Lecture 7: Interaction Effects

Dave Armstrong

University of Wisconsin – Milwaukee
Department of Political Science

e: armstrod@uwm.edu
w: www.quantoid.net/teachuwm/uwm702/

Interaction Effects (1)

- When the partial effect of one variable depends on the value of another variable, those two variables are said to “interact”.
  - For example, we may want to test whether age effects are different for men (coded 1) and women (coded 0).
  - In such cases it is sensible to fit separate regressions for men and women, but this does not allow for a formal statistical test of the differences
  - Specification of interaction effects facilitates statistical tests for a difference in slopes within a single regression

Interaction Effects (2)

- Interaction terms are the *product of the regressors for the two variables*.
- The interaction regressor in the model below is X_iD_i:

$$\begin{aligned} Y_i &= \alpha + \beta X_i + \gamma D_i + \delta(X_iD_i) + \varepsilon_i \\ \text{income}_i &= \alpha + \beta\ \text{age}_i + \gamma\ \text{men}_i + \delta(\text{age}_i \times \text{men}_i) + \varepsilon_i \end{aligned}$$

Ultimately we want to know two things:

1. Is there a statistically significant interactive (i.e., multiplicative or conditional) effect?
2. If the answer to #1 is “yes”, what is the nature of that effect (i.e., what does it look like)?

Below, I will walk you through all of the possible two-way interaction scenarios and we will discuss how to answer these two questions.

Two Categorical Variables

With two categorical variables, essentially you are estimating a different conditional mean for every pair of values across the two categorical variables. You could do that as follows:

```
data(Duncan)
Duncan$inc.cat <- cut(Duncan$income, 3)
mod <- lm(prestige~ inc.cat * type + education,
  data=Duncan)
```


Model Summary

```
summary(mod)

##
## Call:
## lm(formula = prestige ~ inc.cat * type + education, data = Duncan)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -18.617  -5.999  -0.163   4.636  19.037
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   7.8827     3.4364   2.294 0.027915 *
## inc.cat(31.7,56.3]           22.4574     4.8792   4.603 5.30e-05 ***
## inc.cat(56.3,81.1]           51.2807     9.4351   5.435 4.29e-06 ***
## typeprof                     55.6073    11.6800   4.761 3.30e-05 ***
## typewc                        2.5446     8.1162   0.314 0.755746
## education                     0.2799     0.1121   2.496 0.017411 *
## inc.cat(31.7,56.3]:typeprof -41.5789    11.2428  -3.698 0.000740 ***
## inc.cat(56.3,81.1]:typeprof -50.3567    13.3929  -3.760 0.000621 ***
## inc.cat(31.7,56.3]:typewc   -13.0171    10.3130  -1.262 0.215223
## inc.cat(56.3,81.1]:typewc   -33.6407    13.1215  -2.564 0.014806 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.115 on 35 degrees of freedom
```


Anova

Q1: Is there an interaction Effect here?

- An incremental (Type II) F-test will answer that question. We want to test the null hypothesis that all of the interaction dummy regressor coefficients are zero in the population.
- The `inc.cat:type` line of the output gives the results of this test.

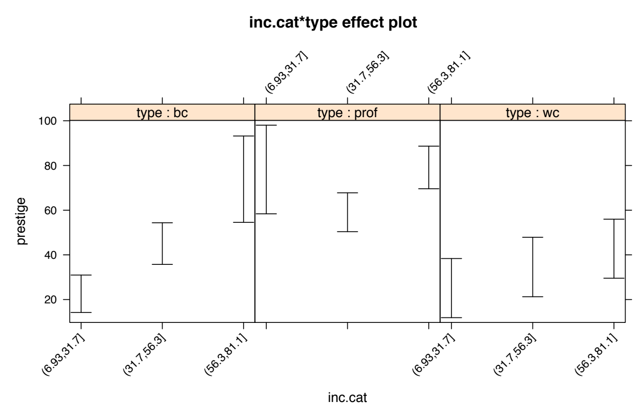
```
Anova(mod)

## Anova Table (Type II tests)
##
## Response: prestige
##              Sum Sq Df F value    Pr(>F)
## inc.cat      3491.9  2 21.0159 1.010e-06 ***
## type         2856.0  2 17.1885 6.308e-06 ***
## education     517.7  1  6.2313  0.017411 *
## inc.cat:type 1644.4  4  4.9484  0.002871 **
## Residuals    2907.7 35
## ---
```


Two Categorical Variables - Plot

Q2: What is the nature of the interaction effect?

