

Statistics for Economists II

Lecture 13: Regression Inference & Goodness of Fit

Güneş A. Aşık

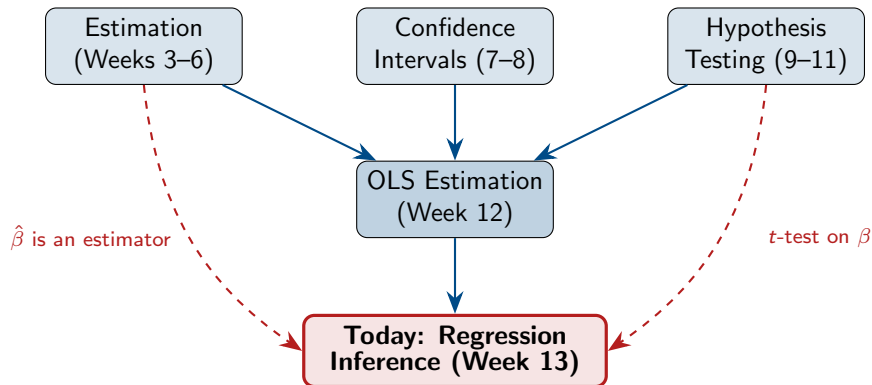
Department of Economics
Middle East Technical University

Spring 2025–2026

Roadmap for Today

- 1 Properties of OLS Estimators
- 2 Estimating σ^2
- 3 Hypothesis Testing for Regression Coefficients
- 4 Goodness of Fit: R^2
- 5 The Consumption Function: A Complete Example
- 6 Asymptotic Properties of OLS
- 7 Practice Problems
- 8 Key Takeaways

Where We Are: Everything Comes Together



OLS gives us *estimates*—but are they reliable? Today we apply the inferential machinery from Weeks 3–11 to regression: unbiasedness, standard errors, t -tests, confidence intervals, and R^2 .

Recall: The Simple Linear Regression Model

Population model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n$$

OLS estimators (from Week 12):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Today's question: **How good are these estimators?**

- Are they unbiased? (Do they hit the target on average?)
- How much do they vary across samples?
- Can we test whether $\beta_1 = 0$?

Classical Assumptions for Regression

Assumptions (SLR.1 – SLR.5)

SLR.1 Linearity: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

SLR.2 Random sampling: (X_i, Y_i) are i.i.d.

SLR.3 Variation in X : $\sum (X_i - \bar{X})^2 > 0$ (not all X_i are equal)

SLR.4 Zero conditional mean: $\mathbb{E}(\varepsilon_i | X_i) = 0$

SLR.5 Homoskedasticity: $\text{Var}(\varepsilon_i | X_i) = \sigma^2$ (constant)

Assumption SLR.4 is the crucial one: it says the error is unrelated to X on average. This rules out **omitted variable bias** and ensures OLS captures the true effect. SLR.5 says the spread of Y around the line is the same for all values of X .

Unbiasedness of OLS

Theorem (Unbiasedness)

Under assumptions SLR.1–SLR.4:

$$\mathbb{E}(\hat{\beta}_0) = \beta_0 \quad \text{and} \quad \mathbb{E}(\hat{\beta}_1) = \beta_1$$

Proof sketch for $\hat{\beta}_1$: Write $\hat{\beta}_1 = \beta_1 + \frac{\sum(X_i - \bar{X})\varepsilon_i}{\sum(X_i - \bar{X})^2}$.

Take expectations conditional on $\mathbf{X}$:

$$\mathbb{E}(\hat{\beta}_1 \mid \mathbf{X}) = \beta_1 + \frac{\sum(X_i - \bar{X})\mathbb{E}(\varepsilon_i \mid X_i)}{\sum(X_i - \bar{X})^2} = \beta_1 + 0 = \beta_1$$

This is just like $\mathbb{E}(\bar{X}) = \mu$ from Week 3! OLS estimators are **centered on the truth**. In repeated samples, they hit the target on average. The key ingredient is $\mathbb{E}(\varepsilon_i \mid X_i) = 0$.

