Single Variable Regression: Inference

EC420 MSU Spring 2021

Justin Kirkpatrick

Last updated February 05, 2021

This lecture

Goal:

1. Review where we are in single-variable regression
2. Review statistical inference
3. Expectation of the estimate $\hat{\beta}$
4. Variance of the estimate, $\hat{\beta}$
5. Homoskedasticity assumption
6. An example

Review

Review single-variable OLS

We have a linear-in-parameters single-variable model:

$$y = \beta_0 + \beta_1 x + u$$

- "In terms of the random sample" (W2.5a): $y_i = \beta_0 + \beta_1 x_i + u_i$
- "Fitting a line"
  - The PRF and the SRF
- $\hat{\beta}_1 = \frac{\widehat{Cov}(x,y)}{\widehat{Var}(x)}$
- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
- SST (Sum of Squares Total) = $\sum_{i=1}^{N} (y_i - \bar{y})^2$
  - SSE (Sum of Squares Explained) = $\sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2$
  - SSR (Sum of Squares Residual) = $\sum_{i=1}^{N} (\hat{u}_i - \bar{\hat{u}})^2$

When we have a random variable with a population characteristic of interest

- X with population mean μ_X

And a sample x_i of observed draws from the RV, then we can make a *hypothesis* about μ_X:

- $H_0 : \mu_X = 0 \quad and \quad H_A : \mu_X \neq 0$

Then, we can develop a sample *test statistic* for the population characteristic:

- $\bar{X} = \frac{1}{N} \sum x_i$

And we know two things about $\bar{X}$:

- $E[\bar{X}] = E[X] = \mu_X$
- $Var(\bar{X}) = \frac{\sigma_X^2}{N}$

Review statistical inference

If we're smart, we make a sample test statistic with a distribution that we know:

$$\frac{\bar{X} - H_0}{\sqrt{\frac{\hat{\sigma}^2}{N}}} \sim N(0,1)$$

or if we don't know σ_X^2

$$\frac{\bar{X} - H_0}{\sqrt{\frac{\hat{s}^2}{N}}} \sim t_{df}$$

We can test our hypothesis by comparing our sample test statistic result to the hypothesized value.

- If observed $\bar{X} = 4$ and observed $\frac{\hat{\sigma}_X}{\sqrt{N}} = 1$, is $H_0 : \mu_X = 0$ likely to be rejected?

We can think of β_1 as the test statistic for the relationship between x and y

What do we need to test a hypothesis?

A distribution

- $E[\hat{\beta}_1]$
- $Var(\hat{\beta}_1)$
- $\hat{\beta}_1 \sim N(?,?)$ (let's assume we know it's Normal for now)

If we did know these three things, we could test any interesting H_0

- Anyone know one that might be interesting?

Expectation of the estimate

Now, remember that we are looking at $\hat{\beta}$, not β itself.

- β is a population parameter,
  - It is unobserved
  - It is a constant
  - Because it is a constant, it can move in and out of Expectations and Variances as a constant would.
- $\hat{\beta}$ depends on the sample. It is therefore a random variable.
  - It has an expected value
  - It has a variance
  - We can use a statistical test on hypothesis about $\hat{\beta}$.

β and $\hat{\beta}$ are two different things, we are interested in whether or not they are the same in $\boldsymbol{E}$

Gauss-Markov

Carl Friedrich Gauss

Andrey Markov

Both images courtesy of Wikimedia Commons

Review statistical inference

MICHIGAN STATE UNIVERSITY

We will need to make the following four assumptions to get $E[\hat{\beta}]$

Gauss-Markov Assumptions

1. SLR.1: In the population, y is a linear function of the parameters, x, and u: $y = \beta_0 + \beta_1 x + u$
2. SLR.2: The sample $(y_i, x_i) : i = 1, 2, \cdots, n$ follows the population model and are independent.
3. SLR.3: "Sample Variation in the Explanatory (X) Variable". That is, x_i is not the same for all i's.
4. SLR.4: "Zero conditional mean". $E[u|x] = 0$ for all x.

File these away for a minute. We'll need them.