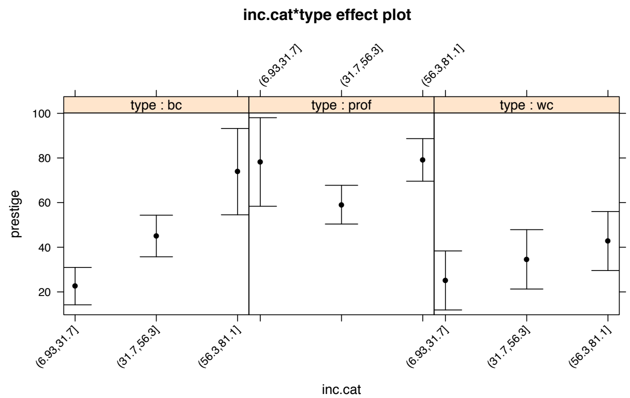
```
e1 <- effect("inc.cat*type", mod)
pe1 <- plot(e1, as.table=T, layout=c(3,1),ci.style="bars", colors=c(NA, "
pe1
```



Two Categorical Variables - Plot (2)

If you would rather have capped error bars rather than confidence lines, you could do the following:

```
update(pe1, panel=panel.2cat)
```



Presenting two categorical variables

```
tab2 <- cat2Table(e1)
noquote(tab2)

##                bc            prof          wc
## (31.7,56.3] 22.59         78.20         25.14
##             (14.22,30.97) (58.36,98.04) (11.91,38.:
## (56.3,81.1] 45.05         59.08         34.58
##             (35.74,54.36) (50.40,67.76) (21.28,47.8
## (6.93,31.7] 73.87         79.12         42.78
##             (54.54,93.20) (69.58,88.66) (29.57,55.!
```


Interpretation

The important points are as follows:

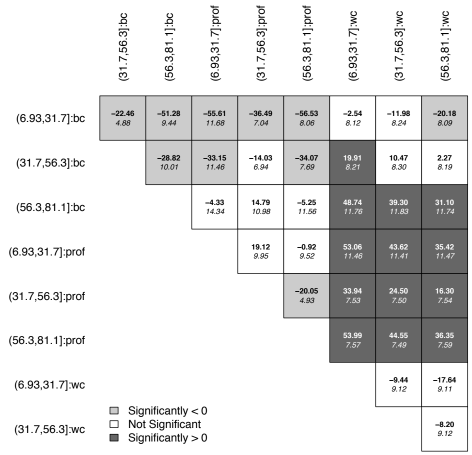
- The interaction term is significant in the F-test, so that indicates a significant interaction effect.
- With no interaction effect, the across each row have the same pattern across the three different tows and down the three different columns.
  - While the trends overall look somewhat different and there are clearly different magnitudes in the differences.
  - This is the same as we look down the rows.

Using Factorplot

You could also use the `factorplot` package to plot the differences based on the output from `effects`.

