

Interpreting and Visualizing Regression Models Using Stata

Second Edition

Michael N. Mitchell



STATA *Press*

A Stata Press Publication
StataCorp LLC
College Station, Texas



Copyright © 2012, 2021 by StataCorp LLC
All rights reserved. First edition 2012
Second edition 2021

Published by Stata Press, 4905 Lakeway Drive, College Station, Texas 77845

Typeset in L^AT_EX 2_ε

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Print ISBN-10: 1-59718-321-0

Print ISBN-13: 978-1-59718-321-5

ePub ISBN-10: 1-59718-322-9

ePub ISBN-13: 978-1-59718-322-2

Mobi ISBN-10: 1-59718-323-7

Mobi ISBN-13: 978-1-59718-323-9

Library of Congress Control Number: 2020950108

No part of this book may be reproduced, stored in a retrieval system, or transcribed, in any form or by any means—electronic, mechanical, photocopy, recording, or otherwise—without the prior written permission of StataCorp LLC.

Stata, **STATA**, Stata Press, Mata, **mata**, and NetCourse are registered trademarks of StataCorp LLC.

Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations.

NetCourseNow is a trademark of StataCorp LLC.

L^AT_EX 2_ε is a trademark of the American Mathematical Society.

Contents

	List of tables	xv
	List of figures	xvii
	Preface to the Second Edition	xxvii
	Preface to the First Edition	xxix
	Acknowledgments	xxxix
1	Introduction	1
1.1	Read me first	1
1.2	The GSS dataset	4
1.2.1	Income	5
1.2.2	Age	6
1.2.3	Education	10
1.2.4	Gender	12
1.3	The pain datasets	12
1.4	The optimism datasets	13
1.5	The school datasets	13
1.6	The sleep datasets	13
1.7	Overview of the book	13
I	Continuous predictors	17
2	Continuous predictors: Linear	19
2.1	Chapter overview	19
2.2	Simple linear regression	19
2.2.1	Computing predicted means using the margins command . .	22
2.2.2	Graphing predicted means using the marginsplot command	24
2.3	Multiple regression	27

2.3.1	Computing adjusted means using the margins command . . .	28
2.3.2	Some technical details about adjusted means	30
2.3.3	Graphing adjusted means using the marginsplot command .	31
2.4	Checking for nonlinearity graphically	32
2.4.1	Using scatterplots to check for nonlinearity	33
2.4.2	Checking for nonlinearity using residuals	33
2.4.3	Checking for nonlinearity using locally weighted smoother .	35
2.4.4	Graphing outcome mean at each level of predictor	36
2.4.5	Summary	39
2.5	Checking for nonlinearity analytically	39
2.5.1	Adding power terms	40
2.5.2	Using factor variables	42
2.6	Summary	46
3	Continuous predictors: Polynomials	49
3.1	Chapter overview	49
3.2	Quadratic (squared) terms	49
3.2.1	Overview	49
3.2.2	Examples	53
3.3	Cubic (third power) terms	59
3.3.1	Overview	59
3.3.2	Examples	60
3.4	Fractional polynomial regression	66
3.4.1	Overview	66
3.4.2	Example using fractional polynomial regression	70
3.5	Main effects with polynomial terms	79
3.6	Summary	81
4	Continuous predictors: Piecewise models	83
4.1	Chapter overview	83
4.2	Introduction to piecewise regression models	84
4.3	Piecewise with one known knot	86

<i>Contents</i>	vii
4.3.1	Overview 86
4.3.2	Examples using the GSS 87
4.4	Piecewise with two known knots 95
4.4.1	Overview 95
4.4.2	Examples using the GSS 95
4.5	Piecewise with one knot and one jump 100
4.5.1	Overview 100
4.5.2	Examples using the GSS 101
4.6	Piecewise with two knots and two jumps 106
4.6.1	Overview 106
4.6.2	Examples using the GSS 106
4.7	Piecewise with an unknown knot 113
4.8	Piecewise model with multiple unknown knots 117
4.9	Piecewise models and the marginsplot command 124
4.10	Automating graphs of piecewise models 127
4.11	Summary 130
5	Continuous by continuous interactions 131
5.1	Chapter overview 131
5.2	Linear by linear interactions 131
5.2.1	Overview 131
5.2.2	Example using GSS data 136
5.2.3	Interpreting the interaction in terms of age 137
5.2.4	Interpreting the interaction in terms of education 139
5.2.5	Interpreting the interaction in terms of age slope 141
5.2.6	Interpreting the interaction in terms of the educ slope . . . 142
5.3	Linear by quadratic interactions 144
5.3.1	Overview 144
5.3.2	Example using GSS data 147
5.4	Summary 152

