# Recitation, Week 12

Ye Wang

New York University

POL-850

Spring 2018

# Outline

1. R-squared
2. Multiple regression
3. **Interaction effects**
4. STATA session (mainly interaction effects)

# R-squared



- The OLS estimation minimizes

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2$$

where $e_i = Y_i - \hat{Y}_i$ (residual) and $\hat{Y}_i = \hat{\alpha} + \hat{\beta} X_i$ (predicted Y)

- Note that we call $\sum_{i=1}^{n} e_i^2$

  Sum of Squared Residuals (SSR)
  Residual Sum of Squares (RSS)

# R-squared

- Proportional Reduction in Error (PRE), does this term sound familiar to you?

# R-squared



- Proportional Reduction in Error (PRE), does this term sound familiar to you?
- Actually, it should be! We saw this term when studying cross tabulations and $\lambda$

$$\lambda = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1}$$

- In fact, this is a PRE measure for cross tabulations

# R-squared



- Proportional Reduction in Error (PRE), does this term sound familiar to you?
- Actually, it should be! We saw this term when studying cross tabulations and $\lambda$
$$\lambda = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1}$$
- In fact, this is a PRE measure for cross tabulations

$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- R-squared ($R^2$) is a PRE measure for linear regression

# R-squared

$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?

# R-squared

$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$?

# R-squared

$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$? $\overline{Y}$
- Error without knowledge?

# R-squared

$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$? $\overline{Y}$
- Error without knowledge? $Y_i - \overline{Y}$

# R-squared



$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$? $\overline{Y}$
- Error without knowledge? $Y_i - \overline{Y} \rightarrow$

# R-squared

$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$? $\overline{Y}$
- Error without knowledge? $Y_i - \overline{Y} \rightarrow \; (Y_i - \overline{Y})^2$

# R-squared



$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$? $\overline{Y}$
- Error without knowledge? $Y_i - \overline{Y} \rightarrow \; (Y_i - \overline{Y})^2$

$$\text{Total Sum of Squares (TSS)} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

- What's our best guess about $Y$ when we know $X$?

# R-squared



$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$? $\overline{Y}$
- Error without knowledge? $Y_i - \overline{Y} \rightarrow \ (Y_i - \overline{Y})^2$

$$\text{Total Sum of Squares (TSS)} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

- What's our best guess about $Y$ when we know $X$? $\hat{Y}_i$
- Error with knowledge?

# R-squared



$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$? $\overline{Y}$
- Error without knowledge? $Y_i - \overline{Y} \rightarrow \ (Y_i - \overline{Y})^2$

$$\text{Total Sum of Squares (TSS)} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

- What's our best guess about $Y$ when we know $X$? $\hat{Y}_i$
- Error with knowledge? $Y_i - \hat{Y}_i$

# R-squared



$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$? $\overline{Y}$
- Error without knowledge? $Y_i - \overline{Y} \rightarrow \ (Y_i - \overline{Y})^2$

$$\text{Total Sum of Squares (TSS)} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

- What's our best guess about $Y$ when we know $X$? $\hat{Y}_i$
- Error with knowledge? $Y_i - \hat{Y}_i \rightarrow$

# R-squared



$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$? $\overline{Y}$
- Error without knowledge? $Y_i - \overline{Y} \rightarrow \ (Y_i - \overline{Y})^2$

$$\text{Total Sum of Squares (TSS)} = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

- What's our best guess about $Y$ when we know $X$? $\hat{Y}_i$
- Error with knowledge? $Y_i - \hat{Y}_i \rightarrow (Y_i - \hat{Y}_i)^2$