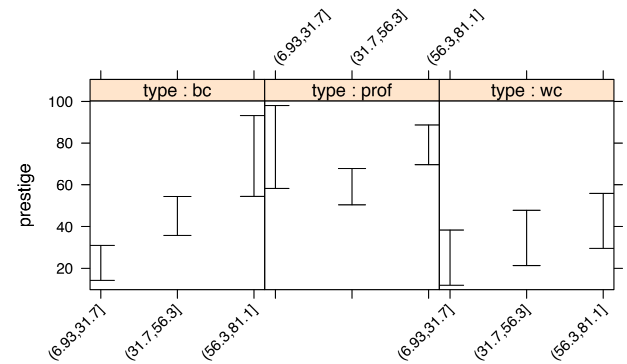
```
e1$fit <- c(e1$fit)
names(e1$fit) <- apply(e1$x, 1, paste, collapse=":")
fp <- factorplot(e1$fit, var=e1$vcov,
    resdf=(nrow(e1$data)-ncol(e1$model.matrix)))
plot(fp, print.square.leg=F, scale.text=.75, abbrev.char=100)
```

Factorplot figure



Testing Differences

```r
e1 <- effect("inc.cat*type", mod)
pe1 <- plot(e1, as.table=T, layout=c(3,1),ci.style="bars",
   colors=c(NA, "black"), rotx=45, xlab="", main="")
pe1
```



Testing Differences

Imagine that you wanted to test whether the effect of moving from middle income to high income was the same for blue collar and white collar occupations.

$$\hat{P} = b_0 + b_1 M + b_2 H + b_3 Pr + b_4 W + b_5 E \\ + b_6 M \times Pr + b_7 H \times Pr + b_8 M \times W + b_9 H \times W$$

The effect for blue collar occupations is:

$$b_2 - b_1$$

And for white collar occupations it is

$$(b_2 + b_9) - (b_1 + b_8)$$

Rearranging, we get:

$$\begin{aligned} b_2 - b_1 &= (b_2 + b_9) - (b_1 + b_8) \\ &= b_2 + b_9 - b_1 - b_8 \\ 0 &= b_9 - b_8 \end{aligned}$$

```r
linearHypothesis(mod,
   "inc.cat(56.3,81.1]:typewc - inc.cat(31.7,56.3]:typewc = 0")

## Linear hypothesis test
##
## Hypothesis:
## - inc.cat(31.7,56.3]:typewc  + inc.cat(56.3,81.1]:typewc = 0
##
## Model 1: restricted model
## Model 2: prestige ~ inc.cat * type + education
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     36 3100.9
## 2     35 2907.7  1    193.19 2.3254 0.1363
```

Two Non-Reference Categories

What if we want to test whether the effect of middle to high income is different for Professional and White Collar occupations? The effect for Professional Occupations is:

$$(b_2 + b_7) - (b_1 + b_6)$$

Thus, the difference in effects is:

$$\begin{aligned} b_2 + b_7 - b_1 - b_6 &= b_2 + b_9 - b_1 - b_8 \\ b_7 - b_6 &= b_9 - b_8 \\ 0 &= b_6 - b_7 + b_9 - b_8 \end{aligned}$$

The test

```
linearHypothesis(mod,
   "inc.cat(31.7,56.3]:typeprof -inc.cat(56.3,81.1]:typeprof +
   inc.cat(56.3,81.1]:typewc - inc.cat(31.7,56.3]:typewc = 0")

## Linear hypothesis test
##
## Hypothesis:
## inc.cat(31.7,56.3]:typeprof - inc.cat(56.3,81.1]:typeprof - inc.cat(31
##
## Model 1: restricted model
## Model 2: prestige ~ inc.cat * type + education
##
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     36 3015.2
## 2     35 2907.7  1    107.52 1.2942  0.263
```


Two Categorical Variables

One Dummy Variable, One Continuous Variable

One Categorical and One Continuous

Two Continuous Variables

One Dummy and One Continuous

Recall the model we talked about briefly above.

$$Y_i = \alpha + \beta X_i + \gamma D_i + \delta(X_i D_i) + \varepsilon_i$$

One way to think about this model is leading to two separate regression lines:

For $D = 0$:

$$\begin{aligned} \hat{Y}_i &= \alpha + \beta X_i + \gamma(0) + \delta(X_i \times 0) \\ &= \alpha + \beta X_i \end{aligned}$$

For D=1:

$$\begin{aligned} \hat{Y}_i &= \alpha + \beta X_i + \gamma(1) + \delta(X_i \times 1) \\ &= (\alpha + \gamma) + (\beta + \delta) X_i \end{aligned}$$

