

450A Political Methodology I

Lecture 4: Simple Regression

2. Statistical Properties of Ordinary Least Squares

Stanford University

Statistical Properties of Ordinary Least Squares

Linear model assumptions

Unbiasedness of OLS Estimators

Variance of OLS Estimators

Estimating Variance for Slope and Intercept

Gauss-Markov Theorem

OLS: Data Summary vs. Statistical Inference

- So far (Lecture 4-1), we have studied the mechanics of OLS: how the estimates are computed from one sample.
- That treats regression as a **data summary** of the sample at hand.
- However, our primary interest is to estimate $E[Y|X]$, the conditional mean/expectation function.
- That is, we want to see linear regression as a tool for **statistical inference**.
- We now therefore turn to:
 - Treating the sample as one draw from a population model, and asking how $(\hat{\beta}_0, \hat{\beta}_1)$ behave across repeated samples
 - A common set of assumptions for the linear regression model
 - Properties of the OLS estimator under those assumptions
 - Standard errors (testing, p-values, and confidence intervals follow in Lecture 4-3)

The **population** model is given by:

$$Y = \beta_0 + \beta_1 X + u$$

- Y : Dependent variable (outcome)
- X : Independent variable (regressor)
- β_0, β_1 : Intercept and slope
- u : **Error term** or **disturbance term**
— a random variable that captures all (unobserved) factors which influence Y other than the regressor X .

The **estimated** model:

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{u}$$

- $\hat{\beta}_0, \hat{\beta}_1$: Estimated intercept and slope
- $\hat{u}$: **Residuals**:
Variation in Y left unexplained by X . Can be viewed as an estimate of u .

Statistical Properties of Ordinary Least Squares

Linear model assumptions

Unbiasedness of OLS Estimators

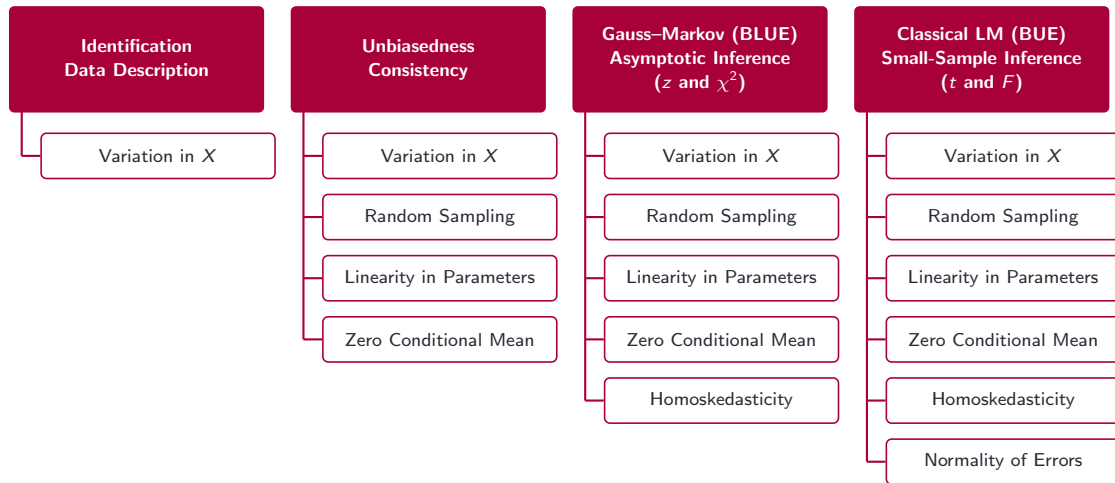
Variance of OLS Estimators

Estimating Variance for Slope and Intercept

Gauss-Markov Theorem

- 1 **Linearity in Parameters/Correct Functional Form:** The population model is linear in its parameters and correctly specified
- 2 **Random Sampling:** The observed data represent a random sample from the population described by the model.
- 3 **Variation in X :** There is variation in the explanatory variable.
- 4 **Zero conditional mean:** Expected value of the error term is zero conditional on all values of the explanatory variable
- 5 **Homoskedasticity:** The error term has the same variance conditional on all values of the explanatory variable.
- 6 **Normality:** The error term is independent of the explanatory variables and normally distributed.

