Gov 2000 - 7. Simple Linear Regression

Matthew Blackwell

October 27, 2015

Where are we? Where are we going?

- Last week:
  - Using the CEF to explore relationships
  - Bias-variance trade-off led us to linear regression.
- This week:
  - Inference for OLS: sampling distribution.
  - Is there really a relationship? Hypothesis tests
  - Can we get a range of plausible slope values? Confidence intervals
  - $\rightsquigarrow$ how to read regression output.

More narrow goal

```r
ajr <- foreign::read.dta("ajr.dta")
summary(lm(logpgp95 ~ logem4, data = ajr))
```

```
##
## Call:
## lm(formula = logpgp95 ~ logem4, data = ajr)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.7130 -0.5333  0.0195  0.4719  1.4467
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.6602     0.3053   34.92  < 2e-16 ***
## logem4       -0.5641     0.0639   -8.83  2.1e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.756 on 79 degrees of freedom
##   (82 observations deleted due to missingness)
## Multiple R-squared:  0.497,  Adjusted R-squared:  0.49
## F-statistic:   78 on 1 and 79 DF,  p-value: 2.09e-13
```

1/ Setup of the OLS estimator

Simple linear regression model

- We are going to assume a linear model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Data:
  - Dependent variable: Y_i
  - Independent variable: X_i
- Population parameters:
  - Population intercept: β_0
  - Population slope: β_1
- Error/disturbance: u_i
  - Represents all unobserved error factors influencing Y_i other than X_i.

What is OLS?

- Ordinary least squares (OLS) is an estimator for the slope and the intercept of the regression line.
- Where does it come from? Minimizing the sum of the squared residuals:

$$(\widehat{\beta}_0, \widehat{\beta}_1) = \underset{b_0, b_1}{\arg\min} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

- Leads to:

$$\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}$$

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_i - \overline{X})^2}$$

Intuition of the OLS estimator

- Regression line goes through the sample means $(\overline{Y}, \overline{X})$:

$$\overline{Y} = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{X}$$

- Slope is the ratio of the covariance to the variance of X_i:

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\widehat{\text{Cov}}(X_i, Y_i)}{\widehat{\mathbb{V}}[X_i]} \\ &= \frac{\text{Sample Covariance between } X \text{ and } Y}{\text{Sample Variance of } X}\end{aligned}$$

The sample linear regression function

- The estimated or sample regression function is:

$$\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_i$$

- Estimated intercept: $\widehat{\beta}_0$
- Estimated slope: $\widehat{\beta}_1$
- Predicted/fitted values: $\widehat{Y}_i$
- Residuals: $\hat{u}_i = Y_i - \widehat{Y}_i$
- You can think of the residuals as the prediction errors of our estimates.

Mechanical properties of OLS

- Some properties are mechanical since they can be derived from the first order conditions of OLS.

1. The residuals will be 0 on average:

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0$$

2. The residuals will be uncorrelated with the predictor:

$$\widehat{\text{Cov}}(X_i, \hat{u}_i) = 0$$

3. The residuals will be uncorrelated with the fitted values:

$$\widehat{\text{Cov}}(\widehat{Y}_i, \hat{u}_i) = 0$$

OLS slope as a weighted sum of the outcomes

- One useful derivation that we'll do moving forward is to write the OLS estimator for the slope as a weighted sum of the outcomes.

$$\widehat{\beta}_1 = \sum_{i=1}^{n} W_i Y_i$$

- Where here we have the weights, W_i as:

$$W_i = \frac{(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

- Estimation error: proof

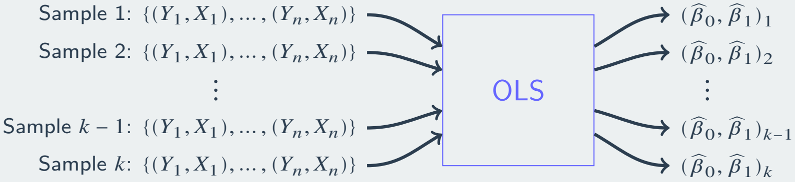
$$\widehat{\beta}_1 - \beta_1 = \sum_{i=1}^{n} W_i u_i$$

- $\rightsquigarrow$ $\widehat{\beta}_1$ is a sum of random variables.

2/ Sampling distribution of the OLS estimator

Sampling distribution of the OLS estimator

- Remember: OLS is an estimator—it's a machine that we plug data into and we get out estimates.



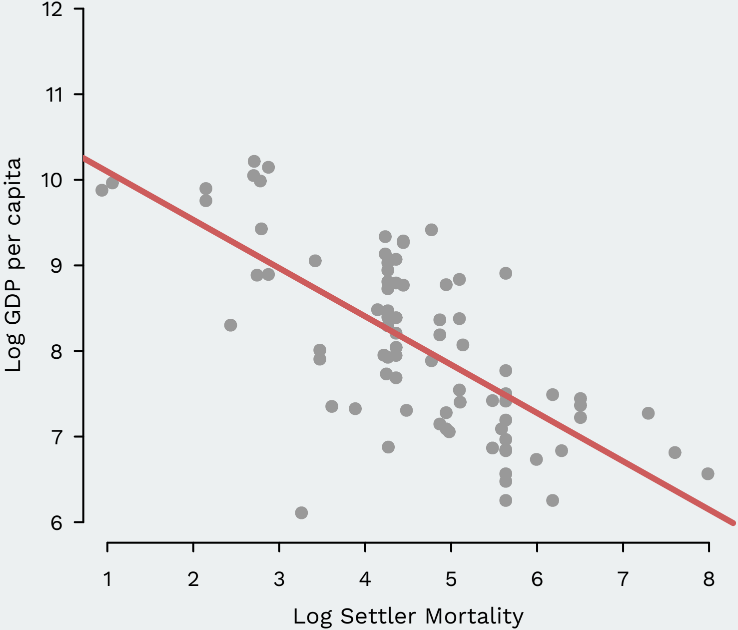
- Just like the sample mean, sample difference in means, or the sample variance
- It has a sampling distribution, with a sampling variance/standard error, etc.