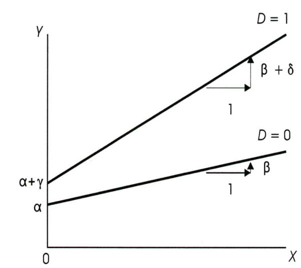


Figure 7.5 from Fox (1997)

Example with one Dummy Variable and One Continuous Variable

```r
library(car)
data(SLID)
mod <- lm(wages ~ age*sex, data=SLID)
summary(mod)
```

```
##
## Call:
## lm(formula = wages ~ age * sex, data = SLID)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -20.928  -4.658  -1.452   3.603  35.359
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.84674    0.50267  15.610  < 2e-16 ***
## age           0.16377    0.01295  12.648  < 2e-16 ***
## sexMale      -1.78986    0.70988  -2.521   0.0117 *
## age:sexMale   0.13625    0.01820   7.485 8.71e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.122 on 4143 degrees of freedom
##   (3278 observations deleted due to missingness)
## Multiple R-squared:  0.1844, Adjusted R-squared:  0.1838
## F-statistic: 312.3 on 3 and 4143 DF,  p-value: < 2.2e-16
```


Assessing Interaction I

Q1: Is there an interaction?

- We want to know whether the lines are parallel or not.
- Note that the coefficient on the interaction term gives the difference in the slope for the $D = 0$ group and the $D = 1$ group.
- The `age:sexMale` line provides the answer to the question.
  - If the coefficient is statistically significant (and it is here), then there is a significant interaction.
  - If the coefficient is not statistically significant, then a purely additive model performs just as well.

Nature of the Interaction

Q2: What is the nature of the interaction?

There are a number of ways we can figure this out. Ultimately, we want to know three things regarding the slope.

1. Is the slope of age for females ($D = 0$) different from zero?
2. Is the slope of age for males ($D = 1$) different from zero?
3. Is the slope of age for men different from the slope of age for women?

Two of these can be answered directly from the coefficient table, one requires a bit of extra work.

Conditional Effect of Age

First, we need to think more generally about the conditional effect of age. If the equation is:

$$\text{wages} = b_0 + b_1\text{age} + b_2\text{male} + b_3\text{age} \times \text{male} + e$$

Then the partial, conditional effect (or what some might call the "marginal effect") of age is:

$$\frac{\partial\widehat{\text{wages}}}{\partial\text{age}} = b_1 + b_3\text{male}$$

Since we will want to test hypotheses about that quantity, we need to know its variance:

$$V(b_1 + b_3\text{male}) = V(b_1) + \text{male}^2 V(b_3) + 2\text{male}V(b_1, b_3)$$

In general, with constants c and d and variables W and Z:

$$V(cW + dZ) = c^2 V(W) + d^2 V(Z) + 2cdV(W, Z)$$

Back to the Questions

1. Is the slope of age for females ($D = 0$) different from zero?
   This amounts to a test of $H_0 : \beta_1 = 0$. This can be evaluated by looking at the `age` line from the output.
2. Is the slope of age for males ($D = 1$) different from zero?
   This amounts to a test of $H_0 : \beta_1 + \beta_3 = 0$. This cannot be directly evaluated by looking at the coefficients. It can be done this way:

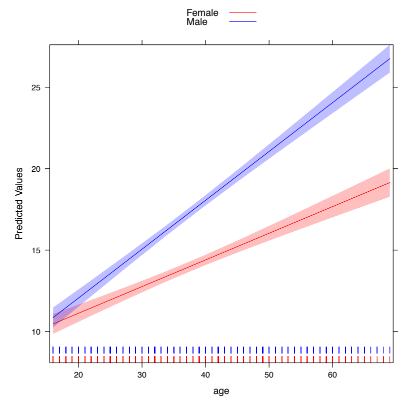
```
intQualQuant(mod, c("age", "sex"), type="slopes", plot=F)

## Conditional effects of  age :
##          B     SE(B) t-stat Pr(>|t|)
## Female 0.164 0.013 12.648 0.000
## Male   0.300 0.013 23.447 0.000
```

3. Is the slope of age for men different from the slope of age for women?
   This amounts to a test of $H_0 : \beta_3 = 0$. This can be evaluated by looking at the `age:sexMale` line from the output.

