Introduction to Multivariate Regression & Program Evaluation

HED 612

Lecture 15

1. End of Semester Logistics
2. Interaction Effects Continued...
3. Interactions between two categorical variables
4. Interactions by Two Continuous Variables
5. THANK YOU AND TAKE CARE OF YOURSELVES

Where are we going....

- Our last Lecture!
  - End of Semester House keeping items
  - Categorical by Categorical interactions
  - Continuous by Continuous interactions
  - Homework: Class Survey!
- Reading Day, "No Class" [5/7/2020]

`AER` R Package and Data!

We're going to try out a textbook I'm considering for HED 613 that comes with an accompanying R package

- Applied Econometrics with R, Christian Kleiber & Achim Zeileis
- `AER` Package
  - Comes with different functions and datasets!

Current Population Survey

- The Current Population Survey (CPS), sponsored jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS), is the primary source of labor force statistics for the population of the United States
- CORRECTION FROM LECTURE 14
  - `earnings` = average hourly earnings
  - `earnings` $\neq$ yearly income in \$000s

End of Semester Logistics

Grading

- Submit whatever homework assignments you can by Friday, May 8th, 2010
  - Your course grade will be based on what you can submit!
  - Don't stress trying to finish them all if you can't!
  - But be sure to download all materials sometime soon after finals!
- For those pursuing final projects...
  - I will absolutely consider final projects as "ongoing" if you'd like to keep working on them!
  - If you would like detailed feedback, submit what you have by May 8th
  - I am happy to work with you all through Summer and Fall 2020 as you continue working on these
- Next steps for those interested in quantitative research!
  - Continue taking classes in both statistics and data management!
    - Fall 2020: HED 696C Data Management and Manipulation in R
    - Spring 2021: HED 613 Regression modeling with non-continuous dependent variables
    - Classes in Sociology and other COE departments are good too!
  - Seek opportunities to practice these skills!

Student Course Survey (formerly Teacher Course Evaluations)

- Student course survey's have been formally canceled!
- But I would love your feedback (the good and the bad!)
- PLEASE TAKE THIS SURVEY
  - It contains same questions as the student course surveys
  - It is anonymous!
  - Your honest feedback will help me improve the course!

Interaction Effects Continued...

What are interaction effects?

- Simple hypothesis
  - X has an effect on Y
  - Ex: Participation in MAS (X) has an effect on graduation (Y)
- Interactions = Conditional Hypothesis
  - The effect of X on Y depends on a third variable
  - Ex: the effect of MAS participation (X) on graduation (Y) differs by race (Z)
- What is an interaction effect? (i.e., "moderators")
  - An interaction effect is when the relationship between two variables (X and Y) depends on the value of a third variable Z

Interaction Effects

- Interaction Effects are difficult!
  - It takes a while (and lots of practice!) to get comfortable thinking about and interpreting interaction effects
- I will provide an "introduction" to interaction effects!
  - We will continue to learn interaction effects in HED 613
  - If you don't take HED 613, seek out more opportunities to learn interaction effects!
  - Read empirical pieces that use/interpret interaction effects!

Three cases of interaction effects

1. **Interaction between a categorical and continuous variable** [last week]
   - Example: the effect of years of schooling (X) on earnings (Y) differs for men and women (Z)
2. **Interaction between two categorical variables** [today]
   - Example: the effect of having more than a high school diploma (X) on earnings (Y) differs for men and women (Z)
3. **Interaction between two continuous variables** [today]
   - Example: the effect of years of schooling (X) on earnings (Y) differs by age (Z)

Interactions between two categorical variables

Interaction between two categorical variables

RQ: Does the effect of having more than a HS diploma (X) on earnings (Y) differ by gender (Z)?