# R-squared

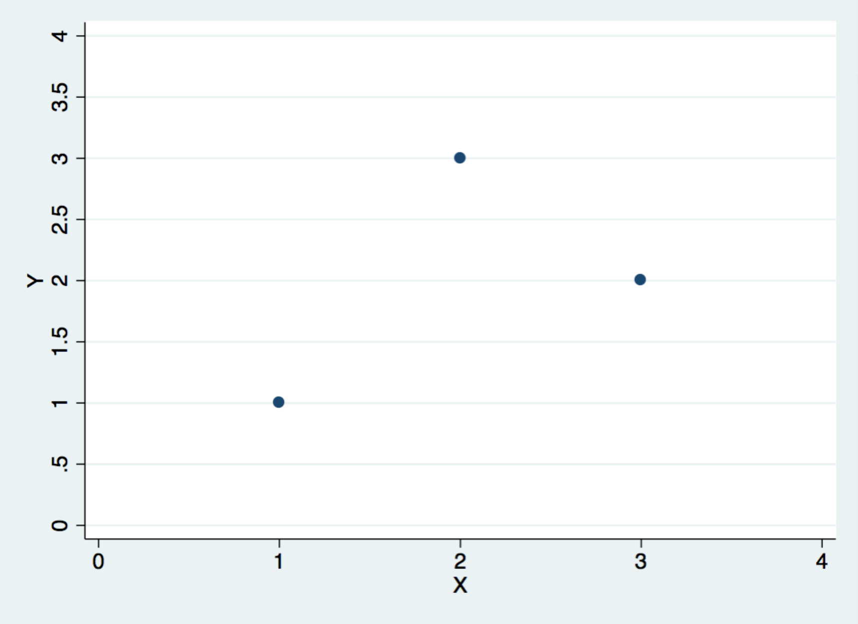
$$\text{PRE} = \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}}$$

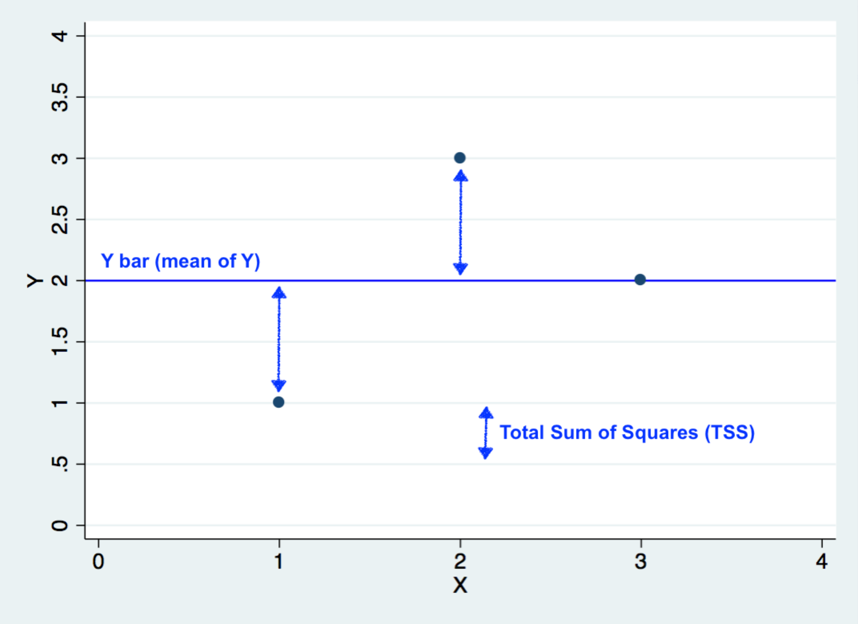
- Error without additional information provided by X?
- What's our best guess about $Y$ when we don't know $X$? $\overline{Y}$
- Error without knowledge? $Y_i - \overline{Y} \rightarrow \; (Y_i - \overline{Y})^2$

