# ISE 315: Engineering Statistics

*Lecture 21: Multiple Linear Regression (Part 2)*

Instructor: Mansur M. Arief, PhD
Industrial and Systems Engineering, KFUPM

Office: 22-219 — Email: mansur.arief@kfupm.edu.sa

*Based on Montgomery & Runger, Applied Statistics and Probability for Engineers, 6th Ed.*

# Lecture 21

Multiple Linear Regression — Part 2 (Chapter 7, cont'd)

## Lecture 21 Outline

- Quick review: MLR, ANOVA, $F$-test (from Lecture 20)

## Lecture 21 Outline

- Quick review: MLR, ANOVA, $F$-test (from Lecture 20)
- Tests and CIs on individual coefficients $\beta_j$ (§7.7)

## Lecture 21 Outline

- Quick review: MLR, ANOVA, $F$-test (from Lecture 20)
- Tests and CIs on individual coefficients $\beta_j$ (§7.7)
- Reading a coefficient table from software (§7.6)

## Lecture 21 Outline

- Quick review: MLR, ANOVA, $F$-test (from Lecture 20)
- Tests and CIs on individual coefficients $\beta_j$ (§7.7)
- Reading a coefficient table from software (§7.6)
- Indicator variables and interaction terms (§7.11, §7.12)

## Lecture 21 Outline

- Quick review: MLR, ANOVA, $F$-test (from Lecture 20)
- Tests and CIs on individual coefficients $\beta_j$ (§7.7)
- Reading a coefficient table from software (§7.6)
- Indicator variables and interaction terms (§7.11, §7.12)
- Multicollinearity and the VIF (§7.10)

## Lecture 21 Outline

- Quick review: MLR, ANOVA, $F$-test (from Lecture 20)
- Tests and CIs on individual coefficients $\beta_j$ (§7.7)
- Reading a coefficient table from software (§7.6)
- Indicator variables and interaction terms (§7.11, §7.12)
- Multicollinearity and the VIF (§7.10)
- Building a model step by step (§7.13)

# Quick Review

What We Did in Lecture 20

## Lecture 20 Recap

**Model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, estimated via $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (7.4).

## Lecture 20 Recap

**Model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, estimated via $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (7.4).

**ANOVA** (7.8): $SS_T = SS_R + SS_E$. Regression df $= k$, residual df $= n - k - 1$, and $\hat{\sigma}^2 = MS_E$ (7.5).

## Lecture 20 Recap

**Model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, estimated via $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (7.4).

**ANOVA** (7.8): $SS_T = SS_R + SS_E$. Regression df $= k$, residual df $= n - k - 1$, and $\hat{\sigma}^2 = MS_E$ (7.5).

**Overall $F$-test** (7.9): $F_0 = MS_R/MS_E \sim F_{k,\,n-k-1}$. Tests whether *any* regressor is useful (Procedure 7.2).

## Lecture 20 Recap

**Model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, estimated via $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (7.4).

**ANOVA** (7.8): $SS_T = SS_R + SS_E$. Regression df $= k$, residual df $= n - k - 1$, and $\hat{\sigma}^2 = MS_E$ (7.5).

**Overall $F$-test** (7.9): $F_0 = MS_R/MS_E \sim F_{k,\,n-k-1}$. Tests whether *any* regressor is useful (Procedure 7.2).

**Our running AI-safety example:** $\hat{y} = 70 + 5x_1 + 8x_2$ (benchmark score vs. coded network size and data quality), $n = 8$, $\hat{\sigma}^2 = 1.6$, $\text{se}(\hat{\beta}_j) = 0.4472$, $F_0 = 222.5$ (significant).

## Lecture 20 Recap

**Model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, estimated via $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (7.4).

**ANOVA** (7.8): $SS_T = SS_R + SS_E$. Regression df $= k$, residual df $= n - k - 1$, and $\hat{\sigma}^2 = MS_E$ (7.5).

**Overall $F$-test** (7.9): $F_0 = MS_R/MS_E \sim F_{k,\,n-k-1}$. Tests whether *any* regressor is useful (Procedure 7.2).

**Our running AI-safety example:** $\hat{y} = 70 + 5x_1 + 8x_2$ (benchmark score vs. coded network size and data quality), $n = 8$, $\hat{\sigma}^2 = 1.6$, $\text{se}(\hat{\beta}_j) = 0.4472$, $F_0 = 222.5$ (significant).

**Today:** Which *individual* regressors matter, categorical inputs, interactions, multicollinearity, and how to build a model.

# Tests on Individual Coefficients

Section 7.7

## The Question Behind the $t$-Test

*Section 7.7*

---

**Question:** Does $x_j$ contribute to the model, *given that* all other regressors are already included?

## The Question Behind the $t$-Test

*Section 7.7*

---

**Question:** Does $x_j$ contribute to the model, *given that* all other regressors are already included?

**Key point:** This is a **marginal** (or partial) test. It measures the contribution of $x_j$ *in the presence of the other regressors*. The same $x_j$ can be significant in one model and insignificant in another, because other regressors may already carry its information.

# The Question Behind the $t$-Test

*Section 7.7*

---

**Question:** Does $x_j$ contribute to the model, *given that* all other regressors are already included?

**Key point:** This is a **marginal** (or partial) test. It measures the contribution of $x_j$ *in the presence of the other regressors*. The same $x_j$ can be significant in one model and insignificant in another, because other regressors may already carry its information.