Hierarchy of OLS Assumptions



Assumption (I. Linearity in Parameters/Correct Functional Form)

The population regression model is linear in its parameters and correctly specified as:

$$Y = \beta_0 + \beta_1 X + u$$

- Note that it can be nonlinear *in variables*
 - OK: $Y = \beta_0 + \beta_1 X + u$ or
 $Y = \beta_0 + \beta_1 X^2 + u$ or
 $Y = \beta_0 + \beta_1 \log(X) + u$
 - Not OK: $Y = \beta_0 + \beta_1^2 X + u$ or
 $Y = \beta_0 + \exp(\beta_1)X + u$
- β_0, β_1 : Population **parameters** — fixed and unknown
- u : Unobserved random variable with $E[u] = 0$ — captures all other factors influencing Y other than X
- We assume this to be the structural model, i.e., the model describing the true process generating Y

Assumption (II. Random Sampling)

The observed data:

$$(y_i, x_i) \text{ for } i = 1, \dots, n$$

represent an i.i.d. random sample of size n following the population model.

Data examples consistent with this assumption:

- A cross-sectional survey where the units are sampled randomly

Potential Violations:

- Time series data (regressor values may exhibit persistence)
- Sample selection problems (sample not representative of the population)

Assumption (III. Variation in X ; a.k.a. No Perfect Collinearity)

The observed data:

$$x_i \text{ for } i = 1, \dots, n$$

are not all the same value.

Satisfied as long as there is some variation in the regressor X in the sample.

Why do we need this?

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

This assumption is needed just to **calculate** $\hat{\beta}_1$: without it the denominator is zero.

In fact, this is the only assumption needed for using OLS as a pure data summary.

Assumption (IV. Zero Conditional Mean)

The expected value of the error term is zero conditional on any value of the explanatory variable:

$$E[u|X] = 0$$

- $E[u|X] = 0$ implies a slightly weaker condition $\text{Cov}(X, u) = 0$
- Given random sampling, $E[u|X] = 0$ also implies $E[u_i|x_i] = 0$ for all i

Violations:

- Recall that u represents all unobserved factors that influence Y
- What factors are in u ? Could these factors be correlated with X , such that $\text{Cov}(X, u) \neq 0$?
- Example: $\text{Wage} = \beta_0 + \beta_1 \text{education} + u$. What is likely to be in u ?
→ It must be assumed $E[\text{ability}|\text{educ} = \text{low}] = E[\text{ability}|\text{educ} = \text{high}]$

Situation where $E[u|X] = 0$ is likely to hold? Experiment where X is randomly assigned.

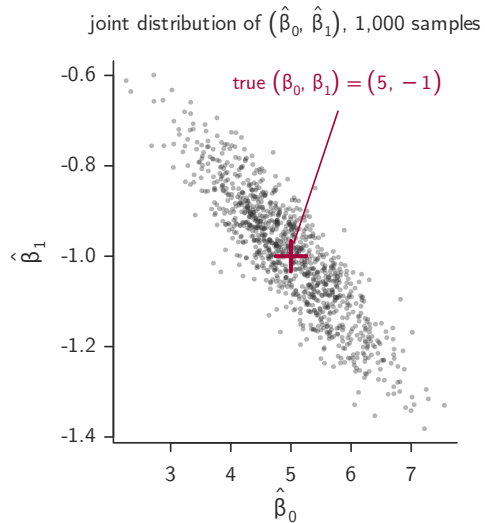
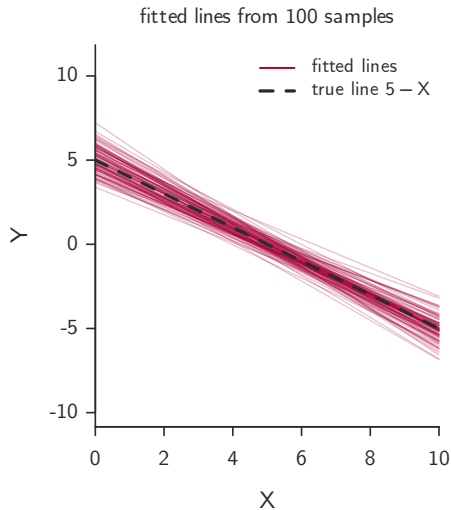
Sampling Distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$