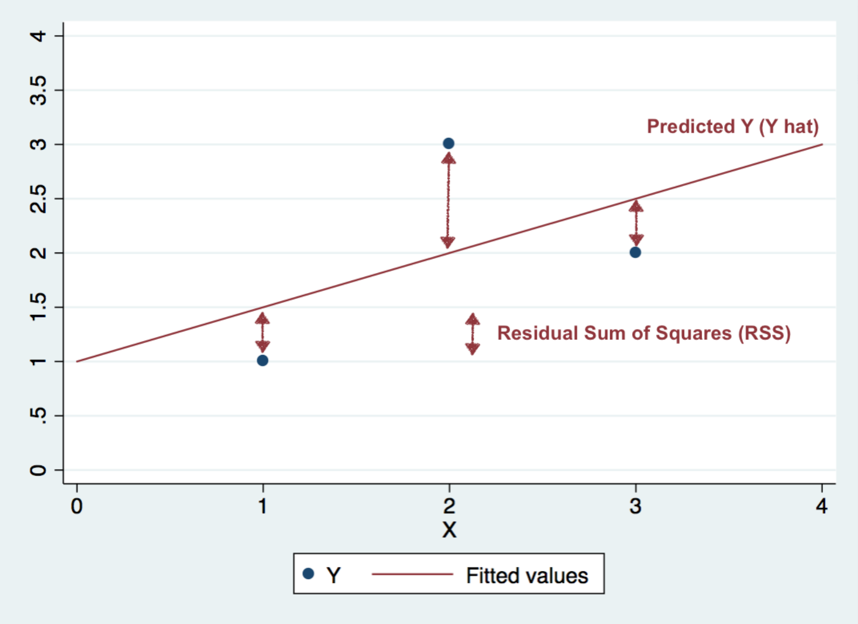
$$\text{Total Sum of Squares (TSS)} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

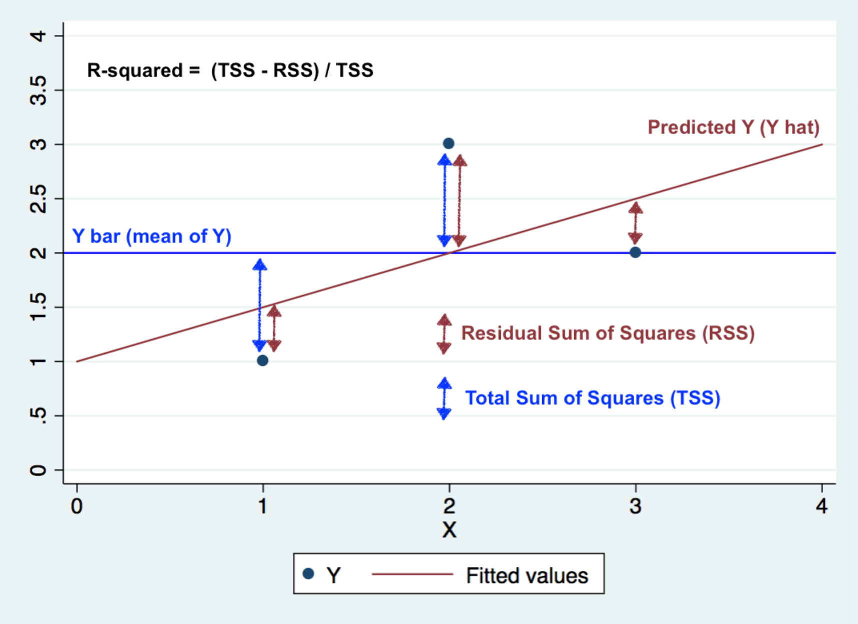
- What's our best guess about $Y$ when we know $X$? $\hat{Y}_i$
- Error with knowledge? $Y_i - \hat{Y}_i \rightarrow (Y_i - \hat{Y}_i)^2$

$$\text{Residual Sum of Squares (RSS)} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$


Y
0
.5
1
1.5
2
2.5
3
3.5
4
X
0
1
2
3
4


Y bar (mean of Y)
Total Sum of Squares (TSS)
Y
X
0
.5
1
1.5
2
2.5
3
3.5
4
0
1
2
3
4


Predicted Y (Y hat)
Residual Sum of Squares (RSS)
0
.5
1
1.5
2
2.5
3
3.5
4
0
1
2
3
4
X
Y
Fitted values



R-squared = (TSS - RSS) / TSS
Predicted Y (Y hat)
Y bar (mean of Y)
Residual Sum of Squares (RSS)
Total Sum of Squares (TSS)
0
.5
1
1.5
2
2.5
3
3.5
4
0
1
2
3
4
X
Y
Fitted values

# R-squared



$$
\begin{aligned}
R^2 &= \frac{\text{Error without knowledge} - \text{Error with knowledge}}{\text{Error without knowledge}} \\
&= \frac{(\text{Total Sum of Squares}) - (\text{Residual Sum of Squares})}{(\text{Total Sum of Squares})} \\
&= \frac{TSS - RSS}{TSS} \\
&= \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 - \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}
\end{aligned}
$$

- Note that $0 \leq R^2 \leq 1$ (when does it become 1?)
- Intuitively, $R^2$ is the proportional reduction in error when we use $X$ to predict $Y$ instead of $\overline{Y}$

# Multiple regression

- So far we've been only talking about bivariate (two variables: X and Y) regression
- When we have multiple independent variables, we use multiple regression
- We still use the OLS estimation

Why do we use multiple regression?

