FLS 6441 - Methods III: Explanation and Causation

Week 1 - Review

Jonathan Phillips

February 2019

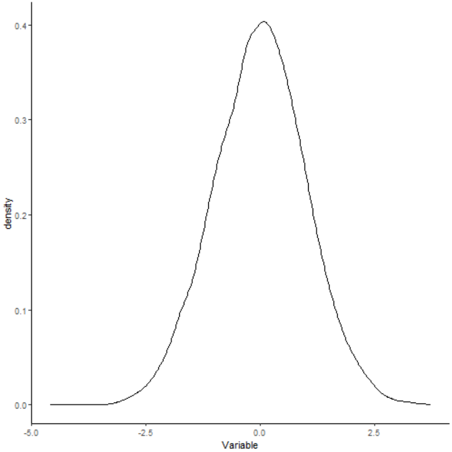
Course Objectives

1. temp

Data

1. We work with variables, which VARY!

	Variable
1	-1.21
2	0.73
3	-0.51
4	-1.41
5	0.62
6	0.33
7	-0.66
8	0.31
9	-0.12
10	0.16



Regression

- Regression is a **Conditional Expectation Function**

Regression

- Regression is a **Conditional Expectation Function**
- Conditional on x, what is our expectation (mean value) of y?

Regression

- Regression is a **Conditional Expectation Function**
- Conditional on x, what is our expectation (mean value) of y?
- $E(y|x)$

Regression

- Regression is a **Conditional Expectation Function**
- Conditional on x, what is our expectation (mean value) of y?
- $E(y|x)$
- When age is 20 ($x = 40$), the average salary is R1.000 ($y = 1.000$)
- When age is 40 ($x = 40$), the average salary is R2.000 ($y = 2.000$)

Regression

- Regression is a **Conditional Expectation Function**

```
## How to cite this model in Zelig:
##   R Core Team. 2007.
##   ls: Least Squares Regression for Continuous Depend
##   in Christine Choirat, Christopher Gandrud, James H
##   "Zelig: Everyone's Statistical Software," http://z
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac
## % Date and time: Thu, Feb 28, 2019 - 8:57:42 AM
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}}
## \cline{2-2}
## \\[-1.8ex] & y \\
```

```
## \hline \\[-1.8ex]
##  x & 0.758$^{***}$ \\
##   & (0.066) \\
##   & \\
##  Constant & $-$0.000 \\
##   & (0.066) \\
##   & \\
## \hline \\[-1.8ex]
## Observations & 100 \\
## R$^{2}$ & 0.575 \\
## Adjusted R$^{2}$ & 0.571 \\
## Residual Std. Error & 0.655 (df = 98) \\
## F Statistic & 132.700$^{***}$ (df = 1; 98) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1;
## \end{tabular}
```

```
## \end{table}
```

Regression

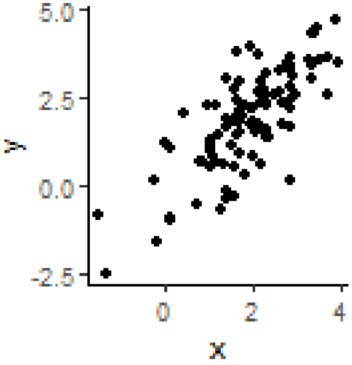
- Regression with two variables is very similar to calculating correlation

Regression

- Regression with two variables is very similar to calculating correlation
- $\hat{\beta} = cor(x, y) * \frac{\sigma_Y}{\sigma_X}$

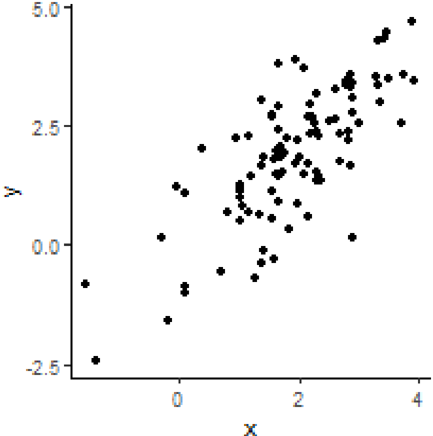
Regression

- Regression with two variables is very similar to calculating correlation
- $\hat{\beta} = cor(x, y) * \frac{\sigma_Y}{\sigma_X}$
- It's *identical* if we standardize both variables first ($\frac{(x - \bar{x})}{\sigma_x}$)