6	Continuous by continuous by continuous interactions	153
6.1	Chapter overview	153
6.2	Overview	153
6.3	Examples using the GSS data	158
6.3.1	A model without a three-way interaction	158
6.3.2	A three-way interaction model	162
6.4	Summary	169
II	Categorical predictors	171
7	Categorical predictors	173
7.1	Chapter overview	173
7.2	Comparing two groups using a t test	174
7.3	More groups and more predictors	175
7.4	Overview of contrast operators	181
7.5	Compare each group against a reference group	182
7.5.1	Selecting a specific contrast	183
7.5.2	Selecting a different reference group	184
7.5.3	Selecting a contrast and reference group	185
7.6	Compare each group against the grand mean	185
7.6.1	Selecting a specific contrast	187
7.7	Compare adjacent means	188
7.7.1	Reverse adjacent contrasts	192
7.7.2	Selecting a specific contrast	193
7.8	Comparing the mean of subsequent or previous levels	194
7.8.1	Comparing the mean of previous levels	198
7.8.2	Selecting a specific contrast	199
7.9	Polynomial contrasts	200
7.10	Custom contrasts	202
7.11	Weighted contrasts	206
7.12	Pairwise comparisons	208

7.13	Interpreting confidence intervals	210
7.14	Testing categorical variables using regression	212
7.15	Summary	215
8	Categorical by categorical interactions	217
8.1	Chapter overview	217
8.2	Two by two models: Example 1	219
8.2.1	Simple effects	224
8.2.2	Estimating the size of the interaction	225
8.2.3	More about interaction	226
8.2.4	Summary	227
8.3	Two by three models	227
8.3.1	Example 2	227
8.3.2	Example 3	232
8.3.3	Summary	237
8.4	Three by three models: Example 4	237
8.4.1	Simple effects	240
8.4.2	Simple contrasts	240
8.4.3	Partial interaction	242
8.4.4	Interaction contrasts	243
8.4.5	Summary	245
8.5	Unbalanced designs	245
8.6	Main effects with interactions: anova versus regress	250
8.7	Interpreting confidence intervals	253
8.8	Summary	255
9	Categorical by categorical by categorical interactions	257
9.1	Chapter overview	257
9.2	Two by two by two models	258
9.2.1	Simple interactions by season	260
9.2.2	Simple interactions by depression status	261
9.2.3	Simple effects	263

9.3	Two by two by three models	263
9.3.1	Simple interactions by depression status	266
9.3.2	Simple partial interaction by depression status	266
9.3.3	Simple contrasts	268
9.3.4	Partial interactions	268
9.4	Three by three by three models and beyond	270
9.4.1	Partial interactions and interaction contrasts	272
9.4.2	Simple interactions	276
9.4.3	Simple effects and simple comparisons	279
9.5	Summary	280
III	Continuous and categorical predictors	281
10	Linear by categorical interactions	283
10.1	Chapter overview	283
10.2	Linear and two-level categorical: No interaction	283
10.2.1	Overview	283
10.2.2	Examples using the GSS	286
10.3	Linear by two-level categorical interactions	291
10.3.1	Overview	291
10.3.2	Examples using the GSS	294
10.4	Linear by three-level categorical interactions	299
10.4.1	Overview	299
10.4.2	Examples using the GSS	301
10.5	Summary	308
11	Polynomial by categorical interactions	311
11.1	Chapter overview	311
11.2	Quadratic by categorical interactions	311
11.2.1	Overview	312
11.2.2	Quadratic by two-level categorical	315
11.2.3	Quadratic by three-level categorical	323

11.3	Cubic by categorical interactions	329
11.4	Summary	334
12	Piecewise by categorical interactions	335
12.1	Chapter overview	335
12.2	One knot and one jump	338
12.2.1	Comparing slopes across gender	342
12.2.2	Comparing slopes across education	343
12.2.3	Difference in differences of slopes	343
12.2.4	Comparing changes in intercepts	344
12.2.5	Computing and comparing adjusted means	344
12.2.6	Graphing adjusted means	347
12.3	Two knots and two jumps	351
12.3.1	Comparing slopes across gender	356
12.3.2	Comparing slopes across education	357
12.3.3	Difference in differences of slopes	358
12.3.4	Comparing changes in intercepts by gender	359
12.3.5	Comparing changes in intercepts by education	360
12.3.6	Computing and comparing adjusted means	361
12.3.7	Graphing adjusted means	364
12.4	Comparing coding schemes	366
12.4.1	Coding scheme #1	367
12.4.2	Coding scheme #2	368
12.4.3	Coding scheme #3	370
12.4.4	Coding scheme #4	372
12.4.5	Choosing coding schemes	373
12.5	Summary	374
13	Continuous by continuous by categorical interactions	375
13.1	Chapter overview	375
13.2	Linear by linear by categorical interactions	376
13.2.1	Fitting separate models for males and females	376