Graphically...

```
trellis.par.set(
        superpose.line=list(col=c("red", "blue")),
        superpose.polygon = list(col=c("red", "blue")))
intQualQuant(mod, c("age", "sex"), type="slopes",
        plot=TRUE, rug=TRUE, ci=TRUE)
```



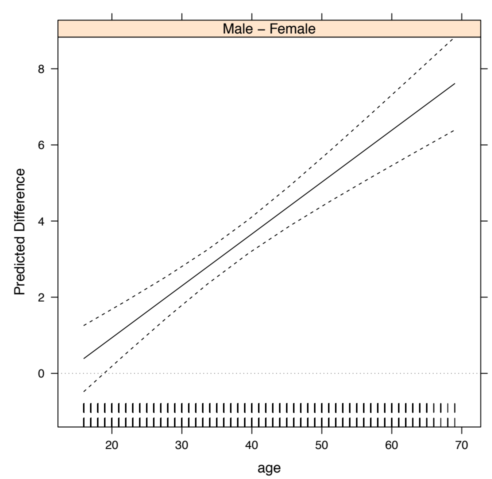
The effect of Gender

Almost always, we are concerned with the results above (i.e., the different slopes for age), but what if we care about the conditional effect of gender?

$$\frac{\partial \widehat{\text{wages}}}{\partial \text{male}} = b_2 + b_3 \text{age}$$

```
trellis.par.set(
        superpose.line=list(col=c("red", "blue")),
        superpose.polygon = list(col=c("red", "blue")))
iq <- intQualQuant(mod, c("age", "sex"),
    type="facs", plot=TRUE)
```


The Figure



Summary

- The interaction is significant (from the `age:sexMale` line of the regression output), so the two variable do have an interactive effect.
- Since the `age` coefficient is positive and the `age:sexMale` coefficient is positive, both men and women have positive slopes of age for wages, but the difference between men and women is significantly bigger than zero, meaning the slope of age for men is bigger than the slope of age for women.
- The results of the `plotSlopes` and `testSlopes` functions (from the `rockchalk` package) provide graphical and numerical results about the two different slopes.
- The above implies that the effect of gender is increasing in age (i.e., the gender gap is growing). The `intQualQuant` function (from the `DAMisc` package) provides numerical and optional graphical results.

One Categorical and One Continuous

With one categorical and one continuous variable, we want to show the conditional coefficients of the continuous variable (probably in a table) and we want to show the conditional coefficients of the dummy variables.

```
Prestige$income <- Prestige$income/1000
mod <- lm(prestige ~ income*type + education,
  data=Prestige)
```


Model Summary

```
summary(mod)

##
## Call:
## lm(formula = prestige ~ income * type + education, data = Prestige)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12.8720  -4.8321   0.8534   4.1425  19.6710
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -6.7273     4.9515  -1.359   0.1776
## income             3.1344     0.5215   6.010 3.79e-08 ***
## typeprof          25.1724     5.4670   4.604 1.34e-05 ***
## typewc             7.1375     5.2898   1.349   0.1806
## education          3.0397     0.6004   5.063 2.14e-06 ***
## income:typeprof   -2.5102     0.5530  -4.539 1.72e-05 ***
## income:typewc     -1.4856     0.8720  -1.704   0.0919 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.455 on 91 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.8663, Adjusted R-squared:  0.8574
## F-statistic: 98.23 on 6 and 91 DF,  p-value: < 2.2e-16
```

Anova

Q1: Is there a significant interaction?

```
Anova(mod)

## Anova Table (Type II tests)
##
## Response: prestige
##              Sum Sq Df F value    Pr(>F)
## income       1058.8  1 25.4132 2.342e-06 ***
## type          591.2  2  7.0947   0.00137 **
## education    1068.0  1 25.6344 2.142e-06 ***
## income:type   890.0  2 10.6814 6.809e-05 ***
## Residuals    3791.3 91
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '
```

Notice that the `income:type` line of the `Anova` output tells us that the interaction is significant. Thus, we should go on to calculate and explain the conditional coefficients.

