Regression III: Advanced Methods

William G. Jacoby
Department of Political Science
Michigan State University

http://polisci.msu.edu/j Jacoby/icpsr/regress3
Interaction Effects (1)

- Dummy variables test for differences in *level* but not in *slope*.
- We may also be interested in whether slopes differ, however.
  - For example, we may want to test whether age affects are different for men and women.
- When the *partial* effect of one variable depends on the value of another, the two variables are said to *interact*
  - 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 $XD$:

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

  $$\text{income}_i = \alpha + \beta \text{age}_i + \gamma \text{men}_i + \delta (\text{age}_i \times \text{men}_i) + \varepsilon_i$$

- The parameters $\alpha$ and $\beta$ are the intercept and slope for the reference group (i.e., the category coded 0 for the dummy regressor for gender, in this case women)
- The intercept for the other group (men) is $\alpha + \gamma$; the slope is $\beta + \delta$
• To clarify this, we can write out the equations as follows:

for $D = 0$ (women):
\[
Y_i = \alpha + \beta X_i \\
+ \gamma(0) + \delta(X_i \times 0) + \varepsilon_i \\
= \alpha + \beta X_i + \varepsilon_i
\]

for $D = 1$ (men):
\[
Y_i = \alpha + \beta X_i \\
+ \gamma(1) + \delta(X_i \times 1) + \varepsilon_i \\
= (\alpha + \gamma) + (\beta + \delta) X_i + \varepsilon_i
\]

• Unlike in the non-interaction model the regression lines are not parallel so we cannot interpret $\gamma$ as the unqualified partial effect of gender controlling for age (it is the effect only at $X=0$). Similarly, $\beta$ is not the unqualified effect of age controlling for gender (it is the effect only for women)
Hypothesis Tests for Interactions

• An **Incremental F-test** gives an hypothesis test for a set of interaction terms
• Using the Duncan data, assume the model

\[
\text{prestige}_i = \alpha + \beta_1 \text{income}_i + \gamma_1 D_{i1} + \gamma_2 D_{i2} + \delta_1 \text{income}_i D_{i1} + \delta_2 \text{income}_i D_{i2}
\]

• Here \( D_1 \) and \( D_2 \) represent dummy regressors for professional and white collar occupations respectively (blue collar is the reference category)
• The terms \textbf{income}*D_1 \textbf{ and income}*D_2 \textbf{ capture the interaction between income and occupation type}

> library(car)
> data(Duncan)
> Duncan.model<-lm(prestige~income+type, data=Duncan)
Output from Duncan model

|                        | Estimate | Std. Error | t value | Pr(>|t|) |
|------------------------|----------|------------|---------|----------|
| (Intercept)            | 2.6828   | 3.8182     | 0.70    | 0.4865   |
| income                 | 0.8450   | 0.1289     | 6.55    | 0.0000   |
| typeprof               | 44.4868  | 10.3641    | 4.29    | 0.0001   |
| typewc                 | 14.8531  | 13.4986    | 1.10    | 0.2779   |
| income:typeprof        | −0.2909  | 0.2017     | −1.44   | 0.1571   |
| income:typewc          | −0.4674  | 0.2736     | −1.71   | 0.0955   |

- From the *t*-tests for the individual coefficients we see that income has a strong and statistically significant effect on prestige.
- Professional jobs are also have higher levels of prestige than blue collar jobs.
- The interaction terms are not statistically significant, but we proceed as if they are for purpose of demonstrating a global test for the interaction using an incremental F-test (compare the RegSS).
Incremental F-tests using the Anova function in \texttt{car}

\begin{verbatim}
> Duncan.model<-lm(prestige~income*type)
> Anova(Duncan.model)

Anova Table (Type II tests)