Expectation of the estimate: Bias

We know how to calculate, from our sample, $\hat{\beta}$

We would hope (and will now prove) that $E[\hat{\beta}] = \beta$

- This is the first step in deriving the distribution of $\hat{\beta}$
- Section 2.5a of Wooldridge
    - If $E[\hat{\beta}] = \beta$, then the estimator is **unbiased**. Let's see if this is the case:

$$\hat{\beta}_1 = \frac{\widehat{Cov}(X,Y)}{\widehat{Var}(X)} = \frac{\frac{1}{N-1}\sum(x_i-\bar{x})(y_i-\bar{y})}{\frac{1}{N-1}\sum(x_i-\bar{x})^2} = \frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2}$$

- The first equality is our derivation of $\hat{\beta}_1$.
- The second uses the definition of Covariance and Variance
- The third cancels out the $\frac{1}{N-1}$ and does some simplification of the numerator (see Appendix A of Wooldridge)

Expectation of the estimate

Let's rewrite, then take expectations to see what the expectation of the estimate is:

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}$$

- Rewrite $\sum(x_i - \bar{x})^2$ as SST_x. After all, it's the total sum of squared deviations from $\bar{x}$.
  - We are just adding that subscript to make sure we remember where it come from.
  - Remember, we originally introduced SST as the *Sum of Squares Total* in a regression and it referred to the total variance in Y, the left-hand-side (LHS) of our regression.
- Substitute our model for y_i: $y_i = \beta_0 + \beta_1 x_i + u_i$
- Rename $x_i - \bar{x}$ as d_i, for **d**eviations from $\bar{x}$.
  - This will make it easier to work with.

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum(x_i - \bar{x})^2} = \frac{\sum(d_i\beta_0) + \sum(d_i\beta_1 x_i) + \sum(d_i u_i)}{SST_x}$$

Let's take a second and make sure everyone is on board here. Remember, $d_i = x_i - \bar{x}$.

Move the β's out as they are constants:

$$\hat{\beta}_1 = \frac{\overbrace{\beta_0 \sum(d_i)}^{\text{First term}} + \overbrace{\beta_1 \sum(d_i x_i)}^{\text{Second term}} + \overbrace{\sum(d_i u_i)}^{\text{Third term}}}{SST_x}$$

In that numerator, $\beta_0 \sum(d_i)$ must be 0 since $\sum(x_i - \bar{x}) = 0$. We can ignore it!

$$\hat{\beta}_1 = \frac{0}{SST_x} + \frac{\beta_1 \sum(d_i x_i)}{SST_x} + \frac{\sum(d_i u_i)}{SST_x}$$

The second term:

$$\frac{\beta_1 \sum(d_i x_i)}{SST_x} = \frac{\beta_1 \sum((x_i - \bar{x})x_i)}{SST_x} = \frac{\beta_1 \sum((x_i - \bar{x})(x_i - \bar{x}))}{SST_x} = \frac{\beta_1 SST_x}{SST_x}$$

And since SST_x is in the denominator and cancels, we will end up with β_1.

This is very important: notice that we now have the true value of beta in there.

β_1 is the true beta. It is *part of* $\hat{\beta}_1$, but there's still the third term:

$$\frac{\sum(d_i u_i)}{SST_x} = \frac{\sum((x_i - \bar{x})u_i)}{SST_x}$$

$$\hat{\beta}_1 = 0 + \beta_1 + \frac{\sum((x_i - \bar{x})u_i)}{SST_x}$$

We will say that the estimate of β_1, $\hat{\beta}_1$ is the true β plus some term.

Expectation of the estimate

$$\hat{\beta}_1 = \beta_1 + \frac{\sum((x_i - \bar{x})u_i)}{SST_x}$$

Conditional on the x_i's (our sample), the entire source of randomness here is in u_i.

Now, we take the last step to show that the $E[\hat{\beta}_1] = \beta_1$.

We will need our four assumptions. Specifically, the fourth.

Our assumptions from before:

Gauss-Markov Assumptions (fancy name for what you already know)

1. SLR.1: In the population, y is a linear function of the parameters, x, and u: $y = \beta_0 + \beta_1 x + u$
2. SLR.2: the sample $(y_i, x_i) : i = 1, 2, \cdots, n$ follows the population model and are independent.
3. SLR.3: "Sample Variation in the Explanatory (X) Variable". That is, x_i is not the same for all i's.
4. SLR.4: "Zero conditional mean". $E[u|x] = 0$ for all x.

