# Single Variable Regression

## EC420 MSU Online

Justin Kirkpatrick
Last updated May 28, 2021

MICHIGAN STATE UNIVERSITY

## Lectures:

1. Population Regression Function: What are we after?
2. Ordinary Least Squares with an example
3. Goodness of fit measures
4. Interpretation of coefficients
5. Rescaling Y and X
6. Non-linear functional forms
7. Regression in R
8. Inference and hypothesis testing: Expectation of $\hat{\beta}_1$
9. Inference and hypothesis testing: Variance of $\hat{\beta}_1$
10. An example

# Population Regression Function

top

# The problem at hand...



We have some data on two (or more, later) variables that we think move together in an interesting way.

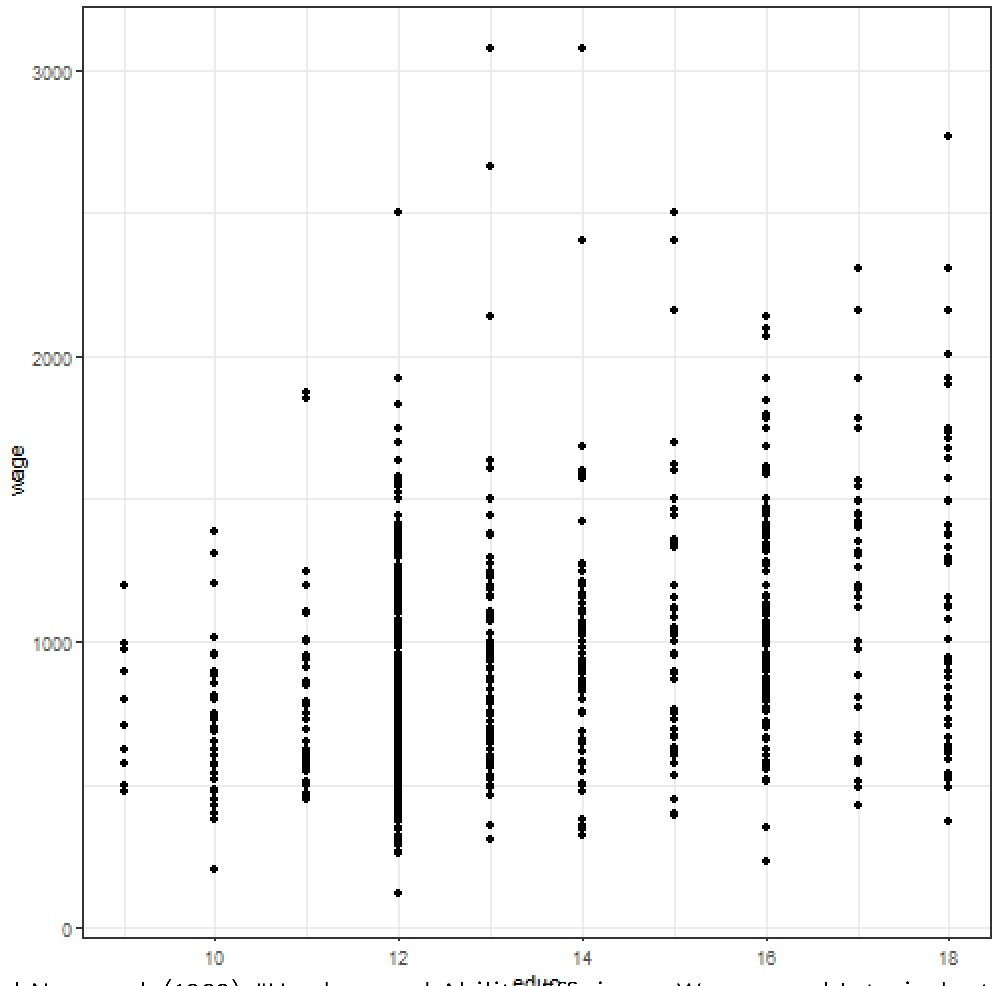
- Wage and education
- Cigarette smoking and life expectancy
- COVID cases and vaccine rates

We want to quantify and test this relationship

- Predict a change
- Test a theory
- Win a bet?

We have a **sample**, but want to predict/test something about the population

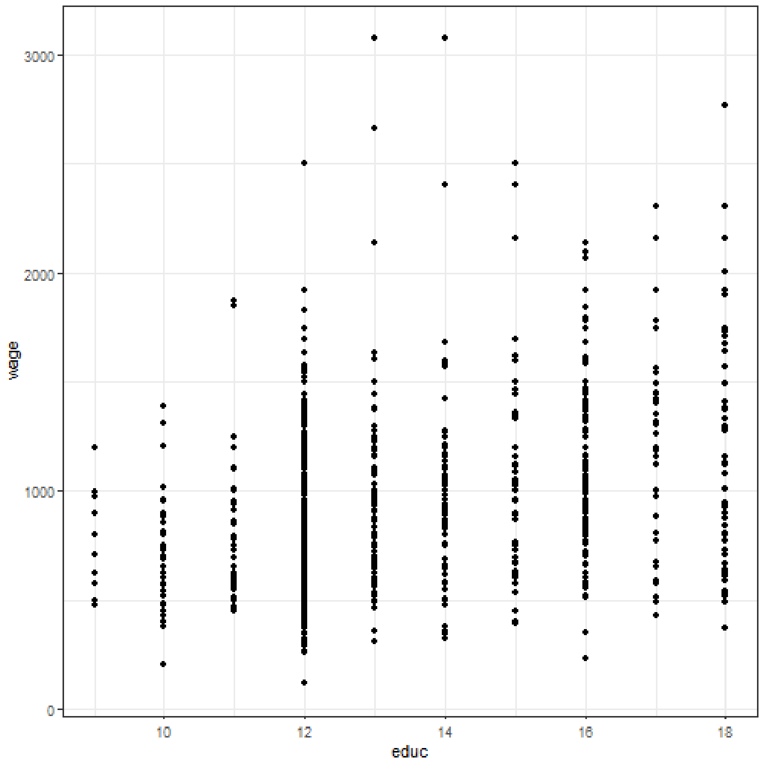
Wage data used in Wooldridge `wage2`




Data from Blackburn and Neumark (1992), "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials" *Quarterly Journal of Economics* 107, 1421-1436

The data looks like this:

| wage | educ |
|---|---|
| 769 | 12 |
| 808 | 18 |
| 825 | 14 |
| 650 | 12 |
| 562 | 11 |
| 1400 | 16 |




$N = 935$, $\overline{wage} = 957.95$ and $\overline{educ} = 13.47$

What we'd like to have is a function that tells us how $wage$ and $educ$ move together in the **population**

## In a perfect world, we would have some function for $X = educ$ and $Y = wage$:

$$g(x) = y$$

Where we give the function any realization of $x$, and it spits out exactly $y$.

## But that isn't going to happen

Think about the data we just looked at - when $educ = 12$ we observed $wage = 769$ and $wage = 650$. The dream function doesn't exist! There are other things not accounted for besides $educ$.

So we settle for something that tells us about the **expectation** of $Y$.
The Population Regression Function

$$E[Y|X] = \beta_0 + \beta_1 X$$

The *Population Regression Function* (PRF) describes the relationship between $X$ and the **conditional expectation** of $Y$.

- $X$ and $Y$ are random variables
- $\beta_0$ and $\beta_1$ are **population parameters**
- We have restricted the $E[Y|X]$ to be a *linear* function of $X$.
  - It can be drawn as a straight line with an intercept and constant slope
  - We will be estimating $\beta_0$ and $\beta_1$

The PRF:

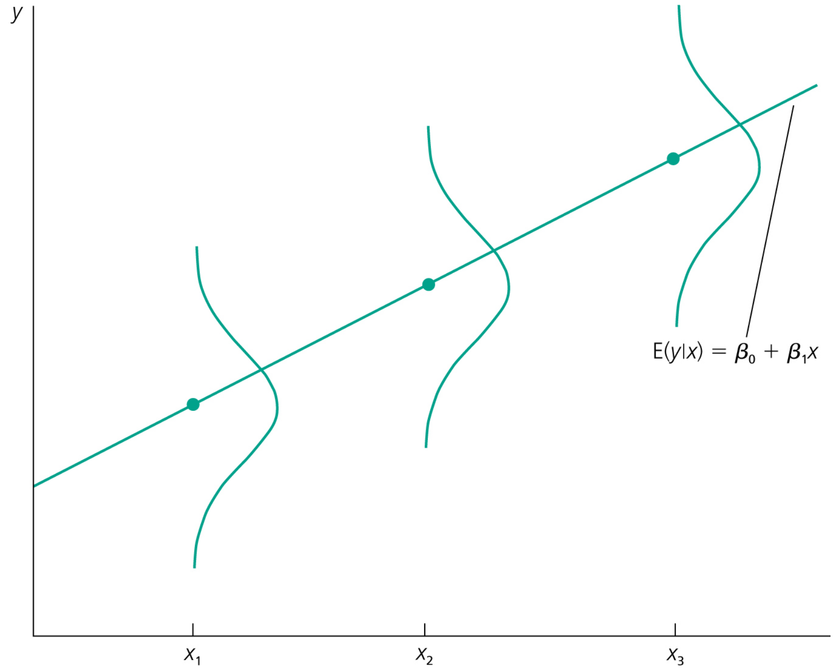
$$E[Y|X] = \beta_0 + \beta_1 X$$

Let $Y = wage$ and $X = educ$

- $E[Y|X = x]$ gives us the expectation of $Y$ (wage) conditional on some realized value of $X = x$ (educ)
- So, if $educ = 16$, then $E[Y|X = 16] = \beta_0 + \beta_1 \times 16$
  - We can plug in any $x_i$ and get the **expected value** of the paired $y_i$

Question: Will the PRF return exactly $y_i$ given a value $x_i$?