Simulation procedure

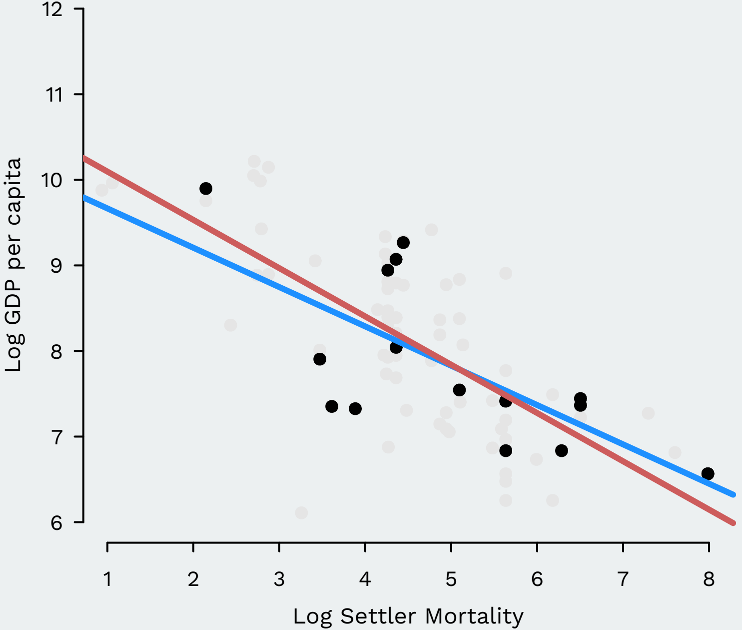
- Let's take a simulation approach to demonstrate:
  - Pretend that the AJR data represents the population of interest
  - See how the line varies from sample to sample

1. Draw a random sample of size $n = 30$ with replacement using `sample()`
2. Use `lm()` to calculate the OLS estimates of the slope and intercept
3. Plot the estimated regression line