# Multiple regression



1. In most of the cases, a theory stipulates **multiple** explanatory variables
2. Multiple regression allows us to estimate **partial regression coefficient**
3. We want to **rule out** a variable(s) which we suspect to have a **spurious association**
4. (*) It may reduce the standard error in experiments

# Multiple regression

Suppose there is a theory which suggests that unemployment and education affects anti-immigration attitudes

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

where

- $Y$: Thermometer toward immigrants
  (0=Coldest, 100=Warmest)
- $X_1$: Education
  (Years of education)
- $X_2$: Unemployment
  (unemployed=1, employed=0)

Note that $\beta$s are **population parameters**!

# Multiple regression



Then we would want to **estimate** $\beta$s with our sample to test the hypothesis

$$Y = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + e$$

A theory of anti-immigration attitudes would suggest:

- $\beta_1$ (Education) is

# Multiple regression



Then we would want to **estimate** $\beta$s with our sample to test the hypothesis

$$Y = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + e$$

A theory of anti-immigration attitudes would suggest:

- $\beta_1$ (Education) is positive
- $\beta_2$ (Unemployment) is

# Multiple regression



Then we would want to **estimate** $\beta$s with our sample to test the hypothesis

$$Y = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + e$$

A theory of anti-immigration attitudes would suggest:

- $\beta_1$ (Education) is positive
- $\beta_2$ (Unemployment) is negative

In the end, we would want to see $\hat{\beta}$s match our expectation!

# Multiple regression



## Regression coefficient (simple regression)

$\hat{\beta}$: average change in Y for a unit change in X

- e.g. $Y = \hat{\alpha} + \hat{\beta}_1 X_1 + e$
- average change in attitudes for a unit change in education ($X_1$) (e.g. one more year of school education)

# Multiple regression



## Partial regression coefficient (multiple regression)

$\hat{\beta}$: average change in Y for a unit change in X,
**holding all the other independent variables constant**

- e.g. $Y = \hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + e$
- average change in attitudes for a unit change in education ($X_1$) **holding** unemployment ($X_2$) **constant**

# Group exercise

```
. reg attitude education unemployed

      Source |       SS           df       MS      Number of obs   =        98
-------------+----------------------------------   F(2, 95)        =    178.65
       Model |  11756.0264         2  5878.01321   Prob > F        =    0.0000
    Residual |  3125.65005        95  32.9015795   R-squared       =    0.7900
-------------+----------------------------------   Adj R-squared   =    0.7855
       Total |  14881.6765        97  153.419345   Root MSE        =     5.736

------------------------------------------------------------------------------
    attitude |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   education |    1.43276   .1950966     7.34   0.000     1.045444    1.820076
  unemployed |  -23.09773   1.254102   -18.42   0.000    -25.58744   -20.60803
       _cons |    57.7035   2.391592    24.13   0.000     52.95559    62.45141
------------------------------------------------------------------------------
```

1. Does the signs of $\hat{\beta}$ match our expectation?
2. What's the substantive interpretation of $\hat{\beta}_1$ (education)?
3. How about $\hat{\beta}_2$ (unemployment)?
4. Which variable is statistically significant?
5. Would you say this result supports the theory?

