Ordinary, Ordinary Least Squares

D. Alex Hughes

October 28, 2014

The Bottom Line

$H_0 : \beta = 0$ $H_A : \beta \neq 0$

Calculate estimated slope and intercept. Could they just be random?

1. State null and alternative hypotheses
2. Calculate standard error of slope and test statistic.
3. How likely is $\hat{\beta}$ if the true value is 0? If there were NO relationship, how likely is it that we would randomly see a slope this big?
4. If very unlikely, reject null hypothesis that no relationship. If rather likely, then maybe it is just random - fail to reject null.

The Bottom Line

difleft	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
octpess	.004647	.0012787	3.63	0.001	.0020454	.0072485
prural	-.0594083	.1155061	-0.51	0.610	-.2944073	.1755906
avgyed	.1135818	.0417256	2.72	0.010	.0286905	.1984731
ppubjobs	.0172879	.0458901	0.38	0.709	-.0760762	.110652
pntmigra	.0008118	.0010028	0.81	0.424	-.0012285	.0028521
m96m92	.003807	.009613	0.40	0.695	-.0157508	.0233647
_cons	-.5324989	.2414704	-2.21	0.035	-1.023774	-.0412238

Regression and Inference

- Let's start with a slightly different perspective. Instead of "what's the best line to draw on this data to summarize a relationship?", let's assume a theory and estimate parameters.
- Nature generates data y according to the model:

$$y = \alpha + \beta x + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

- Our *parameters* of interest are α, β, and σ
- We don't get to see any of them!

Regression and Inference

- Let's start with a slightly different perspective. Instead of "what's the best line to draw on this data to summarize a relationship?", let's assume a theory and estimate parameters.
- Nature generates data y according to the model:

$$y = \alpha + \beta x + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

- Our *parameters* of interest are α, β, and σ
- We don't get to see any of them!

Regression and Inference

- We estimate parameters with statistics a, b, and $\hat{\sigma}$ There are lots of ways to estimate them. (Why is OLS the best? What needs be true?)
- Two things we usually want:
  1. Confidence interval: what could the true value be?
  2. Hypothesis testing: Is the true β different from zero? Or: Does democracy affect P(war), do deficits affect inflation, do electoral rules determine party cohesion?
- More precisely:
  1. "What is the probability of seeing an estimated slope as large as the one we found if the true slope is in fact zero?", **OR**
  2. "If the true slope were zero, what is the probability of seeing data like ours, with an estimated slope like the one we found, just due to random error?"

Regression and Inference

- We estimate parameters with statistics a, b, and $\hat{\sigma}$ There are lots of ways to estimate them. (Why is OLS the best? What needs be true?)
- Two things we usually want:
  1. Confidence interval: what could the true value be?
  2. Hypothesis testing: Is the true β different from zero? Or: Does democracy affect P(war), do deficits affect inflation, do electoral rules determine party cohesion?
- More precisely:
  1. "What is the probability of seeing an estimated slope as large as the one we found if the true slope is in fact zero?", **OR**
  2. "If the true slope were zero, what is the probability of seeing data like ours, with an estimated slope like the one we found, just due to random error?"

To answer these questions, we need to know about the distribution of $\hat{\beta}$. Let's think about this for a minute.

- Recall, β is the true value. It is FIXED.
- $\hat{\beta}$ is a Random variable - it is determined from the data, which includes a systematic component $(\alpha + \beta x)$ and a random component (ϵ).
- So we want to know about the distribution of $\hat{\beta}$
- With enough information, we'll be able to calculate the probability of seeing a $\hat{\beta}$ like ours, assuming that the true β is equal to zero - that is, we'll be able to conduct a hypothesis test.

Simulation

Let's start with a simulation.