Variance of OLS Estimators

Theorem (Variance Formulas)

Under SLR.1–SLR.5:

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sigma^2}{S_{XX}}$$

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

Three things make $\hat{\beta}_1$ more precise:

- 1 **Lower** σ^2 — less noise in the data
- 2 **Larger** S_{XX} — more spread in X (more “leverage”)
- 3 **Larger** n — more data (since S_{XX} grows with n)

This parallels Week 3: $\text{Var}(\bar{X}) = \sigma^2/n$ — more data, less variance!

The Gauss–Markov Theorem

Theorem (Gauss–Markov)

Under assumptions SLR.1–SLR.5, the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are BLUE:

- Best — *smallest variance*
- Linear — *linear functions of $Y_1, \dots, Y_n$*
- Unbiased — $\mathbb{E}(\hat{\beta}) = \beta$
- Estimator

Among *all* estimators of β_1 that are (i) linear in Y and (ii) unbiased, OLS has the **smallest variance**. You cannot do better without either adding bias or using nonlinear methods.

Important Caveat

- BLUE does *not* mean “best among all estimators”—just among *linear unbiased* ones
- If errors are non-normal, nonlinear estimators *could* beat OLS

Estimating the Error Variance σ^2

The variance formulas involve $\sigma^2 = \text{Var}(\varepsilon_i)$, which is **unknown**.

Residuals provide our window into the errors:

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

Unbiased Estimator of σ^2

$$s^2 = \frac{\text{SSE}}{n-2} = \frac{\sum_{i=1}^n e_i^2}{n-2}$$

where $\text{SSE} = \sum_{i=1}^n e_i^2$ is the **Sum of Squared Errors** (residual sum of squares).

Why $n - 2$ and not n ? We lost **2 degrees of freedom** estimating β_0 and β_1 . The residuals satisfy two constraints: $\sum e_i = 0$ and $\sum X_i e_i = 0$, so only $n - 2$ residuals are “free.” This is exactly like dividing by $n - 1$ for S^2 in Week 3—we lost 1 d.f. estimating μ .

Standard Errors of OLS Estimators

Replace σ^2 with s^2 to get **estimated standard errors**:

$$\widehat{SE}(\hat{\beta}_1) = \frac{s}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}} = \frac{s}{\sqrt{S_{XX}}}$$

$$\widehat{SE}(\hat{\beta}_0) = s \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{S_{XX}}}$$

Standard errors tell you **how precisely** you have estimated the regression coefficients. Small SE $\Rightarrow$ the estimate is reliable. Large SE $\Rightarrow$ lots of uncertainty.

Reporting Convention

Regression results are *always* reported with standard errors (or *t*-statistics).
An estimate without an SE is meaningless!

The Key Question: Does X Really Affect Y ?

Does X really affect Y , or is the pattern just noise?

Formal Test

$H_0 : \beta_1 = 0$ (no effect)

$H_a : \beta_1 \neq 0$ (there is an effect)

Even if $\hat{\beta}_1 \neq 0$, this could be **sampling variation**. The t -test asks: is $\hat{\beta}_1$ far enough from zero—relative to its standard error—that we can rule out chance?

This is *exactly* the hypothesis testing framework from Weeks 9–11, now applied to a regression coefficient instead of a population mean!

The t -Test for β_1

Adding Normality: Assumption SLR.6

$\varepsilon_i \mid X_i \sim N(0, \sigma^2)$ (errors are normally distributed)

Under SLR.1–SLR.6:

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{XX}}\right)$$

Test statistic:

$$t = \frac{\hat{\beta}_1 - \beta_1^0}{\widehat{SE}(\hat{\beta}_1)} \sim t_{n-2}$$

For testing $H_0 : \beta_1 = 0$:

$$t = \frac{\hat{\beta}_1}{\widehat{SE}(\hat{\beta}_1)}$$

Decision Rules for the t -Test

Two-sided test ($H_a : \beta_1 \neq 0$):

Reject H_0 if $|t| > t_{\alpha/2, n-2}$

One-sided tests:

- $H_a : \beta_1 > 0$: Reject if $t > t_{\alpha, n-2}$
- $H_a : \beta_1 < 0$: Reject if $t < -t_{\alpha, n-2}$

p -value approach:

The p -value is the probability of observing a test statistic as extreme as t (or more) under H_0 .