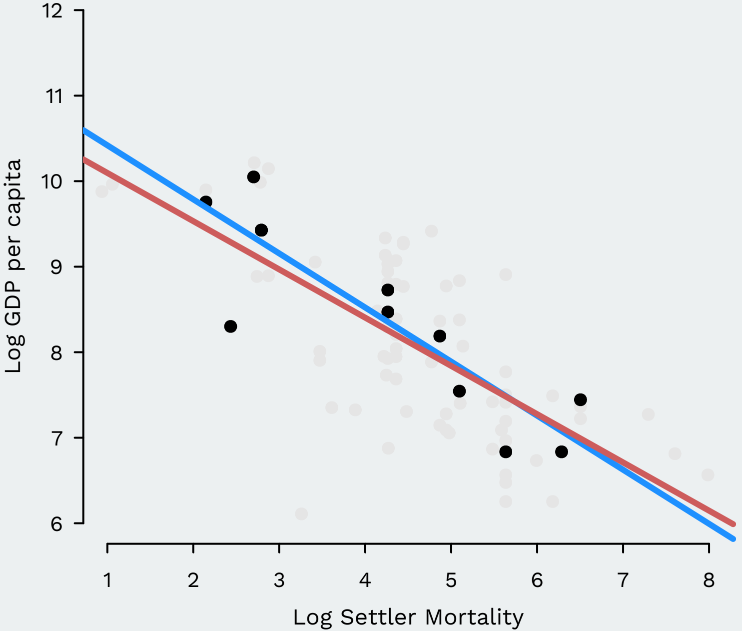
Population Regression



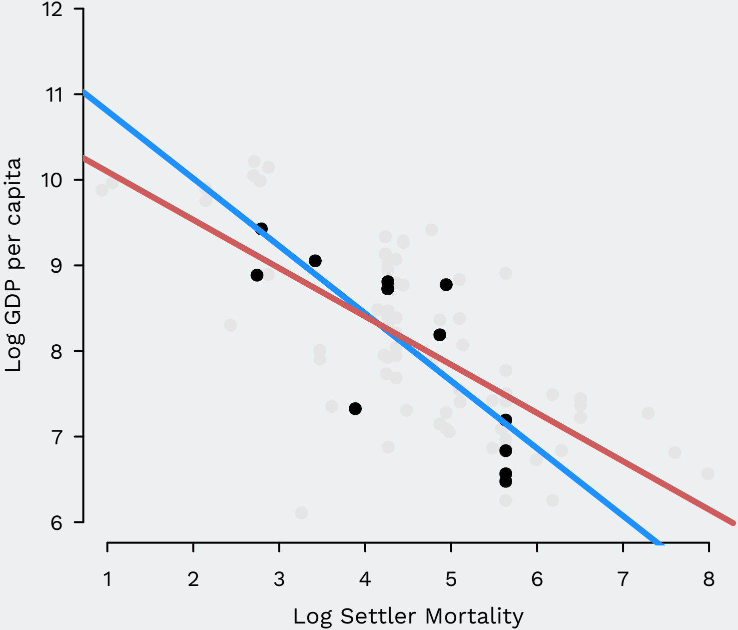
Randomly sample from AJR



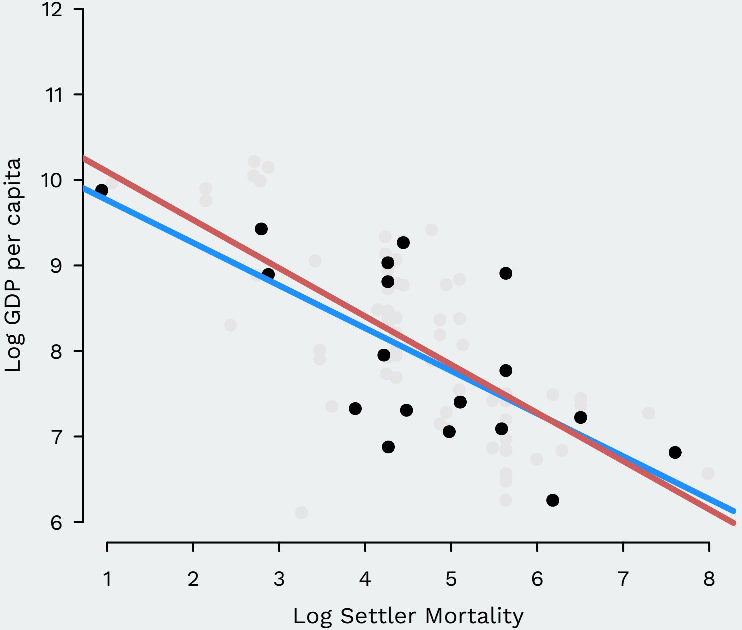
Randomly sample from AJR



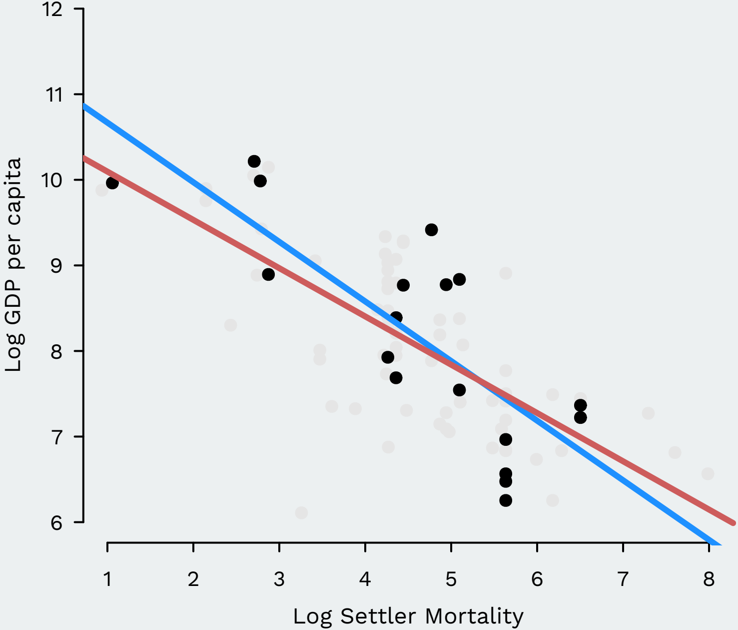
Randomly sample from AJR



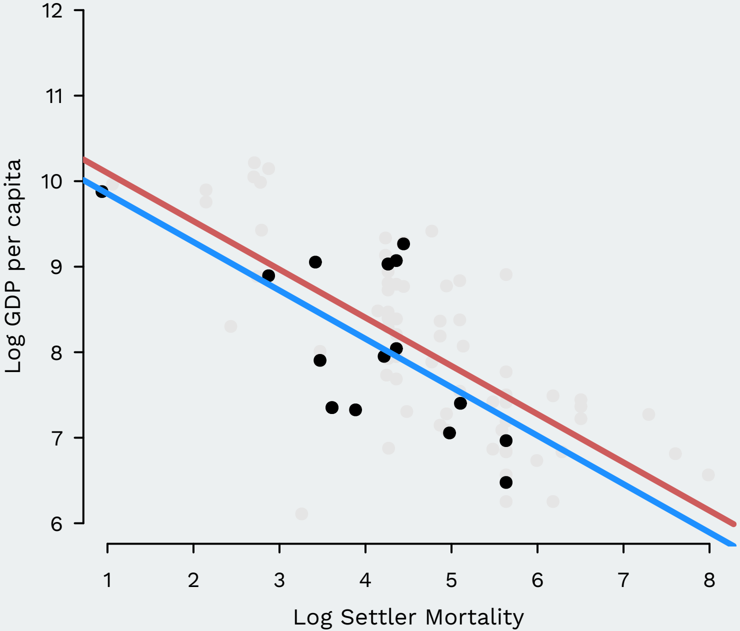
Randomly sample from AJR



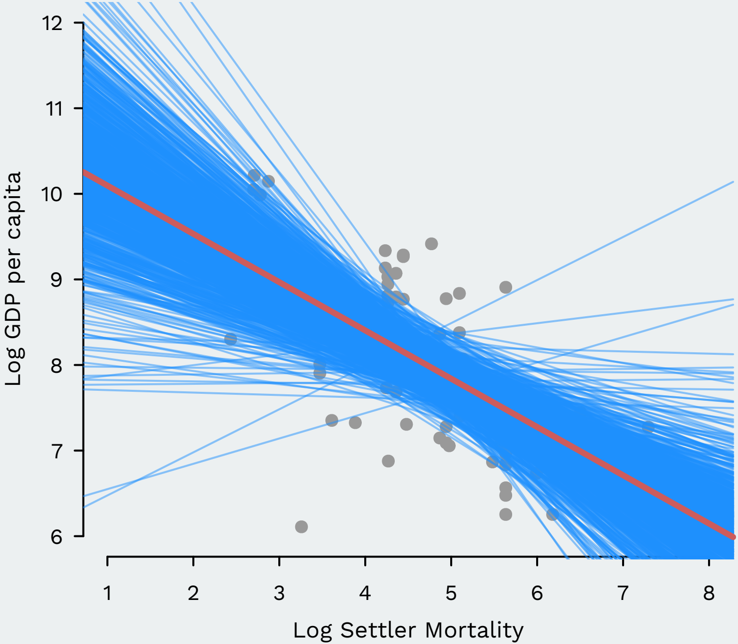
Randomly sample from AJR



Randomly sample from AJR

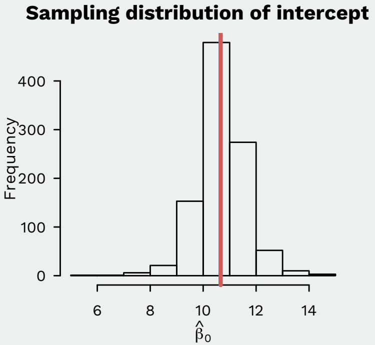


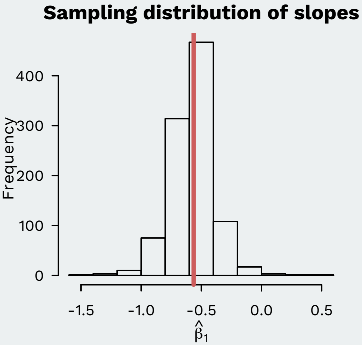
Randomly sample from AJR



Sampling distribution of OLS

- You can see that the estimated slopes and intercepts vary from sample to sample, but that the "average" of the lines looks about right.





Sample mean properties review

- Last couple of weeks we derived the properties of $\overline{X}_n$ under one assumption: i.i.d. random samples.
- In large samples, we derived the sampling distribution:

$$\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Unbiasedness: $\mathbb{E}[\overline{X}_n] = \mu$
- Sampling variance: σ^2/n
- Standard error: $\sigma/\sqrt{n}$
- $\rightsquigarrow$ allows us to do hypothesis tests, calculate confidence intervals.

Our goal

- What is the sampling distribution of the OLS slope?

$$\widehat{\beta}_1 \sim ?(?, ?)$$

- Mean of the sampling distribution: ??