- Given Assumptions I–IV we know where the joint sampling distribution of $(\hat{\beta}_0, \hat{\beta}_1)$ is centered
- Let's use simulations to illustrate what this distribution looks like
- Assume that we have the following population model:

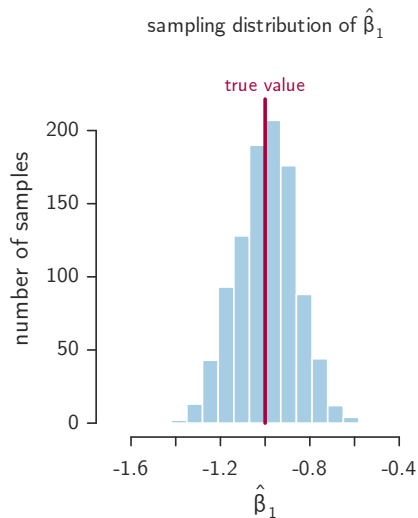
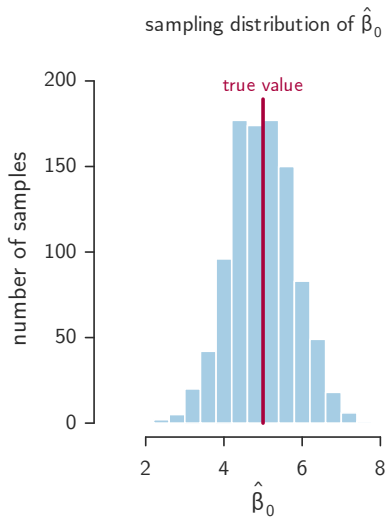
$$Y = \beta_0 + \beta_1 X + u \text{ where } u \sim (\text{some distribution w/ mean } 0)$$

- We are interested in the conditional expectation function ($E[Y|X]$), so we fix values of X for n observations
- For the coefficients, we pick $\beta_0 = 5$ and $\beta_1 = -1$
- We cannot simulate Y_i unless we pick some distribution for u . For simplicity we choose $N(0, 4)$
- Then, repeat the following many times:
 - Draw n independent errors $u_i \sim N(0, 4)$, and compute simulated outcomes $Y = \beta_0 + \beta_1 X + u = 5 - 1x + u$
 - Estimate intercept $\hat{\beta}_0$ and slope $\hat{\beta}_1$ for each simulated data set

Sampling Distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$



Sampling Distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$



Statistical Properties of Ordinary Least Squares

Linear model assumptions

Unbiasedness of OLS Estimators

Variance of OLS Estimators

Estimating Variance for Slope and Intercept

Gauss-Markov Theorem

Theorem (Unbiasedness of OLS)

Given OLS Assumptions I–IV:

$$E[\hat{\beta}_0] = \beta_0 \quad \text{and} \quad E[\hat{\beta}_1] = \beta_1$$

The sampling distributions of the estimators $\hat{\beta}_1$ and $\hat{\beta}_0$ are centered about the true population parameter values β_1 and β_0 .

Worksheet Q1: Prove: $E[\hat{\beta}_1] = \beta_1$.