$p < \alpha \implies$ Reject H_0

Software always reports p -values; compare to your chosen significance level.

Practical Significance vs. Statistical Significance

A slope can be **statistically significant** ($p < 0.05$) but **economically tiny**. Always report and interpret the *magnitude* of $\hat{\beta}_1$, not just whether it is “significant.”

Confidence Intervals for β_1

A $100(1 - \alpha)\%$ confidence interval for β_1 :

$$\hat{\beta}_1 \pm t_{\alpha/2, n-2} \cdot \widehat{SE}(\hat{\beta}_1)$$

Example: $\hat{\beta}_1 = 3.2$, $\widehat{SE}(\hat{\beta}_1) = 0.8$, $n = 22$, $\alpha = 0.05$.

$$t_{0.025, 20} = 2.086$$

$$3.2 \pm 2.086 \times 0.8 = 3.2 \pm 1.669 = (1.531, 4.869)$$

Interpretation: We are 95% confident that the **true slope** β_1 lies between 1.531 and 4.869. Since the interval does *not* contain zero, we would reject $H_0 : \beta_1 = 0$ at $\alpha = 0.05$.

This is the CI from Weeks 7–8, applied to a regression coefficient!

Example: Study Hours and Test Scores

Data: $n = 25$ students. Regress test score (Y) on study hours (X).

	Estimate	Std. Error
Intercept ($\hat{\beta}_0$)	42.5	5.3
Study Hours ($\hat{\beta}_1$)	5.8	1.4
$s = 8.6$	$R^2 = 0.43$	$n = 25$

Example: Study Hours and Test Scores

Data: $n = 25$ students. Regress test score (Y) on study hours (X).

	Estimate	Std. Error
Intercept ($\hat{\beta}_0$)	42.5	5.3
Study Hours ($\hat{\beta}_1$)	5.8	1.4
$s = 8.6$	$R^2 = 0.43$	$n = 25$

Test: $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 > 0$ (more study $\Rightarrow$ higher score)

$$t = \frac{5.8}{1.4} = 4.14, \quad t_{0.05, 23} = 1.714$$

Since $4.14 > 1.714$: **Reject H_0** at $\alpha = 0.05$.

Each additional study hour is associated with a **5.8-point** increase in test score, and this effect is **statistically significant**. But does studying *cause* higher scores? Not necessarily.

Example: Wage Regression

Model: $\text{Wage}_i = \beta_0 + \beta_1 \cdot \text{Education}_i + \varepsilon_i$

	Estimate	Std. Error
Intercept ($\hat{\beta}_0$)	-2.87	1.94
Education in years ($\hat{\beta}_1$)	1.54	0.14
$s = 5.85$	$R^2 = 0.165$	$n = 526$

Test: $H_0 : \beta_1 = 0$ (education has no effect on wages)

$$t = \frac{1.54}{0.14} = 11.0 \quad (\text{d.f.} = 524)$$

Reject H_0 overwhelmingly ($p \approx 0$).

Each year of education is associated with \$1.54/hour higher wages. The effect is **highly significant**. But note: $R^2 = 0.165$, meaning education alone explains only 16.5% of wage variation. Many other factors (experience, ability, industry) matter too.

Decomposing Total Variation: ANOVA for Regression

The Fundamental Decomposition

$$\underbrace{\sum_{i=1}^n (Y_i - \bar{Y})^2}_{\text{TSS (Total SS)}} = \underbrace{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}_{\text{SSR (Regression SS)}} + \underbrace{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}_{\text{SSE (Error SS)}}$$

TSS = Total variation in Y



SSR: Explained by regression **SSE**: Unexplained (residual)

Think of total variation in Y as a **pie**. Regression “explains” part of the pie (SSR), and the rest remains unexplained noise (SSE). R^2 tells you what fraction of the pie your model captures.