**General hypotheses:** $H_0 : \beta_j = \beta_{j,0}$ vs. $H_1 : \beta_j \neq \beta_{j,0}$.
The most common case is $\beta_{j,0} = 0$ ("is $x_j$ needed at all?"), but $\beta_{j,0}$ can be any claimed value.

Textbook §7.7: Chapter 7 ((7.12), (7.13))

## Hypothesis Test on a Single Coefficient $\beta_j$

*Procedure 7.3*

---

**Test statistic:**

$$T_0 = \frac{\hat{\beta}_j - \beta_{j,0}}{\text{se}(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_{j,0}}{\sqrt{MS_E \cdot C_{jj}}} \sim t_{n-p} \tag{7.12}$$

where $C_{jj}$ is the diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$ belonging to $\hat{\beta}_j$ (7.7).

## Hypothesis Test on a Single Coefficient $\beta_j$

*Procedure 7.3*

---

**Test statistic:**

$$T_0 = \frac{\hat{\beta}_j - \beta_{j,0}}{\text{se}(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_{j,0}}{\sqrt{MS_E \cdot C_{jj}}} \sim t_{n-p} \tag{7.12}$$

where $C_{jj}$ is the diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$ belonging to $\hat{\beta}_j$ (7.7).

**Decision:** Reject $H_0$ if $|T_0| > t_{\alpha/2,\, n-p}$, or equivalently if the $p$-value $< \alpha$. Use the one-sided variant when $H_1$ is directional.

## Hypothesis Test on a Single Coefficient $\beta_j$

*Procedure 7.3*

---

**Test statistic:**

$$T_0 = \frac{\hat{\beta}_j - \beta_{j,0}}{\mathrm{se}(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_{j,0}}{\sqrt{MS_E \cdot C_{jj}}} \sim t_{n-p} \tag{7.12}$$

where $C_{jj}$ is the diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$ belonging to $\hat{\beta}_j$ (7.7).

**Decision:** Reject $H_0$ if $|T_0| > t_{\alpha/2,\,n-p}$, or equivalently if the $p$-value $< \alpha$. Use the one-sided variant when $H_1$ is directional.

**Trap 1 — degrees of freedom:** $n - p = n - k - 1$, **not** $n - 2$. The value $n - 2$ is the special case $k = 1$.

## Hypothesis Test on a Single Coefficient $\beta_j$

*Procedure 7.3*

---

**Test statistic:**

$$T_0 = \frac{\hat{\beta}_j - \beta_{j,0}}{\text{se}(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \beta_{j,0}}{\sqrt{MS_E \cdot C_{jj}}} \sim t_{n-p} \tag{7.12}$$

where $C_{jj}$ is the diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$ belonging to $\hat{\beta}_j$ (7.7).

**Decision:** Reject $H_0$ if $|T_0| > t_{\alpha/2,\, n-p}$, or equivalently if the $p$-value $< \alpha$. Use the one-sided variant when $H_1$ is directional.

**Trap 1 — degrees of freedom:** $n - p = n - k - 1$, **not** $n - 2$. The value $n - 2$ is the special case $k = 1$.

**Trap 2 — the subscript:** If the question asks about "number of stops" and stops is $x_2$ in the model, the test is on $\beta_2$. Match the subscript to the regressor's position in the model, not to the order the question mentions it.

## Example: Do Size and Data Quality Each Contribute?

*AI safety evaluation model,* $\alpha = 0.05$

---

**Given** (from Lecture 20): $\hat{y} = 70 + 5x_1 + 8x_2$, $n = 8$, $p = 3$, $\text{se}(\hat{\beta}_j) = 0.4472$ for all $j$.

## Example: Do Size and Data Quality Each Contribute?

*AI safety evaluation model,* $\alpha = 0.05$

---

**Given** (from Lecture 20): $\hat{y} = 70 + 5x_1 + 8x_2$, $n = 8$, $p = 3$, $\text{se}(\hat{\beta}_j) = 0.4472$ for all $j$.

**Step 1: Hypotheses.** For each slope: $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$.

## Example: Do Size and Data Quality Each Contribute?

*AI safety evaluation model,* $\alpha = 0.05$

---

**Given** (from Lecture 20): $\hat{y} = 70 + 5x_1 + 8x_2$, $n = 8$, $p = 3$, $\text{se}(\hat{\beta}_j) = 0.4472$ for all $j$.

**Step 1: Hypotheses.** For each slope: $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$.

**Step 2: Degrees of freedom and critical value.** $\text{df} = n - p = 8 - 3 = 5$; from the $t$-table excerpt (Table 7.7), $t_{0.025,\,5} = 2.571$.

## Example: Do Size and Data Quality Each Contribute?

*AI safety evaluation model, $\alpha = 0.05$*

---

**Given** (from Lecture 20): $\hat{y} = 70 + 5x_1 + 8x_2$, $n = 8$, $p = 3$, $\text{se}(\hat{\beta}_j) = 0.4472$ for all $j$.

**Step 1: Hypotheses.** For each slope: $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$.

**Step 2: Degrees of freedom and critical value.** $\text{df} = n - p = 8 - 3 = 5$; from the $t$-table excerpt (Table 7.7), $t_{0.025,\,5} = 2.571$.

