Regression III
Lecture 2: Effective Presentation of Linear Models

Dave Armstrong

University of Western Ontario
Department of Political Science
Department of Statistics and Actuarial Science (by courtesy)

e: dave.armstrong@uwo.ca
w: www.quantoid.net/teachicpsr/regression3/

Goals of the lecture

Discuss effective ways of testing and presenting effects in linear models

1. Categorical variables
   - Dummy variables
   - Presenting and testing pairwise comparisons
2. Interaction Effects
   - Fitted values and effect displays
3. Relative importance
   - Standardization for quantitative variables
   - Relative importance for sets of effects, including categorical variables

Categorical Explanatory Variables

- *Linear* regression can be extended to accommodate categorical variables (*factors*) using *dummy variable regressors* (or *indicator variables*)
- Below a categorical variable is represented by a dummy regressor D, (coded 1 for one category, 0 for the other):

$$Y_i = \alpha + \beta X_i + \gamma D_i + \varepsilon_i$$

- This fits *two regression lines with the same slope but different intercepts*. In other words, the coefficient γ represents the constant separation between the two regression lines:

$$Y_i = \alpha + \beta X_i + \gamma(0) + \varepsilon_i = \alpha + \beta X_i + \varepsilon_i$$
and
$$Y_i = \alpha + \beta X_i + \gamma(1) + \varepsilon_i = (\alpha + \gamma) + \beta X_i + \varepsilon_i$$

Categorical Explanatory Variables (2)

- In Figure (a) failure to account for a categorical variable (gender) does not produce significantly different results, either in terms of the intercept or the slope
- In Figure (b) the dummy regressor captures a significant difference in intercepts. More importantly, failing to include gender gives a negative slope for the relationship between education and income (dotted line) when in fact it should be positive for both men and women.

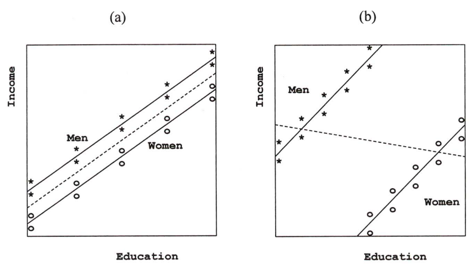


Figure 7.1 from Fox (1997)

Multi-Category Explanatory Variables

- Dummy regressors are easily extended to explanatory variables with more than two categories.
- A variable with m categories has $m - 1$ regressors:
  - As with the two-category case, one of the categories is a reference group (coded 0 for all dummy regressors).

	D_1	D_2
Blue Collar	1	0
Professional	0	1
White Collar	0	0

- The choice of reference category is technically irrelevant, but there are two considerations:
  - Theory may suggest we compare to a particular category
  - The largest category gives the smallest standard errors for each of the regressors

Multi-Category Explanatory Variables (2)

- A model with one quantitative predictor (e.g., income) then takes the following form:

$$Y_i = \alpha + \beta X_i + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \varepsilon_i$$

- This produces three *parallel regression lines*:

$$\text{Blue Collar: } Y_i = (\alpha + \gamma_1) + \beta X_i + \varepsilon_i$$
$$\text{Professional: } Y_i = (\alpha + \gamma_2) + \beta X_i + \varepsilon_i$$
$$\text{White Collar: } Y_i = \alpha + \beta X_i + \varepsilon_i$$

- Again, these lines are different only in terms of their intercepts
  - i.e., the γ coefficients represent the constant distance between the regression lines. γ_1 and γ_2 are the differences between occupation types compared to 'white collar', when holding income constant.

Dummy Variables in **R**

- in **R** , if categorical variables are properly specified as factors, dummy coding is done by default
- To specify a variable as a factor:

```
library(car)
data(Duncan)
contrasts(Duncan$type)
```

```
##      prof wc
## bc      0  0
## prof    1  0
## wc      0  1
```

- It is easy to change the reference category in **R** :

```
type2 <- relevel(Duncan$type, ref="wc")
contrasts(type2)
```

```
##      bc prof
## wc    0    0
## bc    1    0
## prof  0    1
```