`Go to the code, Luke.`

Inference for Bivariate Regression

$$Y = \alpha + \beta X + \epsilon$$

Assumptions:

- Linearity
- Distribution of the errors: $\epsilon \sim N(0, \sigma^2)$
    - Normality
    - Expectation of zero
    - Constant Variance
    - Independence
- Fixed X, or X independent of the error

Inference for Bivariate Regression

OLS Estimates are Unbiased:

$$E(\hat{\alpha}) = \alpha$$

$$E(\hat{\beta}) = \beta$$

The Variances of OLS Bivariate Estimates are:

$$\hat{V}(A) = \frac{\sigma^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}$$

$$\hat{V}(B) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$$

Revisit R simulation for intuition.

Inference for Bivariate Regression

- Problem - we don't observe or know σ. We have to estimate it. How will we estimate it?
- Extent to which we are off in our estimate of β and α means we will also be off in our estimate of σ . Indeed, our estimates of α and β minimize what?
- Means we need to account for this additional uncertainty.

```
# Regression Simulation 2 - Estimating Sigma
# RegSim2.R
```

Inference for Bivariate Regression

$$S_E^2 = \hat{\sigma}^2 = \frac{\sum E_i^2}{n-2}$$

- Why do we divide by $n-2$?
- Remember that our estimates of α and β are certainly not the right ones - they are just the best estimates we can get.
- Furthermore, α and β are chosen to *minimize* Σe_i^2
- So the raw sum of squared residuals (Σe_i^2) is less than the true sum of squared errors ($\Sigma \epsilon_i^2$)
- Said another way, since $\mathbf{e}'\mathbf{e}$ is always smaller than $\epsilon'\epsilon$, we compensate by dividing by $n-k$, or in this case, $n-2$.
- Because our estimate of the sum of squared deviations is biased downward, we divide by a smaller number - inflating the estimate.

- Why do we divide by $n-2$?
- Remember that our estimates of α and β are certainly not the right ones - they are just the best estimates we can get.
- Furthermore, α and β are chosen to *minimize* Σe_i^2
- So the raw sum of squared residuals (Σe_i^2) is less than the true sum of squared errors ($\Sigma \epsilon_i^2$)
- Said another way, since $\mathbf{e}'\mathbf{e}$ is always smaller than $\epsilon'\epsilon$, we compensate by dividing by $n-k$, or in this case, $n-2$.
- Because our estimate of the sum of squared deviations is biased downward, we divide by a smaller number - inflating the estimate.

- Why do we divide by $n-2$?
- Remember that our estimates of α and β are certainly not the right ones - they are just the best estimates we can get.
- Furthermore, α and β are chosen to *minimize* Σe_i^2
- So the raw sum of squared residuals (Σe_i^2) is less than the true sum of squared errors ($\Sigma \epsilon_i^2$)
- Said another way, since $\mathbf{e}'\mathbf{e}$ is always smaller than $\epsilon'\epsilon$, we compensate by dividing by $n-k$, or in this case, $n-2$.
- Because our estimate of the sum of squared deviations is biased downward, we divide by a smaller number - inflating the estimate.

Inference for Bivariate Regression: Intercept

$$S_E^2 = \hat{\sigma}^2 = \frac{\sum E_i^2}{n-2}$$

$$V(A) = \frac{\hat{\sigma}^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}$$

$$SE(\hat{\alpha}) = \sqrt{\frac{\hat{\sigma}^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2}}$$

Inference for Bivariate Regression: Slope

$$S_E^2 = \frac{\sum e_i^2}{n-2}$$

$$V(B) = \frac{\hat{\sigma}^2}{\sum(x_i - \bar{x})^2}$$

$$SE(\hat{\beta}) = \frac{S_E}{\sqrt{\sum(x_i - \bar{x})^2}}$$

Inference for Bivariate Regression: Tests

Confidence Interval for β:

$$\hat{\beta} \pm ME$$

$$\hat{\beta} \pm t_{\alpha/2} * SE\hat{\beta}$$

Test Statistic:

$$t_0 = \frac{\hat{\beta} - \beta_0}{SE(\hat{\beta})}$$

where $t_0 \sim T_{df=n-2}$

Inference for Bivariate Regression

$H_0 : \beta = 0$ $H_A : \beta \neq 0$ Calculate estimated slope and intercept. Could they just be random?