Response: prestige

\begin{tabular}{cccccc}
 & Sum Sq & Df & F value & Pr(>F) \\
income & 5922.4 & 1 & 54.3020 & 6.710e-09 *** \\
type & 8347.6 & 2 & 38.2693 & 6.347e-10 *** \\
income:type & 421.7 & 2 & 1.9331 & 0.1583 \\
Residuals & 4253.5 & 39 & & \\
\end{tabular}
\end{verbatim}

- Here we have better evidence that the interaction between income and occupation type is not statistically significant
The Importance of Fitted Values

- Often reporting coefficients and standard errors is not very helpful
- In such cases, fitted values are very helpful
- **Interaction effects**
  - Especially helpful for logit models, but also for linear models
- **Nonlinear relationships**
  - Polynomial regression, logit models
    - Coefficients are difficult to interpret
  - Nonparametric regression
    - There are no parameters to interpret, so fitted values *must* be plotted on graphs
Fitted Values from Linear Models (1)

1. Select the variable, $X_1$, for which you want to determine the effect.

2. In the *fitted equation*, substitute a “typical” value (e.g., the mean) for all $X$’s except for $X_1$.

   $$\hat{\text{prestige}} = 2.69 \text{constant} + .357 \text{education} + .406 \text{income}$$
   $$\quad - 31.7 \text{bc} - 33.3 \text{wc}$$
   $$\quad + .369 (\text{income} \times \text{bc}) + .009 (\text{income} \times \text{wc})$$

3. Find values of $Y$ at values through the range of $X_1$.
   - If I’m interested in the effects of income for each type of occupation, I’d set education to its mean, but let income and type taken on values through their ranges.
   - I’d find three sets of fitted values for the effects on income: one for each type.
For linear models, the fitted values are easily interpreted because they are in the metric of $Y$.

- If we wish to compare effects for several $X$s that are all categorical, the fitted values for each category can be put in a table. Likewise, interactions between categorical variables can also be displayed in a table.

- If $X_1$ is a quantitative variable, a graph (Effect Display) of $Y$ plotted against $X_1$ is very effective. If interested in an interaction with a quantitative predictor, we will plot several lines.

Confidence envelopes for the line can also be constructed in the same manner from the standard errors.
Fitted Effects in Matrix Form

• Consider the general linear model:

\[ y = X\beta \]

• Let \( X^* \) include all combinations of values of predictors appearing in a higher order term, and “typical” values for all other terms. Here the structure of \( X^* \) is the same as it is for \( X \)

• The fitted values from the equation below represent the effect of interest

\[ \hat{y}^* = X^*b \]

• The standard errors of these fitted values are the square-root of the diagonal entries of

\[ X^*V(\hat{b})X^{*'} \]
Fitted Values and Interactions between Categorical Variables

• Fitted values are easily calculated in R using the `all.effects` command in the `effects` package
  - This function returns the fitted values and 95% confidence band for the fitted values

```r
> attach(Prestige)
> model<-lm(prestige~income+type)
> model.effects<-all.effects(model)
> summary(model.effects)
```

• If we are interested in categorical variables only, we need proceed no further. If we want to understand a quantitative variable (especially an interaction) we should use **effect displays**
Effect Displays for Linear Model Interactions involving quantitative variables

• Plotting with `all.effects` can give two types of effect displays:
  – A display that graphs all effects on a single graph (excludes confidence envelopes)
  – Separate displays for each effect that include confidence envelopes

• Effect displays can also display interactions between two quantitative variables.
  – In these cases, effects for one of the variables are displayed at set levels of the other (by default, `all.effects` places the variable with the largest number of categories on the horizontal axis)
Effect Displays in R (1): Code for effect display

> library(effects)
> lset(col.whitebg()) #makes the background white
> plot(all.effects(mod2))#plot the effects of income

1:education
2:income*type

Selection:
Enter an item from the menu, or 0 to exit
Selection: 2
Effect Displays in R (2)

income*type effect plot

prestige

income

type: wc

type: bc

type: prof
Effect Displays in R (3)

> plot(all.effects(mod2), multiline=TRUE)