R^2 : The Coefficient of Determination

Definition

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

Properties:

- $0 \leq R^2 \leq 1$ (always)
- $R^2 = 0$: the regression line is **flat** ($\hat{\beta}_1 = 0$); X explains nothing
- $R^2 = 1$: **perfect fit**; all points lie exactly on the regression line
- In simple regression: $R^2 = r_{XY}^2$ (square of the sample correlation)

$R^2 = 0.43$ means that 43% of the variation in Y is “explained” by X . The remaining 57% is due to other factors (captured in ε).

Cautions About R^2

Common Misinterpretations

- 1 High R^2 does **not** imply causation
- 2 Low R^2 does **not** mean the model is useless
- 3 R^2 *always* increases when you add more variables (even irrelevant ones)
- 4 R^2 says nothing about whether the *right* model is being estimated

Economics Examples

- **Wage on education:** $R^2 \approx 0.17$ — low, but $\hat{\beta}_1$ is highly significant and economically meaningful
- **Consumption on income:** $R^2 \approx 0.85$ — high, driven by strong Keynesian relationship
- **Stock returns on past returns:** $R^2 \approx 0.02$ — very low, consistent with market efficiency

A model with $R^2 = 0.10$ but a *precisely estimated, economically important* $\hat{\beta}_1$ is far more valuable than a model with $R^2 = 0.90$ where the slope has no causal interpretation.

Example: Keynesian Consumption Function

Model: $\text{Consumption}_i = \beta_0 + \beta_1 \cdot \text{Income}_i + \varepsilon_i$

Economic theory predicts: $0 < \beta_1 < 1$ (marginal propensity to consume).

	Estimate	Std. Error	<i>t</i> -stat	<i>p</i> -value
Intercept ($\hat{\beta}_0$)	1240	385	3.22	0.002
Income ($\hat{\beta}_1$)	0.72	0.04	18.0	< 0.001
$s = 2,450, \quad R^2 = 0.84$			$n = 65$ households	

Example: Keynesian Consumption Function

Model: $\text{Consumption}_i = \beta_0 + \beta_1 \cdot \text{Income}_i + \varepsilon_i$

Economic theory predicts: $0 < \beta_1 < 1$ (marginal propensity to consume).

	Estimate	Std. Error	t-stat	p-value
Intercept ($\hat{\beta}_0$)	1240	385	3.22	0.002
Income ($\hat{\beta}_1$)	0.72	0.04	18.0	< 0.001
$s = 2,450, \quad R^2 = 0.84$			$n = 65$ households	

95% CI for MPC: $0.72 \pm 2.000 \times 0.04 = (0.64, 0.80)$

The MPC is estimated at 0.72: a household receiving 1 TL of additional income consumes about 72 kuruş and saves 28 kuruş. The CI (0.64, 0.80) is entirely within (0, 1), **confirming Keynesian theory**.

Consumption Function: Testing Economic Theory

Can we test whether the MPC equals the classic Keynesian value of $\beta_1 = 0.75$?

$$H_0 : \beta_1 = 0.75 \quad \text{vs.} \quad H_a : \beta_1 \neq 0.75$$

$$t = \frac{0.72 - 0.75}{0.04} = \frac{-0.03}{0.04} = -0.75$$

Critical value: $t_{0.025, 63} \approx 2.000$. Since $|-0.75| < 2.000$:

Do not reject H_0 . The data is consistent with $\beta_1 = 0.75$.

We can test β_1 against *any* hypothesized value, not just zero! Here, the t -test confirms that the estimated MPC (0.72) is **not significantly different** from the textbook Keynesian value of 0.75. The t -test framework from Week 10 applies directly.

Large-Sample Properties: Consistency and Asymptotic Normality

Consistency

Under SLR.1–SLR.4:

$$\hat{\beta}_1 \xrightarrow{P} \beta_1 \quad \text{as } n \rightarrow \infty$$

As the sample grows, OLS converges to the true parameter.