- Sampling variance: ??
- Standard error: ??
- Distribution: ??

Basic assumptions of OLS

- In order to get the mean of the sampling distribution $\mathbb{E}[\widehat{\beta}_1]$, we need to make some assumptions:

1. Linearity
2. Random (iid) sample
3. Variation in X_i
4. Zero conditional mean of the errors

Linearity

Assumption 1: Linearity

The population regression function is linear in the parameters:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Violation of the linearity assumption:

$$Y_i = \frac{1}{\beta_0 + \beta_1 X_i} + u_i$$

- Not a violation of the linearity assumption:

$$Y_i = \beta_0 + \beta_1 X_i^2 + u_i$$

Random sample

Assumption 2: Random Sample

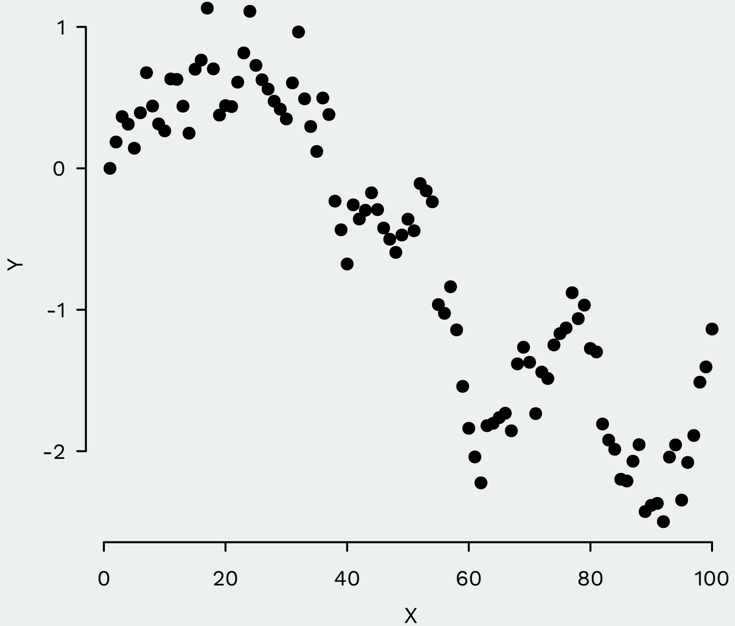
We have a iid random sample of size n, $\{(Y_i, X_i) : i = 1, 2, \ldots, n\}$ from the population regression model above.

- Violation of the random sample assumption: time-series, selected samples.
- Think about the weight example from last week, where Y_i was my weight on a given day and X_i was my number of active minutes the day before:

$$\text{weight}_i = \beta_0 + \beta_1 \text{activity}_i + u_i$$

- What if I only weighed myself on the weekdays?

A non-iid sample



Variation in X

Assumption 3: Variation in X

The in-sample independent variables, $\{X_i : i = 1, \dots, n\}$, are not all the same value.