13.2.2	Fitting a combined model for males and females	378
13.2.3	Interpreting the interaction focusing in the age slope	380
13.2.4	Interpreting the interaction focusing on the educ slope	382
13.2.5	Estimating and comparing adjusted means by gender	384
13.3	Linear by quadratic by categorical interactions	386
13.3.1	Fitting separate models for males and females	386
13.3.2	Fitting a common model for males and females	388
13.3.3	Interpreting the interaction	389
13.3.4	Estimating and comparing adjusted means by gender	390
13.4	Summary	392
14	Continuous by categorical by categorical interactions	393
14.1	Chapter overview	393
14.2	Simple effects of gender on the age slope	398
14.3	Simple effects of education on the age slope	399
14.4	Simple contrasts on education for the age slope	400
14.5	Partial interaction on education for the age slope	400
14.6	Summary	401
IV	Beyond ordinary linear regression	403
15	Multilevel models	405
15.1	Chapter overview	405
15.2	Example 1: Continuous by continuous interaction	406
15.3	Example 2: Continuous by categorical interaction	409
15.4	Example 3: Categorical by continuous interaction	413
15.5	Example 4: Categorical by categorical interaction	417
15.6	Summary	421
16	Time as a continuous predictor	423
16.1	Chapter overview	423
16.2	Example 1: Linear effect of time	424
16.3	Example 2: Linear effect of time by a categorical predictor	428

16.4	Example 3: Piecewise modeling of time	433
16.5	Example 4: Piecewise effects of time by a categorical predictor	438
16.5.1	Baseline slopes	443
16.5.2	Change in slopes: Treatment versus baseline	444
16.5.3	Jump at treatment	445
16.5.4	Comparisons among groups	446
16.6	Summary	448
17	Time as a categorical predictor	449
17.1	Chapter overview	449
17.2	Example 1: Time treated as a categorical variable	450
17.3	Example 2: Time (categorical) by two groups	455
17.4	Example 3: Time (categorical) by three groups	459
17.5	Comparing models with different residual covariance structures	464
17.6	Analyses with small samples	466
17.7	Summary	474
18	Nonlinear models	475
18.1	Chapter overview	475
18.2	Binary logistic regression	476
18.2.1	A logistic model with one categorical predictor	476
18.2.2	A logistic model with one continuous predictor	484
18.2.3	A logistic model with covariates	486
18.3	Multinomial logistic regression	491
18.4	Ordinal logistic regression	497
18.5	Poisson regression	500
18.6	More applications of nonlinear models	503
18.6.1	Categorical by categorical interaction	503
18.6.2	Categorical by continuous interaction	510
18.6.3	Piecewise modeling	516
18.7	Summary	522
19	Complex survey data	523

V	Appendices	529
A	Customizing output from estimation commands	531
A.1	Omission of output	531
A.2	Specifying the confidence level	533
A.3	Customizing the formatting of columns in the coefficient table	534
A.4	Customizing the display of factor variables	536
B	The margins command	545
B.1	The predict() and expression() options	545
B.2	The at() option	548
B.3	Margins with factor variables	551
B.4	Margins with factor variables and the at() option	557
B.5	The dydx() and related options	559
B.6	Specifying the confidence level	563
B.7	Customizing column formatting	564
C	The marginsplot command	567
D	The contrast command	583
D.1	Inclusion and omission of output	584
D.2	Customizing the display of factor variables	586
D.3	Adjustments for multiple comparisons	588
D.4	Specifying the confidence level	588
D.5	Customizing column formatting	589
E	The pwcompare command	591
	References	597
	Author index	601
	Subject index	603

Tables

5.1	The age slope at each level of education	135
5.2	The educ slope at 25, 35, 45, and 55 years of age	136
7.1	Summary of contrast operators	182
12.1	Summary of piecewise regression results with one knot	341
12.2	Summary of piecewise regression results with two knots	355
12.3	Summary of regression results and meaning of coefficients for coding schemes #1 and #2	370
12.4	Summary of regression results and meaning of coefficients for coding schemes #3 and #4	373
14.1	The age slope by level of education and gender	394
14.2	The age slope by level of education and gender	398

(Pages omitted)