**Step 3: Test statistics**, by (7.12):

$$\text{size: } T_0 = \frac{5}{0.4472} = 11.18, \qquad \text{data quality: } T_0 = \frac{8}{0.4472} = 17.89$$

## Example: Do Size and Data Quality Each Contribute?

*AI safety evaluation model, $\alpha = 0.05$*

---

**Given** (from Lecture 20): $\hat{y} = 70 + 5x_1 + 8x_2$, $n = 8$, $p = 3$, $\text{se}(\hat{\beta}_j) = 0.4472$ for all $j$.

**Step 1: Hypotheses.** For each slope: $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$.

**Step 2: Degrees of freedom and critical value.** $\text{df} = n - p = 8 - 3 = 5$; from the $t$-table excerpt (Table 7.7), $t_{0.025,\,5} = 2.571$.

**Step 3: Test statistics**, by (7.12):

$$\text{size: } T_0 = \frac{5}{0.4472} = 11.18, \qquad \text{data quality: } T_0 = \frac{8}{0.4472} = 17.89$$

**Step 4: Decision.** Both $|T_0|$ values exceed 2.571, so both regressors contribute significantly, each given the other is in the model.

## Confidence Interval on an Individual Coefficient

*Section 7.7*

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

$$\hat{\beta}_j - t_{\alpha/2,\,n-p}\,\mathrm{se}(\hat{\beta}_j) \;\leq\; \beta_j \;\leq\; \hat{\beta}_j + t_{\alpha/2,\,n-p}\,\mathrm{se}(\hat{\beta}_j) \tag{7.13}$$

Same pattern as always: estimate $\pm$ critical value $\times$ standard error.

## Confidence Interval on an Individual Coefficient

*Section 7.7*

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

$$\hat{\beta}_j - t_{\alpha/2,\,n-p}\,\mathrm{se}(\hat{\beta}_j) \;\le\; \beta_j \;\le\; \hat{\beta}_j + t_{\alpha/2,\,n-p}\,\mathrm{se}(\hat{\beta}_j) \tag{7.13}$$

Same pattern as always: estimate $\pm$ critical value $\times$ standard error.

**Duality with the test** (Chapter 4 idea):

- If the CI does **not contain zero**, reject $H_0 : \beta_j = 0$: $x_j$ is significant.
- If the CI does **not contain** a claimed value $\beta_{j,0}$, reject $H_0 : \beta_j = \beta_{j,0}$.

## Confidence Interval on an Individual Coefficient

*Section 7.7*

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

$$\hat{\beta}_j - t_{\alpha/2,\,n-p}\,\mathrm{se}(\hat{\beta}_j) \;\le\; \beta_j \;\le\; \hat{\beta}_j + t_{\alpha/2,\,n-p}\,\mathrm{se}(\hat{\beta}_j) \tag{7.13}$$

Same pattern as always: estimate $\pm$ critical value $\times$ standard error.

**Duality with the test** (Chapter 4 idea):

- If the CI does **not contain zero**, reject $H_0 : \beta_j = 0$: $x_j$ is significant.
- If the CI does **not contain** a claimed value $\beta_{j,0}$, reject $H_0 : \beta_j = \beta_{j,0}$.

**Evaluation model:** 95% CI for the data-quality effect $\beta_2$:

$$8 \pm 2.571 \times 0.4472 = 8 \pm 1.150 \;\Rightarrow\; \boxed{(6.850,\ 9.150)}$$

Zero is far outside the interval, in agreement with the $t$ test.

# Worked Example: Which Knobs Move Churn?

*Startup cohorts (Table 7.8 in the textbook)*

A subscription startup models monthly churn $y$ (% of users lost) on $n = 30$ cohorts: $x_1 =$ price (SAR/month), $x_2 =$ median onboarding time (min), $x_3 =$ support tickets per user.

| **Variable** | **Coef** | **SE** | $t$**-Stat** | $p$**-value** |
|---|---|---|---|---|
| Intercept | 4.800 | 1.520 | 3.16 | 0.004 |
| Price ($x_1$) | 0.052 | 0.021 | 2.48 | 0.020 |
| Onboarding ($x_2$) | 0.310 | 0.078 | 3.97 | $< 0.001$ |
| Tickets ($x_3$) | 0.145 | 0.290 | 0.50 | 0.621 |

Residual df $= n - p = 30 - 4 = 26$.

# Worked Example: Which Knobs Move Churn?

*Startup cohorts (Table 7.8 in the textbook)*

A subscription startup models monthly churn $y$ (% of users lost) on $n = 30$ cohorts: $x_1 =$ price (SAR/month), $x_2 =$ median onboarding time (min), $x_3 =$ support tickets per user.

| Variable | Coef | SE | $t$-Stat | $p$-value |
|---|---|---|---|---|
| Intercept | 4.800 | 1.520 | 3.16 | 0.004 |
| Price ($x_1$) | 0.052 | 0.021 | 2.48 | 0.020 |
| Onboarding ($x_2$) | 0.310 | 0.078 | 3.97 | $< 0.001$ |
| Tickets ($x_3$) | 0.145 | 0.290 | 0.50 | 0.621 |

Residual df $= n - p = 30 - 4 = 26$.

**Three questions:**

(a) At $\alpha = 0.05$, which regressors contribute significantly?

(b) Build a 95% CI for the onboarding coefficient $\beta_2$.