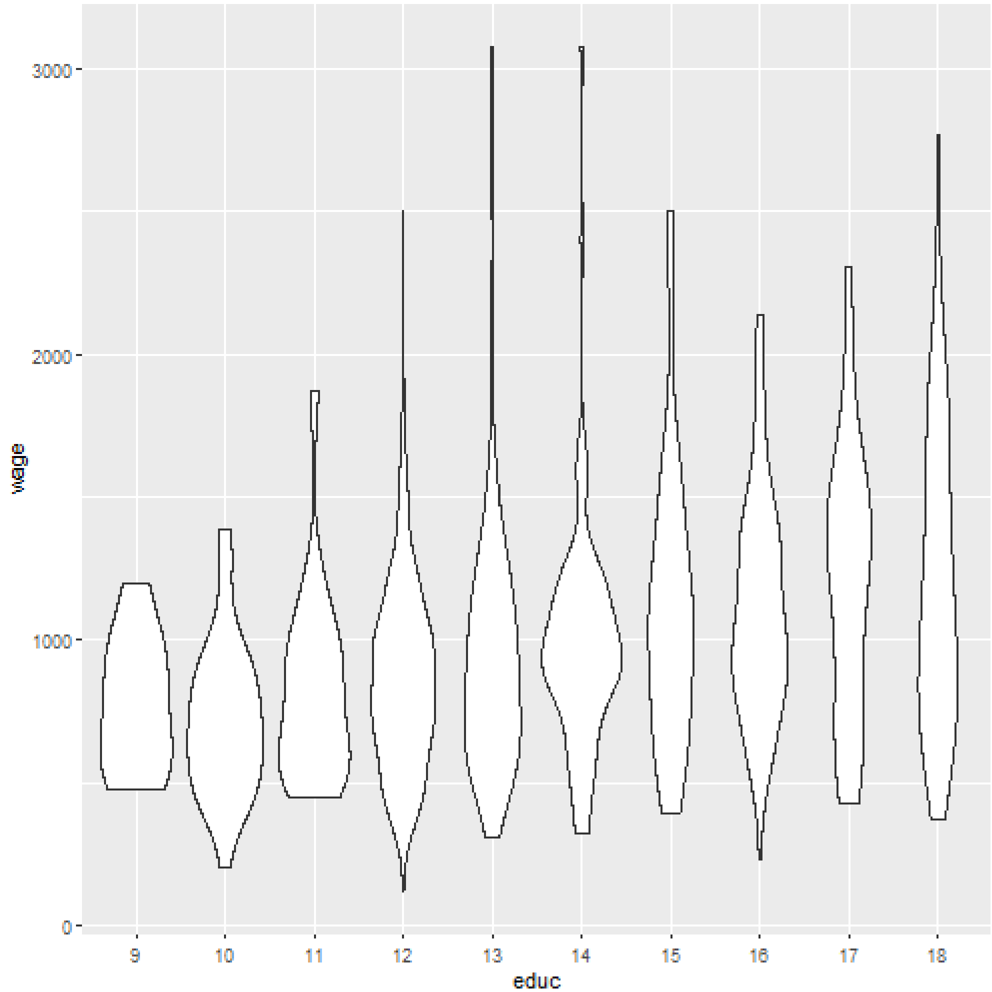
No. It can only give us the expectation.



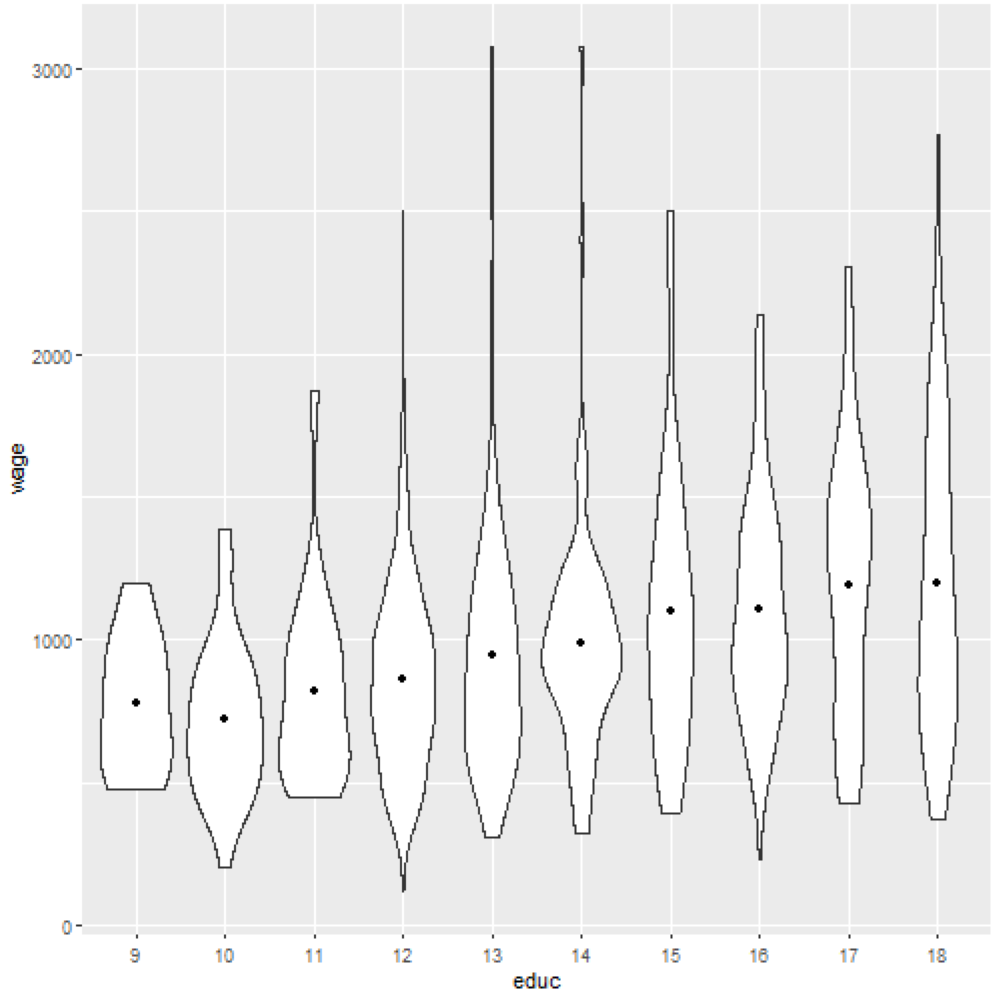

Ch. 2.1 of Wooldridge, example of a conceptual PRF. The line defines the PRF, the expectation of Y conditional on X

# Population Regression Function

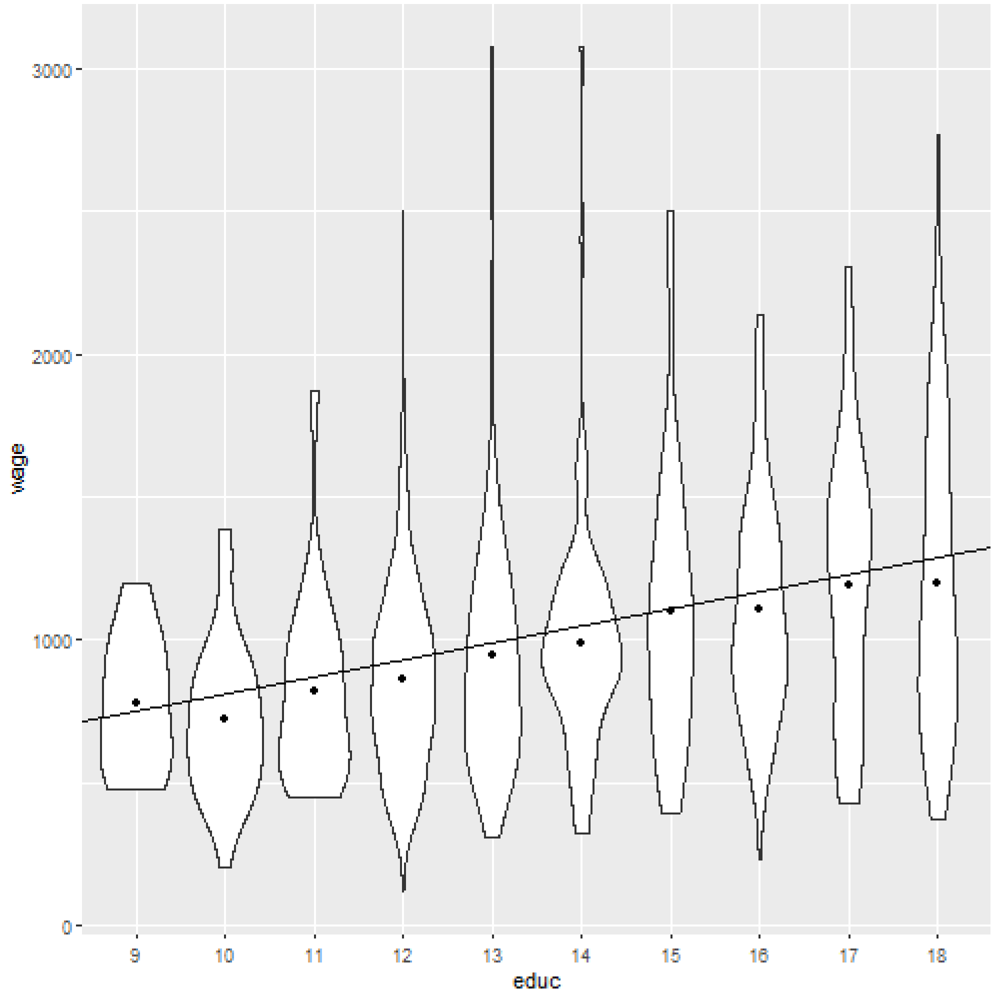





This is the wage data. Each "blob" is an empirical histogram of the data for that value of $educ$ (they are symmetrical). This is called a *violin plot*. It is the empirical counterpart of the previous plot from Wooldridge




Each point is the sample mean for each value of $educ$.




A (linear) PRF would be the straight line that best fits the data. **Regression fits that line**. A brief look at the line shows that it certainly won't be perfect!

## What happens, then, if we want to write $Y$ exactly?

The PRF gives us the *expectation* of $Y$

- So we add a **stochastic error term**, the difference between $E[Y|X]$ and $Y$:

$$Y = E[Y|X] + U = \beta_0 + \beta_1 X + U$$

This is the stochastic population regression function

$U$ is also the **population error term**, and is itself a **random variable**.