Preface to the Second Edition

It was back in March of 2012 that I penned the preface for the first edition of this book. That was over eight years and four Stata versions ago (using Stata 12.1). The techniques illustrated in this book are as relevant today as they were back in 2012. Over this time, Stata has grown considerably. A key change that impacts the interpretation of statistical results (a focus of this book) is that the levels of factor variables are now labeled using value labels (instead of group numbers). For example, a two-level version of marital status might be labeled as `Married` and `Unmarried` instead of using numeric values such as 1 and 2. All the output in this new edition capitalizes on this feature, emphasizing the interpretation of results based on variables labeled using intuitive value labels. Stata now includes features that allow you to customize output in ways that increase the clarity of the results, aiding interpretation. This new edition includes a new appendix (appendix A) that illustrates how you can customize the output of estimation commands for maximum clarity.

The `margins`, `contrast`, and `pwcompare` commands also reflect this new output style, defaulting to labeling groups according to their value labels. Results of these commands are easier to interpret than ever. For instance, a contrast regarding marital status might be labeled as `widowed vs. married`, making it very clear which groups are being compared. This new edition uses this labeling style and also includes appendices that describe how to customize such output. Appendix B is on the `margins` command, appendix D is on the `contrast` command, and appendix E is on the `pwcompare` command—each illustrate how you can customize the display of output produced by these commands. Additionally, appendix C on the `marginsplot` command illustrates new graphical features that have been recently introduced, including using transparency to more clearly visualize overlapping confidence intervals.

Among the other new features introduced since the last edition of this book, the `mixed` and `contrast` commands now include options for computing estimates for small-sample sizes. Chapter 17 describes these techniques and illustrates how the `mixed` and `contrast` commands can use small-sample size methods to analyze a longitudinal dataset with a small-sample size.

As with the first edition, I hope the examples shown in this book help you understand the results of your regression models so you can interpret and present them with clarity and confidence.

*Ventura, California
November 2020*

Michael N. Mitchell

Preface to the First Edition

Think back to the first time you learned about simple linear regression. You probably learned about the underlying theory of linear regression, the meaning of the regression coefficients, and how to create a graph of the regression line. The graph of the regression line provided a visual representation of the intercept and slope coefficients. Using such a graph, you could see that as the intercept increased, so did the overall height of the regression line, and as the slope increased, so did the tilt of the regression line. Within Stata, the `graph twoway lfit` command can be used to easily visualize the results of a simple linear regression.

Over time, we learn about and use fancier and more abstract regression models—models that include covariates, polynomial terms, piecewise terms, categorical predictors, interactions, and nonlinear models such as logistic. Compared with a simple linear regression model, it can be challenging to visualize the results of such models. The utility of these fancier models diminishes if we have greater difficulty interpreting and visualizing the results.

With the introduction of the `marginsplot` command in Stata 12, visualizing the results of a regression model, even complex models, is a snap. As implied by the name, the `marginsplot` command works in tandem with the `margins` command by plotting (graphing) the results computed by the `margins` command. For example, after fitting a linear model, the `margins` command can be used to compute adjusted means as a function of one or more predictors. The `marginsplot` command graphs the adjusted means, allowing you to visually interpret the results.

The `margins` and `marginsplot` commands can be used following nearly all Stata estimation commands (including `regress`, `anova`, `logit`, `ologit`, and `mlogit`). Furthermore, these commands work with continuous linear predictors, categorical predictors, polynomial (power) terms, as well as interactions (for example, two-way interactions, three-way interactions). This book uses the `marginsplot` command not only as an interpretive tool but also as an instructive tool to help you understand the results of regression models by visualizing them.

Categorical predictors pose special difficulties with respect to interpreting regression models, especially models that involve interactions of categorical predictors. Categorical predictors are traditionally coded using dummy (indicator) coding. Many research questions cannot be answered directly in terms of dummy variables. Furthermore, interactions involving dummy categorical variables can be confusing and even misleading. Stata 12 introduces the `contrast` command, a general-purpose command that can be

used to precisely test the effects of categorical variables by forming contrasts among the levels of the categorical predictors. For example, you can compare adjacent groups, compare each group with the overall mean, or compare each group with the mean of the previous groups. The `contrast` command allows you to easily focus on the comparisons that are of interest to you.

The `contrast` command works with interactions as well. You can test the simple effect of one predictor at specific levels of another predictor or form interactions that involve comparisons of your choosing. In the parlance of analysis of variance, you can test simple effects, simple contrasts, partial interactions, and interaction contrasts. These kinds of tests allow you to precisely understand and dissect interactions with surgical precision. The `contrast` command works not only with the `regress` command but also with commands such as `logit`, `ologit`, `mlogit`, as well as random-effects models like `xtmixed`.