Regression

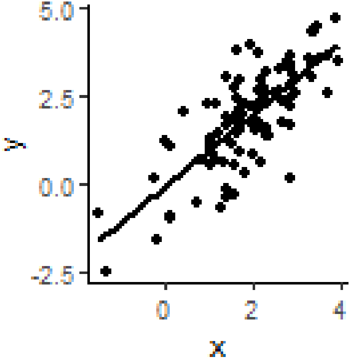
- Regression with two variables is very similar to calculating correlation:
- $\hat{\beta} = cor(x, y) * \frac{\sigma_Y}{\sigma_X}$
- It's *identical* if we standardize both variables first ($\frac{(x - \bar{x})}{\sigma_x}$)



Correlation is 0.758

Regression

- Regression with two variables is very similar to calculating correlation
- $\hat{\beta} = cor(x, y) * \frac{\sigma_Y}{\sigma_X}$
- It's *identical* if we standardize both variables first ($\frac{(x - \bar{x})}{\sigma_x}$)

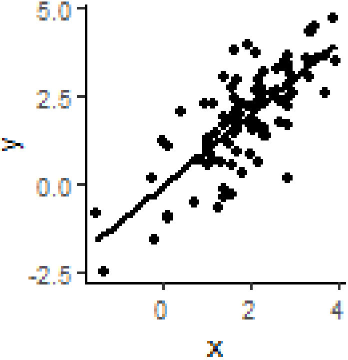


The regression result is:

	Depe
x	
Constant	

Regression

- Regression with two variables is very similar to calculating correlation
- $\hat{\beta} = cor(x, y) * \frac{\sigma_Y}{\sigma_X}$
- It's *identical* if we standardize both variables first ($\frac{(x - \bar{x})}{\sigma_x}$)



The regression result is:

Depe

x

Constant

Regression Guide

1. **Choose variables and measures:** To test a specific hypothesis
2. **Choose a Model/Link Function:** Should match the data type of your outcome variable
3. **Choose Covariates:** To match your strategy of inference
4. **Choose Fixed Effects:** To focus on a specific level of variation
5. **Choose Standard Error Structure:** To match known dependencies/clustering in the data
6. **Interpret the coefficients:** Depending on the type/scale of the explanatory variable

Regression Models

The Regression Model reflects the data type of the outcome variable:

- Continuous -> Ordinary Least Squares

```
zelig(Y   X,data=d,model="ls")
```

- Binary -> Logit

```
zelig(Y   X,data=d,model="logit")
```

- Unordered categories -> Multinomial logit

```
zelig(Y   X,data=d,model="mlogit")
```

- Ordered categories -> Ordered logit

```
zelig(Y   X,data=d,model="ologit")
```

- Count -> Poisson

```
zelig(Y   X,data=d,model="poisson")
```

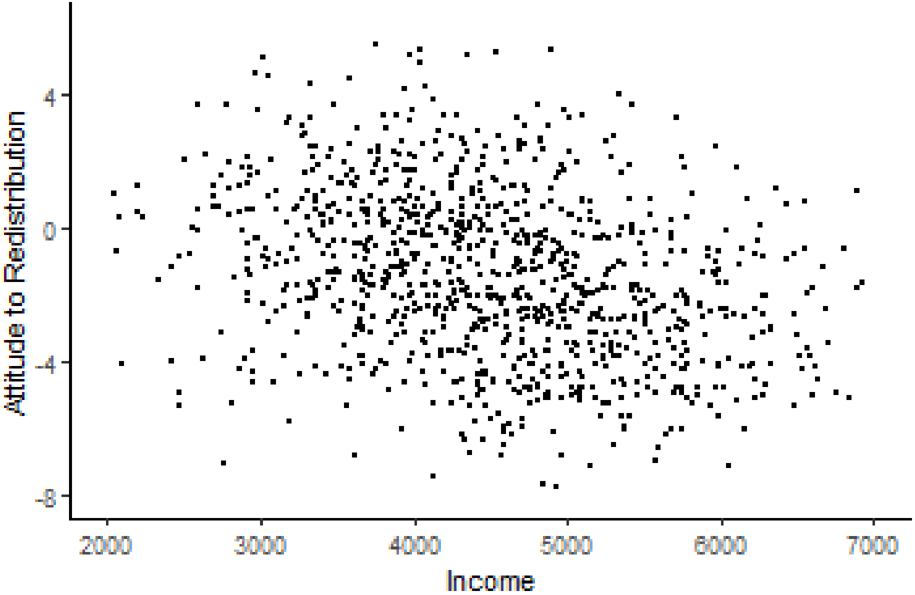
Interpreting Regression Results

- Difficult! It depends on the scale of the explanatory variable, scale of the outcome, the regression model we used, and the presence of any interaction
- Basic OLS:
  - 1 [unit of explanatory variable] change in the explanatory variable is associated with a β [unit of outcome variable] change in the outcome