Assessing the Effects of Dummy Variables in **R** (1)

```
data(Duncan)
mod1<-lm(prestige~income+education+
    type, data=Duncan)
summary(mod1)
```

```
##
## Call:
## lm(formula = prestige ~ income + education + type, data = Duncan)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.890  -5.740  -1.754   5.442  28.972
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.18503    3.71377  -0.050  0.96051
## income        0.59755    0.08936   6.687 5.12e-08 ***
## education     0.34532    0.11361   3.040  0.00416 **
## typeprof     16.65751    6.99301   2.382  0.02206 *
## typewc      -14.66113    6.10877  -2.400  0.02114 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.744 on 40 degrees of freedom
## Multiple R-squared:  0.9131,Adjusted R-squared:  0.9044
## F-statistic:   105 on 4 and 40 DF,  p-value: < 2.2e-16
```


Assessing the Effects of Dummy Variables in **R** (2)

- The `lm` output suggests that the categorical variable `type` has a strong effect on "prestige".
- The incremental F-test confirms this finding

```
Anova(mod1)
```

```
## Anova Table (Type II tests)
##
## Response: prestige
##           Sum Sq Df F value    Pr(>F)
## income    4246.1  1 44.7201 5.124e-08 ***
## education  877.2  1  9.2388  0.004164 **
## type      3708.7  2 19.5302 1.208e-06 ***
## Residuals 3798.0 40
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


The Reference Category Problem

- Typically categorical variables in statistical models are reported in contrast to a reference category
  - It is then difficult to make inferences about differences between categories aside from the reference category
- Typical solutions:
  1. Refit the model with a different reference category
  2. Report the full variance-covariance matrix for the estimated parameters. A *standard error* between any two dummy regressors could then be easily calculated:

$$\mathrm{var}(aX + bY) = a^2\mathrm{var}(X) + b^2\mathrm{var}(Y) + 2ab\mathrm{cov}(X, Y)$$

- For a categorical variable with p levels, this would require reporting $\frac{p(p-1)}{2}$ covariances, making it difficult to do so if only because of space constraints.

Calculating Different Contrasts

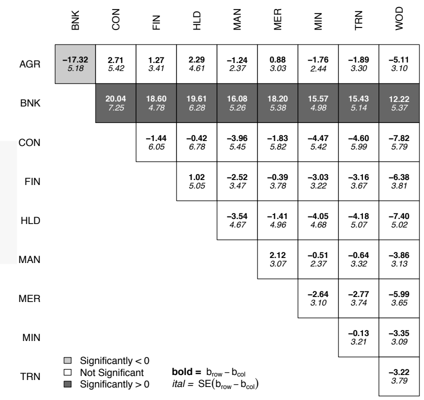
It is straightforward to calculate all pairwise comparisons.

```
data(Ornstein)
omod <- lm(interlocks ~
        nation + sector + log2(assets),
        data=Ornstein)
library(multcomp)
summary(glht(omod, linfct=mcp(nation = "Tukey")))


##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lm(formula = interlocks ~ nation + sector + log2(assets), data = Ornstein)
##
## Linear Hypotheses:
##                  Estimate Std. Error t value Pr(>|t|)
## OTH - CAN == 0   -3.053      3.087  -0.989    0.745
## UK - CAN == 0    -5.329      3.071  -1.735    0.294
## US - CAN == 0    -8.491      1.717  -4.944   <0.001 ***
## UK - OTH == 0    -2.276      3.865  -0.589    0.932
## US - OTH == 0    -5.438      3.018  -1.802    0.262
## US - UK == 0     -3.162      3.028  -1.044    0.711
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```


factorplot

```
library(factorplot)
ofp <- factorplot(
    omod,
    factor.variable="sector")
plot(ofp)
```

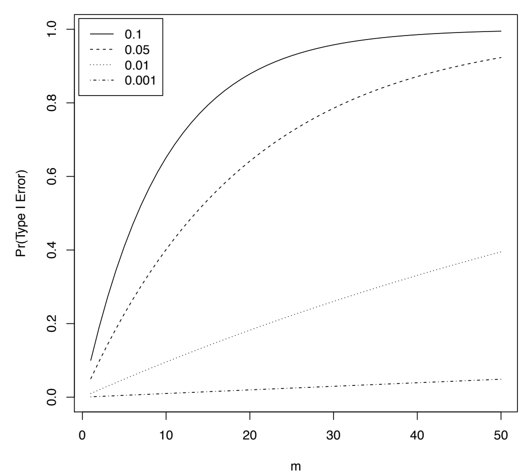


Multiplicity Problem (Bretz, Hothorn and Westfall [2011])

Usually, we choose to control Type I error rates when we test hypotheses, by evaluating a hypothesis, H, at a pre-specified significance level (α).