Expectation of the estimate

MICHIGAN STATE UNIVERSITY

Now, we can go to our equation for $\hat{\beta}_1$:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum((x_i - \bar{x})u_i)}{SST_x}$$

We can take E of each side:

$$E[\hat{\beta}_1] = E[\beta_1] + E\left[\frac{\sum((x_i - \bar{x})u_i)}{SST_x}\right]$$

$E[\beta_1] = \beta_1$.

For any value of x, $E[u|x] = 0$ under SLR.4.

- No matter what x or $(x_i - \bar{x})$ is, once we condition on x, the second term is zero in expectation.

$\Rightarrow E[\hat{\beta}_1] = \beta_1$.

Our estimator, $\hat{\beta}_1$ is unbiased, and we know it is distributed with mean of β_1

Expectation of the estimate

$\boldsymbol{E}[\hat{\beta}_0] = \beta_0$ is shown in Wooldridge 2.5a.

- " $\hat{\beta}_0$ is an unbiased estimator of β_0 "

Now, we simply need to fill in the variance of $\hat{\beta}$ to have a test statistic for β.

Variance of the estimate

A brief interlude about proofs

Question: have you had proofs in your previous classes?

Gauss-Markov Assumptions

1. SLR.1: In the population, y is a linear function of the parameters, x, and u:
   $y = \beta_0 + \beta_1 x + u$
2. SLR.2: the sample $(y_i, x_i) : i = 1, 2, \cdots, n$ follows the population model and are independent.
3. SLR.3: "Sample Variation in the Explanatory (X) Variable". That is, x_i is not the same for all i's.
4. SLR.4: "Zero conditional mean". $E[u|x] = 0$ for all x.

These get us to "$\hat{\beta}$ is unbiased"

Gauss-Markov Assumptions

1. SLR.1: In the population, y is a linear function of the parameters, x, and u: $y = \beta_0 + \beta_1 x + u$
2. SLR.2: the sample $(y_i, x_i) : i = 1, 2, \cdots, n$ follows the population model and are independent.
3. SLR.3: "Sample Variation in the Explanatory (X) Variable". That is, x_i is not the same for all i's.
4. SLR.4: "Zero conditional mean". $E[u|x] = 0$ for all x.

Add one more assumption:

Add SLR.5: $Var[u|x] = \sigma_u^2$ for all x.

- This is similar to the conditional mean, but says that every u_i is drawn from a variable whose distribution has the same value for σ^2.

Variance of the estimate

SLR.5: $Var[u|x] = \sigma_u^2$ for all x

- This is similar to the conditional mean, but says that every u_i is drawn from a variable whose distribution has the same value for σ^2.

- We do **not** need this assumption to show that $\hat{\beta}$ is an unbiased estimator for β

  - But we do need this assumption to calculate the variance of $\hat{\beta}$.

- It does not mean that we know σ_u^2. We don't

Start with where we left off on β_1:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum((x_i - \bar{x})u_i)}{SST_x}$$

Instead of taking the expectation as we did for proving unbiasedness, we take the **variance**:

$$Var(\hat{\beta}_1) = Var(\beta_1) + Var\left[\frac{\sum((x_i - \bar{x})u_i)}{SST_x}\right] + 2Cov\left(\beta_1, \left[\frac{\sum((x_i - \bar{x})u_i)}{SST_x}\right]\right)$$

- Because the variance of any constant (like β_1) is 0, we can drop that 1st term.
- Because $Cov(c, X) = 0$ when c is a constant, we can drop the $2Cov(\cdots)$ term.

Variance of the estimate

MICHIGAN STATE UNIVERSITY

This leaves us with:

$$Var(\hat{\beta}_1) = Var\left[\frac{\sum((x_i - \bar{x})u_i)}{SST_x}\right] = Var\left[\frac{1}{SST_x}\sum((x_i - \bar{x})u_i)\right]$$

We can condition on x_i's again, and make the same argument that, conditional on x_i, we can take them out of the Var term.