1. State null and alternative hypotheses
2. Calculate standard error of slope and test statistic.
3. How likely is $\hat{\beta}$ if the true value is 0? If there were NO relationship, how likely is it that we would randomly see a slope this big?
4. If very unlikely, reject null hypothesis that no relationship. If rather likely, then maybe it is just random - fail to reject null.

A Long Boring Painful Example

Time to go get coffee or check your email...or help me so we cruise through?

Income	Education
9	8
11	8
12	10
13	12
15	12

Multiple Regression

What's wrong with these models?

- Income $= \alpha + \beta$*Female
- P(War) $= \beta_0 + \beta_1$ * Democ Score
- Party Cohesion $= a + b$ * Electoral Rules

Multiple Regression and Matrix Algebra

This implies a more complex set of models:

- Income $= \alpha + \beta$*Female $+ \beta_2$*Market Salary $+ \beta_3$* Publications
- P(War) $= \beta_0 + \beta_1$ * Democ Score $+ \beta_2$ Proximity $+ \beta_3$*Trade
- Party Cohesion $= b_0 + b_1$ * Electoral Rules $+ b_2$*Parliamentary $+ b_3$*Voter Partisanship

Multiple Regression: Interpretation

Party Cohesion $= b_0 + b_1$ * Electoral Rules $+ b_2$*Parliamentary $+$ b_3*Voter Partisanship

- All else equal, a one percent increase in voter partisanship produces a b_3 increase in Party cohesion.
- All else equal, a one unit increase in Electoral Rules Scores produces a b_1 increase in party cohesion.
- All else equal, being a Parliamentary form of government increases party cohesion by b_2.
- If the electoral rules score were 0, the country were a presidential regime, and voter partisanship were zero, the predicted level of party cohesion is b_0.

Graphically

- With two X variables, the formula defines a plane in three dimensions.
- With more than two X variables, the formula defines a "hyperplane" in k+1 dimensions.
- Display options?

```r
require(rgl)
example(plot3d)

n<-2000
x<-rnorm(n)
y<-rnorm(n)
z<-x+.25*y^3+rnorm(n)
plot3d(x,y,z,size=4)

z<-1+2*x+3*y+rnorm(n)
plot3d(x,y,z,size=4)
```

Estimation

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

We want to find the values of β_0, β_1, and β_2, that minimize the sum of squared residuals (least squares regression).
Minimize:

$$\begin{aligned}\sum e_i^2 &= \sum (y_i - \hat{y}_i)^2 \\ &= \sum (y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \hat{\beta}_2 X_{i2}))^2 \\ &= S(B_0, B_1, B_2)\end{aligned}$$

We remember how to do this in live action, right?

Estimation

$$\frac{\partial S(B_0, B_1, B_2}{\partial B_1} = \sum(-1)(2)(Y_i - B_0 - B_1X_1 - B_2X_2$$
$$\frac{\partial S(B_0, B_1, B_2}{\partial B_1} = (-X_{i1})(2)(Y_i - B_0 - B_1X_{i1} - B_2X_{i2}$$
$$\frac{\partial S(B_0, B_1, B_2}{\partial B_3} = -X_{i2}(2)(Y_i - B_0 - B_1X_{i1} - B_2X_{i2}$$

Applying these partial derivatives, then setting equal to zero...

$$An + B_1\sum X_{i1} + B_2X_{i2} = \sum Y_i$$
$$A\sum X_{i1} + B_1\sum X_{i1}^2 + B_2\sum X_{i1}X_{i2} = \sum X_{i1}Y_i$$
$$A\sum X_{i2} + B_1\sum X_{i2}X_{i1} + B_2X_{i2}^2 = \sum X_{i2}Y_i$$