- Assume two hypotheses ($H = \{H_1, H_2\}$), both of which are true, and we are testing them independently, each at level $\alpha = 0.05$.
- The probability of not rejecting either hypothesis is $(1 - \alpha)^2 = 0.9025$
- The probability of falsely rejecting at least one test is $1 - (1 - \alpha)^2 = 0.0975$,
- The probability of falsely rejecting at least one test among a set of m tests $H = \{H_1, \ldots, H_m\}$ is $1 - (1 - \alpha)^m$.

Actual Type I Error Rates with Multiple Testing



Controlling for Multiple Testing

Hypotheses	Not Rejected	Rejected	Total
True	U	V	m_0
False	T	S	$m - m_0$
Total	W	R	m

Extending Type I Error to Multiple Tests

- Per-comparison Error Rate: PCER = $\frac{E(V)}{m}$ is the expected proportion of Type I errors among m comparisons. If tested independently, PCER = $\frac{\alpha m_0}{m} \leq \alpha$
- Family-wise Error Rate: FWER = $P(V > 0)$ is the probability of committing at least one Type I error.
  - Most commonly used measure, good when number of comparisons is moderate or where strong evidence is needed.
  - FWER approaches 1 as number of comparisons increases without a multiplicity adjustment
  - FWER reduces to the Type I error rate α when $m = 1$
  - A less strict version gFWER = $P(V > k)$, where the probability of making some small number (k) of Type I errors is acceptable.
- False Discovery Rate: If $Q = \frac{V}{R}$, the proportion of false rejections among all rejections. $FDR = E\left(\frac{V}{R} R > 0\right) P(R > 0)$. Extensions here abound and is an area of active research.

In general: PCER ≤ FDR ≤ FWER

Strong vs. Weak Control

- Control of Type I error rate is considered *weak* if the Type I error rate is controlled only under the global null hypothesis (i.e., assuming $H_1, \ldots, H_m$ are all true)
- Control of Type I error rate is considered *strong* if the Type I error rate is controlled under any configuration of true null hypotheses (except for the null set).
- Controlling FWER in the strong sense is the most stringent (i.e., conservative) test.

Single-step vs. Stepwise Procedures

- In single-step procedures, the information about rejecting or not rejecting one hypothesis does not enter into the decision for another. (Example: Bonferroni)
- In stepwise procedures (different from and decidedly less controversial than "stepwise regression"), hypotheses are ordered (in a potentially data-dependent fashion) and either:
  - In a step-down procedure, hypotheses are rejected until the first non-rejection and then all others are retained. (Example: Holm)
  - In a setp-up procedure, hypotheses are retained until the first rejection then all others are rejected. (Example: Hochberg)

Adjusted p-values

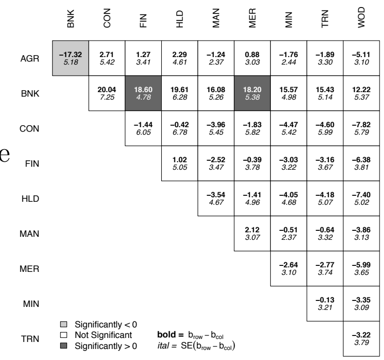
p-values can be calculated adjusting for any multiple comparison procedure mentioned above. The adjusted p-value for test i (call them q_i) take the form:

$$q_i = \inf\{\alpha \in (0,1) | H_i \text{ is rejected at level } \alpha\}$$

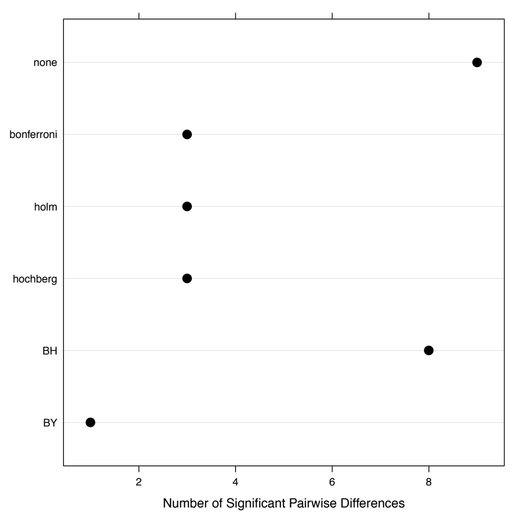
- To control FWER in the strong sense, Bonferroni (single-step), Holm (step-down) and Hochberg (step-up) are options, though Holm's method is known to dominate Bonferroni's under a set of minimally restrictive assumptions.
- To control FDR, Benjamini-Hochberg (BH) works under the assumption of independent tests and Benjamini-Yekuteli (BY) works when independence cannot be assumed.

Multiplicity Correction

- Above, we tested 45 hypotheses simultaneously, so 5% (or ≈ 2) could will be significant "by chance".
- The Holm correction sets the α for the entire set of tests equal to the desired rate by setting the α for each individual test to $\frac{\alpha}{n}$ where n is the number of comparisons.



Different Corrections



Factorplot Summary

```
summary(ofp2)

##      sig+ sig- insig
## AGR    0    1     8
## BNK    3    0     6
## CON    0    0     9
## FIN    0    1     8
## HLD    0    0     9
## MAN    0    0     9
## MER    0    1     8
## MIN    0    0     9
## TRN    0    0     9
## WOD    0    0     9
```

Factorplot Print

```
print(ofp2, sig=T)

##             Difference    SE p.val
## AGR - BNK    -17.323 5.185 0.042
## BNK - FIN     18.597 4.784 0.006
## BNK - MER     18.203 5.377 0.037
```


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:

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


Model Summary

```r
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.

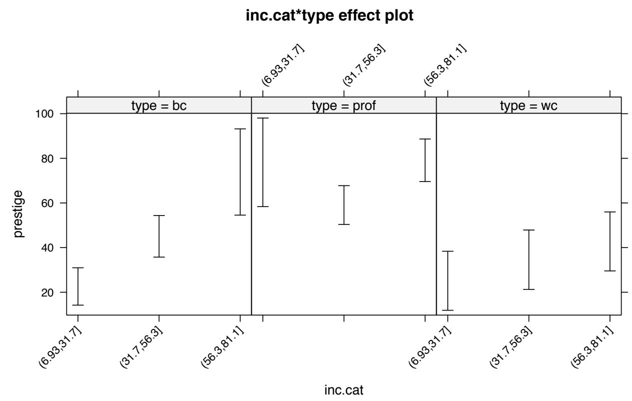
```r
Anova(mod)
```

```
## 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
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Two Categorical Variables - Plot

Q2: What is the nature of the interaction effect?

```r
library(effects)
e1 <- effect("inc.cat*type", mod)
pe1 <- plot(e1, as.table=T, layout=c(3,1),rotx=45,
     colors=c("black", "black"), lwd=NA)
pe1
```



Presenting two categorical variables

```r
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) (11.91,38.36)
## (56.3,81.1] 45.05         59.08         34.58
##             (35.74,54.36) (50.40,67.76) (21.28,47.88)
## (6.93,31.7] 73.87         79.12         42.78
##             (54.54,93.20) (69.58,88.66) (29.57,55.99)
```


Interpretation

The important points are as follows:

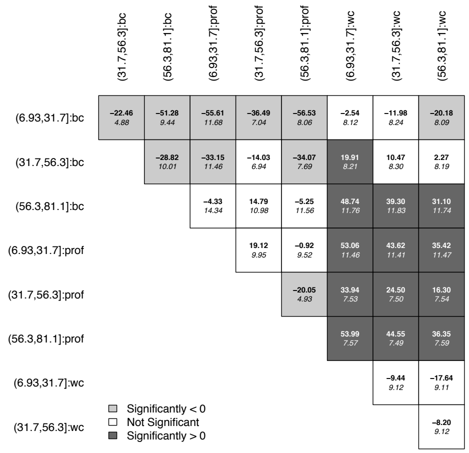
- The interaction term is significant in the F-test, so that indicates a significant interaction effect.
- With no interaction effect, the across each row have the same pattern across the three different tows and down the three different columns.
    - While the trends overall look somewhat different 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`.

```r
fp <- factorplot(e1)
plot(fp, print.square.leg=F,
  scale.text=.75, abbrev.char=100)
```

Factorplot figure



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.