## Churn Example (a): Individual Tests

**Step 1: Critical value.** df $= 26$, so $t_{0.025,\,26} = 2.056$ (Table 7.7).

## Churn Example (a): Individual Tests

**Step 1: Critical value.** df $= 26$, so $t_{0.025,\,26} = 2.056$ (Table 7.7).

**Step 2: Compare each $|t|$-Stat** (equivalently, each $p$-value to 0.05):

- Price: $|2.48| > 2.056 \;\Rightarrow$ **significant**
- Onboarding: $|3.97| > 2.056 \;\Rightarrow$ **significant**
- Tickets: $|0.50| < 2.056 \;\Rightarrow$ **not significant**

## Churn Example (a): Individual Tests

**Step 1: Critical value.** df $= 26$, so $t_{0.025,\,26} = 2.056$ (Table 7.7).

**Step 2: Compare each $|t|$-Stat** (equivalently, each $p$-value to 0.05):

- Price: $|2.48| > 2.056 \;\Rightarrow$ **significant**
- Onboarding: $|3.97| > 2.056 \;\Rightarrow$ **significant**
- Tickets: $|0.50| < 2.056 \;\Rightarrow$ **not significant**

**Conclusion in context:** Given price and onboarding are in the model, the ticket rate adds no detectable explanatory power.

## Churn Example (a): Individual Tests

**Step 1: Critical value.** df $= 26$, so $t_{0.025,\,26} = 2.056$ (Table 7.7).

**Step 2: Compare each $|t|$-Stat** (equivalently, each $p$-value to 0.05):

- Price: $|2.48| > 2.056 \;\Rightarrow$ **significant**
- Onboarding: $|3.97| > 2.056 \;\Rightarrow$ **significant**
- Tickets: $|0.50| < 2.056 \;\Rightarrow$ **not significant**

**Conclusion in context:** Given price and onboarding are in the model, the ticket rate adds no detectable explanatory power.

**Careful:** This does *not* say tickets are irrelevant to churn. It says tickets add nothing *once price and onboarding are known* — perhaps because slow onboarding generates the tickets. Marginal tests rank knobs given the model, not in isolation.

# Churn Example (b), (c): CI and Testing a Claim

**(b) 95% CI for** $\beta_2$, by (7.13):

$$0.310 \pm 2.056 \times 0.078 = 0.310 \pm 0.1604 \;\Rightarrow\; \boxed{(0.150,\ 0.470)}$$

(% churn per minute of onboarding). Zero is outside, matching the $t$-test in part (a).

# Churn Example (b), (c): CI and Testing a Claim

**(b) 95% CI for** $\beta_2$, by (7.13):

$$0.310 \pm 2.056 \times 0.078 = 0.310 \pm 0.1604 \;\Rightarrow\; \boxed{(0.150,\ 0.470)}$$

(% churn per minute of onboarding). Zero is outside, matching the $t$-test in part (a).

**(c) Test the consultant's claim** $H_0 : \beta_2 = 1$ vs. $H_1 : \beta_2 \neq 1$, by (7.12) with $\beta_{2,0} = 1$:

$$T_0 = \frac{0.310 - 1}{0.078} = \frac{-0.690}{0.078} = -8.846$$

# Churn Example (b), (c): CI and Testing a Claim

**(b) 95% CI for** $\beta_2$, by (7.13):

$$0.310 \pm 2.056 \times 0.078 = 0.310 \pm 0.1604 \;\Rightarrow\; \boxed{(0.150,\ 0.470)}$$

(% churn per minute of onboarding). Zero is outside, matching the $t$-test in part (a).

**(c) Test the consultant's claim** $H_0 : \beta_2 = 1$ vs. $H_1 : \beta_2 \neq 1$, by (7.12) with $\beta_{2,0} = 1$:

$$T_0 = \frac{0.310 - 1}{0.078} = \frac{-0.690}{0.078} = -8.846$$

Since $|-8.846| > 2.056$, **reject** $H_0$. Equivalently: $1 \notin (0.150, 0.470)$.

## Churn Example (b), (c): CI and Testing a Claim

**(b) 95% CI for** $\beta_2$, by (7.13):

$$0.310 \pm 2.056 \times 0.078 = 0.310 \pm 0.1604 \;\Rightarrow\; \boxed{(0.150,\ 0.470)}$$

(% churn per minute of onboarding). Zero is outside, matching the $t$-test in part (a).

**(c) Test the consultant's claim** $H_0 : \beta_2 = 1$ vs. $H_1 : \beta_2 \neq 1$, by (7.12) with $\beta_{2,0} = 1$:

$$T_0 = \frac{0.310 - 1}{0.078} = \frac{-0.690}{0.078} = -8.846$$

Since $|-8.846| > 2.056$, **reject** $H_0$. Equivalently: $1 \notin (0.150, 0.470)$.

**Conclusion:** Onboarding friction matters, but at roughly a third of a point per minute, not a full point. The data contradict the claim.