Asymptotic Normality

Under mild conditions (no normality of ε required):

$$\frac{\hat{\beta}_1 - \beta_1}{\widehat{SE}(\hat{\beta}_1)} \xrightarrow{D} N(0, 1)$$

For large n , the t -test is valid even without normal errors.

These results connect directly to Weeks 1–2:

- **Consistency** is the WLLN applied to regression: more data $\Rightarrow \hat{\beta}_1 \rightarrow \beta_1$
- **Asymptotic normality** is the CLT applied to regression: standardized $\hat{\beta}_1$ becomes normal

So assumption SLR.6 (normal errors) is a convenience for small samples. With large n , the CLT does the work for us

Summary of OLS Properties

Property	Requires	Connection
Unbiasedness	SLR.1–SLR.4	Week 3: $\mathbb{E}(\bar{X}) = \mu$
BLUE (Gauss–Markov)	SLR.1–SLR.5	Minimum variance
Exact t -distribution	SLR.1–SLR.6	Week 10: t -test
Consistency	SLR.1–SLR.4	Week 1: WLLN
Asymptotic normality	Mild conditions	Week 1: CLT

Notice the progression: more assumptions buy you stronger results. The **minimum** you need for OLS to be useful is SLR.1–SLR.4 (unbiasedness). Add SLR.5 for efficiency (BLUE), SLR.6 for exact finite-sample inference.

Practice Problem 1: Testing a Slope

Problem: A researcher regresses hourly wage (Y , in TL) on years of experience (X) using $n = 42$ workers and obtains:

$$\hat{Y} = 18.5 + 1.25X, \quad \widehat{SE}(\hat{\beta}_1) = 0.45, \quad s = 6.2$$

- (a) Test $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 > 0$ at $\alpha = 0.05$.
- (b) Construct a 95% CI for β_1 .
- (c) Interpret $\hat{\beta}_1 = 1.25$ in economic terms.

Practice Problem 1: Testing a Slope

Problem: A researcher regresses hourly wage (Y , in TL) on years of experience (X) using $n = 42$ workers and obtains:

$$\hat{Y} = 18.5 + 1.25X, \quad \widehat{SE}(\hat{\beta}_1) = 0.45, \quad s = 6.2$$

- (a) Test $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 > 0$ at $\alpha = 0.05$.
- (b) Construct a 95% CI for β_1 .
- (c) Interpret $\hat{\beta}_1 = 1.25$ in economic terms.

Solution:

- (a) $t = 1.25/0.45 = 2.78$. $t_{0.05, 40} = 1.684$. Since $2.78 > 1.684$: **Reject H_0** .
- (b) $t_{0.025, 40} = 2.021$. CI: $1.25 \pm 2.021 \times 0.45 = (0.34, 2.16)$.
- (c) Each additional year of experience is associated with a **1.25 TL/hour** increase in wages, on average, holding other factors constant (implicitly).

Practice Problem 2: R^2 and ANOVA Decomposition

Problem: A regression of firm profits (Y , thousands TL) on advertising expenditure (X , thousands TL) with $n = 30$ firms yields:

$$\text{TSS} = 500, \quad \text{SSE} = 175$$

- (a) Calculate SSR and R^2 .
- (b) Interpret R^2 .
- (c) Calculate s^2 and s .

Practice Problem 2: R^2 and ANOVA Decomposition

Problem: A regression of firm profits (Y , thousands TL) on advertising expenditure (X , thousands TL) with $n = 30$ firms yields:

$$\text{TSS} = 500, \quad \text{SSE} = 175$$

- (a) Calculate SSR and R^2 .
- (b) Interpret R^2 .
- (c) Calculate s^2 and s .

Solution:

- (a) $\text{SSR} = \text{TSS} - \text{SSE} = 500 - 175 = 325$.
 $R^2 = \text{SSR}/\text{TSS} = 325/500 = 0.65$.
- (b) Advertising explains **65%** of the variation in firm profits across the sample.
- (c) $s^2 = \text{SSE}/(n - 2) = 175/28 = 6.25$. $s = \sqrt{6.25} = 2.5$ (thousand TL).