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

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:

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

```
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(31.7,56.3]:t
##
## 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
```


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:

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

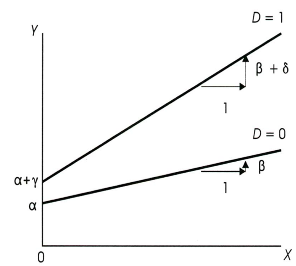


Figure 7.5 from Fox (1997)

Example with one Dummy Variable and One Continuous Variable

```
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:

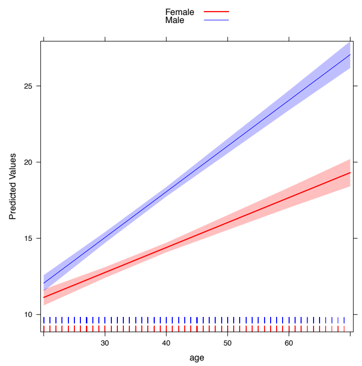
```
intQualQuant(mod, c("age", "sex"), type="slopes", plot=F)

## 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...

```r
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)
```



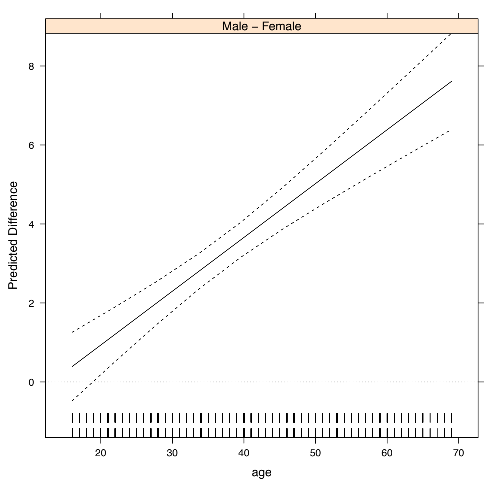
The effect of Gender

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?

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

```r
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 `intQualQuant` function (from the `DAMisc` 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 '.' 0.1 ' ' 1
```

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?

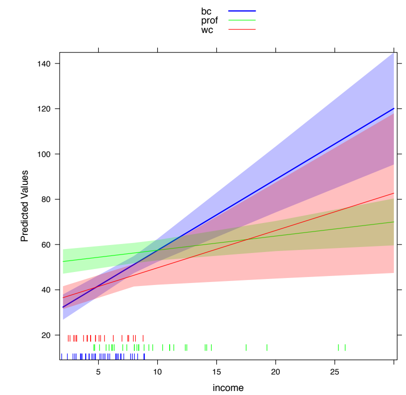
- The nature of the interaction has to be considered both for `income` and for `type`.
- We can calculate the conditional effects and variances of `income` as follows:

```
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`)?

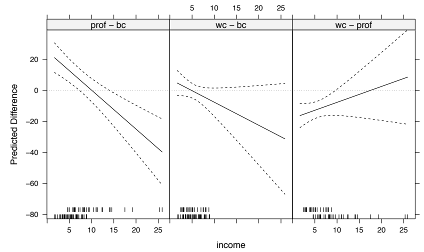
- The conditional effect of type (as we saw) is a bit more difficult. Here, We would presumably have to test each pairwise difference: BC vs Prof, BC vs WC and Prof vs WC for different values of education. First, let's think about what we need.

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.

Interaction Example

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

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

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_2 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.

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,

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

Testable Hypotheses

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

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?

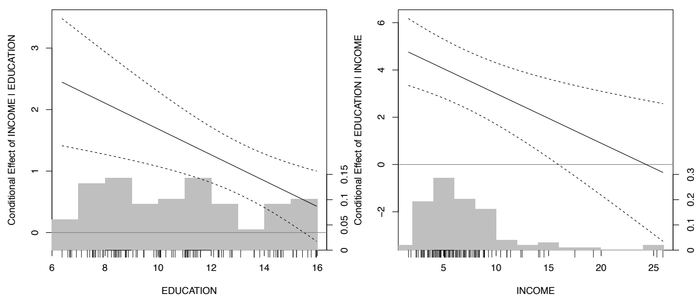
- The `income:education` line answers this question. If it is significant, then there is a significant interaction, otherwise there is not.
- This is counter to a minor, though still influential, point in Brambor, Clark and Golder (2006), but is consistent with Berry, Golder and Milton (2012).
- In this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



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.

When Confidence Bounds Equal Zero

You may want to know when the confidence bounds are equal to zero. Consider the equation:

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

- We know that the conditional effect of X_1 is $\beta_1 + \beta_4 X_2$ and that the lower bound is $(\beta_1 + \beta_4 X_2) - 1.96 \times SE(\beta_1 + \beta_4 X_2)$.
- Since those are all quantities that we know (or estimate), we could set it equal to zero and solve.
- This is what the `changeSig` function does.

Change in Significance

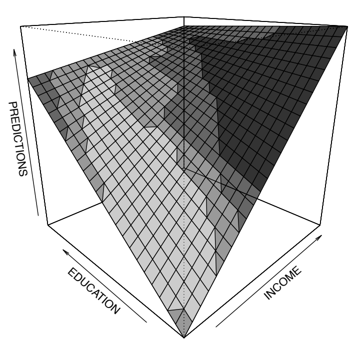
```
changeSig(mod, c("income", "education"))

