

PLSC 503 – Spring 2017

Multivariate Regression I

February 9, 2017

The Model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

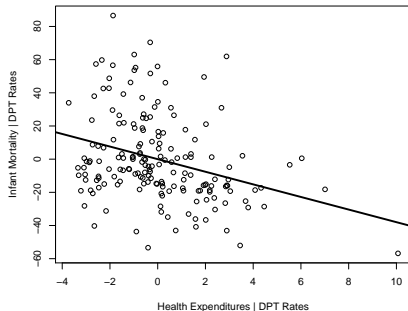
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + u_i$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{K1} \\ 1 & X_{12} & X_{22} & \cdots & X_{K2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1N} & X_{2N} & \cdots & X_{KN} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}.$$

Diversion: “Added Variable Plots”

- Regress Y on X_1 and save the residuals \hat{u}_i ,
- Regress X_2 on X_1 and save the residuals (call these \hat{e}_i),
- Plot \hat{u}_i (conventionally on the y-axis) vs. \hat{e}_i (conventionally on the x-axis).

Example: Infant Mortality and Health Expenditures Given DPT Immunization Rates



Residuals:

$$\mathbf{u} = \mathbf{Y} - \mathbf{X}\beta$$

The inner product of \mathbf{u} :

$$\begin{aligned} \mathbf{u}'\mathbf{u} &= \begin{bmatrix} u_1 & u_2 & \cdots & u_N \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \\ &= u_1^2 + u_2^2 + \dots + u_N^2 \\ &= \sum_{i=1}^N u_i^2 \end{aligned}$$

Estimating β

$$\begin{aligned}\mathbf{u}'\mathbf{u} &= (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta) \\ &= \mathbf{Y}'\mathbf{Y} - 2\beta'\mathbf{X}'\mathbf{Y}' + \beta'\mathbf{X}'\mathbf{X}\beta\end{aligned}$$

Now get:

$$\frac{\partial \mathbf{u}'\mathbf{u}}{\partial \beta} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta$$

Solve:

$$\begin{aligned}-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\beta &= 0 \\ -\mathbf{X}'\mathbf{Y} + \mathbf{X}'\mathbf{X}\beta &= 0 \\ \mathbf{X}'\mathbf{X}\beta &= \mathbf{X}'\mathbf{Y} \\ (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\ \beta &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\end{aligned}$$

OLS Assumptions

1. Zero Expectation Disturbances

$$E(\mathbf{u}) = \mathbf{0}$$

OLS Assumptions

2. Homoscedasticity / No Error Correlation

$$\begin{aligned}\mathbf{uu}' &= \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_N \end{bmatrix} \\ &= \begin{bmatrix} u_1^2 & u_1 u_2 & \cdots & u_1 u_N \\ u_2 u_1 & u_2^2 & \cdots & u_2 u_N \\ \vdots & \vdots & \ddots & \vdots \\ u_N u_1 & u_N u_2 & \cdots & u_N^2 \end{bmatrix}\end{aligned}$$

Expectation must be:

$$E(\mathbf{uu}') = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

OLS Assumptions

3. "Fixed" \mathbf{X} ...

- No *measurement error* in the \mathbf{X} s, and
- $\text{Cov}(\mathbf{X}, \mathbf{u}) = \mathbf{0}$.

4. \mathbf{X} is full column rank.

Means:

- no exact linear relationship among \mathbf{X} , and
- $K < N$.

5. Normal Disturbances

$$\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \sigma^2 \mathbf{I})$$

OLS: Unbiasedness

Unbiasedness:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Substitute OLS $\hat{\boldsymbol{\beta}}$:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\end{aligned}$$

and so:

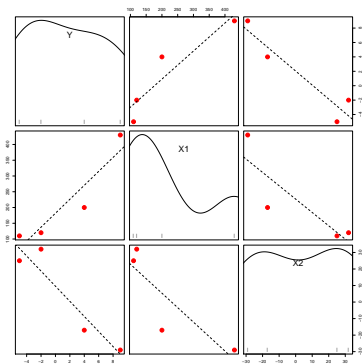
$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}.$$

By $\text{Cov}(\mathbf{X}, \mathbf{u}) = \mathbf{0}$, we have $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$.

An Example

$$\mathbf{Y} = \begin{bmatrix} 4 \\ -2 \\ 9 \\ -5 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & 200 & -17 \\ 1 & 120 & 32 \\ 1 & 430 & -29 \\ 1 & 110 & 25 \end{bmatrix}$$



Correlation

```
Y<-c(4,-2,9,-5)
X1<-c(200,120,430,110)
X2<-c(-17,32,-29,25)
data<-cbind(Y,X1,X2)
```

```
cor(data)
```

	Y	X1	X2
Y	1.0000	0.9285	-0.9425
X1	0.9285	1.0000	-0.8613
X2	-0.9425	-0.8613	1.0000

Regression

```
fit<-lm(Y~X1+X2)
summary(fit)
```

```
Call:
lm(formula = Y ~ X1 + X2)
```

Residuals:

1	2	3	4
0.531	1.639	-0.201	-1.970

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.2643	4.7284	-0.48	0.72
X1	0.0190	0.0200	0.95	0.52
X2	-0.1141	0.0985	-1.16	0.45

Residual standard error: 2.62 on 1 degrees of freedom

Multiple R-Squared: 0.941, Adjusted R-squared: 0.823

F-statistic: 7.99 on 2 and 1 DF, p-value: 0.243

“Do not compute the least squares estimates using $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$!”

– Sanford Weisberg (p. 61)

Most software uses:

$$\mathbf{X} = \mathbf{Q}\mathbf{R}$$

where \mathbf{Q} is orthogonal ($\mathbf{Q}'\mathbf{Q} = \mathbf{I}$) and \mathbf{R} is upper-triangular.

Why???

Estimation Example

```
options(digits=16)
options(scipen=99)
```

```
z<-c(-10000000000000,0.0000000000000001,10000000000000)
x<-c(-50000,0.000007,5000000)
fit<-lm(z~x)
fit
```

Call:

```
lm(formula = z ~ x)
```

Coefficients:

(Intercept)	x
-494950994952.3740845	299970.2999707

Estimation Example (continued)

```
X<-as.matrix(x)
Z<-as.matrix(z)
beta.hat <- solve(t(X) %*% X) %*% t(X) %*% Z
beta.hat
      [,1]
[1,] 201979.802019798

(fit$coefficients[2] / beta.hat) * 100
      [,1]
[1,] 148.5149985152023
```