- It must be that $E[U] = 0$

Now we can write our **simple linear regression model**:

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

This is a statement about the relationship between observed realizations $(y_i, x_i)$ based on the population parameters $\beta_0, \beta_1$

We will call $u$ the **error term** - it is the difference between the conditional expected mean and the observed $y_i$ given a value of $x_i$.

- It might be different for two identical realizations of $x_i$

Naturally, we would think that the "right" value of the population parameters, $\beta = \{\beta_0, \beta_1\}^*$, minimizes all of the $u_i$ values in a sample. That's where Ordinary Least Squares comes in.

* A parameter vector is just a list of numbers.

## The Sample Regression Function

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

## The "hats" are important

They mean we have a *sample estimate* of the population parameters.

- $\beta_0, \beta_1$ are the population
- $\hat{\beta}_0, \hat{\beta}_1$ are the sample estimates and will change when the sample changes
  - So they are random variables!

## Where did $u$ go?

Since we have a hat on $y_i$, there is no $u$, but $\hat{y}_i \neq y_i$.

- Define $\hat{u}_i = \hat{y}_i - y_i$.
- $\hat{u}_i$ is the *residual*.

## To summarize:

The $PRF$ is

$$E[Y|X] = \beta_0 + \beta_1 X$$

The simple linear regression model is:

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

The SRF is:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

And if we want to write the sample regression model:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$$

# Ordinary Least Squares

top

## We have a linear PRF

$$E[Y|X] = \beta_0 + \beta_1 X$$

## What happens, then, if we want to write $Y$ exactly?

The PRF gives us the *expectation* of $Y$

- So we add that **stochastic error term**, the difference between $E[Y|X]$ and $Y$:

$$Y = E[Y|X] + U = \beta_0 + \beta_1 X + U$$

This is the stochastic population regression function

$U$ is also the **population error term**, and is itself a **random variable**.

- It must be that $E[U] = 0$

## And we can write our SRF:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

## How do we get those $\hat{\beta}$'s in the SRF?

We make two assumptions:

**First**, if the expectation of $Y$ equals $\beta_0 + \beta_1 X$, then *in expectation*, $E[U] = 0$. Because:

$$E[Y|X] = \beta_0 + \beta_1 X \quad \textbf{and} \quad Y = \beta_0 + \beta_1 X + U$$

**Second**, our first assumption should hold no matter what $x$ is. So, it should be true that $E[U|X] = 0$ for **all** possible values of $X$.

There are very important assumptions as they will define our Sample Regression Function (SRF).

## Let's make these assumptions formal:

1. $E[U] = 0$.
   - As long as there is a $\beta_0$ (regardless of $\beta_1$), this is true. We call this assumption **trivial**.
2. $E[U|X] = E[U]$
   - **Mean independence**. The **mean** of $U$ is the same, regardless of the value of $X$:

These are **population moments**

- A **moment** is a specific attribute of a distribution

Economists spend a lot of time showing mean independence $E[U|X] = E[U]$.

Two quick math reminders before we introduce the Ordinary Least Squares (OLS) estimator for $\beta$:

$$Cov(Y, X) = E[YX] - E[Y]E[X]$$

and

$$\text{If} \quad E[U] = 0$$

then

$$Cov(U, X) = E[UX] - E[U]E[X] = E[UX] - 0$$

And note that the simple linear regression model $\boldsymbol{y = \beta_0 + \beta_1 x + u}$ implies that:

$$\boldsymbol{u = y - \beta_0 - \beta_1 x}$$

## Since $u = y - \beta_0 - \beta_1 x$:

Let's write Assumption 1 and Assumption 2 using expectations of the regression model from before

- $E[U] = 0 \Rightarrow E[(y - \beta_0 - \beta_1 x)] = 0$
- $E[U|X] = 0 \Rightarrow E[x(y - \beta_0 - \beta_1 x)] = 0$
  - To see this, picture any expected value of $x$. Now, multiply it by $0$.

- How many equations?

- How many unknowns?

# Let's solve for $\beta$. To the "board"!

These are *moments*, and this way of deriving $\beta$ is known as "method of moments".

What we just derived on the board depends on **population** moments: $Cov(X,Y)$ and $Var(X)$.

But, just as before when we didn't know $\mu$ but we could calculate $\bar{y}$ (and we even know something about the distribution of $\bar{y}$), we can calculate sample values for $Cov(X,Y)$ and $Var(X)$

## Sample estimates get a $\hat{}$

So if we take a sample and calculate, from that sample, $Var(X)$, we would call it $\widehat{Var}(X)$

First, let's tackle the estimate of $\hat{\beta}_1$.

We know how to calculate the sample covariance:

- $\widehat{Cov}(Y, X) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$

We know how to calculate the sample variance:

- $\widehat{Var}(X) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$

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

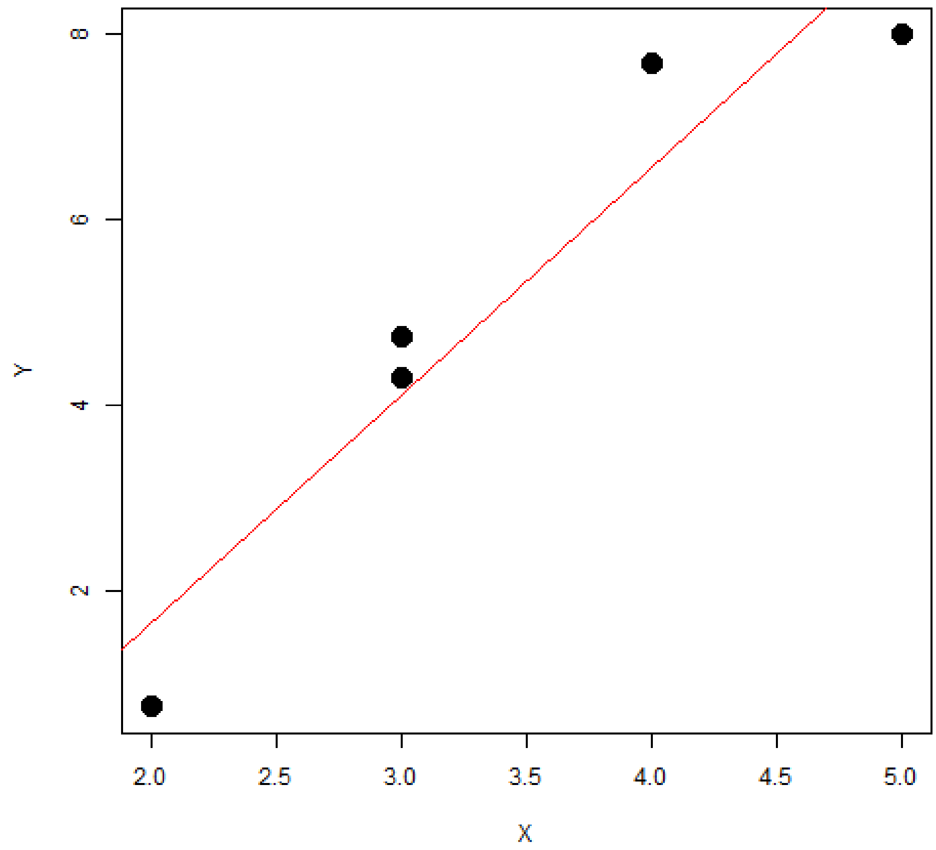
What is important here is that **these are all observable in the data, and you know how to calculate them**. You know how to calculate $\bar{x}$ and $\bar{y}$, you know how to sum things, and you know $x_i$ and $y_i$ in the data.

**As long as your assumptions hold**, you have an estimate of the PRF.

Second, let's tackle the *estimate* of $\beta_0$.

- We know, from the board, that $\beta_0 = E[Y] - \beta_1 E[X]$
- We have a good, unbiased sample estimator for $E[Y]$: $\bar{y}$.
- And we have a good, unbiased sample estimator for $E[X]$: $\bar{x}$
  - Plugging in: $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$

We don't observe $\beta_0$, but we can estimate it by taking sample mean of $y$ and $x$ (if we knew $\hat{\beta}_1$.




The red line is the *sample regression function*, or *SRF*.

A couple important terms:

- The **fitted value**, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$
- The **residual**, $\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

And note that:

- $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i$
  - The $\hat{u}_i$ "trues up" the fitted value.

Note that the residual is not the same as the error term.

- The residual is an empirical estimate from the sample
- The error term, $u_i$, is different

That error term is doing a lot of work here - it is covering *everything* that isn't $x_1$. So, what's inside the error term?

