Conditional Relationships in the Data

Marco Morales
mam2519@columbia.edu

GR5069
Topics in Applied Data Science
for Social Scientists

Spring 2018
Columbia University

Conditional Relationships in the Data

- when analyzing people and behaviors, we're not only concerned about **levels**
- we typically care about behaviors **conditional** on something else happening
  - do incumbent presidents lose elections when shark attacks increase?
- note that this is **different from "holding the rest constant"**
- can be easily computed through **multiplicative interactions**

Conditional Relationships in the Data

Use multiplicative interaction terms to model conditional relationships

- a typical case of **describing the data generating mechanism** through a statistical model
- we start with a simple model...

$$Y = \beta_0 + \beta_X \mathbf{X} + \beta_Z \mathbf{Z} + +\epsilon$$

- ... and add the **multiplicative interaction** term

$$Y = \beta_0 + \beta_X \mathbf{X} + \beta_Z \mathbf{Z} + \beta_{XZ} \mathbf{XZ} + \epsilon$$

- that now accounts for the conditional relationship between X and Z

Conditional Relationships in the Data

Use multiplicative interaction terms to model conditional relationships

Hypothesis H_1: An increase in X is associated with an increase in Y when condition Z is met, but not when condition Z is absent.

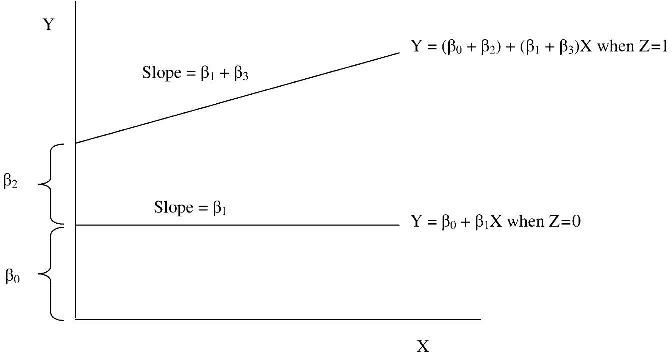

Fig. 1 A graphical illustration of an interaction model consistent with hypothesis H_1.

Figure: Brambor et al. (2006)

Conditional Relationships in the Data

Differences between additive and conditional models

- a linear **additive model** assumes a **constant effect** of X on Y

$$Y = \beta_0 + \beta_X \mathbf{X} + \beta_Z \mathbf{Z} + + \epsilon$$

- an **interactive model** assumes that the effect of X on Y **depends on the value of $\mathbf{Z}$**

$$Y = \beta_0 + \beta_X \mathbf{X} + \beta_Z \mathbf{Z} + \beta_{XZ} \mathbf{XZ} + \epsilon$$

Conditional Relationships in the Data

Include all constitutive terms

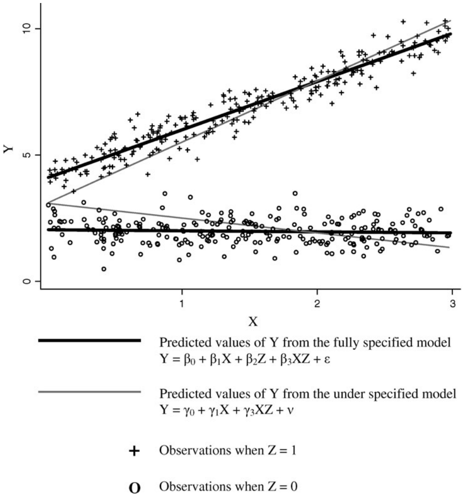

Fig. 2 An illustration of the consequences of omitting a constitutive term.

Figure: Brambor et al. (2006)

Conditional Relationships in the Data

Computing and interpreting marginal effects

- from the (interactive) model

$$Y = \beta_0 + \beta_X \mathbf{X} + \beta_Z \mathbf{Z} + \beta_{XZ} \mathbf{XZ} + \epsilon$$

- we could be interested in the marginal effect of X given Z on Y

$$\frac{\partial E[Y|X,Z]}{\partial \mathbf{X}} = \beta_X + \beta_{XZ} \mathbf{Z}$$

Conditional Relationships in the Data

Computing and interpreting marginal effects

$$\frac{\partial E[Y|X,Z]}{\partial \mathbf{X}} = \beta_X + \beta_{XZ}\mathbf{Z}$$

- it is **wrong** to assume that β_{XZ} is the **marginal effect** of X given Z on Y
  - β_{XZ} is the effect of Z on Y when $X = 0$
  - β_X is the effect of X on Y when $Z = 0$
- **marginal effects** are **composite quantities**

Conditional Relationships in the Data

Always compute meaningful standard errors

- interactions are have also an **associated uncertainty**
- so, in addition to the marginal effect

$$\frac{\partial E[Y|X,Z]}{\partial \mathbf{X}} = \beta_X + \beta_{XZ}\mathbf{Z}$$