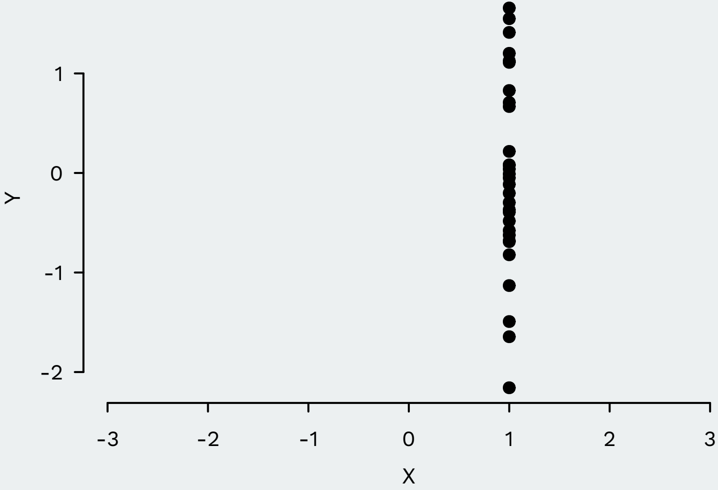
- Also remember the formula for the OLS slope estimator:

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

- What happens here when X_i doesn't vary?

Stuck in a moment

- Why does this matter? How would you draw the line of best fit through this scatterplot, which is a violation of this assumption?



Zero conditional mean

Assumption 4: Zero conditional mean of the errors

The error, u_i, has expected value of 0 given any value of the independent variable:

$$\mathbb{E}[u_i | X_i = x] = 0 \quad \forall x.$$

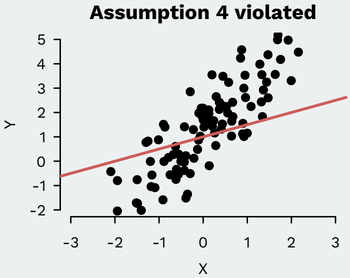
- $\rightsquigarrow$ u_i and X_i uncorrelated: $\text{Cov}[u_i, X_i] = \mathbb{E}[u_i X_i] = 0$
- $\rightsquigarrow$ $\mathbb{E}[Y_i | X_i] = \beta_0 + \beta_1 X_i$ is the CEF

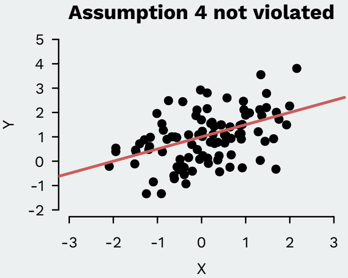
Violating the zero conditional mean assumption

- How does this assumption get violated? Let's generate data from the following model:

$$Y_i = 1 + 0.5X_i + u_i$$

- But let's compare two situations:
    1. Where the mean of u_i depends on X_i (they are correlated)
    2. No relationship between them (satisfies the assumption)





More examples of zero conditional mean in the error

- Think about the weight example from last week, where Y_i was my weight on a given day and X_i was my number of active minutes the day before:

$$\text{weight}_i = \beta_0 + \beta_1 \text{activity}_i + u_i$$

- What might in u_i here? Amount of food eaten, workload, etc etc.
- We have to assume that all of these factors have the same mean, no matter what my level of activity was. Plausible?
- When is this assumption most plausible? When X_i is randomly assigned.

Unbiasedness

- Used 4 assumptions:

1. Linearity: $Y_i = \beta_0 + \beta_1 X_i + u_i$
2. Random (iid) sample
3. Variation in X_i
4. Zero conditional mean of the errors: $\mathbb{E}[u_i | X_i = x] = 0$

- Letting $X = (X_1, \ldots, X_n)$

Unbiasedness of OLS

Under assumptions 1-4, the OLS estimator is unbiased:

$$\mathbb{E}[\widehat{\beta}_1] = \beta_1$$

Unbiasedness proof

- Remember the estimation error:

$$\widehat{\beta}_1 - \beta_1 = \sum_{i=1}^{n} W_i u_i$$

- $W_i = (X_i - \overline{X})/(\sum_{i=1}(X_i - \overline{X})^2)$.
- Use this to prove conditional unbiasedness:

$$\begin{aligned}\mathbb{E}[\widehat{\beta}_1 - \beta_1|X] &= \mathbb{E}\left[\sum_{i=1}^{n} W_i u_i \middle| X\right] = \sum_{i=1}^{n} \mathbb{E}[W_i u_i|X] \\ &= \sum_{i=1}^{n} W_i \mathbb{E}[u_i|X] \\ &= \sum_{i=1}^{n} W_i \times 0 = 0\end{aligned}$$

- Unconditionally unbiased: $\mathbb{E}[\widehat{\beta}_1] = \mathbb{E}[\mathbb{E}[\widehat{\beta}_1|X]] = \mathbb{E}[\beta_1] = \beta_1$
- Law of Large Numbers $\rightsquigarrow$ $\widehat{\beta}_1$ consistent

3/ Sampling variance of OLS

Where are we?

- Now we know that, under Assumptions 1-4, we know that

$$\widehat{\beta}_1 \sim ?(\beta_1, ?)$$

- That is we know that the sampling distribution is centered on the true population slope, but we don't know the population sampling variance.

$$\mathbb{V}[\widehat{\beta}_1] = ??$$

Sampling variance of estimated slope

- In order to derive the sampling variance of the OLS estimator,

1. Linearity
2. Random (iid) sample
3. Variation in X_i
4. Zero conditional mean of the errors
5. Homoskedasticity

Homoskedasticity

Assumption 5

The conditional variance of Y_i given X_i is constant:

$$\mathbb{V}(Y_i|X_i = x) = \mathbb{V}(u_i|X_i = x) = \sigma_u^2.$$

- $\mathbb{V}[Y_i|X_i = x]$ sometimes called the skedastic function, thus the name homoskedasticity.
- Under homoskedasticity proof:

$$\mathbb{V}[\widehat{\beta}_1|X] = \frac{\sigma_u^2}{\sum_{i=1}^n (X_i - \overline{X})^2}$$

- Standard error:

$$SE[\widehat{\beta}_1|X] = \sqrt{\mathbb{V}[\widehat{\beta}_1|X]} = \frac{\sigma_u}{\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2}}$$

- Violations: magnitude of u_i differ at different levels of X_i.