- Categorical by Categorical Interaction
  - Y = earnings
  - X = 0/1 more than a HS diploma
  - Z (interaction variable) = 0/1 Women (0=men)
- Simple hypothesis
  - Having more than a HS diploma affects earnings
- Conditional hypothesis (interaction effect)
  - The effect of having more than a HS diploma (X) on earnings (Y) differs for women and men

Simple Regression (no interaction effect)

- Simple Regression
  - What is the having more than a HS diploma on earnings?
  - Population regression model:
    - $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$
    - where Y= earnings, $X_1 = 0/1$ having more than a HS diploma
- Run model in R
- $\widehat{earnings} = \hat{\beta}_0 + \hat{\beta}_1 * morehs_i$
  - $\widehat{earnings} = 14.48726 + 6.54299 * morehs_i$
- $\hat{\beta}_1 = 6.54299$
  - On average, having more than a high school diploma as opposed to having a high school diploma or less is associated with a \$6.54 increase in hourly earnings
  - $\hat{\beta}_1$ is significant at the 0.000 level

Multivariate Regression (no interaction effect)

- Multivariate Regression
  - What is the effect of having more than a HS diploma on earnings controlling for gender?
  - Population regression model:
    - $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$
    - where Y= earnings, $X_1 = 0/1$ more than HS, $X_2 = 0/1$ women
- Run model in R
- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i}$
  - $\hat{Y}_i = 16.16394 + 6.81100 * X_{1i} - 4.17302 * X_{2i}$
- $\hat{\beta}_1 = 6.81100$
  - On average, having more than a high school diploma as opposed to having a high school diploma or less is associated with a $6.81 increase in hourly earnings, controlling for gender
  - $\hat{\beta}_1$ is significant at the 0.000 level
- $\hat{\beta}_2$ = -4.17302
  - On average, identifying as a woman as opposed to identifying as a man, is associated with a $4.17 decrease in hourly earnings, controlling for whether or not participant has more than a HS diploma
  - $\hat{\beta}_2$ is significant at the 0.000 level

Multivariate Regression (no interaction effect)

- The multivariate regression which only controls for gender...
  - Assumes that the effect of having more than a HS diploma is the same for men and women!
  - We can extend the model by allowing the effect of having more than a high school diploma to depend on gender by including an interaction between these two variables!
- Same steps as last week... [refresher!]
  - We will run separate models by different groups of Z
  - Then we will run interaction model

Separate Models for each group of Z

- What is the effect having more than a HS Diploma on earnings for women?
  - Run model in R
  - $\widehat{Y}_i = 12.31394 + 6.30489 * X_{1i}$
  - $\widehat{\beta_0}$: predicted earnings for women with a HS diploma or less
  - $\widehat{\beta_1}$: On average, having more than a HS diploma as opposed to having a HS diploma or less is associated with a \$6.31 increase in hourly earnings *for women*
  - Y| X=1: $12.31394 + 6.30489 * 1 =$ \$18.61883 hourly earnings
- What is the effect having more than a HS Diploma on earnings for men?
  - Run model in R
  - $\widehat{Y}_i = 15.94698 + 7.18773 * X_{1i}$
  - $\widehat{\beta_0}$: predicted earnings for men with a HS diploma or less
  - $\widehat{\beta_1}$: On average, having more than a HS diploma as opposed to having a HS diploma or less is associated with a \$7.19 increase in hourly earnings *for men*
  - Y| X=1: $15.94698 + 7.18773 * 1 =$ \$23.13471 hourly earnings

Interaction Model for Two Categorical Variables

- Let's run an interaction effect model to investigate whether the effect of having more than a HS diploma (X) on earnings (Y) differs by gender (Z)
- Population regression model
  - $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 Z_i + \beta_3 (X_{1i} * Z_i) + u_i$
    - where Y= earnings, X_1 = 0/1 More than HS, Z_i = 0/1 Women
    - $X_{1i} * Z_i$ = interaction for more than HS and gender

- OLS Prediction line without estimates
  - $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 Z_i + \hat{\beta}_3 (X_{1i} * Z_i)$

- $\hat{\beta}_0$ = $\hat{Y}_i$ **when X1=0 and Z=0**
  - Predicted earnings for men (Z=0) with a high school diploma or less (X1=0)
- $\hat{\beta}_1$ = **change in** $\hat{Y}_i$ **for X1=1 as opposed to X1=0, when Z=0**
  - Change in earnings for having more than a HS diploma (X1=1_) as opposed to having a HS diploma or less (X1=0) for men (Z=0)
- $\hat{\beta}_2$ = **change in** $\hat{Y}_i$ **for Z1=1 as opposed to Z1=0, when X1=0**
  - Change in earnings for women (Z=1) as opposed to men (Z=0) with a HS diploma or less (X1=0)
- $\hat{\beta}_3$ = **interaction term: how much the effect of X1 on** $\hat{Y}_i$ **changes when Z "increases by one unit" or when Z=1 as opposed to Z=0**
  - change in the effect of having more than a HS diploma on earnings for for Z=1 as opposed to Z=0

Interaction Model for Categorical by Categorical Interaction

- Run in R
- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 Z_i + \hat{\beta}_3 (X_{1i} * Z_i)$
- $\hat{Y}_i = 15.94698 + 7.18773 * X_{1i} - 3.63304 * Z_i - 0.88283 * (X_{1i} * Z_i)$

What do we want to know from interactions?