Practice Problem 3: Prediction

Problem: Using the wage–experience regression from Problem 1: $\hat{Y} = 18.5 + 1.25X$, $s = 6.2$, $n = 42$, $\bar{X} = 10$, $S_{XX} = 1,200$.

Predict the wage for a worker with $X^* = 15$ years of experience.

Practice Problem 3: Prediction

Problem: Using the wage–experience regression from Problem 1: $\hat{Y} = 18.5 + 1.25X$, $s = 6.2$, $n = 42$, $\bar{X} = 10$, $S_{XX} = 1,200$.

Predict the wage for a worker with $X^* = 15$ years of experience.

Solution:

$$\hat{Y} = 18.5 + 1.25 \times 15 = 37.25 \text{ TL/hour}$$

Practice Problem 4: Complete Regression Analysis

Problem: A development economist estimates:

$$\widehat{\text{LifeExp}}_i = 52.3 + 0.0038 \cdot \text{GDPperCapita}_i$$

$n = 85$ countries, $\widehat{SE}(\hat{\beta}_1) = 0.0006$, $R^2 = 0.62$.

- (a) Is per-capita GDP significantly related to life expectancy at $\alpha = 0.01$?
- (b) What is the predicted life expectancy for a country with GDP/capita = \$5,000?
- (c) The researcher claims: “GDP causes 62% of life expectancy.” Is this correct?

Practice Problem 4: Complete Regression Analysis

Problem: A development economist estimates:

$$\widehat{\text{LifeExp}}_i = 52.3 + 0.0038 \cdot \text{GDPperCapita}_i$$

$n = 85$ countries, $\widehat{SE}(\hat{\beta}_1) = 0.0006$, $R^2 = 0.62$.

- (a) Is per-capita GDP significantly related to life expectancy at $\alpha = 0.01$?
- (b) What is the predicted life expectancy for a country with GDP/capita = \$5,000?
- (c) The researcher claims: “GDP causes 62% of life expectancy.” Is this correct?

Solution:

- (a) $t = 0.0038/0.0006 = 6.33$. $t_{0.005, 83} \approx 2.636$. Since $6.33 > 2.636$: **Reject H_0** at 1%.
- (b) $\hat{Y} = 52.3 + 0.0038 \times 5000 = 52.3 + 19.0 = 71.3$ years.
- (c) **No!** R^2 measures *association*, not causation. GDP/capita is correlated with many other factors (healthcare, nutrition, education) that also affect life expectancy.

Key Takeaways from Week 13

- 1 Under classical assumptions, OLS is **unbiased**: $\mathbb{E}(\hat{\beta}) = \beta$
- 2 The **Gauss–Markov Theorem** guarantees OLS is BLUE (best among linear unbiased)
- 3 $s^2 = \text{SSE}/(n - 2)$ is the unbiased estimator of σ^2 (we lose 2 d.f.)
- 4 The ***t*-test** for β_1 : $t = \hat{\beta}_1 / \widehat{SE}(\hat{\beta}_1) \sim t_{n-2}$ under H_0
- 5 **Confidence intervals**: $\hat{\beta}_1 \pm t_{\alpha/2, n-2} \cdot \widehat{SE}(\hat{\beta}_1)$
- 6 R^2 measures goodness of fit but **not** causation and **not** model quality alone
- 7 **Prediction intervals** are wider than CIs because of individual-level noise
- 8 **Asymptotic properties**: OLS is consistent and asymptotically normal (CLT!)

Everything in this course—estimation, CIs, hypothesis testing, CLT—**converges** in regression. OLS is estimation. The *t*-test on β_1 is hypothesis testing. The CI for β_1 is a confidence interval. Consistency and asymptotic normality come from the LLN and CLT. This is the payoff!

Further Reading

- Wackerly, Mendenhall & Scheaffer, Ch. 11: Linear Models and Estimation by Least Squares
- Wooldridge, *Introductory Econometrics*, Ch. 2–3: Simple and Multiple Regression
- Devore & Berk, Ch. 12: Simple Linear Regression and Correlation
- Stock & Watson, *Introduction to Econometrics*, Ch. 4–5: Linear Regression