As you can see, the scope of the application of the `margins`, `marginsplot`, and `contrast` commands is broad. Likewise, so is the scope of this book. It covers continuous variables (modeled linearly, using polynomials, and piecewise), interactions of continuous variables, categorical predictors, interactions of categorical predictors, as well as interactions of continuous and categorical predictors. The book also illustrates how the `margins`, `marginsplot`, and `contrast` commands can be used to interpret results from multilevel models, models where time is a continuous predictor, models with time as a categorical predictor, nonlinear models (such as logistic regression or ordinal logistic regression), and analyses that involve complex survey data. However, this book does not contain information about the theory of these statistical models, how to perform diagnostics for the models, the formulas for the models, and so forth. The summary section concluding each chapter includes references to books and articles that provide background for the techniques illustrated in the chapter.

My goal for this book is to provide simple and clear examples that illustrate how to interpret and visualize the results of regression models. To that end, I have selected examples that illustrate large effects generally combined with large sample sizes to create patterns of effects that are easy to visualize. Most of the examples are based on real data, but some are based on hypothetical data. In either case, I hope the examples help you understand the results of your regression models so you can interpret and present them with clarity and confidence.

(Pages omitted)

14 Continuous by categorical by categorical interactions

14.1	Chapter overview	393
14.2	Simple effects of gender on the age slope	398
14.3	Simple effects of education on the age slope	399
14.4	Simple contrasts on education for the age slope	400
14.5	Partial interaction on education for the age slope	400
14.6	Summary	401

14.1 Chapter overview

This chapter considers models that involve the interaction of two categorical predictors with a linear continuous predictor. Such models blend ideas from chapter 10 on categorical by continuous interactions and ideas from chapter 8 on categorical by categorical interactions. As we saw in chapter 10, interactions of categorical and continuous predictors describe how the slope of the continuous variable differs as a function of the categorical variable. In chapter 8, we saw models that involve the interaction of two categorical variables. This chapter blends these two modeling techniques by exploring how the slope of the continuous variable varies as a function of the interaction of the two categorical variables.

Let's consider a hypothetical example of a model with income as the outcome variable. The predictors include gender (a two-level categorical variable), education (treated as a three-level categorical variable), and age (a continuous variable). Income can be modeled as a function of each of the predictors, as well as the interactions of all the predictors. A three-way interaction of age by gender by education would imply that the effect of age interacts with gender by education. One way to visualize such an interaction would be to graph age on the x axis, with separate lines for the levels of education and separate graphs for gender. Figure 14.1 shows such an example, illustrating how the slope of the relationship between income and age varies as a function of education and gender.

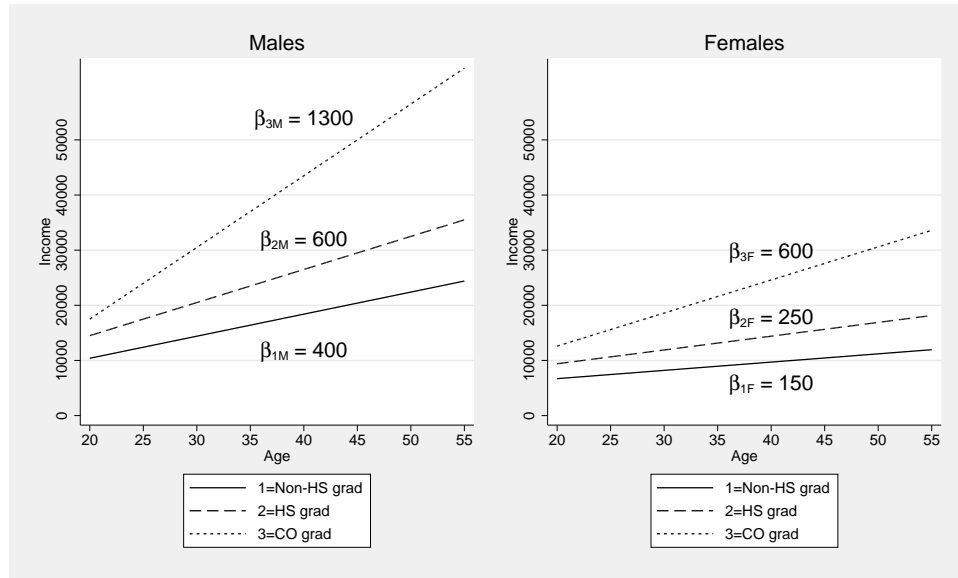


Figure 14.1. Fitted values of income as a function of age, education, and gender

The graph can be augmented by a table that shows the **age** slope broken down by education and gender. Such a table is shown in 14.1. The **age** slope shown in each cell of table 14.1 reflects the slope of the relationship between income and age for each of the lines illustrated in figure 14.1. For example, β_{3M} represents the **age** slope for male college graduates, and this slope is 1,300.