Proof for the intercept is analogous.

Example: Binary Independent Variable

Suppose **Assumptions I–IV** hold for a simple regression model with $X = 0$ or $X = 1$, which may be the indicator of an experimental treatment.

$$Y = \beta_0 + \beta_1 X + u$$

In this case, the OLS slope estimator $\hat{\beta}_1$ from the regression of Y on X is unbiased for β_1 , the expected effect of the treatment on Y .

$$\begin{aligned} E[Y|X = 1] &= \beta_0 + \beta_1 E[X|X = 1] + E[u|X = 1] \\ &= \beta_0 + \beta_1 \end{aligned}$$

$$\begin{aligned} E[Y|X = 0] &= \beta_0 + \beta_1 E[X|X = 0] + E[u|X = 0] \\ &= \beta_0 \end{aligned}$$

$$E[Y|X = 1] - E[Y|X = 0] = \beta_1$$

Example: Binary Independent Variable

Suppose we drop Assumption IV ($E[u|X] = 0$) and so only **Assumptions I–III** hold for our regression model with $X = 0$ or $X = 1$.

$$Y = \beta_0 + \beta_1 X + u$$

In this case, the OLS slope estimator $\hat{\beta}_1$ from the regression of Y on X is generally biased for β_1 , the expected effect of the treatment on Y .

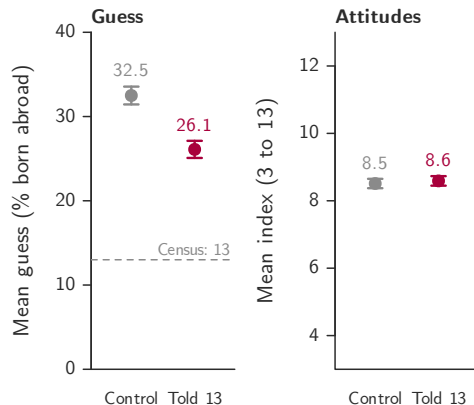
$$\begin{aligned} E[Y|X = 1] &= \beta_0 + \beta_1 E[X|X = 1] + E[u|X = 1] \\ &= \beta_0 + \beta_1 + E[u|X = 1] \end{aligned}$$

$$\begin{aligned} E[Y|X = 0] &= \beta_0 + \beta_1 E[X|X = 0] + E[u|X = 0] \\ &= \beta_0 + E[u|X = 0] \end{aligned}$$

$$E[Y|X = 1] - E[Y|X = 0] = \underbrace{\beta_1}_{\text{Effect}} + \underbrace{\{E[u|X = 1] - E[u|X = 0]\}}_{\text{BIAS}}$$

Example: Telling People How Many Immigrants There Are

- An online survey in March 2017. A random draw put each respondent in a group.
- $X = 1$: told that the Census Bureau estimates about 13 of every 100 people in the U.S. were born abroad.
 $X = 0$: told nothing.
- Then everyone answered three questions about immigration and guessed again how many of every 100 people in the U.S. were born abroad.
- A random draw set X , so the two groups are alike on average in everything else, and $E[u|X = 1] = E[u|X = 0]$.
→ **Assumption IV holds by design**, and the bias term on the last slide is zero.



Group means with 95% intervals. The attitude index adds up the three answers, 3 to 13 (higher = more restrictive).

Source: Hopkins, Sides & Citrin (2019), *Journal of Politics* 81(1); their 2017 Morning Consult survey (non-Hispanic respondents; control vs. information groups).