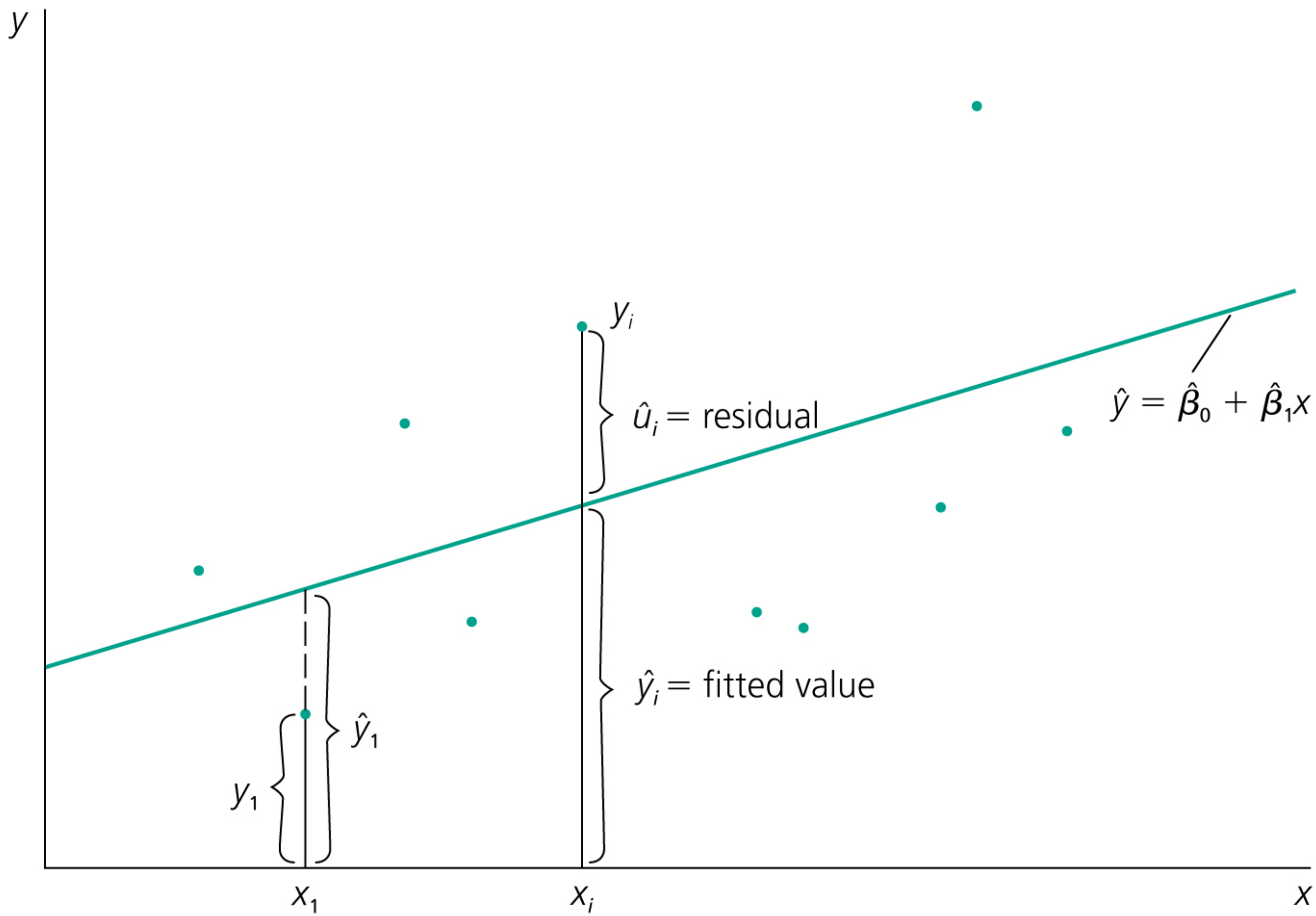
In $u_i$

- Omitted variables
  - There might be another covariate, $x_2$, that is missing.
- Measurement error
  - That $x$ might not be correctly measured.
- Non-linearities
  - Maybe there are some non-linear effects included in there.

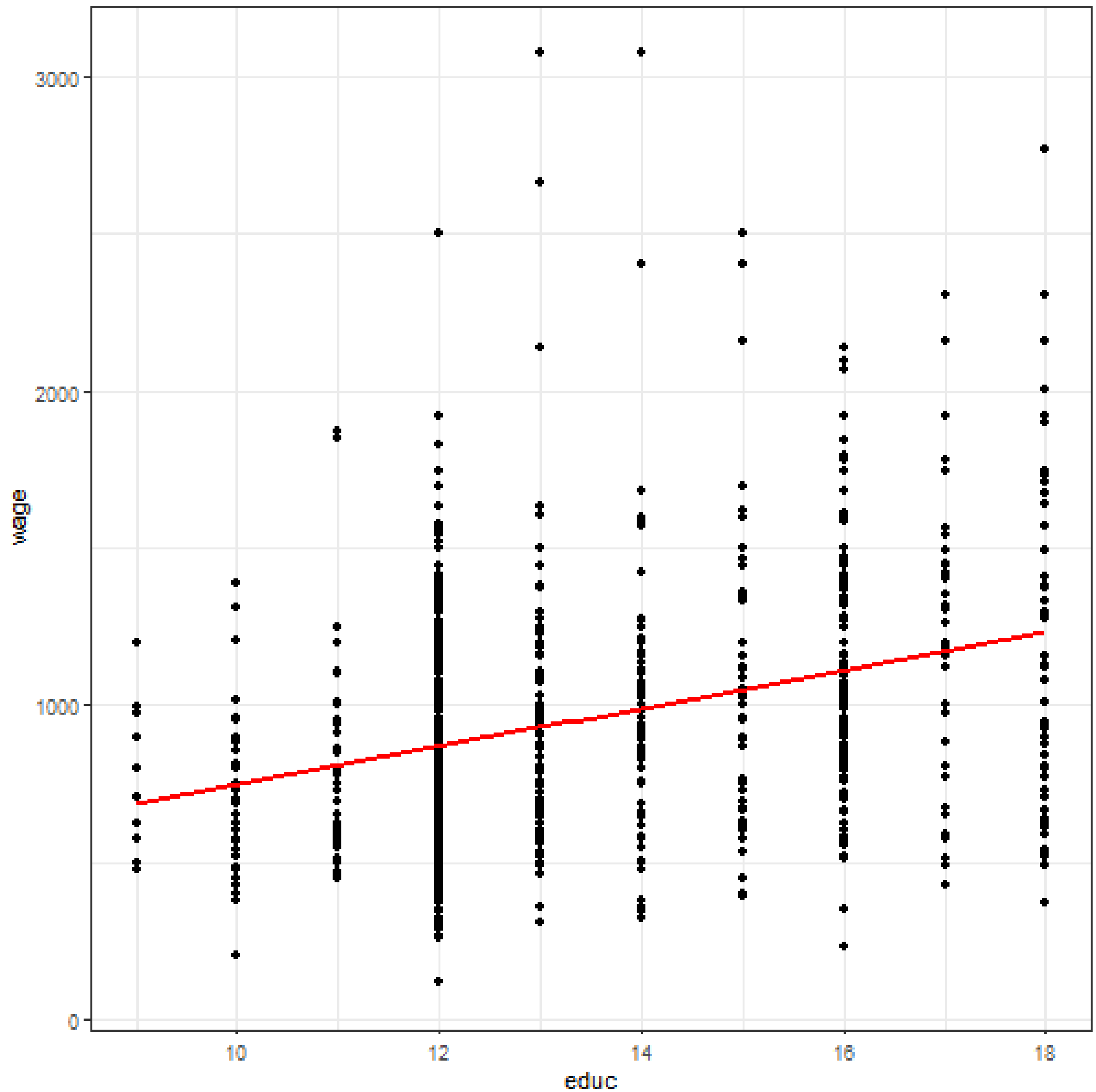
These are all in $u_i$.

$$y_i = \beta_0 + \beta_1 x_1 + \underbrace{\beta_2 x_{omitted} + \beta_3(x_1^* - x_1) + f(nonlinears) + \tilde{u}_i}_{\text{other things, u}}$$

Our estimator, $\hat{\beta}$ assumes allllll these things are 0 in expectation, no matter the value of $x$




Wooldridge Fig. 2.4




Regression line for wage2 data

It will always be the case that, for any estimates $\beta$ from a sample:

- $\sum_{i=1}^{N}(\hat{u}_i) = 0$
- $\sum_{i=1}^{N}(x_i\hat{u}_i) = 0$
- The point $(\bar{y}, \bar{x})$ is always on the regression line

The "squares" part refers to the squaring of the error term.

The "least" part refers to a minimzation of the (squared) error term.

Let's define the **sum of squared residuals** as:

$$SSR = \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

And $\beta$ is the "Least Squares" estimate if it minimizes $SSR$. How?

Take the derivative and set it equal to zero:

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = 2 \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

and

$$\frac{\partial SSR}{\partial \hat{\beta}_1} = 2 \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0$$

## Terminology

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

$y$ is called

- The dependent variable (DV)
- The "left hand side" (LHS)
- The outcome variable
- The response variable
- The target variable (in ML)

$u$ is called

- The residual (when $\hat{u}$)
- The error term (when $u$)

$x$ is called

- The independent variable
- The "right hand side" (RHS)
- The explanatory variable
- The control variable
- A covariate or a regressor

# Goodness of fit measures

top

## SSR, SSE, and SST

We know that $\beta_{OLS}$ minimizes the sum of squares. How do we measure how good of a fit we get?

We previously defined $SSR = \sum_{i=1}^{N} \hat{u}_i^2$

Define two more in addition to $SSR$:

- Sum of Squares Total: $SST = \sum_{i=1}^{N}(y_i - \bar{y})^2$
  - $SST$ is a total sum of squares (notice no hats).
- Sum of Squares Explained: $SSE = \sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2$
  - $SSE$ can be thought of as how much is *explained* by $\hat{y}_i$, *relative to just guessing the obvious:* $\bar{y}$

$$SST = SSR + SSE$$

The total variance is the sum of the variance of the residuals ("what isn't explained by your model") and the $SSE$ ("the variance that is explained").

This is a *decomposition* of variance. It is important.