# Group exercise

```
. reg attitude education unemployed

      Source |       SS           df       MS      Number of obs   =        98
-------------+----------------------------------   F(2, 95)        =    178.65
       Model |  11756.0264         2  5878.01321   Prob > F        =    0.0000
    Residual |  3125.65005        95  32.9015795   R-squared       =    0.7900
-------------+----------------------------------   Adj R-squared   =    0.7855
       Total |  14881.6765        97  153.419345   Root MSE        =     5.736

------------------------------------------------------------------------------
    attitude |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   education |    1.43276   .1950966     7.34   0.000     1.045444    1.820076
  unemployed |  -23.09773   1.254102   -18.42   0.000    -25.58744   -20.60803
       _cons |    57.7035   2.391592    24.13   0.000     52.95559    62.45141
------------------------------------------------------------------------------
```

# Interaction effects

- Two independent variables **interact** if the effect caused by the change of one variable depends on the level of the other variable
- What does this mean? Let's compare an interaction case with a non-interaction case
- Suppose we are interested in the effect of education on attitude towards immigrants

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

# Interaction effects



- Two independent variables **interact** if the effect caused by the change of one variable depends on the level of the other variable
- What does this mean? Let's compare an interaction case with a non-interaction case
- Suppose we are interested in the effect of education on attitude towards immigrants

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

$$\frac{\Delta Y}{\Delta X_1} = \frac{\text{changes in } \texttt{attitude}}{\text{one unit change in } \texttt{education}} = \beta_1$$

# Interaction effects



$$\frac{\Delta Y}{\Delta X_1} = \frac{\text{changes in } \texttt{attitude}}{\text{one unit change in } \texttt{education}} = \beta_1$$

- Note that the sign and magnitude of $X_1$'s effect on $Y$ is constant across all observations
- Note also $X_2$ has nothing to do with $X_1$'s effect on $Y$

# Interaction effects



- Now suppose the theory predicts the **interaction** effects between education and unemployment
- Equivalent to adding an **interaction** term $\beta_3 X_1 X_2$ to the previous model

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

# Interaction effects

- Now suppose the theory predicts the **interaction** effects between education and unemployment
- Equivalent to adding an **interaction** term $\beta_3 X_1 X_2$ to the previous model

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

$$\frac{\Delta Y}{\Delta X_1} = \frac{\text{changes in } \texttt{attitude}}{\text{one unit change in } \texttt{education}} = \beta_1 + \beta_3 X_2$$

# Interaction effects

$$\frac{\Delta Y}{\Delta X_1} = \frac{\text{changes in } \texttt{attitude}}{\text{one unit change in } \texttt{education}} = \beta_1 + \beta_3 X_2$$

- Note that the sign and magnitude of $X_1$'s effect on $Y$ is not constant across all observations

# Interaction effects



$$\frac{\Delta Y}{\Delta X_1} = \frac{\text{changes in } \texttt{attitude}}{\text{one unit change in } \texttt{education}} = \beta_1 + \beta_3 X_2$$

- Note that the sign and magnitude of $X_1$'s effect on $Y$ is not constant across all observations
- Why?