Derive the sampling variance

$$\mathbb{V}[\widehat{\beta}_1|X] = \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

- What drives the sampling variability of the OLS estimator?
  - The higher the variance of Y_i, the higher the sampling variance
  - The lower the variance of X_i, the higher the sampling variance
  - As we increase n, the denominator gets large, while the numerator is fixed and so the sampling variance shrinks to 0.

Estimating the sampling variance/standard error

- But we don't observe σ_u^2—it is the variance of the errors.
- Estimate with the residuals:

$$\widehat{\sigma}_u^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2$$

- Why $n-2$ instead of n or $n-1$? To correct for OLS slightly underestimating the variance.
  - We already used the data twice to estimate $\widehat{\beta}_0$ and $\widehat{\beta}_1$
- Estimated standard error of the OLS slope:

$$\widehat{SE}[\widehat{\beta}_1|X] = \frac{\sqrt{\widehat{\sigma}_u^2}}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}} = \frac{\widehat{\sigma}_u}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}}$$

Where are we?

- Under Assumptions 1-5, we know that

$$\widehat{\beta}_1 \sim ?\left(\beta_1, \frac{\sigma_u^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\right)$$

- Now we know the mean and sampling variance of the sampling distribution.
- How does this compare to other estimators for the population slope?

OLS is BLUE :(

Gauss-Markov Theorem

Under assumptions 1-5, the OLS estimator is BLUE, or the Best Linear Unbiased Estimator, where by "best" we mean it lowest sampling variance.

- Assumptions 1-5: the "Gauss Markov Assumptions"
- The proof is very detailed, so we'll skip it. See Wooldridge, Appendix 3A.6 for details.
- Fails to hold when the assumptions are violated!

Where are we?

- Under Assumptions 1-5, we know that

$$\widehat{\beta}_1 \sim ?\left(\beta_1, \frac{\sigma_u^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\right)$$

- And we know that $\frac{\sigma_u^2}{\sum_{i=1}^n (X_i - \overline{X})^2}$ is the lowest variance of any linear estimator of β_1
- What about the last question mark? What's the form of the distribution? Uniform? t? Normal? Exponential? Hypergeometric?

Large-sample distribution of OLS estimators

- OLS estimator is the sum of independent r.v.'s:

$$\widehat{\beta}_1 = \sum_{i=1}^{n} W_i Y_i$$

- Weighted sum of r.v.s $\rightsquigarrow$ central limit theorem:

$$\widehat{\beta}_1 \overset{d}{\rightarrow} N\left(\beta_1, \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right)$$

- True here as well, so we know that in large samples:

$$\frac{\widehat{\beta}_1 - \beta_1}{SE[\widehat{\beta}_1]} \sim N(0, 1)$$

- Can also replace SE with an estimate:

$$\frac{\widehat{\beta}_1 - \beta_1}{\widehat{SE}[\widehat{\beta}_1]} \sim N(0, 1)$$

Where are we?

Under Assumptions 1-5 and in large samples, we know that

$$\widehat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right)$$

Sampling distribution in small samples

- What if we have a small sample? What can we do then? Back here:

$$\widehat{\beta}_1 \sim ?\left(\beta_1, \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\right)$$

- Can't get something for nothing, but we can make progress if we make another assumption:

1. Linearity
2. Random (iid) sample
3. Variation in X_i
4. Zero conditional mean of the errors
5. Homoskedasticity
6. Errors are conditionally normal

Normal errors

Assumption 6: Conditionally Normal Errors

The conditional distribution of u_i given X_i is normal with mean 0 and variance σ_u^2.

- This implies that the distribution of Y_i given X_i is: $N(\beta_0 + \beta_1 X_i, \sigma_u^2)$.

Sampling distribution of OLS slope

- If we have Y_i given X_i is distributed $N(\beta_0 + \beta_1 X_i, \sigma_u^2)$, then we have the following at any sample size:

$$\frac{\widehat{\beta}_1 - \beta_1}{SE[\widehat{\beta}_1]} \sim N(0, 1)$$

- Furthermore, if we replace the true standard error with the estimated standard error, then we get the following:

$$\frac{\widehat{\beta}_1 - \beta_1}{\widehat{SE}[\widehat{\beta}_1]} \sim t_{n-2}$$

- The standardized coefficient follows a t distribution $n - 2$ degrees of freedom. We take off an extra degree of freedom because we had to one more parameter than just the sample mean.
- All of this depends on normal errors! We can check to see if the residuals do look normal.

Where are we?

- Under Assumptions 1-5 and in large samples, we know that

$$\frac{\widehat{\beta}_1 - \beta_1}{\widehat{SE}[\widehat{\beta}_1]} \sim N(0, 1)$$

- Under Assumptions 1-6 and in any sample, we know that

$$\frac{\widehat{\beta}_1 - \beta_1}{\widehat{SE}[\widehat{\beta}_1]} \sim t_{n-2}$$

4/ Hypothesis tests and confidence intervals

Null and alternative hypotheses review

- Null: $H_0 : \beta_1 = 0$
  - The null is the straw man we want to knock down.
  - With regression, almost always null of no relationship
- Alternative: $H_a : \beta_1 \neq 0$
  - Claim we want to test
  - Almost always "some effect"
  - Could do one-sided test, but you shouldn't, for reasons we've already discussed
- Notice these are statements about the population parameters, not the OLS estimates.