## If your model fits well...

Then $SSR = \sum_{i=1}^{N} \hat{u}_i^2$ is very small

- Because if your model fits well, then $\hat{y}_i - y_i = \hat{u}_i$ is very small over all $i$

Then $SSE = \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2$ should be large

- Because if your mdoel fits well, then you are explaining a lot of the observed deviations from $\bar{y}$

## The $R^2$

$R^2$ ('r-square') is the comparison of $SSE$ to $SST$. Since $SSE < SST$ always, and both are always positive, $0 < R^2 \leq 1$

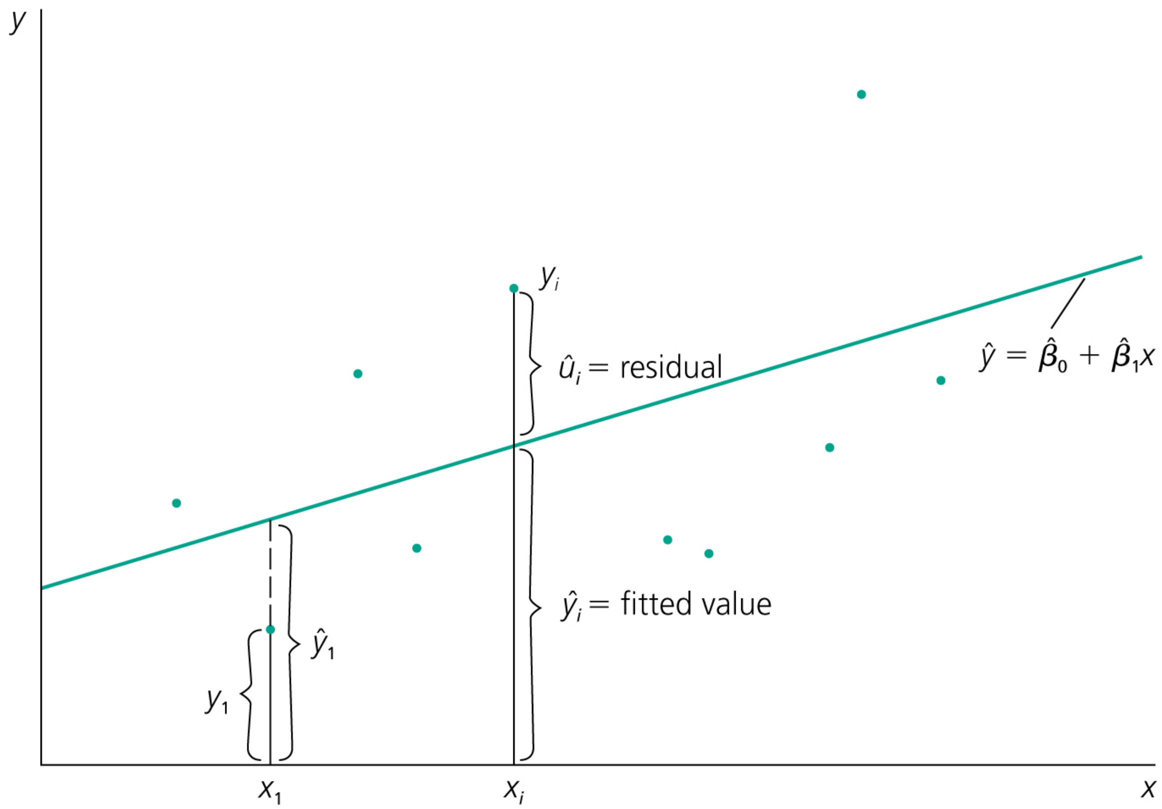
$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

## Interpretation of $R^2$

The $R^2$ is interpreted as the "fraction of variance in $y$ explained by the model"

- Your regression, the SRF, is a model
- The variance being explained is the variance in the outcome, $y$.
- If it is 0, then $SSE$ is zero. If $SSE$ is zero, then $y_i = \bar{y}$ for all $i$.
    - And in that case, your model isn't explaining *any* of the variance in $y$.

From earlier:



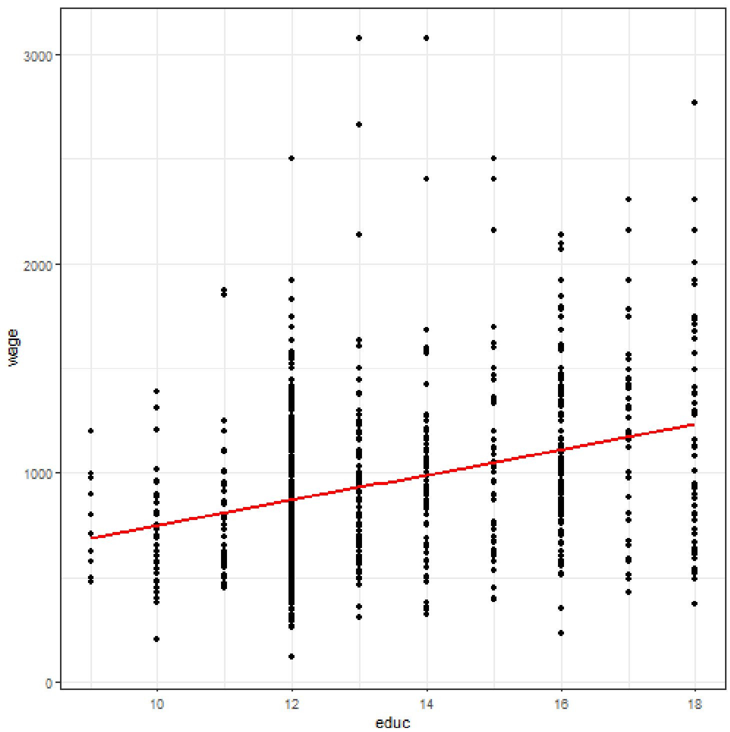

Wooldridge Fig. 2.4

# Interpretation of coefficients

top

Last time, we discussed a single variable regression from Wooldridge `wage2` where $Y$ is $wage$ and $X$ is $educ$:

$$wage = \beta_0 + \beta_1 educ + u$$




This resulted in a $\hat{\beta}_1 = 60.21$. How do we interpret this coefficient?

## Let's start with our simple linear regression model:

where $wage$ and $educ$ are random variables

$$wage = \beta_0 + \beta_1 educ + u$$

Our PRF is:

$$E[wage|educ] = \beta_0 + \beta_1 educ$$

> "One additional year of education is associated with a $\beta_1 = 60.21$ increase in *expected* monthly earnings, **all else held equal**"

- Why "expected"? We are estimating the PRF, so we are looking for the relationship between *expected* monthly earnings and education.
- Why "all else held equal"? Because we have assumed that $E[U|X] = \mathbf{0}$, so our estimate tells us how $E[Y]$ changes as $X$ *and not* $U$ changes.]
  - $U$ is held at zero, no matter the $X$

## Ceteris Paribus

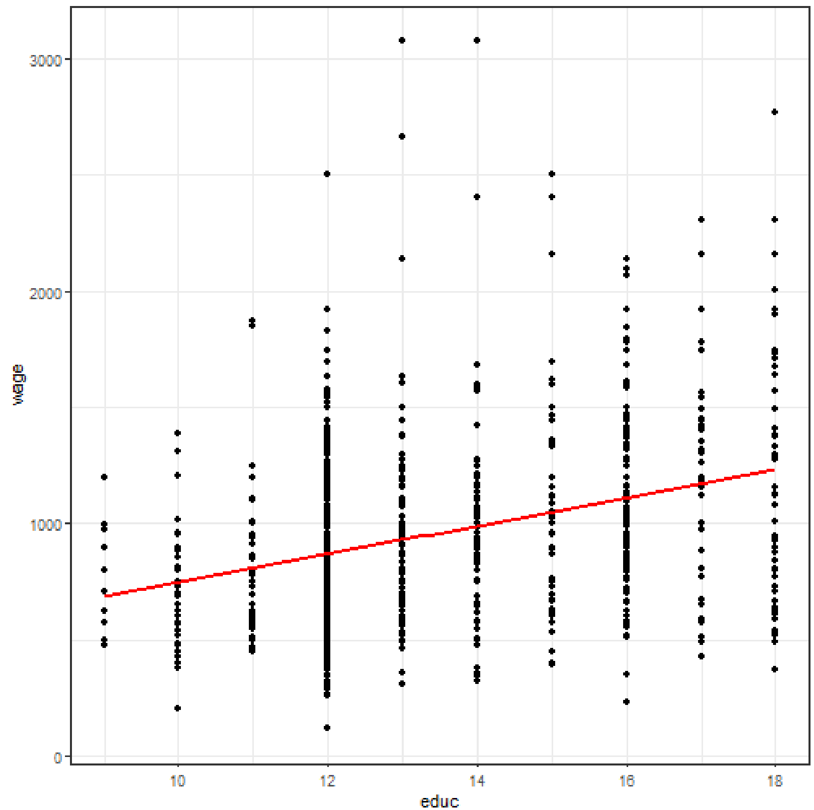
is Latin for "all else held equal"

---

So the interpretation of $\hat{\beta}_1$ is:

> "The (estimated) increase in the expectation of $wage$ associated with a 1-unit increase in $educ$, ceteris paribus"

The "all else held equal" part is very important.




- $\hat{\beta}_1$ is $\frac{\Delta wage}{\Delta educ}$
- $\hat{\beta}_1$ is the slope of the line
    - The line is $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, the $SRF$

## Regression in R

```
wage2 = wooldridge::wage2
myRegression = lm(wage ~ educ, data=wage2)
summary(myRegression)
```

```
##
## Call:
## lm(formula = wage ~ educ, data = wage2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -877.38 -268.63  -38.38  207.05 2148.26
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  146.952     77.715   1.891   0.0589 .
## educ          60.214      5.695  10.573   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 382.3 on 933 degrees of freedom
## Multiple R-squared:  0.107,    Adjusted R-squared:  0.106
## F-statistic: 111.8 on 1 and 933 DF,  p-value: < 2.2e-16
```

We see the $\hat{\beta}_0$ labeled "intercept", and $\hat{\beta}_1$, the "coefficient on *educ*" is labeled with the variable name, *educ*.