# Choosing the Right Hypothesis

*Table 7.9 in the textbook*

| **If the question says...** | **Use...** |
|---|---|
| "Is the model significant?" | $F$-test (7.9): $H_0 : \beta_1 = \cdots = \beta_k = 0$ |
| "Is $x_j$ significant?" | $t$-test (7.12): $H_0 : \beta_j = 0$ |
| "Is the effect of $x_j$ equal to $c$?" | $t$-test (7.12): $H_0 : \beta_j = c$ |
| "Is the relationship inverse?" | one-sided $t$: $H_0 : \beta_j = 0$, $H_1 : \beta_j < 0$ |
| "Does a *group* of $x$'s contribute?" | partial $F$-test (7.14) |

## Choosing the Right Hypothesis

*Table 7.9 in the textbook*

| **If the question says...** | **Use...** |
|---|---|
| "Is the model significant?" | $F$-test (7.9): $H_0 : \beta_1 = \cdots = \beta_k = 0$ |
| "Is $x_j$ significant?" | $t$-test (7.12): $H_0 : \beta_j = 0$ |
| "Is the effect of $x_j$ equal to $c$?" | $t$-test (7.12): $H_0 : \beta_j = c$ |
| "Is the relationship inverse?" | one-sided $t$: $H_0 : \beta_j = 0$, $H_1 : \beta_j < 0$ |
| "Does a *group* of $x$'s contribute?" | partial $F$-test (7.14) |

The partial $F$-test (Procedure 7.4) compares a full model against a reduced one — it is covered in the textbook, §7.8, and worth a read.

## Choosing the Right Hypothesis

*Table 7.9 in the textbook*

| **If the question says...** | **Use...** |
|---|---|
| "Is the model significant?" | $F$-test (7.9): $H_0 : \beta_1 = \cdots = \beta_k = 0$ |
| "Is $x_j$ significant?" | $t$-test (7.12): $H_0 : \beta_j = 0$ |
| "Is the effect of $x_j$ equal to $c$?" | $t$-test (7.12): $H_0 : \beta_j = c$ |
| "Is the relationship inverse?" | one-sided $t$: $H_0 : \beta_j = 0$, $H_1 : \beta_j < 0$ |
| "Does a *group* of $x$'s contribute?" | partial $F$-test (7.14) |

The partial $F$-test (Procedure 7.4) compares a full model against a reduced one — it is covered in the textbook, §7.8, and worth a read.

**Exam tip:** Identify the question type first, then pick the statistic.

# Reading Software Output

Section 7.6

# Practice: Filling in a Coefficient Table

*Delivery model (Table 7.6)*

The delivery-time model from Lecture 20 ($n = 25$, df $= 21$), with entries blanked out:

| **Variable** | **Coef** | **SE** | $t$-**Stat** | $p$-**value** |
|---|---|---|---|---|
| Intercept | 3.147 | 0.448 | 7.03 | $< 0.001$ |
| Distance ($x_1$) | 0.093 | ? | 6.37 | $< 0.001$ |
| Stops ($x_2$) | 0.633 | 0.132 | ? | $< 0.001$ |
| Load ($x_3$) | $-0.302$ | 0.146 | ? | 0.051 |

# Practice: Filling in a Coefficient Table

*Delivery model (Table 7.6)*

The delivery-time model from Lecture 20 ($n = 25$, df $= 21$), with entries blanked out:

| Variable | Coef | SE | $t$-Stat | $p$-value |
|---|---|---|---|---|
| Intercept | 3.147 | 0.448 | 7.03 | $< 0.001$ |
| Distance ($x_1$) | 0.093 | ? | 6.37 | $< 0.001$ |
| Stops ($x_2$) | 0.633 | 0.132 | ? | $< 0.001$ |
| Load ($x_3$) | $-0.302$ | 0.146 | ? | 0.051 |

**Rule:** $t$-Stat = Coef / SE. Given any two of the three, recover the third.

# Practice: Filling in a Coefficient Table

*Delivery model (Table 7.6)*

The delivery-time model from Lecture 20 ($n = 25$, df $= 21$), with entries blanked out:

| Variable | Coef | SE | $t$-Stat | $p$-value |
|---|---|---|---|---|
| Intercept | 3.147 | 0.448 | 7.03 | $< 0.001$ |
| Distance ($x_1$) | 0.093 | ? | 6.37 | $< 0.001$ |
| Stops ($x_2$) | 0.633 | 0.132 | ? | $< 0.001$ |
| Load ($x_3$) | $-0.302$ | 0.146 | ? | 0.051 |

**Rule:** $t$-Stat = Coef / SE. Given any two of the three, recover the third.
**Distance SE:** $\text{se}(\hat{\beta}_1) = 0.093/6.37 = 0.0146$.

# Practice: Filling in a Coefficient Table

*Delivery model (Table 7.6)*

The delivery-time model from Lecture 20 ($n = 25$, df $= 21$), with entries blanked out:

| Variable | Coef | SE | $t$-Stat | $p$-value |
|---|---|---|---|---|
| Intercept | 3.147 | 0.448 | 7.03 | $< 0.001$ |
| Distance ($x_1$) | 0.093 | ? | 6.37 | $< 0.001$ |
| Stops ($x_2$) | 0.633 | 0.132 | ? | $< 0.001$ |
| Load ($x_3$) | $-0.302$ | 0.146 | ? | 0.051 |

**Rule:** $t$-Stat = Coef / SE. Given any two of the three, recover the third.
**Distance SE:** $\text{se}(\hat{\beta}_1) = 0.093/6.37 = 0.0146$.
**Stops $t$-Stat:** $t_0 = 0.633/0.132 = 4.80$.