Estimation

Multiplying everything out, differentiating and setting equal to zero and substituting will produce:

$$\hat{\beta}_0 = \bar{Y} - \beta_1 \bar{X}_1 - \beta_2 \bar{X}_2$$

$$\hat{\beta}_1 = \frac{\sum[(X_1 - \bar{X}_1)(Y - \bar{Y})] \sum(X_2 - \bar{X}_2)^2 - \sum[(X_2 - \bar{X}_2)(Y - \bar{Y})] \sum[(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)]}{\sum(X_1 - \bar{X}_1)^2 \sum(X_2 - \bar{X}_2)^2 - \sum[(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)]}$$

$$\hat{\beta}_2 = \frac{\sum[(X_2 - \bar{X}_2)(Y - \bar{Y})] \sum(X_1 - \bar{X}_1)^2 - \sum[(X_1 - \bar{X}_1)(Y - \bar{Y})] \sum[(X_2 - \bar{X}_2)(X_1 - \bar{X}_1)]}{\sum(X_2 - \bar{X}_2)^2 \sum(X_1 - \bar{X}_1)^2 - \sum[(X_2 - \bar{X}_2)(X_1 - \bar{X}_1)]}$$

You can do this. But it would be a real pain.

Standard Error of the Regression

Reasonably, the **Standard Error of the Regression** is related to the SE_R from the single variable case, and is

$$SE_R = S_E = \sqrt{\frac{\sum E_i^2}{n - k - 1}}$$

where n is the number of obs; k is the number of substantive regressors, and 1 is for the intercept.
Heuristically: we "lose" $k+1$ degrees of freedom by calculating the $k+1$ regression coefficients.

Sums of Squares

The sums of squares are calculated in exactly the same way as simple regression

$$TSS = \sum (Y_i - \bar{Y})^2$$

$$RegSS = \sum (\hat{Y}_i - \bar{Y})^2$$

$$RSS = \sum (Y_i - \hat{Y}_i)^2 = \sum E_i^2$$

Estimation

- If it's a pain once...Alex will make us do it.
- But what about when we have 3 substantive RHS variables?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3))^2$$

- Four?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4))^2$$

- Five?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 X_5))^2$$

Problem:

- A different formula for every possible number of variables
- A pain to calculate by hand

Estimation

- If it's a pain once...Alex will make us do it.
- But what about when we have 3 substantive RHS variables?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3))^2$$

- Four?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4))^2$$

- Five?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 X_5))^2$$

Problem:

- A different formula for every possible number of variables
- A pain to calculate by hand

Estimation

- If it's a pain once...Alex will make us do it.
- But what about when we have 3 substantive RHS variables?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3))^2$$

- Four?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4))^2$$

- Five?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 X_5))^2$$

Problem:

- A different formula for every possible number of variables
- A pain to calculate by hand

Estimation

- If it's a pain once...Alex will make us do it.
- But what about when we have 3 substantive RHS variables?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3))^2$$

- Four?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4))^2$$

- Five?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 X_5))^2$$

Problem:

- A different formula for every possible number of variables
- A pain to calculate by hand

Estimation

- If it's a pain once... Alex will make us do it.
- But what about when we have 3 substantive RHS variables?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3))^2$$

- Four?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4))^2$$

- Five?

$$\sum(y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_5 X_5))^2$$

Solution:

- There *must* be a better way...
- There is, and it is in linear algebra.

Why Linear Algebra

- Multiple, and varied estimating equations are, well, hairy.
- Basis of more advanced work for the rest of the course
- It's elegant. Like Dame Judy Dench.

Because we did this in math boot camp, I'm taking a *very serious risk* that you can do this on your own. If you want more, see `http://polisci2.ucsd.edu/dhughes/204bLecture4LinearAlgebra.pdf`.