Binary X in R: The Slope Is a Difference in Means

R Code

```
> lm.guess <- lm(guess ~ info, data = hsc)
> summary(lm.guess)

~~~~ output omitted ~~~~

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  32.4943     0.5258   61.796  <2e-16 ***
info         -6.3941     0.7484   -8.544  <2e-16 ***

Residual standard error: 23.09 on 3806 degrees of freedom
```

- $\hat{\beta}_0 = 32.5$ is the mean guess when $X = 0$, and $\hat{\beta}_0 + \hat{\beta}_1 = 26.1$ is the mean guess when $X = 1$. So $\hat{\beta}_1 = -6.39$ (SE 0.75, $t = -8.54$, $N = 3,808$) is the difference in means.
- The guesses are skewed. The median falls from 25 to 13, but the mean only from 32.5 to 26.1.
- Attitudes barely move. The same regression with the attitude index gives $\hat{\beta}_1 = +0.08$ ($t = 0.77$).

Statistical Properties of Ordinary Least Squares

Linear model assumptions

Unbiasedness of OLS Estimators

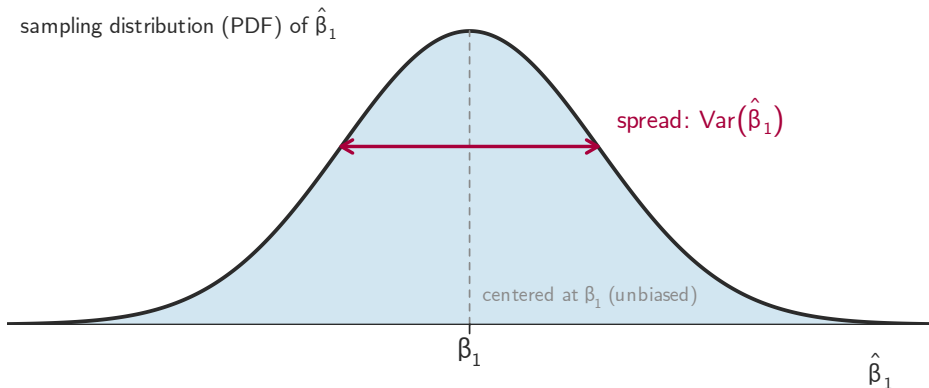
Variance of OLS Estimators

Estimating Variance for Slope and Intercept

Gauss-Markov Theorem

Variance of OLS Estimators

- Given unbiasedness we now know the means of the sampling distributions of our estimators $\hat{\beta}_1$ and $\hat{\beta}_0$.
- What about the variances of the sampling distributions, $\text{Var}[\hat{\beta}_0]$ and $\text{Var}[\hat{\beta}_1]$?



Variance of OLS Estimators

How can we derive $\text{Var}[\hat{\beta}_0]$ and $\text{Var}[\hat{\beta}_1]$? Let's make the following additional assumption:

Assumption (V. Homoskedasticity)

The conditional variance of the error term is constant and does not vary as a function of the explanatory variable:

$$\text{Var}[u|X] = \sigma_u^2$$

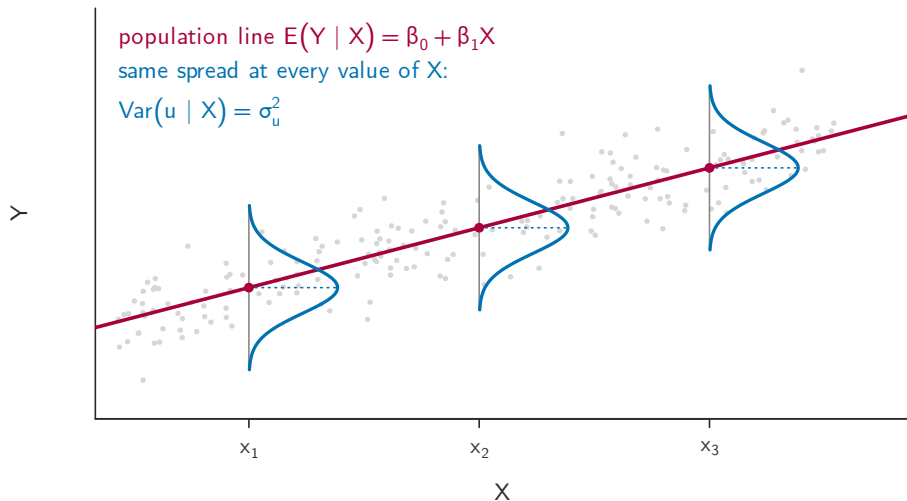
- This implies $\text{Var}[u] = \sigma_u^2$
→ all errors have an identical **error variance** ($\sigma_{u_i}^2 = \sigma_u^2$ for all i)

- Taken together, Assumptions I–V imply:

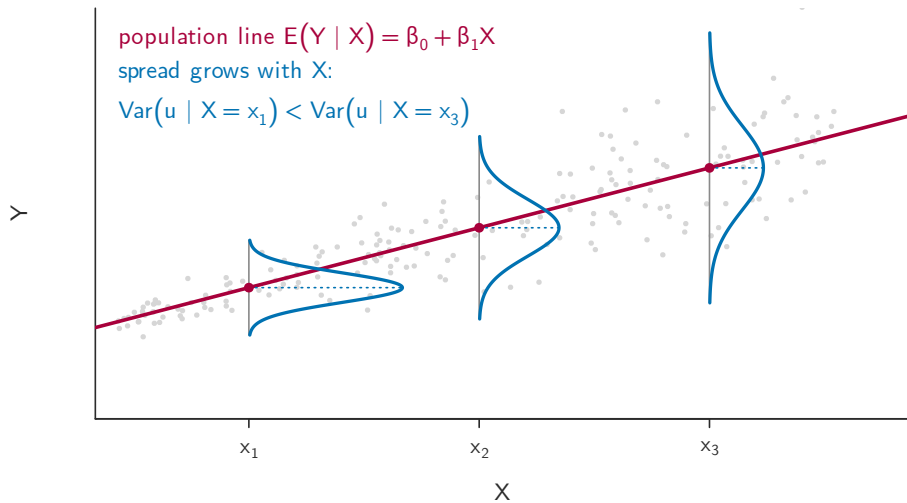
$$\begin{aligned} E[Y|X] &= \beta_0 + \beta_1 X \\ \text{Var}[Y|X] &= \sigma_u^2 \end{aligned}$$

- Violation: $\text{Var}[u|X = x_1] \neq \text{Var}[u|X = x_2]$ called **heteroskedasticity**.
- Assumptions I–V are collectively known as the **Gauss-Markov assumptions**