- we need to compute **appropriate standard errors** as well

$$Var\left(\frac{\partial \hat{E}[Y|X,Z]}{\partial \mathbf{X}}\right) = Var[\hat{\beta}_X] + \mathbf{Z}^2 Var[\hat{\beta}_{XZ}] + 2\mathbf{Z}Cov[\hat{\beta}_X, \hat{\beta}_{XZ}]$$

Conditional Relationships in the Data

Conditional effects: an example

- going back to our example:
  - **are there more expected deaths when combat is heavier?**
    - let's look at the case of events where the Navy is involved
    - we'd need to assume that more seized heavy weapons indicate heavier combat and compute

$$\beta_{navy} + \beta_{navy,long\_guns\_seized} * long\_guns\_seized$$

  - **are there less expected number of deaths when no weapons are seized?**
    - let's look at the case of the Army
    - we maintain the same assumption and compute

$$\beta_{army}$$

Conditional Relationships in the Data

Conditional effects: an example

```
Call:
lm(formula = organized.crime.dead ~ organized.crime.wounded +
    afi * long.guns.seized + army * long.guns.seized + navy *
    long.guns.seized + federal.police * long.guns.seized + afi *
    cartridge.sezied + army * cartridge.sezied + navy * cartridge.sezied +
    federal.police * cartridge.sezied + small.arms.seized + clips.seized,
    data = AllData)

Residuals:
    Min      1Q  Median      3Q     Max
-8.6509 -0.7385 -0.4189  0.1933 27.2187

Residual standard error: 1.714 on 5378 degrees of freedom
Multiple R-squared:  0.1587,Adjusted R-squared:  0.156
F-statistic: 59.67 on 17 and 5378 DF,  p-value: < 2.2e-16
```

Conditional Relationships in the Data

Conditional effects: an example

```
Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      0.4188645  0.0336777  12.437  < 2e-16 ***
organized.crime.wounded          0.3624050  0.0237796  15.240  < 2e-16 ***
afi                             -0.0419271  0.5040535  -0.083   0.9337    
long.guns.seized                 0.1713811  0.0172327   9.945  < 2e-16 ***
army                             0.4244453  0.0556353   7.629 2.78e-14 ***
navy                             0.2772627  0.1567621   1.769   0.0770 .  
federal.police                  -0.1113463  0.0801781  -1.389   0.1650    
cartridge.sezied                 0.0002292  0.0000968   2.368   0.0179 *  
small.arms.seized               -0.0452969  0.0186014  -2.435   0.0149 *  
clips.seized                     0.0003127  0.0003146   0.994   0.3202    
afi:long.guns.seized             0.0229013  0.0784035   0.292   0.7702    
long.guns.seized:army           -0.0459567  0.0181403  -2.533   0.0113 *  
long.guns.seized:navy            0.1761160  0.0421782   4.176 3.02e-05 ***
long.guns.seized:federal.police -0.0253811  0.0190541  -1.332   0.1829    
afi:cartridge.sezied            -0.0050516  0.0031231  -1.617   0.1058    
army:cartridge.sezied           -0.0003911  0.0000981  -3.987 6.78e-05 ***
navy:cartridge.sezied           -0.0006909  0.0001728  -3.998 6.47e-05 ***
federal.police:cartridge.sezied -0.0001518  0.0001102  -1.377   0.1685    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1  ' ' 1
```

Conditional Relationships in the Data

Conditional effects: an example

- marginal effect of 5 seized long guns on the expected number of dead on events that involve the Navy $(\beta_{navy} + \beta_{navy,long\_guns\_seized} * 5)$

  1.15
  [0.74, 1.56]

- marginal effect on the expected number of dead of events that involve the Army when no long guns (zero) are seized $(\beta_{army} + \beta_{army,long\_guns\_seized} * 0)$

  0.42
  [0.31, 0.53]

Conditional Relationships in the Data

- Always, always, always remember (Brambor et al. 2006):
    1. Use multiplicative interaction models **whenever one's hypothesis is conditional** in nature.
    2. Include **all constitutive terms** in the model specification.
    3. **Do not interpret the coefficients on constitutive terms as if they are unconditional marginal effects.**
    4. Do not forget to **calculate substantively meaningful marginal effects and standard errors**.
- ... or face the wrath of the stats gods!

Conditional Relationships in the Data

A note on interactions and classifiers

- interactions and interactive effects are a **lesser concern for prediction/classification**
  - relevant to inferential methods that seek to describe **mechanics** of a process
- most **classifiers** can **identify interactions automatically**
  - interactions are already included in their predictions/classifications

Conditional Relationships in the Data

Marco Morales
mam2519@columbia.edu

GR5069
Topics in Applied Data Science
for Social Scientists

Spring 2018
Columbia University