POLS/CS&SS 503:
Advanced Quantitative Political Methodology

LINEAR REGRESSION ESTIMATOR

April 14, 2015

Jeffrey B. Arnold

UNIVERSITY of WASHINGTON
DEPARTMENT OF POLITICAL SCIENCE

CENTER for STATISTICS and the SOCIAL SCIENCES

Overview

Regression Coefficient Anatomy

Coeficients of a simple regression

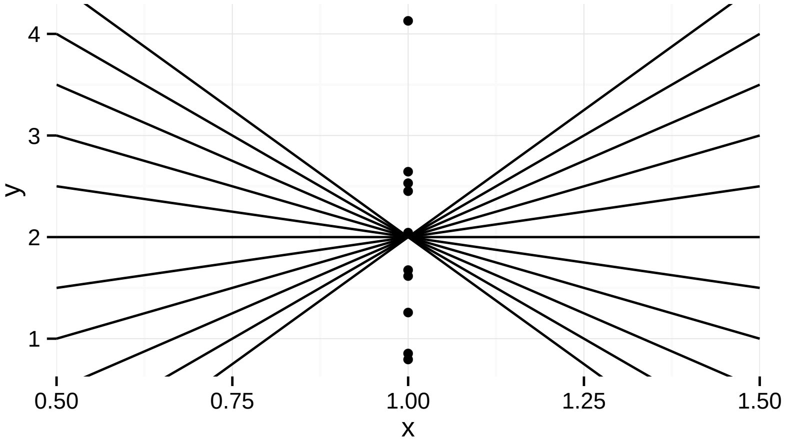
$$Y_i = A + BX_i + E_i$$

The least squares coefficients are

$$A = \bar{Y} + B\bar{X}$$
$$B = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_i \left(X_i - \bar{X}\right)^2}$$

- State B in terms of covariance of X and Y and variance?
- State B in terms of correlation of X and Y and standard deviations?
- What values can B take if sd(X) $=$ sd(Y) $= 1$?
- What is $\hat{Y}$ for $X = \bar{X}$?
- What happens to B as V(X) decreases? V(Y) decreases? If V$(X) = 0$

Least squares when V $X = 0$



Least squares coefficients are unidentified if V $x = 0$

- If V $x = 0$ then least squares solution is unidentified
- There is no unique value of A, B that $\arg\min_{A,B} \sum_i E_i^2$

```
y <- c(1, 2, 3, 4, 5)
x <- 1
ybar <- mean(y)
ybar
```

```
## [1] 3
```

```
# A = 2, B = 1
sum((y - 2 - 1 * x) ^ 2)
```

```
## [1] 10
```

```
# A = -7, B = 10
sum((y + 7 - 10 * x) ^ 2)
```

```
## [1] 10
```

Coefficients of a multiple regression

$$\vec{Y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}$$

- $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$. *Not that intuitive!*
- Coefficient $\boldsymbol{b}_j$ is

$$\boldsymbol{b}_k = \frac{\mathsf{C}(\boldsymbol{y}, \tilde{\boldsymbol{x}}_j)}{\mathsf{V}(\tilde{\boldsymbol{x}}_k)}$$

- Where $\tilde{\boldsymbol{x}}_j$ are the residulals of $\boldsymbol{x}_j$ on all X_h where $h \neq j$

$$\tilde{X}_{j,i} = X_{j,i} - \tilde{A} - \sum_{h \neq j} \tilde{B}_h X_h$$

Regression example

See `multiple_regression_anatomy.R`