# Practice: Filling in a Coefficient Table

*Delivery model (Table 7.6)*

The delivery-time model from Lecture 20 ($n = 25$, df $= 21$), with entries blanked out:

| Variable | Coef | SE | $t$-Stat | $p$-value |
|---|---|---|---|---|
| Intercept | 3.147 | 0.448 | 7.03 | $< 0.001$ |
| Distance ($x_1$) | 0.093 | ? | 6.37 | $< 0.001$ |
| Stops ($x_2$) | 0.633 | 0.132 | ? | $< 0.001$ |
| Load ($x_3$) | $-0.302$ | 0.146 | ? | 0.051 |

**Rule:** $t$-Stat = Coef / SE. Given any two of the three, recover the third.
**Distance SE:** $\text{se}(\hat{\beta}_1) = 0.093/6.37 = 0.0146$.
**Stops $t$-Stat:** $t_0 = 0.633/0.132 = 4.80$.
**Load $t$-Stat:** $t_0 = -0.302/0.146 = -2.07$.

# Practice: Filling in a Coefficient Table

*Delivery model (Table 7.6)*

The delivery-time model from Lecture 20 ($n = 25$, df $= 21$), with entries blanked out:

| Variable | Coef | SE | $t$-Stat | $p$-value |
|---|---|---|---|---|
| Intercept | 3.147 | 0.448 | 7.03 | $< 0.001$ |
| Distance ($x_1$) | 0.093 | ? | 6.37 | $< 0.001$ |
| Stops ($x_2$) | 0.633 | 0.132 | ? | $< 0.001$ |
| Load ($x_3$) | $-0.302$ | 0.146 | ? | 0.051 |

**Rule:** $t$-Stat = Coef / SE. Given any two of the three, recover the third.
**Distance SE:** $\text{se}(\hat{\beta}_1) = 0.093/6.37 = 0.0146$.
**Stops $t$-Stat:** $t_0 = 0.633/0.132 = 4.80$.
**Load $t$-Stat:** $t_0 = -0.302/0.146 = -2.07$.

Behind the scenes, each SE is $\sqrt{MS_E \cdot C_{jj}}$ (7.7) with $MS_E = 0.2068$ from the

## Acting on the Output: Which Regressors Matter?

*Delivery model,* $\alpha = 0.05$

---

**Compare each** $p$**-value to** $\alpha = 0.05$ (or each $|t|$ to $t_{0.025,\,21} = 2.080$):

- Distance: $p < 0.001 \;\Rightarrow$ **significant**
- Stops: $p < 0.001 \;\Rightarrow$ **significant**
- Load: $p = 0.051 > 0.05 \;\Rightarrow$ **not significant** (barely — $|-2.07| < 2.080$)

## Acting on the Output: Which Regressors Matter?

*Delivery model,* $\alpha = 0.05$

---

**Compare each** ***p*-value to** $\alpha = 0.05$ (or each $|t|$ to $t_{0.025,\,21} = 2.080$):

- Distance: $p < 0.001 \;\Rightarrow$ **significant**
- Stops: $p < 0.001 \;\Rightarrow$ **significant**
- Load: $p = 0.051 > 0.05 \;\Rightarrow$ **not significant** (barely — $|-2.07| < 2.080$)

**The negative load coefficient looks wrong** — heavier trucks should be slower. Operations knowledge resolves it: heavily loaded routes are assigned to senior drivers on highway corridors. The coefficient is the effect of load *holding distance and stops fixed*, and within that slice the driver-assignment policy dominates.

## Acting on the Output: Which Regressors Matter?

*Delivery model, $\alpha = 0.05$*

---

**Compare each *p*-value to** $\alpha = 0.05$ (or each $|t|$ to $t_{0.025,\,21} = 2.080$):

- Distance: $p < 0.001 \;\Rightarrow$ **significant**
- Stops: $p < 0.001 \;\Rightarrow$ **significant**
- Load: $p = 0.051 > 0.05 \;\Rightarrow$ **not significant** (barely — $|-2.07| < 2.080$)

**The negative load coefficient looks wrong** — heavier trucks should be slower. Operations knowledge resolves it: heavily loaded routes are assigned to senior drivers on highway corridors. The coefficient is the effect of load *holding distance and stops fixed*, and within that slice the driver-assignment policy dominates.

**Lesson:** Always reconcile coefficient signs with how the system actually operates before trusting or deleting a regressor.

# Indicator Variables & Interactions

Sections 7.11 and 7.12

# Indicator Variables: Categories in a Regression

*Section 7.11*

A categorical input with two levels — annual vs. monthly plan, curated vs. uncurated data, highway vs. city route — enters the model as an **indicator variable**, coded $z = 1$ for one level and $z = 0$ for the other:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 z + \varepsilon \tag{7.18}$$

# Indicator Variables: Categories in a Regression

*Section 7.11*

A categorical input with two levels — annual vs. monthly plan, curated vs. uncurated data, highway vs. city route — enters the model as an **indicator variable**, coded $z = 1$ for one level and $z = 0$ for the other:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 z + \varepsilon \tag{7.18}$$

**One model, two parallel lines:**

- Group $z = 0$: mean response $= \beta_0 + \beta_1 x_1$
- Group $z = 1$: mean response $= (\beta_0 + \beta_2) + \beta_1 x_1$

$\beta_2$ is the vertical gap between the groups at any fixed $x_1$. The $t$-test of $H_0 : \beta_2 = 0$ (7.12) asks whether the groups differ once $x_1$ is accounted for.