- When we do this, we must **square** what we remove:

$$Var(\hat{\beta}_1) = \frac{1}{SST_x^2} \times Var\left[\sum(x_i - \bar{x})u_i\right] = \frac{1}{SST_x^2} \times \left[\sum(x_i - \bar{x})^2\right] Var(u_i)$$
$$= \frac{SST_x}{SST_x^2}\sigma_u^2 = \frac{1}{SST_x}\sigma_u^2$$

Variance of the estimate

So variance is:

$$Var(\hat{\beta}_1) = \frac{\sigma_u^2}{SST_x}$$

For any realization of x

- Variance of the estimator is increasing in σ_u^2.
- Variance of the estimator is decreasing in SST_x, variation in X.

Good, but we don't know σ_u^2, do we?

- $\hat{u}$ seems like a good start.
- In our model, u_i is the *error*, but we observe $\hat{u}$, which is the residual.
  - $\hat{u}_i = u_i - (\hat{\beta}_0 - \beta_0) - (\hat{\beta}_1 - \beta_1)x_i$
  - So $E[\hat{u}_i] = u_i$

As Wooldridge states: "the *error*, u, shows up in the equation containing the *population parameters*, β. The residual shows up in the *estimated* equation with $\hat{\beta}$.

- Remember, u_i is not observed.
- But $\hat{u}_i$ is observed.

Variance of the estimate

We can use $\sum_{i=1}^{N} \hat{u}_i^2$ as an estimator for σ_u^2 if we make this small adjustment.

- $\hat{\sigma}_u^2 = \frac{1}{(N-2)} \sum_{i=1}^{N} \hat{u}_i^2 = \frac{SSR}{N-2}$
- This is because we know two things about $\hat{u}$:

$$\sum \hat{u} = 0$$

  and

$$\sum x_i \hat{u}_i = 0$$

- We lose two **degrees of freedom**.
  - If we know all but two u_i's, we could calculate the last two knowing these.
- **degrees of freedom** will be very important when we get to multiple regression.

This is the Standard Error of the Regression, SER

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{\sum \hat{u}_i^2}{(N-2)}}$$

We have used all five assumptions, but we can now say we know the distribution of $\hat{\beta}$:

$$\hat{\beta}_1 \sim N(\beta_1, \frac{\hat{\sigma}_u^2}{SST_x})$$

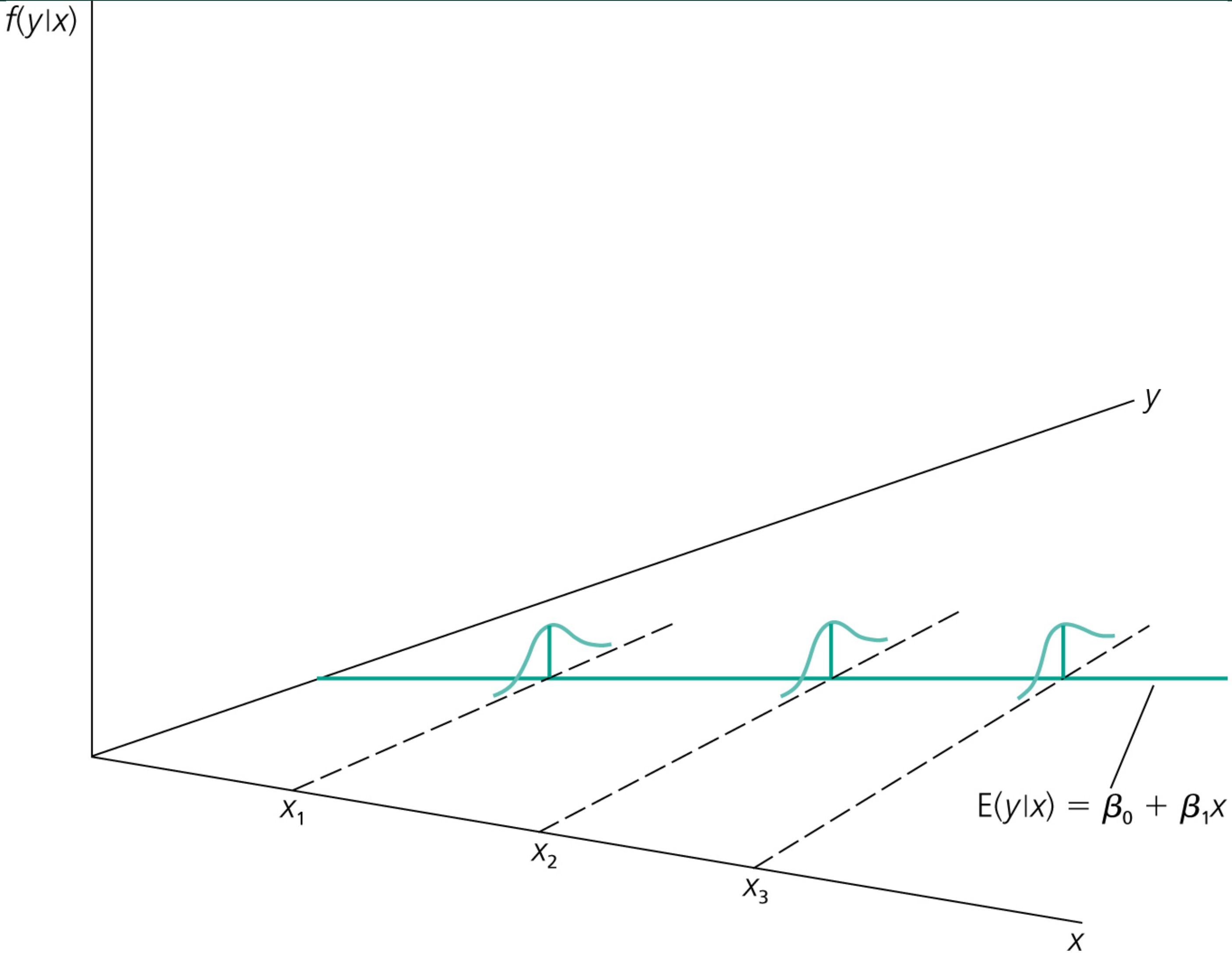
If we want to test a hypothesis about β_1, we now can.

But only **if** we assume homoskedasticity - that $Var(u|x) = Var(u) = \sigma_u^2$.

Let's take a look at this assumption briefly.