Table 14.1. The **age** slope by level of education and gender

	Non-HS grad	HS grad	CO grad
Male	$\beta_{1M} = 400$	$\beta_{2M} = 600$	$\beta_{3M} = 1,300$
Female	$\beta_{1F} = 150$	$\beta_{2F} = 250$	$\beta_{3F} = 600$

The age by education by gender interaction described in table 14.1 can be understood and dissected like the two by three interactions illustrated in chapter 8. The key difference is that table 14.1 is displaying the slope of the relationship between income and age, and the three-way interaction refers to the way that the slope varies as a function of education and gender.¹

If there were no three-way interaction of age by gender by education, we would expect (for example) that the gender difference in the **age** slope would be approximately the

1. More precisely, how the slope varies as a function of the interaction of age and gender.

same at each level of education. But consider the differences in the `age` slopes between females and males at each level of education. This difference is -250 ($150 - 400$) for non-high school graduates, whereas this difference is -350 ($250 - 600$) for high school graduates, and the difference is -700 ($600 - 1300$) for college graduates. The difference in the `age` slopes between females and males seems to be much larger for college graduates than for high school graduates and non-high school graduates. This pattern of results appears consistent with a three-way interaction of age by education by gender.

Let's explore this in more detail with an example using the GSS dataset. To focus on the linear effect of `age`, we will keep those who are 22 to 55 years old.

```
. use gss_ivrm
. keep if age>=22 & age<=55
(18,936 observations deleted)
```

In this example, let's predict income as a function of gender² (`gender`), a three-level version of education (`educ3`), and `age`. The `regress` command below predicts `realrinc` from `i.gender`, `i.educ3`, and `c.age` (as well as all interactions of the predictors). The variable `i.race` is also included as a covariate.³

2. I use the variable `gender` (coded: 1 = Male and 2 = Female), because it leads to output that clearly distinguishes the variable name (that is, `gender`) and its values (that is, `Male`, and `Female`).

3. The `noci` option is included to make the output more readable for this example. See the callout titled *Using the noci option for clearer output* in section 2.5.1 for more details.

```
. regress realrinc i.gender##i.educ3##c.age i.race, vce(robust) noci
Linear regression          Number of obs   =   25,718
                          F(13, 25704)       =   411.30
                          Prob > F         =   0.0000
                          R-squared        =   0.1839
                          Root MSE     =   23556
```

realrinc	Coef.	Robust Std. Err.	t	P> t
gender				
Male	0	(base)		
Female	1337.125	1693.694	0.79	0.430
educ3				
not hs	0	(base)		
HS	550.476	1782.192	0.31	0.757
Coll	-11156.1	2618.976	-4.26	0.000
gender#educ3				
Female#HS	783.0991	2021.654	0.39	0.698
Female#Coll	7657.907	3164.299	2.42	0.016
age	413.8695	45.62015	9.07	0.000
gender#c.age				
Female	-264.9842	50.65695	-5.23	0.000
educ3#c.age				
HS	175.8497	54.7504	3.21	0.001
Coll	897.3326	77.47101	11.58	0.000
gender#educ3#c.age				
Female#HS	-80.30545	60.94575	-1.32	0.188
Female#Coll	-414.6562	93.26714	-4.45	0.000
race				
white	0	(base)		
black	-2935.138	273.3294	-10.74	0.000
other	185.3956	956.338	0.19	0.846
_cons	2691.23	1495.778	1.80	0.072

Let's test the interaction of gender, education, and age using the `contrast` command below. The three-way interaction is significant.

```
. contrast i.gender#i.educ3#c.age
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	df	F	P>F
gender#educ3#c.age	2	10.17	0.0000
Denominator	25704		

To begin the process of interpreting the three-way interaction, let's create a graph of the adjusted means as a function of age, education, and gender. First, the `margins` command below is used to compute the adjusted means by gender and education for ages 22 and 55 (the output is omitted to save space). Then, the `marginsplot` command is used to graph the adjusted means, as shown in figure 14.2.