More on Regression in R later

# Rescaling Y and X

top

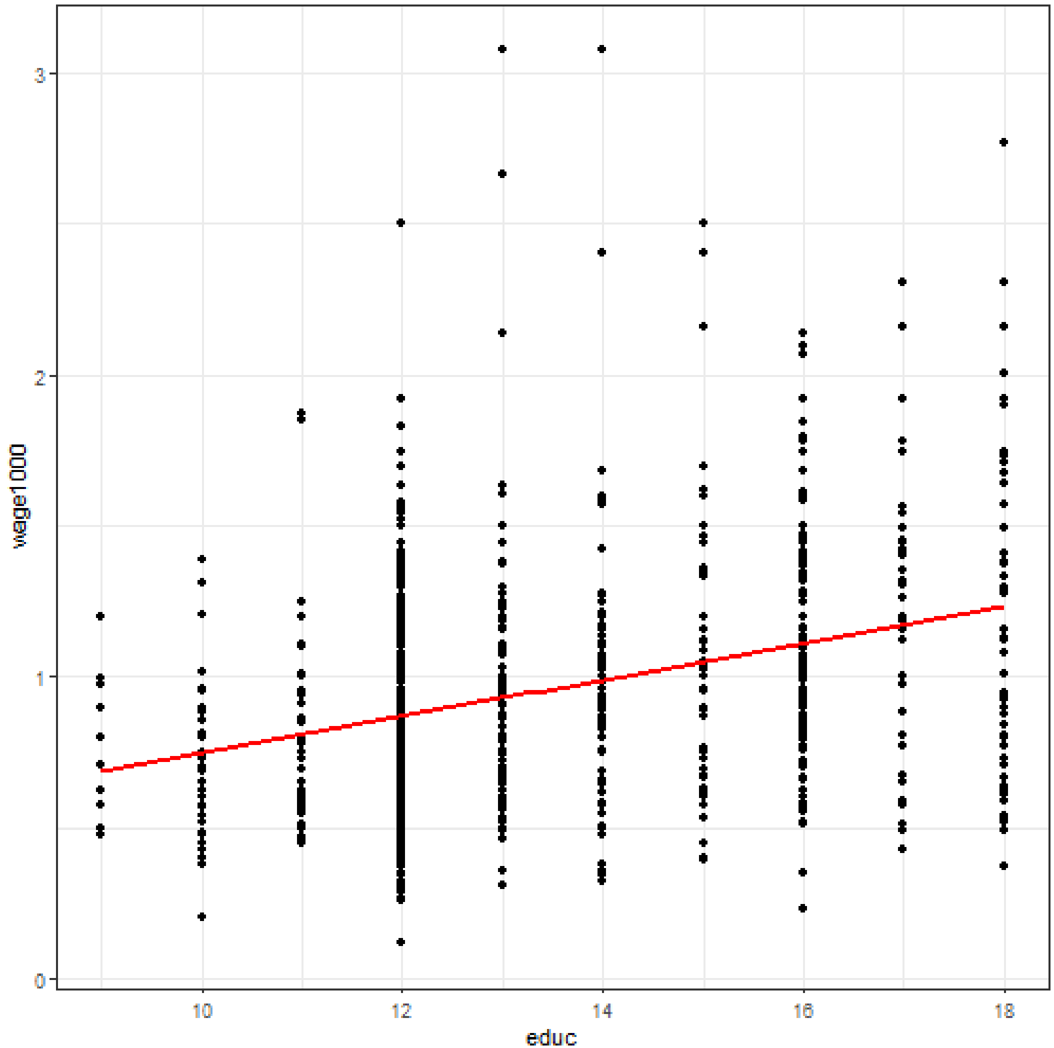
## What happens if we re-scale the dependent variable, wage?

Maybe we have $wage$ in dollars, but want it in thousands of dollars

## We hope that it still gives us the same relationship

Define $wage1000 = .001 \times wage$

- Any ideas what will happen to our coefficient?




Looks pretty similar, right? But the y-axis scale is very different.

MICHIGAN STATE UNIVERSITY

A regression of:

$$wage1000 = \beta_0 + \beta_1 educ + u$$

```
##
## Call:
## lm(formula = wage1000 ~ educ, data = wage2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.87738 -0.26863 -0.03838  0.20705  2.14826
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.146952   0.077715   1.891   0.0589 .
## educ        0.060214   0.005695  10.573   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3823 on 933 degrees of freedom
## Multiple R-squared:  0.107,	Adjusted R-squared:  0.106
## F-statistic: 111.8 on 1 and 933 DF,  p-value: < 2.2e-16
```

$\hat{\beta}_1 = 0.06$ when we use $wage1000$

$\hat{\beta}_1 = 60.21$ when we use $wage$.

Re-scaling the dependent variable, *wage*, results in an equal rescaling of the coefficient.

The relationship predicted by the *SRF* stays the same.

## Now, let's re-scale the *independent* variable

- That's the "right hand side" variable, $educ$.
- Let's do education in months: $educMonths = educ \times 12$
- Any predictions on what will result?

```
##
## Call:
## lm(formula = wage ~ educMonths, data = wage2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -877.38 -268.63  -38.38  207.05 2148.26
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 146.9524    77.7150   1.891   0.0589 .
## educMonths    5.0179     0.4746  10.573   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 382.3 on 933 degrees of freedom
## Multiple R-squared:  0.107,    Adjusted R-squared:  0.106
## F-statistic: 111.8 on 1 and 933 DF,  p-value: < 2.2e-16
```

What was the result?

Re-scaling the independent variable simply rescales the coefficient by the *inverse* amount:

- $12 \times educ \Rightarrow \hat{\beta}_1^{new} = \frac{\hat{\beta}_1}{12}$

Re-scaling the dependent variable simply rescales the coefficient on it by an equal amount:

- $\hat{\beta}_1^{new} = \hat{\beta}_1 \times .001$

The relationship always remains the same

Let's take a look at the $R^2$ of the original regression:

```
##
## Call:
## lm(formula = wage ~ educ, data = wage2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -877.38 -268.63  -38.38  207.05 2148.26
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  146.952     77.715   1.891   0.0589 .
## educ          60.214      5.695  10.573   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 382.3 on 933 degrees of freedom
## Multiple R-squared:  0.107,    Adjusted R-squared:  0.106
## F-statistic: 111.8 on 1 and 933 DF,  p-value: < 2.2e-16
```

Now, the re-scaled dependent variable:

```
##
## Call:
## lm(formula = wage1000 ~ educ, data = wage2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.87738 -0.26863 -0.03838  0.20705  2.14826
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.146952   0.077715   1.891   0.0589 .
## educ        0.060214   0.005695  10.573   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3823 on 933 degrees of freedom
## Multiple R-squared:  0.107,    Adjusted R-squared:  0.106
## F-statistic: 111.8 on 1 and 933 DF,  p-value: < 2.2e-16
```

And the re-scaled independent variable:

```
##
## Call:
## lm(formula = wage ~ educMonths, data = wage2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -877.38 -268.63  -38.38  207.05 2148.26
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 146.9524    77.7150   1.891   0.0589 .
## educMonths    5.0179     0.4746  10.573   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 382.3 on 933 degrees of freedom
## Multiple R-squared:  0.107,    Adjusted R-squared:  0.106
## F-statistic: 111.8 on 1 and 933 DF,  p-value: < 2.2e-16
```

Heck, let's rescale both and look at the $R^2$

```
##
## Call:
## lm(formula = wage1000 ~ educMonths, data = wage2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.87738 -0.26863 -0.03838  0.20705  2.14826
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.1469524  0.0777150   1.891   0.0589 .
## educMonths  0.0050179  0.0004746  10.573   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3823 on 933 degrees of freedom
## Multiple R-squared:  0.107,    Adjusted R-squared:  0.106
## F-statistic: 111.8 on 1 and 933 DF,  p-value: < 2.2e-16
```

The $R^2$ is the same in every single one!

The "fraction of variance explained by the model" does not change.

Intuitively, you shouldn't be able to explain more variance simply by re-scaling a variable. The relationship that holds for wages and years of education must hold for 12 x years of education as well.

Since rescaling linearly doesn't matter, we can use a scale that is easiest to interpret and to read.

- $wage1000$ in thousands of dollars is a lot easier to look at than the larger number we get using $wage$.
- You often don't want to have very extreme numbers of decimal places (e.g. a coefficient of .00000051 will be a lot easier to talk about if it's in millions: 5.1)

Now that we've seen an example, can we derive this result from the definition of $\beta_1$ ?

$$\beta_1 = \frac{Cov(X,Y)}{Var(X)}$$

and if we rescale $X$ by $a$:

$$\begin{aligned}
\beta_1^{rescaled} &= \frac{Cov(aX,Y)}{Var(aX)} \\
&= \frac{aCov(X,Y)}{a^2Var(X)} \quad \text{(by rules of Cov and Var)} \\
&= \frac{Cov(X,Y)}{aVar(X)} \quad \text{(canceling)} \\
&= \frac{1}{a}\beta_1
\end{aligned}$$

# Non-linear Functional Forms

top

## What do we mean by "non-linear" function?

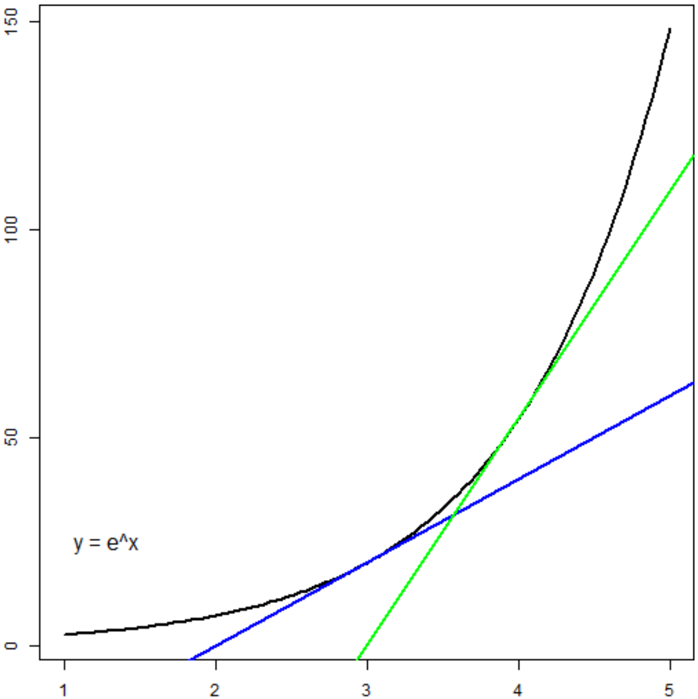
**A function** here is any mathematical operation or transformation that takes an input (usually called $x$ ) and returns an output (usually called $y$ ).

A non-linear function is any function where the graph is not a straight line.