Homoskedasticity



Heteroskedasticity



Theorem (Variance of OLS Estimators)

Given OLS Assumptions I–V (Gauss-Markov Assumptions):

$$\text{Var}[\hat{\beta}_1 | X] = \frac{\sigma_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sigma_u^2}{SST_x}$$

$$\text{Var}[\hat{\beta}_0 | X] = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\}$$

where $\text{Var}[u | X] = \sigma_u^2$ (the error variance).

Worksheet Q2: Show $\text{Var}[\hat{\beta}_1 | X] = \frac{\sigma_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$

Intuition for the Variance of Slope Estimator

$$\text{Var}[\hat{\beta}_1 | X] = \frac{\sigma_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

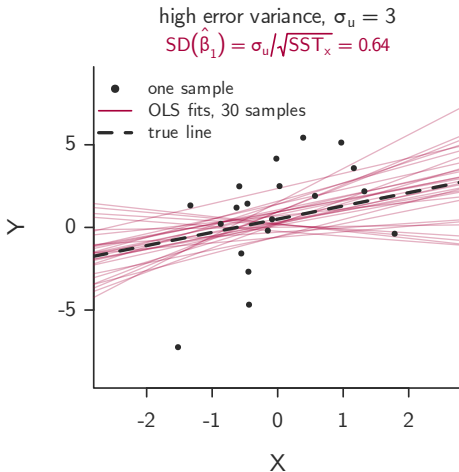
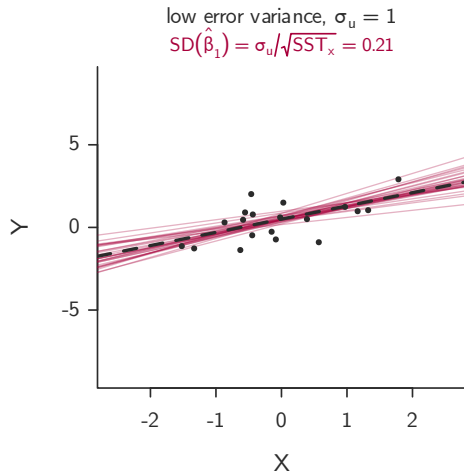
Do we know σ_u^2 ? No.

But let's assume we do for now and ask: What factors affect $\text{Var}[\hat{\beta}_1 | X]$?

Variance is low if

- 1 Error variance $\text{Var}[u] = \sigma_u^2$ is low.

Error Variance and $\text{Var}[\hat{\beta}_1]$



Approximation of true slope is better if $\text{Var}[u] = \sigma_u^2$ is lower (for fixed x , β_1 , β_0)

$$\text{Var}[\hat{\beta}_1 | X] = \frac{\sigma_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Variance is low if

① Error variance $\text{Var}[u] = \sigma_u^2$ is low.

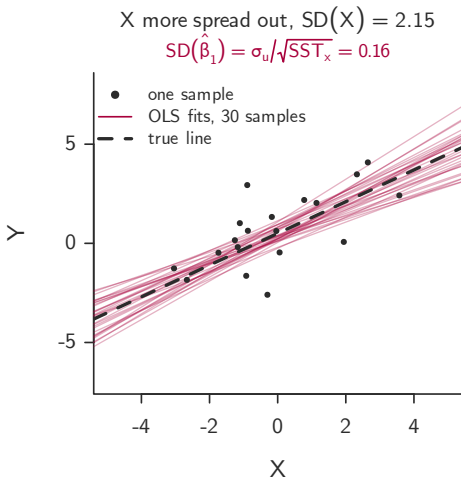
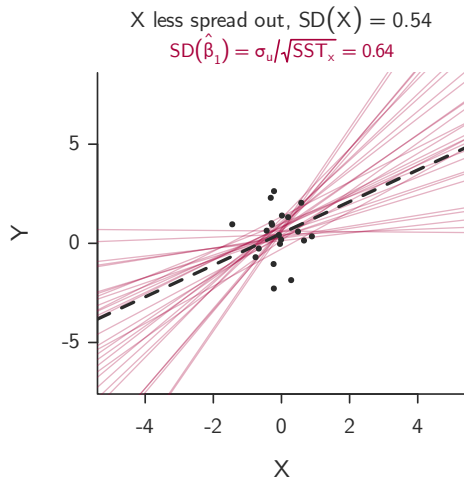
② Note that

$$\sum_{i=1}^n (x_i - \bar{x})^2 = (n-1)S_x^2$$

where S_x^2 is the sample variance of X . Therefore,

- $\text{Var}[\hat{\beta}_1 | X]$ decreasing in n (holding S_x^2 fixed)
- $\text{Var}[\hat{\beta}_1 | X]$ decreasing in S_x^2 (holding n fixed)

$\text{Var}[x]$ and $\text{Var}[\hat{\beta}_1]$



Approximation of true slope is better if $\text{Var}[x]$ is higher (for fixed σ_u , β_1 , β_0)

Statistical Properties of Ordinary Least Squares

Linear model assumptions

Unbiasedness of OLS Estimators

Variance of OLS Estimators

Estimating Variance for Slope and Intercept

Gauss-Markov Theorem

Estimating the Variance of OLS Estimators

How can we estimate the unobserved error variance $\text{Var}[u] = \sigma_u^2$?

We can derive an estimator based on the **residuals**:

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

Recall: The **errors** u_i are NOT the same as the residuals $\hat{u}_i$.

Intuitively, the scatter of the residuals around the fitted regression line should reflect the unseen scatter about the true population regression line.

We can measure scatter with the mean squared deviation:

$$MSD(\hat{u}) \equiv \frac{1}{n} \sum_{i=1}^n (\hat{u}_i - \bar{\hat{u}})^2 = \frac{1}{n} \sum_{i=1}^n \hat{u}_i^2$$

Intuitively, which line is likely to be closer to the observed sample values on X and Y , the true line $E[Y | X = x_i] = \beta_0 + \beta_1 x_i$ or the fitted regression line $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$?

Estimating the Variance of OLS Estimators

- By construction, the regression line is closer since it is drawn to fit the actual sample we have
- Specifically, the regression line is drawn so as to minimize the sum of the squares of the distances between it and the observations
- So the spread of the residuals $MSD(\hat{u})$ will slightly *underestimate* the error variance $\text{Var}[u] = \sigma_u^2$ on average
- In fact, we can show that with a single regressor X we have:

$$E[MSD(\hat{u})] = \frac{n-2}{n} \sigma_u^2 \text{ (degrees of freedom adjustment)}$$

- Thus, an **unbiased estimator** for the error variance is:

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

We plug this estimate into the variance estimators for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Estimating the Variance of OLS Estimators

Given Assumptions I–V (Gauss-Markov Assumptions), we can then estimate the variance of the slope and intercept as:

$$\begin{aligned}\widehat{V}[\hat{\beta}_1 | X] &= \frac{\hat{\sigma}_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \widehat{V}[\hat{\beta}_0 | X] &= \hat{\sigma}_u^2 \left\{ \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right\}\end{aligned}$$

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

The estimated standard errors of the slope and the intercept are $\sqrt{\widehat{V}[\hat{\beta}_1 | X]}$ and $\sqrt{\widehat{V}[\hat{\beta}_0 | X]}$, respectively.

College Degrees and the 2024 Trump Vote, 50 States

R Code