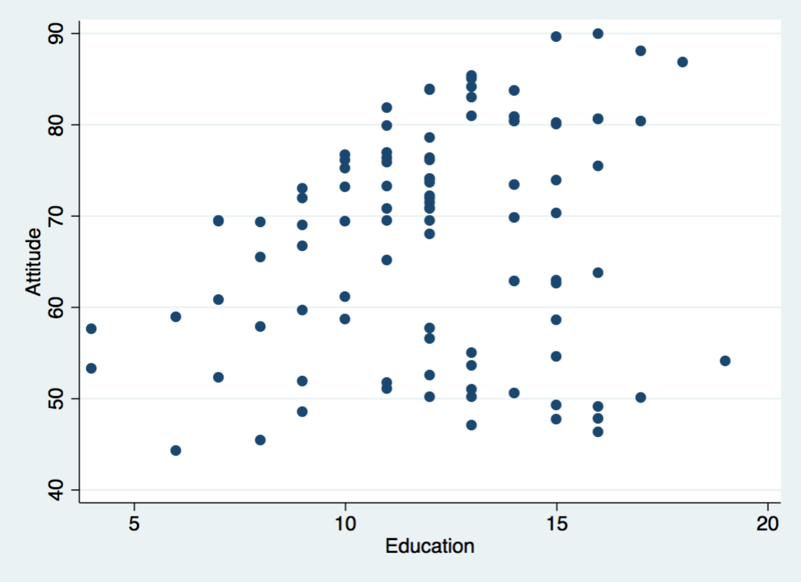
# Interaction effects



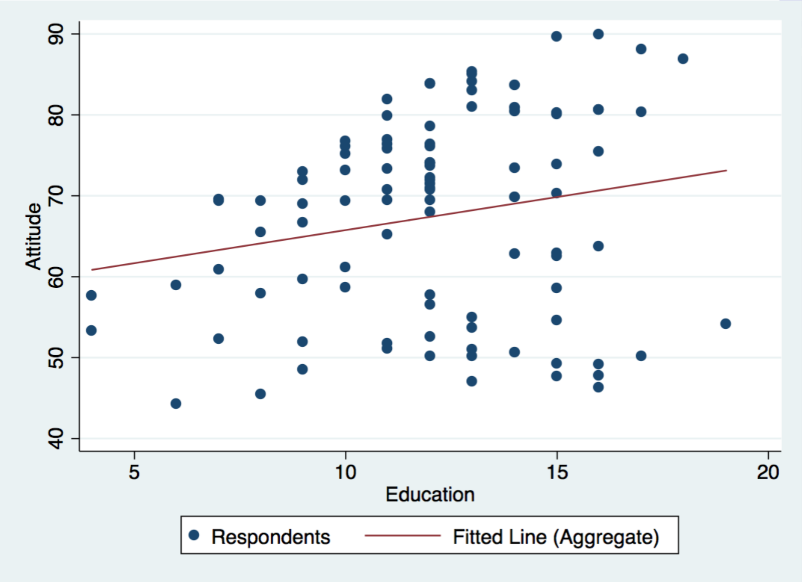
$$\frac{\Delta Y}{\Delta X_1} = \frac{\text{changes in } \texttt{attitude}}{\text{one unit change in } \texttt{education}} = \beta_1 + \beta_3 X_2$$

- Note that the sign and magnitude of $X_1$'s effect on $Y$ is not constant across all observations
- Why? $X_2$ (`unemployment`) varies!
- Thus, the sign and the magnitude of $X_1$'s effect on Y changes depending on the value of $X_2$