Why Linear Algebra

- Multiple, and varied estimating equations are, well, hairy.
- Basis of more advanced work for the rest of the course
- It's elegant. Like Dame Judy Dench.

Because we did this in math boot camp, I'm taking a *very serious risk* that you can do this on your own. If you want more, see `http://polisci2.ucsd.edu/dhughes/204bLecture4LinearAlgebra.pdf`.

General Linear Form

In the single regression case we wrote our estimating equation as:

$$Y_i = A + BX_i + e_i$$

Without loss of generality, for any set of regression coefficients, $1 \dots k$, we can write the following regression equation

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + \epsilon_i$$

Collect those regressors (X's) into a row vector, and the regression coefficients into a column vector:

$$Y_i = [1, x_{1i}, x_{2i}, \cdots, x_{ki}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \epsilon_i$$

That was for one observation

For several (N) observations, we are just adding rows to the $Y_i s$,$X_i s$, & $\epsilon_i s$.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Which we can just *VERY* compactly write as

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

Definition

The **Model Matrix** is the matrix of X's stacked and aligned.

The Vector of Errors

The ϵ is a vector of errors, and so we can again, very succinctly restate the assumptions of OLS:

1. $E(\epsilon) = 0$
2. $V(\epsilon) = \sigma^2 < \infty$
3. $COV(\epsilon_i, \epsilon_j) = 0, \forall i \neq j$
   - $V(\epsilon) = E(\epsilon\epsilon') = \sigma^2_\epsilon \mathbf{I}_n$
4. And if Y is a linear combination of Xs

OLS Estimator

To find the OLS coefficients, we rewrite the linear model as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

where

- $\mathbf{b} = [B_0 + B_1 + \cdots + B_k)'$ is the vector of fitted coefficients; and
- $\mathbf{e} = [E_1 + E_2 + \cdots + E_n]'$ is the vector or residuals.

And so, following our standard optimization...

$$S(b) = \sum E_i^2$$

OLS Estimator

$$S(b) = \sum E_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} = \mathbf{X}\mathbf{b})$$
$$= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b} - \mathbf{b}'\mathbf{X}'\mathbf{y} + \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b}$$
$$= \mathbf{y}'\mathbf{y} - (2\mathbf{y}'\mathbf{X})\mathbf{b} + \mathbf{b}'(\mathbf{X}'\mathbf{X})\mathbf{b}$$

Here, we take the partial derivative with respect to **b**

$$\frac{\partial S(\mathbf{b})}{\partial \mathbf{b}} = 0 - 2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}$$

Setting equal to zero and solving

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}$$

And then, if we can invert $\mathbf{X}'\mathbf{X}$,

$$b = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Properties of Least Squares Estimator

What are properties we care about?

1. Unbiasedness
2. Efficiency
3. Consistent
4. Sufficient

Unbiasedness

Fix the model matrix, **X**. Then,

- **b** is a linear transformation of the response variable

$$b = (X'X)^{-1}X'y = My$$

- Where we have defined $M \equiv (X'X)^{-1}X'$
- And the expected value can be calculated easily

$$\begin{aligned} E(b) &= E(My) \\ &= ME(y) \\ &= (X'X)^{-1}X'(X\beta) \\ &= \beta \end{aligned}$$

And so the OLS estimator **b** is unbiased. Sweet!

Unbiasedness

Fix the model matrix, **X**. Then,

- **b** is a linear transformation of the response variable

$$b = (X'X)^{-1}X'y = My$$

- Where we have defined $M \equiv (X'X)^{-1}X'$
- And the expected value can be calculated easily

$$\begin{aligned} E(b) &= E(My) \\ &= ME(y) \\ &= (X'X)^{-1}X'(X\beta) \\ &= \beta \end{aligned}$$

And so the OLS estimator **b** is unbiased. Sweet!