## LB for B(income | education) = 0 when education=15.4979 (95th pctile)
## UB for B(income | education) = 0 when education=27.9396 (> Maximum Value in Data
## LB for B(education | income) = 0 when income=15.9273 (96th pctile)
## UB for B(education | income) = 0 when income=59.5175 (> Maximum Value in Data)
```


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.

```
##                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
```

Categorical Explanatory Variables
- Dumy Variables
- Multi-Category EV's

Interaction Effects
- Two Categorical Variables
- One Dummy Variable, One Continuous Variable
- One Categorical and One Continuous
- Two Continuous Variables
- Centering and Interactions

Relative Importance

Centering and Interactions

- Let's assume we have the following model:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon_i$$

Where $\beta = \begin{bmatrix} 2 \\ 3 \\ -4 \\ 3 \end{bmatrix}$ and $X \sim \mathcal{N}_2(\boldsymbol{\mu}, \Sigma)$

$\boldsymbol{\mu} = \begin{bmatrix} 10 \\ 10 \end{bmatrix}$ and $\Sigma = \begin{bmatrix} 1.0 & 0.4 \\ 0.4 & 1.0 \end{bmatrix}$

- Both X variables are always positive and correlated at a reasonable level. Let's see what happens to the fitted values and coefficients when we mean-center them.

Mean Centering (1)

- Let's see what mean centering does to the interaction:

No Mean Centering			Mean Centering		
X_1	X_2	$X_1 \times X_2$	X_1	X_2	$X_1 \times X_2$
20	20	400	9	9	81
10	20	200	-1	9	-9
1	20	20	-10	9	-90
20	10	200	9	-1	-9
10	10	100	-1	-1	1
1	10	10	-10	-1	10
20	1	20	9	-10	-90
10	1	10	-1	-10	10
1	1	1	-10	-10	100

- So, mean centering is definitely changing the ordering of your variables. Let's see if it actually gives us different answers

Mean Centering (2)

	Not Cent	Cent
(Intercept)	23.55*	−1.29*
	(11.75)	(0.13)
x1	0.69	32.90*
	(1.18)	(0.13)
x2	−6.07*	26.14*
	(1.18)	(0.13)
x1:x2	3.22*	3.22*
	(0.12)	(0.12)
N	1000	1000
R^2	0.99	0.99
adj. R^2	0.99	0.99
Resid. sd	3.91	3.91

Standard errors in parentheses
* indicates significance at $p < 0.05$

	No Cent	Cent
x1	90.54	1.19
x2	90.73	1.19
x1:x2	251.45	1.00

Table: VIF Statistics

Conditional Effect of X

- Since we've moved the X's around, we need to consider not the effects in the model, but the conditional effects holding the X's at the same places *relative* to their respective distributions, for instance:

	x1	x1 (cent)	x2	x2 (cent)
25th	9.34	-0.66	9.31	-0.69
50th	10.00	-0.00	10.02	0.02
75th	10.64	0.64	10.71	0.71

- Now, we can look at the conditional effects of X_1 and X_2 at the given values above:

	eff x1	eff x1 (cent)	eff x2	eff x2 (cent)
25th	30.68	30.68	24.03	24.03
	0.16	0.16	0.15	0.15
50th	32.96	32.96	26.13	26.13
	0.14	0.14	0.14	0.14
75th	35.18	35.18	28.21	28.21
	0.16	0.16	0.15	0.15

Table: Conditional Effects of x1 and x2

Determining Relative Importance

- If two explanatory variables are measured in exactly the same units, we can asses their relative importance in their effect on y quite simply
  - The larger the coefficient, the stronger the effect
- Often, however, our explanatory variables are not all measured in the same units, making it difficult to assess relative importance
- This problem can be overcome for quantitative variables by using standardized variables
- We can generalize standardization to include sets of variables, thus incorporating factors, interactions, and multiple effects

Standardized Regression Coefficients

- Standardized coefficients enable us to compare the relative effects of two or more explanatory variables that have different units of measurement
- Standardized coefficients convert all the variables into standard deviation units:

$$\frac{Y_i - \bar{Y}}{S_y} = \left(B_1 \frac{S_1}{S_y}\right) \frac{X_{i1} - \bar{X}_1}{S_1} + \cdots + \left(B_k \frac{S_k}{S_y}\right) \frac{X_{ik} - \bar{X}_k}{S_k} + \frac{E_i}{S_y}$$

