate regressions

An alternative but little-used approach to interactions is to run separate regressions for each level of the interacting variable. There are, however, objections to this approach. While separate regressions are feasible for a single variable such as gender, running a male and a female regression, it is not the best approach for two reasons. In practical terms, if there are multiple categorical variables each with multiple categories (levels), the number of needed regressions may become unwieldy and the sample size for individual regressions may become too small. Smaller sample size means the researcher will lose statistical power compared to one overall regression. Loss of statistical power means the researcher will be more likely to make Type II errors (false negatives, thinking there is no relationship when in reality there is).

Predicted values

Predicted values, also called fitted values, are the values of each case based on using the regression equation for all cases selected for analysis. SPSS, SAS, and Stata all support generating fitted values and saving these as new variables to the researcher's working dataset, which can then be saved to file with the command File > Save As.

SPSS

Saving estimates is accomplished in SPSS under the "Output" button. SPSS will save the predicted values under the term PRED to refer to predicted values and ZPRED to refer to standardized predicted values. Click the Save button in SPSS to add and save these as new variables in your dataset.

SAS

In SAS, saving is accomplished by defining a library and using the OUTPUT statement, as in the syntax below:

```
LIBNAME in "C:\Data";
```

PROC REG syntax supports an OUTPUT statement. The syntax below creates a file called mydata.sas7bdat in the C:\Data folder. This file contains all the variables from the original dataset plus new variables: (1) "fitted", containing the predicted values; and (2) "resid", containing residuals for each observation. The optional PROC PRINT statement simply prints out the first 10 records for verification of what has been saved to the working dataset in memory.

```
OUTPUT
  OUT=in.mydata
  PREDICTED=fitted
  RESIDUAL=resid;
PROC PRINT DATA=in.mydata (OBS=10);
```

Note that if no library name is declared and used (here, "in" is the library defining the file folder), then the output datafile (here, mydata, without the "in" prefix) will be saved to memory only and not to file.

SAS supports many other options in addition to PREDICTED under the “Output” statement:

Keyword	Description
COOKD= <i>names</i>	Cook’s <i>D</i> influence statistic
COVRATIO= <i>names</i>	standard influence of observation on covariance of betas
DFFITs= <i>names</i>	standard influence of observation on predicted value
H= <i>names</i>	leverage, $x_i(\mathbf{X}'\mathbf{X})^{-1}x_i'$
LCL= <i>names</i>	lower bound of a $100(1 - \alpha)\%$ confidence interval for an individual prediction. This includes the variance of the error, as well as the variance of the parameter estimates.
LCLM= <i>names</i>	lower bound of a $100(1 - \alpha)\%$ confidence interval for the expected value (mean) of the dependent variable
PREDICTED P= <i>names</i>	predicted values
PRESS= <i>names</i>	<i>i</i> th residual divided by $(1 - h)$, where <i>h</i> is the leverage, and where the model has been refit without the <i>i</i> th observation
RESIDUAL R= <i>names</i>	residuals, calculated as ACTUAL minus PREDICTED
RSTUDENT= <i>names</i>	a studentized residual with the current observation deleted
STDI= <i>names</i>	standard error of the individual predicted value
STDP= <i>names</i>	standard error of the mean predicted value
STDR= <i>names</i>	standard error of the residual
STUDENT= <i>names</i>	studentized residuals, which are the residuals divided by their standard errors
UCL= <i>names</i>	upper bound of a $100(1 - \alpha)\%$ confidence interval for an individual prediction
UCLM= <i>names</i>	upper bound of a $100(1 - \alpha)\%$ confidence interval for the expected value (mean) of the dependent variable

Stata

Following a successful `regress` command, the Stata postestimation `predict` command can be used to create new variables containing the fitted values (here labeled “fitted”) and their standard errors (here labeled “sterror”).

```
predict fitted
predict sterror, stdp
```

After the desired variables are created, from the Stata menu system select File > Save As, to save to file the data with the new variables added.

END OF PREVIEW OF FIRST 30 PAGES OF 410 PAGES

To buy the Kindle version for \$9.99, click [here](#).

To buy the entire Statistical Associates “Regression Models” library of 10 statistics books in no-password pdf format on DVD plus one year of free updates for \$50, click [here](#).

To buy the entire Statistical Associates library of 50 statistics books in no-password pdf format on DVD plus one year of free updates for \$120, click [here](#).

To register for a password-protected pdf version when available, go to <http://www.statisticalassociates.com> .

Copyright 1998, 2008, 2010, 2012, 2013, 2014 by G. David Garson and Statistical Associates Publishers. Do not copy, distribute, or post on any medium. Last update: 1 Oct. 2014.

Statistical Associates Publishing

Blue Book Series

NEW! For use by a single individual, our entire current library is available at Amazon in no-password pdf format on DVD for \$120 plus shipping. Click on <http://www.amazon.com/dp/1626380201> . Includes one year of free updates when email address is provided.

NEW! For use by a single individual, our "Regression Models" library of 10 titles is available at Amazon in no-password pdf format on DVD for \$50 plus shipping. Click on <http://www.amazon.com/dp/1626380252>

NEW! For use by a single individual, our "Qualitative Methods" library of 10 titles is available at Amazon in no-password pdf format on DVD for \$50 plus shipping. Click on <http://www.amazon.com/dp/B00JJ2JZYM>

NEW FOR CLASS USE! If you are requesting this for class use, consider recommending site licensing so the ebook is free for everyone at your institution and is always available. For class use, see our new low-cost site license policy for university libraries and others at <http://statisticalassociates.com/FAQ.htm#sales> . Site

Association, Measures of
Canonical Correlation
Case Studies
Cluster Analysis
Content Analysis
Correlation
Correlation, Partial
Correspondence Analysis
Cox Regression
Creating Simulated Datasets
Crosstabulation
Curve Estimation & Nonlinear Regression
Delphi Method in Quantitative Research
Discriminant Function Analysis
Ethnographic Research
Evaluation Research
Factor Analysis
Focus Group Research
Game Theory
Generalized Linear Models/Generalized Estimating Equations
GLM (Multivariate), MANOVA, and MANCOVA
GLM (Univariate), ANOVA, and ANCOVA
Grounded Theory
Life Tables & Kaplan-Meier Survival Analysis
Literature Review in Research and Dissertation Writing
Logistic Regression: Binary & Multinomial

Log-linear Models,
Longitudinal Analysis
Missing Values & Data Imputation
Multidimensional Scaling
Multiple Regression
Narrative Analysis
Network Analysis
Neural Network Models
Nonlinear Regression
Ordinal Regression
Parametric Survival Analysis
Partial Correlation
Partial Least Squares Regression
Participant Observation
Path Analysis
Power Analysis
Probability
Probit and Logit Response Models
Research Design
Scales and Measures
Significance Testing
Social Science Theory in Research and Dissertation Writing
Structural Equation Modeling
Survey Research & Sampling
Testing Statistical Assumptions
Two-Stage Least Squares Regression
Validity & Reliability
Variance Components Analysis
Weighted Least Squares Regression

Statistical Associates Publishing
<http://www.statisticalassociates.com>
sa.publishers@gmail.com