```
. margins gender#educ3, at(age=(22 55))
(output omitted)
. marginsplot, bydimension(gender) noci
Variables that uniquely identify margins: age gender educ3
```

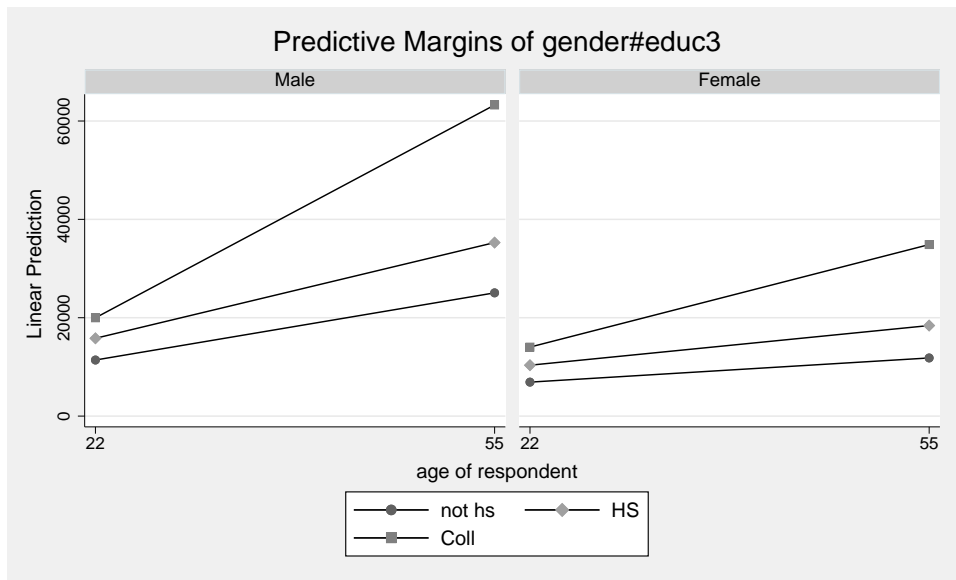


Figure 14.2. Fitted values of income as a function of age, education, and gender

The graph in figure 14.2 illustrates how the `age` slope varies as a function of gender and education. Let's compute the `age` slope for each of the lines shown in this graph. The `margins` command is used with the `dydx(age)` and `over()` options to compute the `age` slopes separately for each combination of gender and education.

```
. margins, noci dydx(age) over(gender educ3)
Average marginal effects           Number of obs   =   25,718
Model VCE       : Robust
Expression      : Linear prediction, predict()
dy/dx w.r.t.    : age
over            : gender educ3
```

	Delta-method		t	P> t
	dy/dx	Std. Err.		
<code>age</code>				
<code>gender#educ3</code>				
Male#not hs	413.8695	45.62015	9.07	0.000
Male#HS	589.7192	30.37993	19.41	0.000
Male#Coll	1311.202	62.88374	20.85	0.000
Female#not hs	148.8854	22.09037	6.74	0.000
Female#HS	244.4296	15.25412	16.02	0.000
Female#Coll	631.5618	46.90854	13.46	0.000

Let's reformat the output of the `margins` command to emphasize how the `age` slope varies as a function of the interaction of gender and education (see table 14.2). Each cell of table 14.2 shows the `age` slope for the particular combination of gender and education. For example, the `age` slope for males with a college degree is 1,311.20 and is labeled as β_{3M} .

Table 14.2. The `age` slope by level of education and gender

	Non-HS grad	HS grad	CO grad
Male	$\beta_{1M} = 413.87$	$\beta_{2M} = 589.72$	$\beta_{3M} = 1,311.20$
Female	$\beta_{1F} = 148.89$	$\beta_{2F} = 244.43$	$\beta_{3F} = 631.56$

We can dissect the three-way interaction illustrated in table 14.2 using the techniques from section 8.3 on two by three models. Specifically, we can use simple effects analysis, simple contrasts, and partial interactions.

14.2 Simple effects of gender on the `age` slope

We can use the `contrast` command to test the simple effect of gender on the `age` slope. This is illustrated below.

```
. contrast gender#c.age@educ3, nowald pveffects
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	Contrast	Std. Err.	t	P> t
gender@educ3#c.age				
(Female vs base) not hs	-264.9842	50.65695	-5.23	0.000
(Female vs base) HS	-345.2896	33.98931	-10.16	0.000
(Female vs base) Coll	-679.6404	78.4498	-8.66	0.000

Each of these tests represents the comparison of females versus males in terms of the `age` slope. The first test compares the `age` slope for females versus males among non-high school graduates. Referring to table 14.2, this test compares β_{1F} with β_{1M} . The difference in these `age` slopes is -264.98 ($148.89 - 413.87$), and this difference is significant. The `age` slope for females who did not graduate high school is 264.98 units smaller than the `age` slope for males who did not graduate high school. The second test is similar to the first, except the comparison is made among high school graduates, comparing β_{2F} with β_{2M} from table 14.2. This test is also significant. The third test compares the `age` slope between females and males among college graduates (that is, comparing β_{3F} with β_{3M}). This test is also significant. In summary, the comparison of the `age` slope for females versus males is significant at each level of education.