- We interpret the effects of a standardized variable as the average number of standard deviation units Y changes with an increase in one standard deviation in X
- Since they don't have a standard deviation, standardized coefficients for factors are meaningless

Standardized Coefficients using Matrices

- Recall that the matrix equation for the least-squares slopes is:

$$\boldsymbol{B} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

  where $\boldsymbol{X}'\boldsymbol{X}$ is the variance-covariance matrix
- The matrix equation for the standardized coefficients is then:

$$\boldsymbol{b*} = \boldsymbol{R}_{XX}^{-1}\boldsymbol{r}_{X\boldsymbol{y}}$$

  where $\boldsymbol{R}_{XX}$ is the correlation matrix between the explanatory variables and $\boldsymbol{r}_{X\boldsymbol{y}}$ is the vector of correlations between the explanatory variables and the outcome variable.

Standardized Variables: Cautions

- It makes little sense to standardized dummy variables:
  - It cannot be increased by a standard deviation so the regular interpretation for standardized coefficients does not apply
  - Moreover, the standard interpretation of the dummy variable showing differences in level between two categories is lost
- We cannot standardize interaction effects
  - They are dependent on the main effects
  - We can, however, standardize quantitative variables beforehand and construct interaction terms afterwards.

Standardized Variables in **R**

- Unlike some statistical packages, **R** does not automatically return standardized coefficients
- A separate model must be fitted to a dataset for which all quantitative variables have been standardized
  - This is done using the scale function.
  - Variables can be standardized individually:

    ```
    > gini.std <- scale(gini)
    ```
  - Alternatively, all the quantitative variables can be standardized at the same time by creating a new scaled dataset (from the `DAMisc` package):

    ```
    > scaled.data <- scaleDataFrame(Duncan)
    ```


Relative Importance of a Set of Predictors (1)

- In the standardized variables case, we can easily determine relative importance by the ratio of the two standardized coefficients
  - In other words, we assess the ratio of the standard deviations of the two contributions to the linear predictor
- Imagine now that we are interested in the relative effects of two sets of variables (e.g., a set of dummy regressors for a single variables versus the effects of another variable)
  - Instead of individual standardize variables, we explore the relative contributions that the set of variables make to the dispersion of the fitted values

Relative Importance of a Set of Predictors (2)

- Following from Silber et al. (1995) the ratio of variances of the contributions of two sets of variables, ω, can be determined by:

$$\omega = \frac{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}}{\boldsymbol{\gamma}'\boldsymbol{H}'\boldsymbol{H}\boldsymbol{\gamma}}$$

  Where β is the coefficient vector and $\boldsymbol{X}$ is the model matrix for the *set1 predictors*; γ is the coefficient vector and $\boldsymbol{H}$ is the model matrix for the *set2 predictors*
- If $\omega = 1$, then both sets of predictors contribute the same amount of variation in the outcome variable
- MLE also provides an approximate test of $H_0 : \omega = 1$ which refers to the standard normal distribution, yielding the standard confidence intervals, thus making the test generalizable to GLMs

The `relimp` Package in **R**

- The `relimp` package for **R** implements the ω measure of relative importance of Silber et al.
- The variables of interest can be specified in a command line, with each effect given the number corresponding to its column(s) in the model matrix (or row in the regression output)