- "Affine transformation" is the technical term for $y = ax + b$.
- "Non-affine transformation" is non-linear

Another way of thinking about non-linear functions is that $\frac{\Delta y}{\Delta x}$ depends on the value of $x$

- The slope of the graph changes as $x$ changes.
- The slope at $x_1$ (blue) is different than the slope at $x_2$ (green)



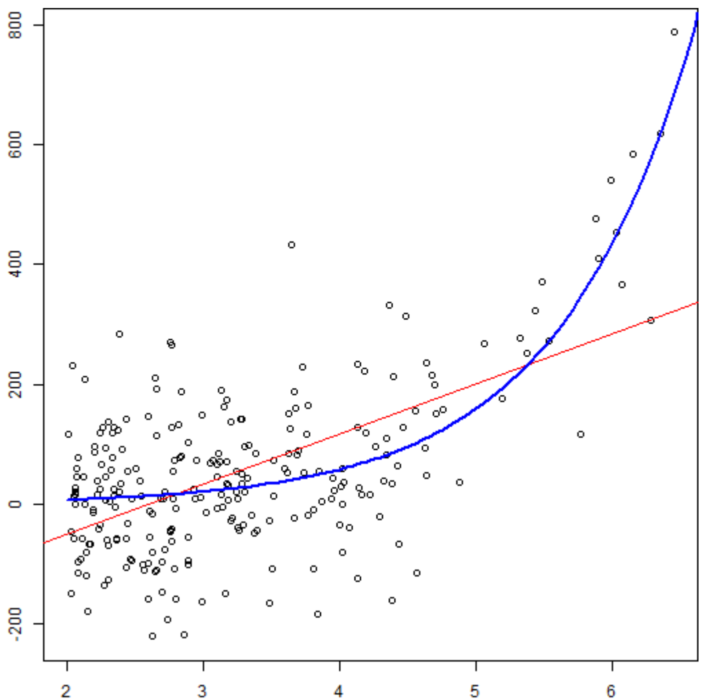
In the previous slide, we saw a non-linear function, the exponential function, $e^x$. If we wanted a model to use in a regression that includes an exponential function, we could use:

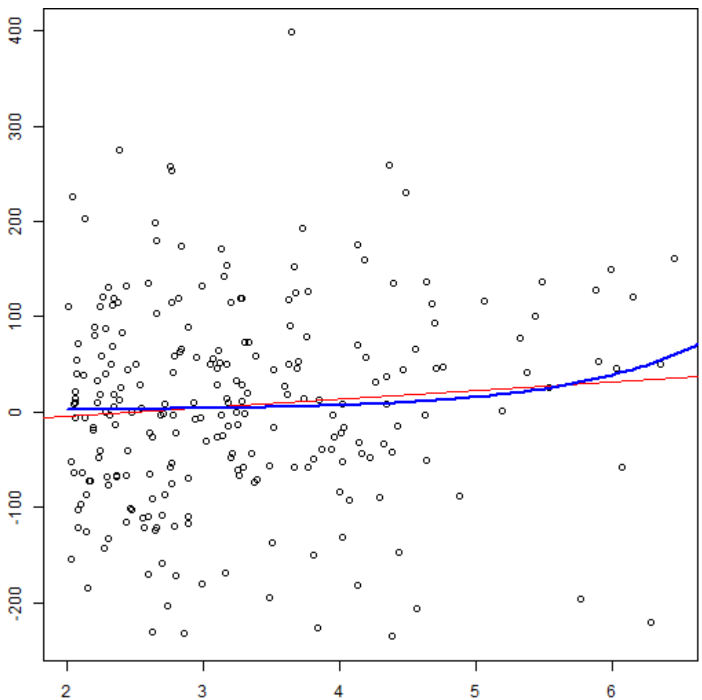
$$y_i = \beta_0 + \beta_1 e^{x_i} + u_i$$

Note that the value of $x_i$ is exponentiated.

- So this model has a non-linear term.
- It lets $y$ respond to changes in $x$ more flexibly

- but imposes that relationship whether it is appropriate (left) or not (right).



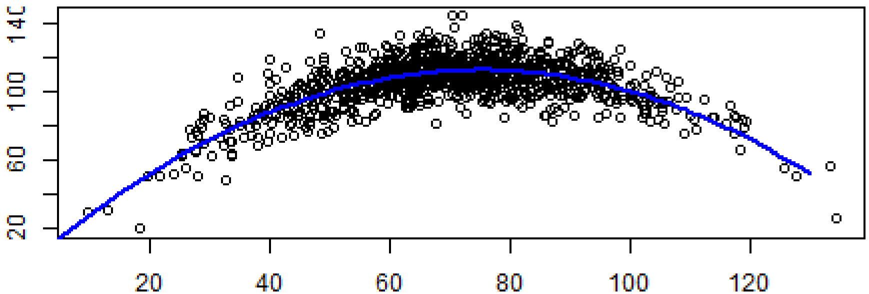
The most common non-linear transformation is the **polynomial**

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + u$$

For instance, plant growth rates over temperatures may be quadratic

- The *marginal effect* of an increase in temperature will be big and positive at lower temperatures.
- The *marginal effect* of an increase in temperature will be negative at very high temperatures.
- And somewhere in the middle, the *marginal effect* will be around zero.

The *marginal effect* is saying "the change in $y$ per change in $x$", or $\frac{dy}{dx}$.



If we have a polynomial relationship:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

Then we can obtain the slope, $\frac{dy}{dx}$ as the derivative of the relationship:

$$\frac{\partial y}{\partial x} = \beta_1 + 2\beta_2 x$$

If we propose a "higher order polynomial" relationship like:

$$y = \beta_0 + \beta 1 x + \beta_2 x^2 + \beta_3 x^3$$

Then we get a more complicated function for the slope at any $x$:

$$\frac{\partial y}{\partial x} = \beta_1 + 2\beta_2 x + 3\beta_3 x^2$$

There are other possible non-linear forms: $\sqrt{x}$, the natural log, $log_{10}$, the inverse hyperbolic sine...

## Even though these specifications are non-linear transformations, the regression is still **linear-in-parameters**

That is, all of the transformations we have discussed are still in the category of "linear models" because they are linear in the parameters.

So, our $PRF$ (population regression function) is still linear, even with one of these transformations.

The quadratic specification, $y = \beta_0 + \beta_1 x + \beta_2 x^2$ is particularly useful anytime you have an effect of $x$ on $y$ that dissipates or declines with increasing values of $x$.

Quick question: if the *effect* of $x$ on $y$ **declines** as $x$ increases, then is the slope *increasing* or *decreasing* as $x$ gets larger?

## An example:

In many cases, the effect of household income on some behavior may change as income increases.

- A low-income person may spend more on food when income increases
- But a high-income person may not spend much more on food when their income increases
    - But of course, the high-income will spend more on food than the low-income person.

We see these declining effects in many economic situations, but we also see increasing effects.

- Installing solar panels
- Others?

The quadratic "specification" can capture these phenomon.

## The natural log, $ln(x)$

The natural log is the most common transformation. It is particularly useful because of the following:

$$ln(1+x) \approx x \quad \textbf{when} \quad x \approx 0$$

Let's say $x^1 = x^0 + \Delta x$.

$$ln(x^1) - ln(x^0) = ln\left(\frac{x^1}{x^0}\right) = ln\left(\frac{x^0 + \Delta x}{x^0}\right) = ln\left(1 + \frac{\Delta x}{x^0}\right) \approx \frac{\Delta x}{x^0}$$

- This is the percent change in $x$: $\frac{\Delta x}{x}$
- $100 \times [ln(x^1) - ln(x^0)] \approx \%\Delta x$

## The natural log, $ln(x)$

Recall the formula for *elasticity*: $\frac{\%\Delta y}{\%\Delta x} = \frac{\Delta y}{\Delta x} \times \frac{x}{y}$

And recall that, in a linear model ( $y = \beta_0 + \beta_1 x$ ), this elasticity is **not** constant:

$$\frac{\Delta y}{\Delta x} \times \frac{x}{y} = \beta_1 \times \frac{x}{y} = \beta_1 \times \frac{x}{\beta_0 + \beta_1 x + u}$$

But, when a model takes the form: $ln(y) = \beta_0 + \beta_1 ln(x)$

$$\frac{\%\Delta y}{\%\Delta x} \approx \frac{ln(y^1) - ln(y^0)}{ln(x^1) - ln(x^0)} = \frac{\beta_1[ln(x^1) - ln(x^0)]}{ln(x^1) - ln(x^0)} = \beta_1$$

## The coefficient on a log-log model is the elasticity

$ln(y) = \beta_0 + \beta_1 ln(x)$ results in $\beta_1$ being the elasticity of y, or "percent change in y from a 1 percent change in x".

Econometrics is frequently about estimating that elasticity.

# Regression in R

top

## First, data

You should have already installed `wooldridge`. If not, type `install.packages('wooldridge')` directly in your console. Then, we can use R's built-in "data" function to load `wage2`

```
library(wooldridge)
wage2 = wooldridge::wage2 # creates a wage2 object
print(wage2[1:5,1:9]) # first 5 rows; first 9 columns
```

```
##   wage hours  IQ KWW educ exper tenure age married
## 1  769    40  93  35   12    11      2  31       1
## 2  808    50 119  41   18    11     16  37       1
## 3  825    40 108  46   14    11      9  33       1
## 4  650    40  96  32   12    13      7  32       1
## 5  562    40  74  27   11    14      5  34       1
```

## Second, run the regression

We will use the `lm()` function. You will provide the regression formula and the name of the data to use.

The formula will be of the form *y* ~ *x*. You'll specify the data with `data = wage2`

```
MyRegression = lm(wage ~ educ, data=wage2)
print(MyRegression)
```

```
##
## Call:
## lm(formula = wage ~ educ, data = wage2)
##
## Coefficients:
## (Intercept)         educ
##      146.95        60.21
```

## Finally, we want a little more detail.

`MyRegression` is an R object. We can ask R to summarize it, and R will know to give us information about the regression:

```
summary(MyRegression)
```

```
##
## Call:
## lm(formula = wage ~ educ, data = wage2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -877.38 -268.63  -38.38  207.05 2148.26
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  146.952     77.715   1.891   0.0589 .
## educ          60.214      5.695  10.573   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 382.3 on 933 degrees of freedom
## Multiple R-squared:  0.107,	Adjusted R-squared:  0.106
## F-statistic: 111.8 on 1 and 933 DF,  p-value: < 2.2e-16
```

# Inference and hypothesis testing:

Expectation of $\hat{\beta}_1$

top

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

- For example, $X$ with population mean $\mu_X$

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

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

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 $\bar{X}$:

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

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 $\sigma_X^2$

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

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?