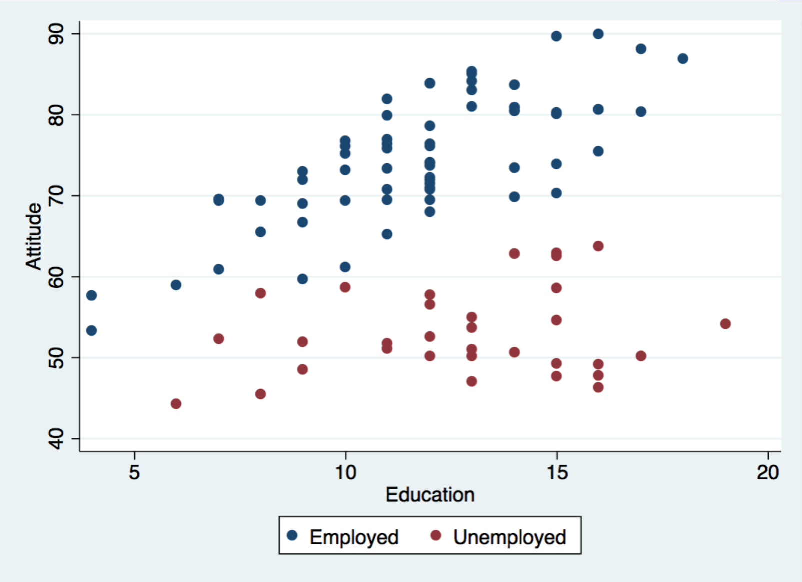
# Interaction Effects



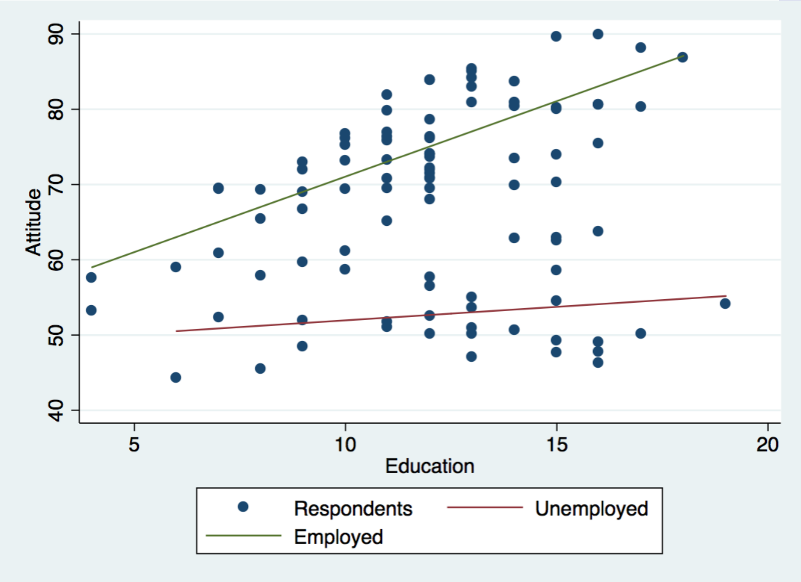
# Interaction Effects



# Interaction Effects



# Interaction Effects



# Group exercise



$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$$

| attitude | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| education | 2.006247 | .2212553 | 9.07 | 0.000 | 1.566939 | 2.445554 |
| 1.unemployed | -2.618776 | 4.80073 | -0.55 | 0.587 | -12.15074 | 6.913186 |
| unemployed#c.education | | | | | | |
| 1 | -1.647611 | .3750243 | -4.39 | 0.000 | -2.39223 | -.9029911 |
| _cons | 50.98676 | 2.670802 | 19.09 | 0.000 | 45.68382 | 56.2897 |

1. Calculate the marginal effect of education on attitude for the employed/unemployed
2. What's the substantive interpretation of $\hat{\beta}_3$?

# STATA



- How to do these things in STATA?
- Multiple regression is extremely easy! You just add more variables
- For example, `regress Y X1 X2 X3` is identical to:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

- You estimate $\beta_1$, $\beta_2$ and $\beta_3$ using OLS
- Interaction effects is a bit more complicated

# STATA



- Suppose you want to add an interaction term between two independent variables X1 and X2
- In other words, your model is:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \color{red}{\beta_3 X_1 X_2} + \varepsilon$$

- To distinguish a variable is whether binary or continuous is important
- Binary variables take either 0 or 1 (e.g. employment status)
- Continuous variables take values in a range (e.g. income or education level measured by years)

# STATA



- Note that every STATA syntax below indicates the OLS estimation of: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$
- **Case 1**: Binary (X1) $\times$ Binary (X2):
  `regress Y i.X1##i.X2`
- **Case 2**: Binary (X1) $\times$ Continuous (X2):
  `regress Y i.X1##c.X2`
- **Case 3**: Continuous (X1) $\times$ Continuous (X2):
  `regress Y c.X1##c.X2`
- **DIY**: Alternatively, you can generate your own interaction variable (this works regardless of a type of variables)

```
generate X1X2 = X1 * X2
regress Y X1 X2 X1X2
```

# Key takeaways

You should be able to answer these questions:

- What does $R^2$ represent?

# Key takeaways

You should be able to answer these questions:

- What does $R^2$ represent?
  How much the model can explain the variation in Y
- Why do we use **multiple regression**?

# Key takeaways



You should be able to answer these questions:

- What does $R^2$ represent?
  How much the model can explain the variation in Y
- Why do we use **multiple regression**?
  (1) Theory, (2) Partial regression coefficient, and (3) Ruling out a spurious association
- What does a **partial regression coefficient** mean?

# Key takeaways

You should be able to answer these questions:

- What does $R^2$ represent?
  How much the model can explain the variation in Y
- Why do we use **multiple regression**?
  (1) Theory, (2) Partial regression coefficient, and (3) Ruling out a spurious association
- What does a **partial regression coefficient** mean?
  Change in Y for a unit change in X, holding all the other independent variables constant
- What are **interaction effects**?

# Key takeaways



You should be able to answer these questions:

- What does $R^2$ represent?
  How much the model can explain the variation in Y
- Why do we use **multiple regression**?
  (1) Theory, (2) Partial regression coefficient, and (3) Ruling out a spurious association
- What does a **partial regression coefficient** mean?
  Change in Y for a unit change in X, holding all the other independent variables constant
- What are **interaction effects**?
  The effect of X on Y depends on the other independent variable Z