1. Is there an interaction effect?
2. What is the predicted value of Y for Z=0 and Z=1 at different values of X?
3. What is the effect of X on Y for different values of Z?

Is there an interaction effect?

- Is the effect of years having more than a high school degree significantly different for women vs men?
  - $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 Z_i + \beta_3 (X_{1i} * Z_i) + u_i$
- Hypothesis test
  - $H_0 : \beta_3 = 0$ vs $H_0 : \beta_3 \neq 0$
  - In other words, test whether the beta coefficient for the interaction term is significantly different from zero. If so, then there is an interaction!
- $\hat{\beta_3}$ = -0.88283 , p-value of 0.000
  - We reject H_0 at α of 0.05
  - There is a statistically significant interaction!
- Magnitude
  - If $\hat{\beta_3}$ is greater than zero (and statistically significant!) the effect of X on Y is larger when Z=1 than when Z=0
  - If $\hat{\beta_3}$ is less than zero (and statistically significant!) the effect of X on Y is smaller when Z=1 than when Z=0

What is the predicted value of Y for Z=0 and Z=1 at different values of X

- What is the predicted earnings (Y) for men (Z=0) with more than a HS diploma (X=1)?
  - $\hat{Y}_i = 15.94698 + 7.18773 * X_{1i} - 3.63304 * Z_i - 0.88283 * (X_{1i} * Z_i)$
  - Z=0 & X=1: $15.94698 + (7.18773 * 1) - (3.63304 * 0) - (0.88283 * (1 * 0))$
  - $\$23.13 = 15.94698 + (7.18773)$
- What is the predicted earnings (Y) for women (Z=1) with more than a HS diploma (X=1)?
  - $\hat{Y}_i = 15.94698 + 7.18773 * X_{1i} - 3.63304 * Z_i - 0.88283 * (X_{1i} * Z_i)$
  - Z=1 & X=1: $15.94698 + (7.18773 * 1) - (3.63304 * 1) - (0.88283 * (1 * 1))$
  - $\$18.62 = 15.94698 + (7.18773) - (3.63304) - (0.88283)$

- Same result as when we ran model separately by sample!
  - Men: $15.94698 + 7.18773 * 1 = 23.13471$
  - Women: $12.31394 + 6.30489 * 1 = 18.61883$
- But running separately by sample does not test whether there is a statistically significant interaction!

What is the effect of X on Y for different values of Z

- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 * 1 + \hat{\beta}_3 (X_{1i} * 1)$
- $\hat{Y}_i = 15.94698 + 7.18773 * X_{1i} - 3.63304 * Z_i - 0.88283 * (X_{1i} * Z_i)$
  - Remember that:
  - $\hat{\beta}_1$ = change in $\hat{Y}_i$ for a one-unit increase in X1, when Z=0 – $\hat{\beta}_1$ = how much effect of X1 on Y changes when Z increases by one unit

- $\hat{\beta}_1$: **Average effect of X on Y when Z=0**
  - $\hat{\beta}_1 = 7.18773$
  - change in $\hat{Y}_i$ for X1=1 as opposed to X1=0, when Z=0___
  - Change in earnings for having more than a HS diploma (X1=1_) as opposed to having a HS diploma or less (X1=0) for men (Z=0)
  - On average, having more than a HS diploma (X1=1) as opposed to having a HS diploma or less (X1=0) is associated with $7.19 increase in hourly earnings for men (Z=0)

- $(\hat{\beta}_1 + \hat{\beta}_3)$: **Average effect of X on Y when Z=1**
  - $\hat{\beta}_1 = 7.18773$; $\hat{\beta}_3 =$ - 0.88283
  - 7.18773 + - 0.88283 = 6.3049
  - On average, having more than a HS diploma (X1=1) as opposed to having a HS diploma or less (X1=0) is associated with $6.30 increase in hourly earnings for women (Z=1)

- Same result as when we ran model separately by sample!