```
> lm.out <- lm(trump ~ ba, data = states)
> summary(lm.out)

~~~~ output omitted ~~~~

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  102.8376      5.9818  17.192  < 2e-16 ***
ba           -1.4537      0.1738  -8.366  6.2e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.589 on 48 degrees of freedom
Multiple R-squared:  0.5932,    Adjusted R-squared:  0.5847
F-statistic: 69.99 on 1 and 48 DF,  p-value: 6.199e-11
```

- By hand, $\hat{\sigma}_u^2 = \sum_i \hat{u}_i^2 / (n - 2) = 2,083.9 / 48 = 43.4$. With $\sum_i (x_i - \bar{x})^2 = 1,437.9$, the standard error of $\hat{\beta}_1$ is $\sqrt{43.4 / 1,437.9} = 0.174$.
- Assumption II is a model here. The 50 states are the whole population of states. We treat each state's vote as one draw of Y given its X .
- Assumption V looks reasonable. The residuals spread about equally across X (a Breusch–Pagan test, taught in Lecture 6-3, gives $p = .75$).

Statistical Properties of Ordinary Least Squares

Linear model assumptions

Unbiasedness of OLS Estimators

Variance of OLS Estimators

Estimating Variance for Slope and Intercept

Gauss-Markov Theorem

Theorem (Gauss-Markov)

Given OLS Assumptions I–V, the OLS estimator is **BLUE**, i.e. the

- ① **Best**: Lowest variance in class
- ② **Linear**: Among Linear estimators
- ③ **Unbiased**: Among Linear Unbiased estimators
- ④ **Estimator**.

Proof.

This proof is not terribly insightful and low priority so we take it on trust (read up on it in Wooldridge, *Introductory Econometrics*, 3rd ed., Appendix 3A.6!).



Gauss-Markov Theorem