Efficiency

We want the most efficient estimate, *based on the data we have*.

$$
\begin{aligned}
V(b) &= V(My) \\
&= MV(y)M' = [(X'X)^{-1}X']\sigma_\epsilon^2 \mathbf{I}_n[(X'X)^{-1}X']' \\
&= \sigma_\epsilon^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma_\epsilon^2(X'X)^{-1}
\end{aligned}
$$

Efficiency

We want the most efficient estimate, *based on the data we have*.

$$
\begin{aligned}
V(b) &= V(My) \\
&= MV(y)M' = [(X'X)^{-1}X']\sigma_\epsilon^2 \mathbf{I}_n[(X'X)^{-1}X']' \\
&= \sigma_\epsilon^2(X'X)^{-1}X'X(X'X)^{-1} \\
&= \sigma_\epsilon^2(X'X)^{-1}
\end{aligned}
$$

Distributions of β

To derive $E(\mathbf{b})$ and $V(\mathbf{b})$, then "all" we needed were:

1. Linearity: $E(\mathbf{y}) = \mathbf{X}\beta$
2. Known, constant variance: $V(b) = \sigma^2_\epsilon(\mathbf{X}'\mathbf{X})^{-1}$, and
3. Independence of Xs: $V(\mathbf{y}) = \sigma^2_\epsilon \mathbf{I}_n$

But, if we make the additional assumption that conditional on $\mathbf{X}\beta$ then Y is normally distributed, then soo tooo is β normally distributed:

$$\mathbf{b} \sim N_{k+1}[\beta, \sqrt{\sigma^2_\epsilon(\mathbf{X}'\mathbf{X})^{-1}}]$$

Inference for Multiple Regression

We know (from the last slide) that **b** is distributed normal.

- $E(b) = \beta$
- $VCOV(b) = \sigma_\epsilon^2(\mathbf{X}'\mathbf{X}^{-1})$

Theorem

Then, for some individual coefficient, B_j, the coefficient is also normally distributed with expectation β_j and sampling variance $\sigma_\epsilon^2 v_{jj}$, where v_{jj} is the jth diagonal entry of the VCOV matrix.

Then, most awesomely, the ratio

$$\frac{(B_j - \beta_j)}{\sigma_\epsilon \sqrt{v_{jj}}}$$

is on the unit-normal distribution ($N(0, 1)$). So, most, most awesomely, we can calculate the following test statistic for the hypothesis:

$$H_0 : \beta_j = \beta_j$$

$$Z_0 = \frac{B_j - \beta_j}{\sigma_\epsilon \sqrt{v_{jj}}}$$

Inference for Multiple Regression

- Just like in the simple OLS case, we don't actually *know* σ_ϵ^2.
- But, we do know an unbiased estimate (estimates on estimates..., for days...):

$$S_E^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k-1}$$

- And so, we can estimate the VCOV of the least squares coefficients as:

$$\begin{aligned} \hat{V}(\mathbf{b}) &= S_E^2(\mathbf{X}'\mathbf{X})^{-1} \\ &= \frac{\mathbf{e}'\mathbf{e}}{n-k-1}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Multiple Regression Inference

$H_0 : \beta_j = 0$ $H_A : \beta_j \neq 0$
Calculate estimated slopes and intercept. Could they just be random?

1. State Null and alternative hypotheses
2. Calculate test statistic.
3. How likely is $\hat{\beta}_j$ if the true value is 0? If there were NO relationship, how likely is it that we would randomly see a slope this big?
4. If very unlikely, reject null hypothesis that no relationship. If rather likely, then maybe it is just random - fail to reject null.

Multiple Regression Inference

$$\hat{\beta} \sim N(\beta, \sigma_E^2(X'X)^{-1})$$

Standard error for a slope under multiple regression:

$$\mathbf{V}\hat{(\beta)} = S_E^2(X'X)^{-1}$$

Square root of diagonals are the standard errors for each coefficient.

$$\frac{\hat{\beta} - \beta_0}{SE\hat{(\beta_j)}} \sim t_{n-k-1}$$