```
> library(relimp)
> relimp(model, set1=1:3, set2=4:5)
```


Relative Importance: An Example (1)

```
summary(mod1)$coefficients
```

```
##               Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) -28.4429670  4.9271875 -5.7726578 2.465024e-08
## log(assets)   5.9907825  0.6813797  8.7921354 3.235865e-16
## sectorBNK    17.3227304  5.1846800  3.3411378 9.710771e-04
## sectorCON    -2.7126874  5.4241073 -0.5001168 6.174628e-01
## sectorFIN    -1.2744881  3.4121039 -0.3735197 7.090998e-01
## sectorHLD    -2.2916036  4.6132359 -0.4967454 6.198350e-01
## sectorMAN     1.2440168  2.3665722  0.5256619 5.996209e-01
## sectorMER    -0.8801086  3.0346472 -0.2900201 7.720577e-01
## sectorMIN     1.7566138  2.4447619  0.7185214 4.731527e-01
## sectorTRN     1.8888418  3.3023169  0.5719747 5.678882e-01
## sectorWOD     5.1056070  3.0990366  1.6474820 1.008012e-01
## nationOTH    -3.0533129  3.0872167 -0.9890180 3.236759e-01
## nationUK     -5.3294006  3.0714272 -1.7351544 8.403005e-02
## nationUS     -8.4912938  1.7174063 -4.9442544 1.458432e-06
```


Relative Importance: An Example (2)

```
# load relimp package
library(relimp)
relimp(mod1, set1=3:11, set2=12:14)
```

```
##
## Relative importance summary for model
##     lm(formula = interlocks ~ log(assets) + sector + nation, data = Ornstein)
##
##        Numerator effects ("set1")      Denominator effects ("set2")
## 1                       sectorBNK                          nationOTH
## 2                       sectorCON                           nationUK
## 3                       sectorFIN                           nationUS
## 4                       sectorHLD
## 5                       sectorMAN
## 6                       sectorMER
## 7                       sectorMIN
## 8                       sectorTRN
## 9                       sectorWOD
##
## Ratio of effect standard deviations: 0.858
## Log(sd ratio):                 -0.153   (se 0.314)
##
## Approximate 95% confidence interval for log(sd ratio): (-0.768,0.461)
## Approximate 95% confidence interval for sd ratio:      (0.464,1.586)
```

In this instance, the ω measure suggests these two sets of predictors contribute equally to the variation in the dependent variable.

Recap

Categorical Explanatory Variables
- Dumy Variables
- Multi-Category EV's

Interaction Effects
- Two Categorical Variables
- One Dummy Variable, One Continuous Variable
- One Categorical and One Continuous
- Two Continuous Variables
- Centering and Interactions

Relative Importance

Readings

Today:

* Firth (2003)
– Firth and Menzes (2004)
* Brambor, Clark and Golder (2006)
* Braumoeller (2004)
– Kam and Franzese (2007)
* Silber, Rosenbaum and Ross (1995)

Brambor, Thomas, William Clark and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14(1):63–82.

Braumoeller, Bear F. 2004. "Hypothesis Testing and Multiplicative Interaction Terms." *International Organization* 58(4):807–820.

Firth, David. 2003. "Overcoming the Reference Category Problem in the Presentation of Statistical Models." *Sociological Methodology* 33:1–18.

Firth, David and Renee X. De Menzes. 2004. "Quasi-Variances." *Biometrika* 91(1):65–80.

Kam, Cindy and Robert J. Franzese. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analyses.* Ann Arbor: University of Michigan Press.

Silber, Jeffrey H., Paul R. Rosenbaum and Richard N. Ross. 1995. "Comparing the Contributions of Groups of Predictors: Which Outcomes Vary with Hospital Rather Than Patient Characteristics." *Journal of the American Statistical Association* 90(429):7–18.