# Indicator Variables: Categories in a Regression

*Section 7.11*

A categorical input with two levels — annual vs. monthly plan, curated vs. uncurated data, highway vs. city route — enters the model as an **indicator variable**, coded $z = 1$ for one level and $z = 0$ for the other:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 z + \varepsilon \tag{7.18}$$

**One model, two parallel lines:**

- Group $z = 0$: mean response $= \beta_0 + \beta_1 x_1$
- Group $z = 1$: mean response $= (\beta_0 + \beta_2) + \beta_1 x_1$

$\beta_2$ is the vertical gap between the groups at any fixed $x_1$. The $t$-test of $H_0 : \beta_2 = 0$ (7.12) asks whether the groups differ once $x_1$ is accounted for.

**Rule:** A category with $m$ levels needs $m - 1$ indicators (one level is the baseline).

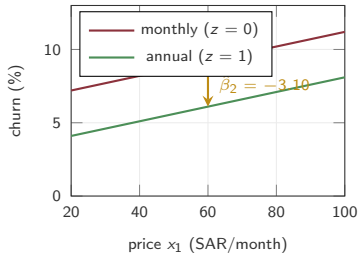
# Example: Churn for Monthly vs. Annual Plans

*Figure 7.5 in the textbook*

The startup fits churn (%) on price $x_1$ (SAR/month) and an indicator $z$ ($1 =$ annual, $0 =$ monthly), $n = 24$:

$$\hat{y} = 6.20 + 0.050\,x_1 - 3.10\,z$$

with $\mathrm{se}(\hat{\beta}_2) = 0.85$ and $n - p = 21$ df.



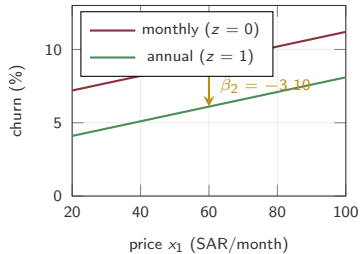
# Example: Churn for Monthly vs. Annual Plans

*Figure 7.5 in the textbook*

The startup fits churn (%) on price $x_1$ (SAR/month) and an indicator $z$ ($1 =$ annual, $0 =$ monthly), $n = 24$:

$$\hat{y} = 6.20 + 0.050\,x_1 - 3.10\,z$$

with $\mathrm{se}(\hat{\beta}_2) = 0.85$ and $n - p = 21$ df.




**Step 1: Two parallel lines.** Monthly ($z = 0$): $\hat{y} = 6.20 + 0.050\,x_1$. Annual ($z = 1$): $\hat{y} = 3.10 + 0.050\,x_1$. At any given price, annual cohorts churn 3.10 points less; the price slope is common to both.

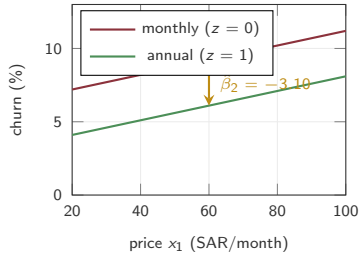
# Example: Churn for Monthly vs. Annual Plans

*Figure 7.5 in the textbook*

The startup fits churn (%) on price $x_1$ (SAR/month) and an indicator $z$ ($1 =$ annual, $0 =$ monthly), $n = 24$:

$$\hat{y} = 6.20 + 0.050\,x_1 - 3.10\,z$$

with $\mathrm{se}(\hat{\beta}_2) = 0.85$ and $n - p = 21$ df.




**Step 1: Two parallel lines.** Monthly ($z = 0$): $\hat{y} = 6.20 + 0.050\,x_1$. Annual ($z = 1$): $\hat{y} = 3.10 + 0.050\,x_1$. At any given price, annual cohorts churn 3.10 points less; the price slope is common to both.

**Step 2: Test the gap** at $\alpha = 0.05$, by (7.12): $T_0 = -3.10/0.85 = -3.647$. Critical value $t_{0.025,\,21} = 2.080$. Since $|-3.647| > 2.080$, **reject** $H_0$: annual plans churn significantly less.

## Interaction Terms: When One Effect Depends on Another

*Section 7.12*

If the effect of $x_1$ on the response depends on the level of $x_2$, add their **product** as a new regressor:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12}\, x_1 x_2 + \varepsilon \tag{7.19}$$

## Interaction Terms: When One Effect Depends on Another

*Section 7.12*

If the effect of $x_1$ on the response depends on the level of $x_2$, add their **product** as a new regressor:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12}\, x_1 x_2 + \varepsilon \tag{7.19}$$

Setting $x_3 = x_1 x_2$ turns this into an ordinary three-regressor model, fitted by (7.4) as usual — no new theory needed. "Linear" means linear in the $\beta$'s.

# Interaction Terms: When One Effect Depends on Another

*Section 7.12*

If the effect of $x_1$ on the response depends on the level of $x_2$, add their **product** as a new regressor:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12}\, x_1 x_2 + \varepsilon \tag{7.19}$$

Setting $x_3 = x_1 x_2$ turns this into an ordinary three-regressor model, fitted by (7.4) as usual — no new theory needed. "Linear" means linear in the $\beta$'s.