So what if we want to test something about $\beta_1$?

## We can think of $\beta_1$ 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?

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

- $\beta$ 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 $\hat{\beta}$.

$\beta$ 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

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)

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 $\beta$'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 $\beta_1$.

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

$\beta_1$ 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 $\beta_1$, $\hat{\beta}_1$ is the true $\beta$ plus some term.

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

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 $x$.

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 $\beta_1$

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

- " $\hat{\beta}_0$ is an unbiased estimator of $\beta_0$ "

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

# Inference and hypothesis testing:

Variance of $\hat{\beta}_1$

top

## 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 $\sigma^2$.

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 $\sigma^2$.
- We do **not** need this assumption to show that $\hat{\beta}$ is an unbiased estimator for $\beta$
    - But we do need this assumption to calculate the variance of $\hat{\beta}$.
- It does not mean that we know $\sigma_u^2$. We don't

Start with where we left off on $\beta_1$:

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

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 $\beta_1$) 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.

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:

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

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 $\sigma_u^2$.
- Variance of the estimator is decreasing in $SST_x$, variation in $X$.

## Good, but we don't know $\sigma_u^2$, 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*, $\beta$. The residual shows up in the *estimated* equation with $\hat{\beta}$.

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

We can use $\sum_{i=1}^{N} \hat{u}_i^2$ as an estimator for $\sigma_u^2$ 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 $\hat{\beta}$:

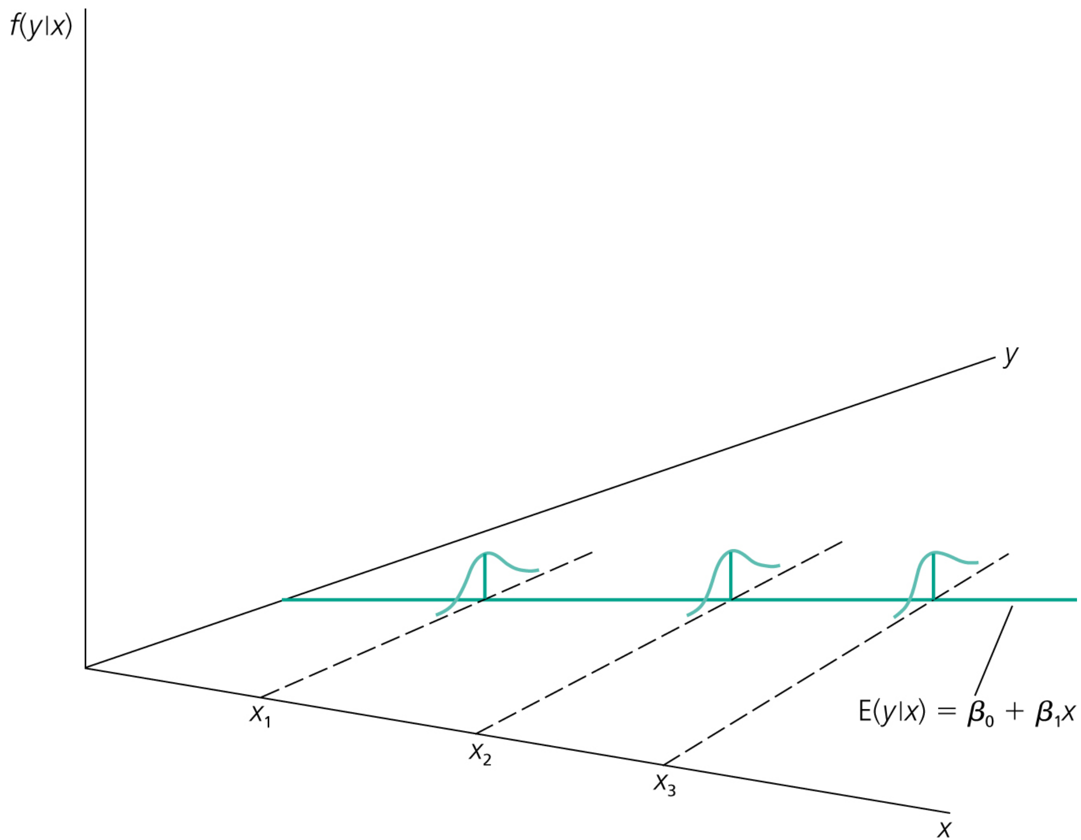
$$\hat{\beta}_1 \sim N(\beta_1, \frac{\hat{\sigma}_u^2}{SST_x})$$

If we want to test a hypothesis about $\beta_1$, we now can.

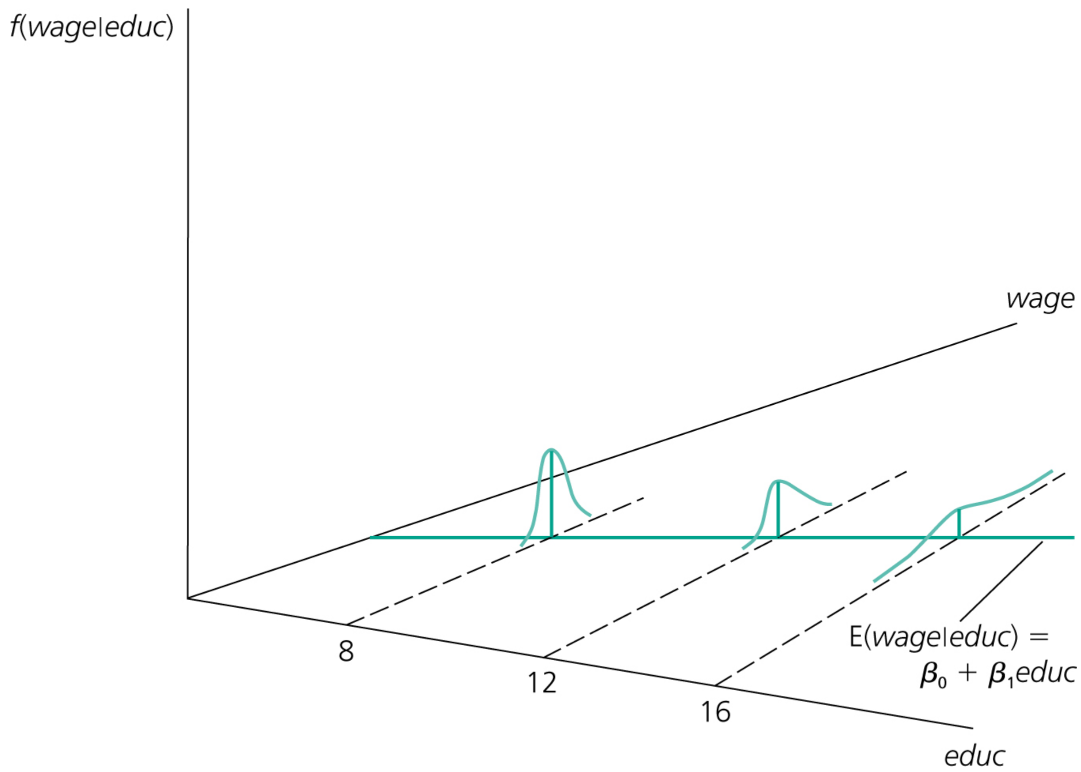
But only **if** we assume homoskedasticity - that $Var(u|x) = Var(u) = \sigma_u^2$.

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.




Homoskedasticity (from Wooldridge)




Heteroskedasticity (from Wooldridge)

# Single variable inference: an example

top

# An example



Let's work through an example with real data. Our goal is to take the data, calculate $\beta_1$, $\hat{se}(\hat{\beta}_1)$, and test a hypothesis $H_0 : \beta_1 = 0$.

# An example



| ID | Outcome | Dose |
|---|---|---|
| 1 | 4.9 | 3 |
| 2 | 19.7 | 6 |
| 3 | 9.7 | 3 |
| 4 | 15.6 | 6 |
| 5 | 18.5 | 7 |

| Statistic | Value |
|---|---|
| $\bar{y}$ | 13.68 |
| $\bar{x}$ | 5 |
| $SST_y = \sum(y_i - \bar{y})^2$ | 156.088 |
| $SST_x = \sum(x_i - \bar{x})^2$ | 14 |
| $\sum(y_i - \bar{y})(x_i - \bar{x})$ | 43.1 |

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

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

| ID | Outcome | Dose | Fitted | Residual |
|---|---|---|---|---|
| 1 | 4.9 | 3 | | |
| 2 | 19.7 | 6 | | |
| 3 | 9.7 | 3 | | |
| 4 | 15.6 | 6 | | |
| 5 | 18.5 | 7 | | |

- Calculate $\hat{y}$ using $\beta_0$ and $\beta_1$
- Calculate $\hat{u}$ using $y_i - \hat{y}$
- Calculate $\hat{\sigma}_u^2$
  - Remember to divide by $(n-2)$ here 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 $\hat{\beta}_0$?

Using $\hat{\beta}_1$ and the distribution of $\hat{\beta}_1$, what is the t-statistic under the null: $H_0 : \beta_1 = 0$:

$$\hat{t} = \frac{\hat{\beta}_1 - 0}{\hat{se}(\hat{\beta}_1)}$$

- Our $t$ is normally distributed*, so we can check the p-value using the back of the Wooldridge book
  - Or, our "rule of thumb" of $|t| > 1.96$

* Since s^2 is estimated, this follows a t-distribution. However, for large enough N, the t distribution and the standard normal are very similar.

Check your work here:

```
##
## Call:
## lm(formula = Outcome ~ Dose, data = df)
##
## Residuals:
##      1      2      3      4      5
## -2.623  2.941  2.177 -1.159 -1.337
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.7129     3.9357  -0.435   0.6928
## Dose          3.0786     0.7464   4.124   0.0258 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.793 on 3 degrees of freedom
## Multiple R-squared:  0.8501,    Adjusted R-squared:  0.8001
## F-statistic: 17.01 on 1 and 3 DF,  p-value: 0.02584
```