Conditional Coefficients of Income

Q2: What is the nature of the interaction effect?

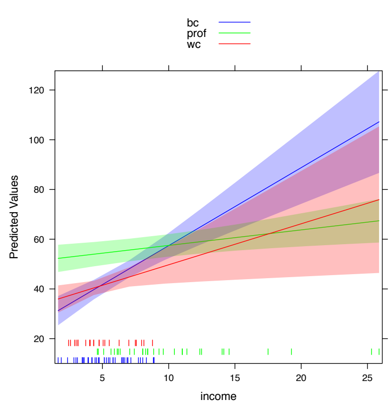
- The nature of the interaction has to be considered both for `income` and for `type`.
- We can calculate the conditional effects and variances of `income` as follows:

```
intQualQuant(mod, c("income", "type"),
    type="slopes", plot=F)

## Conditional effects of  income :
##          B     SE(B) t-stat Pr(>|t|)
## bc   3.134 0.522 6.010  0.000
## prof 0.624 0.222 2.816  0.006
## wc   1.649 0.709 2.326  0.022
```


Plotting the Conditional Effects of Income

```
cols <- c("blue", "green", "red")
trellis.par.set(
        superpose.line = list(col=cols),
        superpose.polygon = list(col=cols))
intQualQuant(mod, c("income", "type"),
    type="slopes", plot=TRUE)
```



Interpretation

- The slope is significant for all occupation types and is the biggest for blue collar.
- Confidence bounds for both blue collar and white collar occupation lines are very big at high levels of income (lack of data density).
- The only valid places where professional occupations can be compared to the others is between around $5,000 and $8,000.

Conditional Effect of Type

Q2: What is the nature of the interaction effect (this time for `type`)?

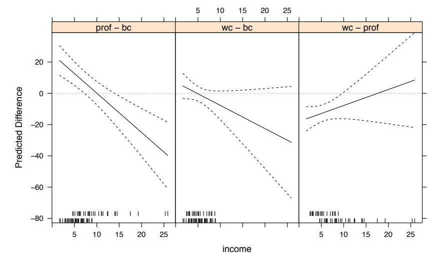
- The conditional effect of type (as we saw) is a bit more difficult. Here, We would presumably have to test each pairwise difference: BC vs Prof, BC vs WC and Prof vs WC for different values of education. First, let's think about what we need.

BC vs Prof	$\frac{\partial \text{Prestige}}{\partial \text{Prof}}$	$b_2 + b_5\text{Income}$
BC vs WC	$\frac{\partial \text{Prestige}}{\partial \text{WC}}$	$b_3 + b_6\text{Income}$
Prof vs WC	$\frac{\partial \text{Prestige}}{\partial \text{Prof}} - \frac{\partial \text{Prestige}}{\partial \text{WC}}$	$(b_2 - b_3) + (b_5 - b_6)\text{Income}$

Conditional Effect of Type

The conditional effect of type is a bit more difficult, luckily a function exists to help. Here, We would want to test each pairwise difference: BC vs Prof, BC vs WC and Prof vs WC.

```
mod.out <- intQualQuant(mod, c("income", "type"),
    type="facs", n=25, plot=T)
update(mod.out, layout=c(3,1))
```