Test statistic

- Under the null of $H_0 : \beta_1 = b$, we can use the following familiar test statistic:

$$T = \frac{\widehat{\beta}_1 - b}{\widehat{SE}[\widehat{\beta}_1]}$$

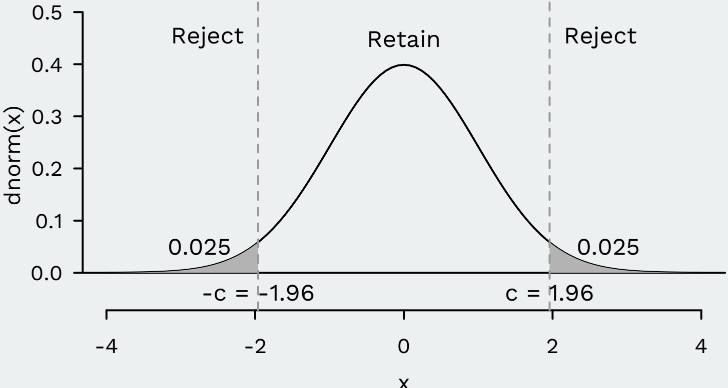
- Under then null hypothesis:
    - Large samples: $T \sim N(0, 1)$.
    - Small samples, plus conditionally normal errors: $T \sim t_{n-2}$
    - Safe to use t_{n-2} in either case since $t_{n-2} \rightsquigarrow N(0, 1)$
- Thus, under the null, we know the distribution of T and can use that to formulate a critical value and calculate p-values.

Critical values

- Choose a level of the test, α, and find the critical value:

$$\mathbb{P}(|T| > c) = \alpha \iff \mathbb{P}(-c < T < c) = 1 - \alpha$$

- This is exactly the same as with sample means.
- In large samples with an $\alpha = 0.05$, find the values so that we reject 5% of the time under the null:



p-value

- The interpretation of the p-value is the same: the probability of seeing a test statistic at least this extreme if the null hypothesis were true
- Mathematically:

$$\mathbb{P}\left(\left|\frac{\widehat{\beta}_1 - b}{\widehat{SE}[\widehat{\beta}_1]}\right| \geq |T_{obs}|\right)$$

- If the p-value is less than α we would reject the null at the α level.

R output

- By default, R shows you the T_{obs} for the test statistic with the null of $\beta_1 = 0$, which is just the estimate divided by the standard error:

$$T_{obs} = \frac{\widehat{\beta}_1 - 0}{\widehat{SE}[\widehat{\beta}_1]} = \frac{\widehat{\beta}_1}{\widehat{SE}[\widehat{\beta}_1]}$$

- R also calculates the p-values for you.
- In the AJR data:

```
out <- lm(logpgp95 ~ logem4, data = ajr)
coef(summary(out))
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  10.6602    0.30528   34.92 8.759e-50
## logem4       -0.5641    0.06389   -8.83 2.094e-13
```

Confidence intervals

- Very similar to the approach with sample means. By the sampling distribution of the OLS estimator, we know that we can find t-values such that:

$$\mathbb{P}\left(-t_{\alpha/2,n-2} \leq \frac{\widehat{\beta}_1 - \beta_1}{\widehat{SE}[\widehat{\beta}_1]} \leq t_{\alpha/2,n-2} \right) = 1 - \alpha$$

- If we rearrange this as before, we can get an expression for confidence intervals:

$$\mathbb{P}\left(\widehat{\beta}_1 - t_{\alpha/2,n-2}\widehat{SE}[\widehat{\beta}_1] \leq \beta_1 \leq \widehat{\beta}_1 + t_{\alpha/2,n-2}\widehat{SE}[\widehat{\beta}_1] \right) = 1 - \alpha$$

- Thus, we can write the confidence intervals as:

$$\widehat{\beta}_1 \pm t_{\alpha/2,n-2}\widehat{SE}[\widehat{\beta}_1]$$

- "In 95% of repeated samples, the confidence interval for β_1 will cover the true value."

Confidence intervals in R

- Confidence intervals are not outputted by default, but you grab them for any regression using the `confint()` function:

```
confint(lm(logpgp95 ~ logem4, data = ajr))
```

```
##               2.5 % 97.5 %
## (Intercept) 10.0526 11.268
## logem4      -0.6913 -0.437
```

5/ Goodness of fit

Prediction error

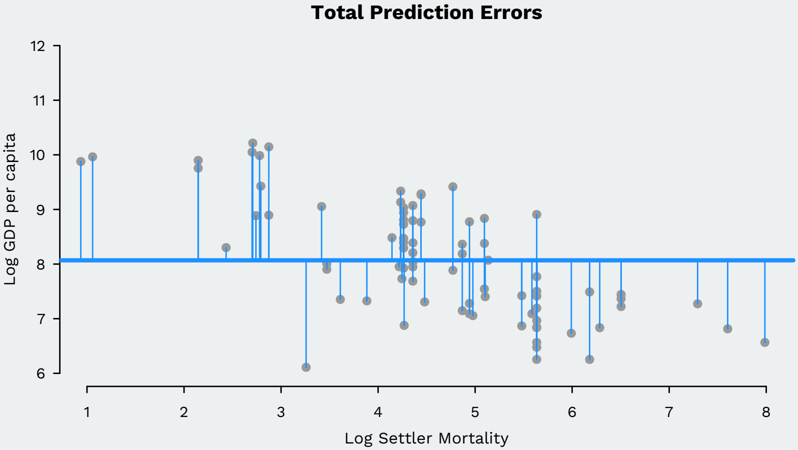
- How do we judge how well a line fits the data? Is there some way to judge?
- One way is to find out how much better we do at predicting Y_i once we include X_i into the regression model.
- Prediction errors without X_i: best prediction is the mean, so our squared errors, or the total sum of squares (SS_{tot}) would be:

$$SS_{tot} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

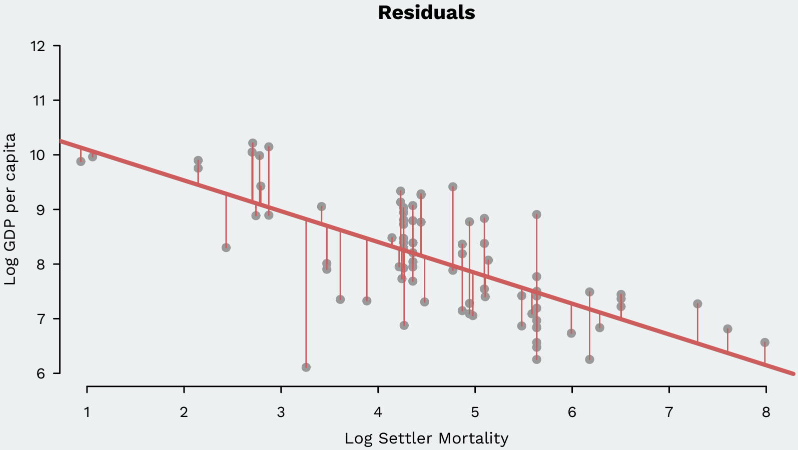
- Prediction errors with X_i: the sum of the squared residuals or SS_{res}:

$$SS_{res} = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2$$

Total SS vs SSR



Total SS vs SSR



R-square

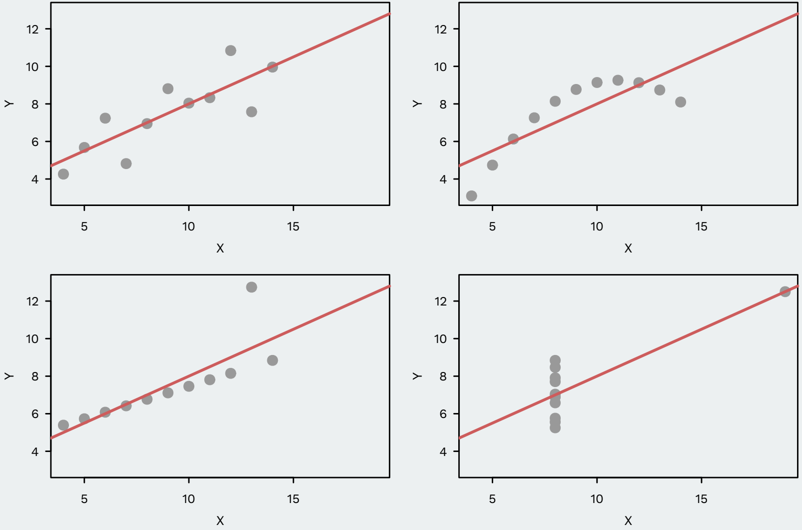
- By definition, the residuals have to be smaller than the deviations from the mean, so we might ask the following: how much lower is the SS_{res} compared to the SS_{tot}?
- We quantify this question with the coefficient of determination or R^2. This is the following:

$$R^2 = \frac{SS_{tot} - SS_{res}}{SS_{tot}} = 1 - \frac{SS_{res}}{SS_{tot}}$$

- This is the fraction of the total prediction error eliminated by providing information on X_i.
- Common interpretation: R^2 is the fraction of the variation in Y_i is "explained by" X_i.
  - $R^2 = 0$ means no relationship
  - $R^2 = 1$ implies perfect linear fit

Is R-squared useful?

- Can be very misleading. Each of these samples have the same R^2 even though they are vastly different:



Review of Assumptions

- What assumptions do we need to make what claims with OLS?
  1. Data description: variation in X_i
  2. Unbiasedness/Consistency: linearity, iid, variation in X_i, zero conditional mean error.
  3. Large-sample inference: linearity, iid, variation in X_i, zero conditional mean error, homoskedasticity.
  4. Small-sample inference: linearity, iid, variation in X_i, zero conditional mean error, homoskedasticity, Normal errors.
- Can we weaken these? In some cases, yes.
- Next week: adding another variable to regression.

Estimation error proof

Return

- Key facts:
  - $\sum_{i=1}^{n} W_i = 0$ because $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$
  - $\sum_{i=1}^{n} W_i X_i = 1$ because $\sum_{i=1}^{n} X_i(X_i - \overline{X}) = \sum_{i=1}^{n}(X_i - \overline{X})^2$
- Proof:

$$\begin{aligned}
\widehat{\beta}_1 &= \sum_{i=1}^{n} W_i Y_i \\
&= \sum_{i=1}^{n} W_i(\beta_0 + \beta_1 X_i + u_i) \\
&= \beta_0 \left(\sum_{i=1}^{n} W_i\right) + \beta_1 \left(\sum_{i=1}^{n} W_i X_i\right) + \sum_{i=1}^{n} W_i u_i \\
&= \beta_1 + \sum_{i=1}^{n} W_i u_i
\end{aligned}$$

Variance proof

Return

- Proof:

$$
\begin{aligned}
\mathbb{V}[\widehat{\beta}_1|X] &= \mathbb{V}\left[\sum_{i=1}^{n} W_i Y_i \middle| X\right] \\
&= \sum_{i=1}^{n} \mathbb{V}[W_i Y_i|X] \\
&= \sum_{i=1}^{n} W_i^2 \mathbb{V}[Y_i|X] \\
&= \sum_{i=1}^{n} W_i^2 \sigma_u^2 \\
&= \sigma_u^2 \sum_{i=1}^{n} W_i^2 \\
&= \sigma_u^2 \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{(\sum_{i=1}^{n}(X_i - \overline{X})^2)^2} = \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}
\end{aligned}
$$