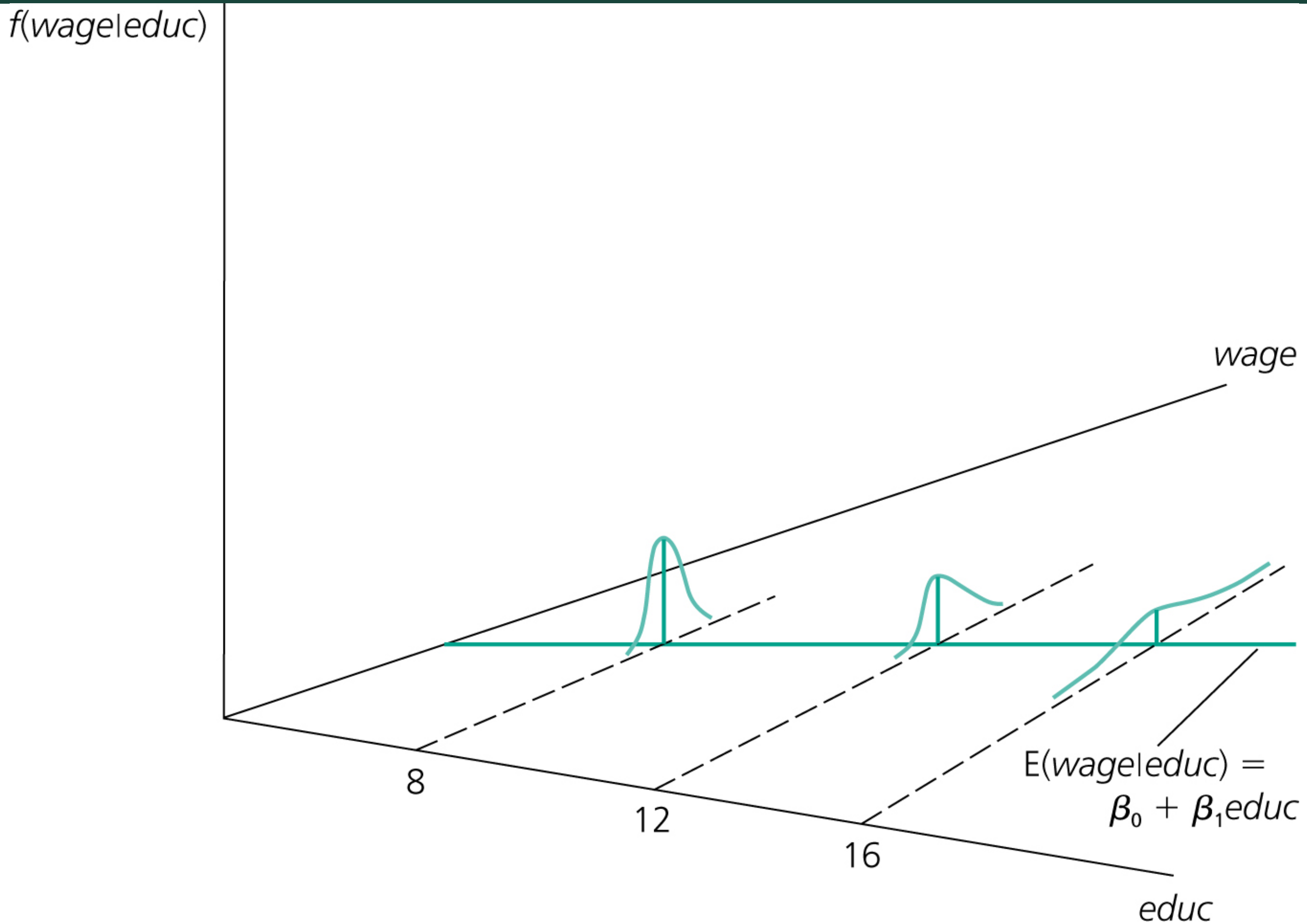
- Later on, we'll talk about how to adjust the Standard Error of the Regression for heteroskedasticity.

Variance of the estimate



Homoskedasticity (from Wooldridge)

Variance of the estimate



Heteroskedasticity (from Wooldridge)

ID	Outcome	Dose
1	11.4	4
2	5.5	1
3	9.6	1
4	17.2	7
5	6.8	1

Statistic	**Value**
$\bar{y}$	10.1
$\bar{x}$	2.8
$SST_y = \sum(y_i - \bar{y})^2$	84.4
$SST_x = \sum(x_i - \bar{x})^2$	28.8
$\sum(y_i - \bar{y})(x_i - \bar{x})$	46.5

What is $\hat{\beta}_1$?

What is $\hat{\beta}_0$?

ID	Outcome	Dose	Fitted	Residual
1	11.4	4		
2	5.5	1		
3	9.6	1		
4	17.2	7		
5	6.8	1		

- Calculate $\hat{y}$ using β_0 and β_1
- Calculate $\hat{u}$ using $y_i - \hat{y}$
- Calculate $\hat{\sigma}_u^2$
  - Remember to divide by $(n-2)$ for correct degrees of freedom

---

The formula for $Var(\hat{\beta}_1)$ is $\frac{\hat{\sigma}_u^2}{SST_x}$

- What is the distribution of $\hat{\beta}_1$?

The formula for $Var(\hat{\beta}_0)$ is $\hat{\sigma}_u^2 \left[\frac{1}{N} + \frac{\bar{x}^2}{SST_x}\right]$ (from Wooldridge)

- What is the distribution of $\hat{\beta}_0$?

Check your work here:

```
##
## Call:
## lm(formula = Outcome ~ Dose, data = df)
##
## Residuals:
##       1       2       3       4       5
## -0.6375 -1.6937  2.4062  0.3188 -0.3938
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.5792     1.2113   4.606   0.0192 *
## Dose          1.6146     0.3285   4.915   0.0161 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.763 on 3 degrees of freedom
## Multiple R-squared:  0.8896,    Adjusted R-squared:  0.8527
## F-statistic: 24.16 on 1 and 3 DF,  p-value: 0.01613
```