Numerical Results

If you would rather have numbers than a figure, you could look at:

```
intQualQuant(mod, c("income", "type"),
    vals=c(1,6.8,25), plot=F, type="facs")

##       fit se.fit    x  contrast   lower   upper
## 1  22.662  5.057  1.0 prof - bc  12.618  32.707
## 2   8.103  3.547  6.8 prof - bc   1.057  15.150
## 3 -37.582 10.267 25.0 prof - bc -57.977 -17.187
## 4   5.652  4.523  1.0   wc - bc  -3.332  14.636
## 5  -2.964  2.606  6.8   wc - bc  -8.140   2.212
## 6 -30.001 17.207 25.0   wc - bc -64.181   4.178
## 7 -17.010  4.267  1.0 wc - prof -25.486  -8.535
## 8 -11.068  2.815  6.8 wc - prof -16.660  -5.475
## 9   7.580 14.667 25.0 wc - prof -21.553  36.714
```


Interpretation

In the previous graph, we see the following:

- From its lowest values through the mean of income, professional occupations are expected to have more prestige than blue collar occupations. However, when income is highest, blue collar occupations are expected to have more prestige than professional occupations (first row of table)
- The difference between white collar and blue collar is never significantly different from zero (second row of table).
- From its lowest values through the mean of income, professional occupations are expected to have more prestige than white collar occupations. When income is high, however, there is no expected difference between professional and white collar occupations as regards prestige.

Two Categorical Variables

One Dummy Variable, One Continuous Variable

One Categorical and One Continuous

Two Continuous Variables

Interaction Example

With two continuous variables the interpretation gets a bit trickier. For example, consider the following model:

$$\hat{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i3}$$

We want to know the partial conditional effect of both X_1 and X_2, but unlike above, neither can be boiled down to a small set of values. Just think about the equation:

$$\frac{\partial \hat{Y}}{\partial X_1} = \beta_1 + \beta_4 X_2 \quad (1)$$

$$\frac{\partial \hat{Y}}{\partial X_2} = \beta_2 + \beta_4 X_1 \quad (2)$$

$$(3)$$

- Note, that β_4 is the amount by which the *effect* of X_1 goes up for every additional unit of X_2 and the amount by which the *effect* of $X_@$ goes up for every additional unit of X_1.

Variance of a Linear Combination

Ultimately, we will want to know when conditional effects are significantly different from zero. This requires us to be able to calculate the variance of the conditional effects.

- Since these are linear combinations of random variables ($\hat{\beta}_1, \hat{\beta}_2$ and $\hat{\beta}_4$) and constants X_1 and X_2, its variance can be easily calculated.

For any linear combination of variables (W and Z) and coefficients g and h

$$\text{var}(gW + hZ) = g^2\text{var}(W) + h^2\text{var}(Z) + 2gh\text{cov}(W, Z)$$

Variance of Conditional Effects in Matrix Form

The results above are useful, but these terms get complicated to calculate "by hand" if there is are more than 2 terms for which you want to calculate the variance.

- The variance is the sum of all the variance and 2 times all of the pairwise covariances

If we think about it in matrix terms, it is easier:

$$\boldsymbol{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} \quad V(\boldsymbol{W}) = \begin{bmatrix} V(w_1) & V(w_1, w_2) & \cdots & V(w_1, w_k) \\ V(w_2, w_1) & V(w_2) & \cdots & V(w_2, w_k) \\ \vdots & \vdots & \ddots & \vdots \\ V(w_k, w_1) & V(w_k, w_2) & \cdots & V(w_k) \end{bmatrix}$$

Then,

$$V(\boldsymbol{A'W}) = \boldsymbol{A'}V(\boldsymbol{W})\boldsymbol{A}$$

Testable Hypotheses

$$\hat{Y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i1} X_{i3}$$

Berry, Golder and Milton (2012) suggest that we should be able to test 5 hypotheses:

- $\boldsymbol{P}_{X_1|X_2=\text{min}}$ The marginal effect of X_1 is [positive, zero, negative] when X_2 takes its lowest value.
- $\boldsymbol{P}_{X_1|X_2=\text{max}}$ The marginal effect of X_1 is [positive, zero, negative] when X_2 takes its highest value.
- $\boldsymbol{P}_{X_2|X_1=\text{min}}$ The marginal effect of X_2 is [positive, zero, negative] when X_1 takes its lowest value.
- $\boldsymbol{P}_{X_2|X_1=\text{max}}$ The marginal effect of X_2 is [positive, zero, negative] when X_1 takes its highest value.
- $\boldsymbol{P}_{X_1 X_2}$ The marginal effect of each of X_1 and X_2 is [positively, negatively] related to the other variable.