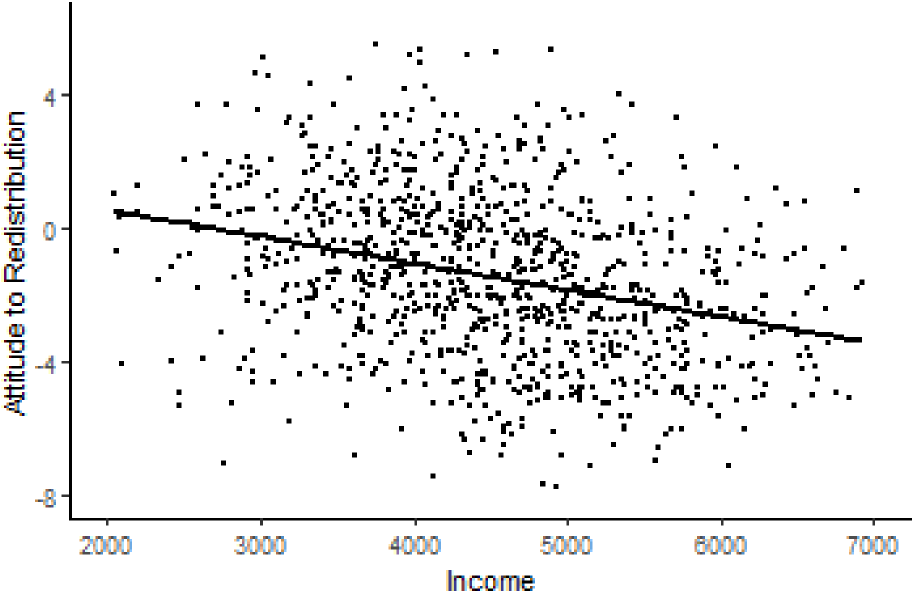
Predictions from Regressions

- temp

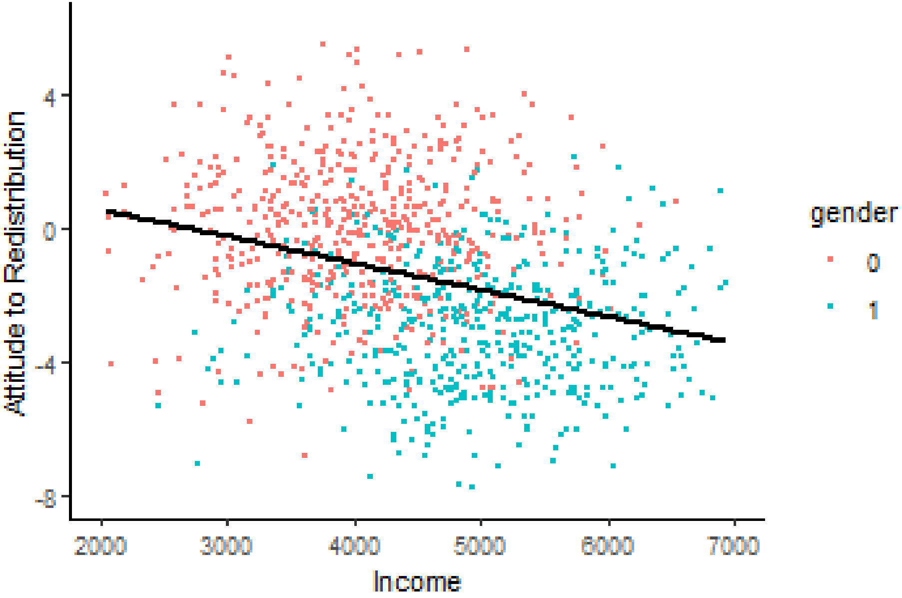
Omitted Variable Bias



Omitted Variable Bias



Omitted Variable Bias



Omitted Variable Bias