14.3 Simple effects of education on the age slope

We can also look at the simple effects of education on the `age` slope at each level of gender. This test is performed using the `contrast` command below.

```
. contrast educ3#c.age@gender
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	df	F	P>F
educ3@gender#c.age			
Male	2	70.96	0.0000
Female	2	43.37	0.0000
Joint	4	57.21	0.0000
Denominator	25704		

The first test compares the `age` slope among the three levels of education for males. Referring to table 14.2, this tests the following null hypothesis.

$$H_0: \beta_{1M} = \beta_{2M} = \beta_{3M}$$

This test is significant. The `age` slope significantly differs as a function of education among males.

The second test is like the first test, except that the comparisons are made for females. This tests the following null hypothesis.

$$H_0: \beta_{1F} = \beta_{2F} = \beta_{3F}$$

This test is also significant. Among females, the `age` slope significantly differs among the three levels of education.

14.4 Simple contrasts on education for the age slope

We can further dissect the simple effects tested above by applying contrast coefficients to the education factor. For example, say that we used the `ar.` contrast operator to form reverse adjacent group comparisons. This would yield comparisons of group 2 versus 1 (high school graduates with non-high school graduates) and group 3 versus 2 (college graduates with high school graduates). Applying this contrast operator yields simple contrasts on education at each level of gender, as shown below.

```
. contrast ar.educ3#c.age@gender, nowald peffects
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	Contrast	Std. Err.	t	P> t
educ3@gender#c.age				
(HS vs not hs) Male	175.8497	54.7504	3.21	0.001
(HS vs not hs) Female	95.54426	26.83611	3.56	0.000
(Coll vs HS) Male	721.4829	69.74939	10.34	0.000
(Coll vs HS) Female	387.1322	49.38976	7.84	0.000

The first test compares the `age` slope for male high school graduates with the `age` slope for males who did not graduate high school. In terms of table 14.2, this is the comparison of β_{2M} with β_{1M} . The difference in these `age` slopes is 175.85 and is significant. The second test is the same as the first test, except the comparison is made for females, comparing β_{2F} with β_{1F} . The difference is 95.54 and is significant. The third and fourth tests compare college graduates with high school graduates. The third test forms this comparison among males and is significant, and the fourth test forms this comparison among females and is also significant.

14.5 Partial interaction on education for the age slope

The three-way interaction can be dissected by forming contrasts on the three-level categorical variable. Say that we use reverse adjacent group comparisons on education, which compares high school graduates with non-high school graduates and college graduates with high school graduates. We can interact that contrast with gender and age, as shown in the `contrast` command below.

```
. contrast ar.educ3#r.gender#c.age, nowald pveffects
Contrasts of marginal linear predictions
Margins      : asbalanced
```

	Contrast	Std. Err.	t	P> t
educ3#gender#c.age				
(HS vs not hs) (Female vs Male)	-80.30545	60.94575	-1.32	0.188
(Coll vs HS) (Female vs Male)	-334.3507	85.48932	-3.91	0.000

The first comparison tests the interaction of the contrast of high school graduates versus non-high school graduates by gender by `age`. The difference in the `age` slope between high school graduates and non-high school graduates for females is 244.43 minus 148.89 (95.54). For males, this difference is 589.72 minus 413.87 (175.85). The difference in these differences is -80.31 , which is not significant (see the first comparison from the `margins` command). The difference in the `age` slope comparing high school graduates with non-high school graduates is not significantly different for males and females.

The second test forms the same kind of comparison, but compares college graduates with high school graduates. The difference in the `age` slope comparing female college graduates with female high school graduates is 631.56 minus 244.43 (387.13). This difference for males is 1,311.20 minus 589.72 (721.48). The difference of these differences is -334.35 and is statistically significant (see the second comparison from the `margins` output). The increase in the `age` slope comparing college graduates with high school graduates is greater for males than it is for females.

14.6 Summary

This chapter has illustrated models that involve the interaction of a continuous variable with two categorical variables. When such an interaction is significant, the slope associated with the continuous variable varies as a function of the interaction of the two categorical variables. As we saw, the interaction can be dissected and understood using the methods covered in chapter 8 on categorical by categorical interactions, forming contrasts with respect to the slope term.

I am not aware of any books that directly cover this kind of interaction, so I recommend the references provided in chapter 10 (linear by categorical interactions) and chapter 8 (categorical by categorical interactions).

(Pages omitted)