Example

```
mod <- lm(prestige ~ income*education + type, data=Prestige)
summary(mod)

##
## Call:
## lm(formula = prestige ~ income * education + type, data = Prestige)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.6084  -5.0225   0.3531   4.9346  17.9175
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -17.80359    7.59424  -2.344 0.021212 *
## income             3.78593    0.94453   4.008 0.000124 ***
## education          5.10432    0.77665   6.572 2.93e-09 ***
## typeprof           5.47866    3.71385   1.475 0.143574
## typewc            -3.58387    2.42775  -1.476 0.143303
## income:education  -0.21019    0.06977  -3.012 0.003347 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.806 on 92 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8415
## F-statistic:   104 on 5 and 92 DF,  p-value: < 2.2e-16
```


Example (2)

Q1: Is there a significant interaction?

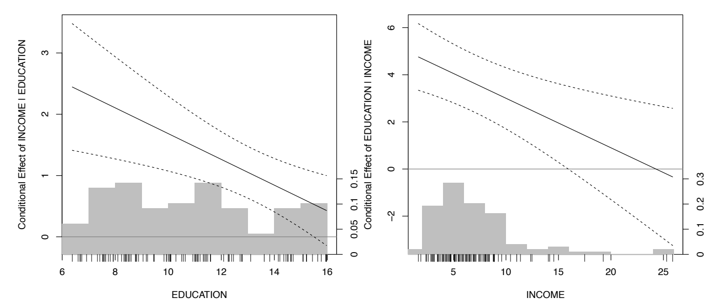
- The `income:education` line answers this question. If it is significant, then there is a significant interaction, otherwise there is not.
- This is counter to a minor, though still influential, point in Brambor, Clark and Golder (2006), but is consistent with Berry, Golder and Milton (2012).
- In this case, the interaction is significant, so we can move on to the next question

Example (3)

Q2: What is the nature of the interaction?

- This needs to be shown visually, since there are an infinite number of possibilities.

```
DAintfun2(mod, c("income", "education"), hist=T,
    scale.hist=.3, plot.type="pdf")
```



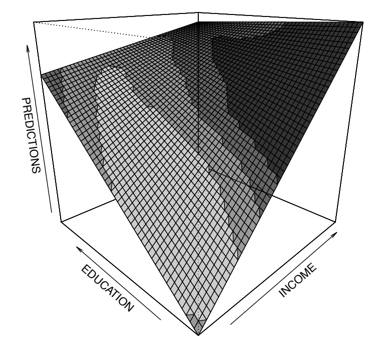
Interpretation

- The effect of income is nearly always significant, though it gets smaller as income gets bigger. That is, as income increases, smaller expected increases in prestige obtain from increasing education.
- The effect of education is significant and positive until around $16,000, which is around 2/3 the range of `income`, but is the 100^{th} percentile because of the skewness of income.
- This suggests that people tend to derive prestige from either higher incomes or higher education, but not really both.

Alternate Visualization

An alternate way to visualize the information is with a three-dimensional surface.

```
DAintfun(mod, c("income","education"),
        theta=-45, phi=20)
```



BGM Test for Prestige model

Here is the set of tests that Berry, Golder and Milton (2012) suggest. In the input to the function, the first variable in the `vars` argument is considered `X` and the second variable is considered `Z` for the purposes of the function.

```
BGMtest(mod, vars=c("income", "education"))

##               est    se      t p-value
## P(X|Zmin)  2.445 0.520  4.698 0.000
## P(X|Zmax)  0.429 0.287  1.495 0.138
## P(Z|Xmin)  4.756 0.712  6.681 0.000
## P(Z|Xmax) -0.335 1.466 -0.229 0.820
## P(XZ)     -0.210 0.070 -3.012 0.003
```