Interactions by Two Continuous Variables

Interactions by Two Continuous Variables

RQ: Does the effect of years of schooling (X) on earnings (Y) differ by age (Z)?

- Continuous by Categorical Interaction
  - Y = earnings
  - X = years of schooling
  - Z (interaction variable) = age
- Simple hypothesis
  - Years of schooling affects earnings
- Conditional hypothesis (interaction effect)
  - The effect of years of schooling (X) on earnings (Y) depends on age

Simple and Multivariate Regression (no interaction effect)

Simple Regression

- Population Regression Model
  - $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$
  - Where Y= earnings, X_1 = years of schooling
- Run in R
- OLS Prediction line with estimates
  - $Y_i = -5.37626 + 1.74515 * X_{1i}$
  - On average, a one-year increase in years of schooling is associated with a \$1.75 increase in hourly earnings

Multivariate Regression

- Population Regression Model
  - $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$
  - Where Y= earnings, X_1 = years of schooling, X_2 = age
- Run in R
- OLS Prediction line with estimates
  - $Y_i = -11.177774 + 1.706443 * X_{1i} + 0.153515 * X_{2i}$
  - On average, a one-year increase in years of schooling is associated with a \$171 increase in hourly earnings, holding age constant
- But this assumes the effect of years of schooling is the same across all ages!

Interaction Model for Two Categorical Variables

- Let's run an interaction effect model to investigate whether the effect of years of school (X) on earnings (Y) differs by age (Z)
- Population regression model
  - $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 Z_i + \beta_3 (X_{1i} * Z_i) + u_i$
    - where Y= earnings, X_1 = years of school, Z_i = age
    - $X_{1i} * Z_i$ = interaction for years of schooling and age

- OLS Prediction line without estimates
  - $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 Z_i + \hat{\beta}_3 (X_{1i} * Z_i)$

- $\hat{\beta}_0$ = $\hat{Y}_i$ **when X1=0 and Z=0**
  - Predicted earnings for observations with zero years of schooling (X=0) and age zero (Z=0)
- $\hat{\beta}_1$ = **change in** $\hat{Y}_i$ **one-unit increase in X1, when Z=0**
  - Change in earnings for one year increase in schooling when age is zero
- $\hat{\beta}_2$ = **change in** $\hat{Y}_i$ **for one-unit increase in Z, when X1=0**
  - Change in earnings for one year increase in age when years of schooling is zero
- $\hat{\beta}_3$ = **interaction term: how much the effect of X1 on** $\hat{Y}_i$ **changes when Z increases by one unit**
  - change in the effect of years of schooling for a one-year increase in age

Interaction Model for Continuous by Continuous Interaction

- Run in R
- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 Z_i + \hat{\beta}_3 (X_{1i} * Z_i)$
- $\hat{Y}_i = -7.032907 + 1.398881 * X_{1i} + 0.055153 * Z_i + 0.007281 * (X_{1i} * Z_i)$

- **Is there an interaction effect?**
- β_3 is significant at the 0.000
  - Yes, there is a statistically significant interaction between years of schooling and age!
- Positive β_3 coefficient means that the effect of years of schooling gets stronger as age increases!
- Negative β_3 coefficient means that the effect of years of schooling gets weaker as age increases!

What is the predicted value of Y for different values of X and Z

- X = 16 years of schooling (BA) and Z= 24 years old
  - Y |X=16, Z=24:
    $\widehat{Y}_i = -7.032907 + (1.398881 * 16) + (0.055153 * 24) + (0.007281 * (16 * 24))$
  - $\widehat{Y}_i = -7.032907 + 22.3821 + 1.323672 + 0.007281 * (384)$
  - $19.46877 = -7.032907 + 22.3821 + 1.323672 + 2.795904$

- X = 16 years of schooling (BA) and Z = 45 years old
  - Y |X=16, Z=45:
    $\widehat{Y}_i = -7.032907 + (1.398881 * 16) + (0.055153 * 45) + (0.007281 * (16 * 45))$
  - $\widehat{Y}_i = -7.032907 + 22.3821 + 2.481885 + 0.007281 * (720)$
  - $21.91519 = -7.032907 + 22.3821 + 1.323672 + 5.24232$

THANK YOU AND TAKE CARE OF YOURSELVES