**How to read it — group the terms in $x_1$:**

$$\text{slope of } y \text{ w.r.t. } x_1 = \beta_1 + \beta_{12}\, x_2$$

a slope that *changes* with $x_2$.

# Interaction Terms: When One Effect Depends on Another

*Section 7.12*

If the effect of $x_1$ on the response depends on the level of $x_2$, add their **product** as a new regressor:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12}\, x_1 x_2 + \varepsilon \tag{7.19}$$

Setting $x_3 = x_1 x_2$ turns this into an ordinary three-regressor model, fitted by (7.4) as usual — no new theory needed. "Linear" means linear in the $\beta$'s.

**How to read it — group the terms in $x_1$:**

$$\text{slope of } y \text{ w.r.t. } x_1 = \beta_1 + \beta_{12}\, x_2$$

a slope that *changes* with $x_2$.

When $x_2 = z$ is an indicator, the model draws two lines with different intercepts **and** different slopes — it relaxes the parallel-lines assumption. The $t$-test on $\beta_{12}$

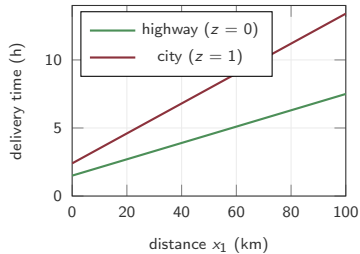
# Example: Distance $\times$ Route Type in Delivery Times

*Figure 7.6 in the textbook*

A carrier fits delivery time (h) on distance $x_1$ (km), route indicator $z$ ($1 =$ city, $0 =$ highway), and their product, $n = 30$:

$$\hat{y} = 1.50 + 0.060\,x_1 + 0.90\,z + 0.050\,x_1 z$$

with $\text{se}(\hat{\beta}_{12}) = 0.012$, $\text{df} = 30 - 4 = 26$.



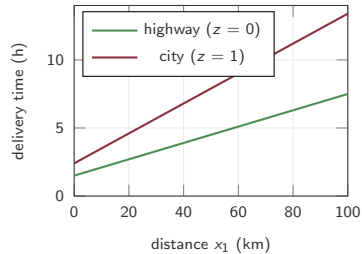
# Example: Distance $\times$ Route Type in Delivery Times

*Figure 7.6 in the textbook*

A carrier fits delivery time (h) on distance $x_1$ (km), route indicator $z$ ($1 =$ city, $0 =$ highway), and their product, $n = 30$:

$$\hat{y} = 1.50 + 0.060\,x_1 + 0.90\,z + 0.050\,x_1 z$$

with $\text{se}(\hat{\beta}_{12}) = 0.012$, $\text{df} = 30 - 4 = 26$.




**Step 1: One line per group.** Highway ($z = 0$): $\hat{y} = 1.50 + 0.060\,x_1$. City ($z = 1$): $\hat{y} = (1.50 + 0.90) + (0.060 + 0.050)\,x_1 = 2.40 + 0.110\,x_1$.

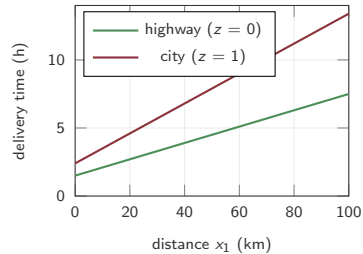
# Example: Distance × Route Type in Delivery Times

*Figure 7.6 in the textbook*

A carrier fits delivery time (h) on distance $x_1$ (km), route indicator $z$ ($1 =$ city, $0 =$ highway), and their product, $n = 30$:

$$\hat{y} = 1.50 + 0.060\,x_1 + 0.90\,z + 0.050\,x_1 z$$

with $\text{se}(\hat{\beta}_{12}) = 0.012$, $\text{df} = 30 - 4 = 26$.




**Step 1: One line per group.** Highway ($z = 0$): $\hat{y} = 1.50 + 0.060\,x_1$. City ($z = 1$): $\hat{y} = (1.50 + 0.90) + (0.060 + 0.050)\,x_1 = 2.40 + 0.110\,x_1$.

**Step 2: Read the slopes.** Each extra kilometer costs 0.060 h on the highway but 0.110 h in the city — the effect of distance depends on route type.

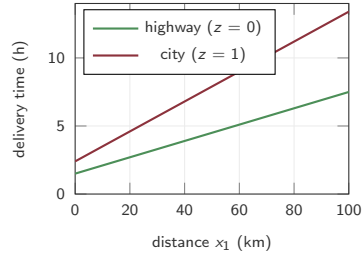
# Example: Distance × Route Type in Delivery Times

*Figure 7.6 in the textbook*

A carrier fits delivery time (h) on distance $x_1$ (km), route indicator $z$ ($1 =$ city, $0 =$ highway), and their product, $n = 30$:

$$\hat{y} = 1.50 + 0.060\,x_1 + 0.90\,z + 0.050\,x_1 z$$

with $\text{se}(\hat{\beta}_{12}) = 0.012$, $\text{df} = 30 - 4 = 26$.




**Step 1: One line per group.** Highway ($z = 0$): $\hat{y} = 1.50 + 0.060\,x_1$. City ($z = 1$): $\hat{y} = (1.50 + 0.90) + (0.060 + 0.050)\,x_1 = 2.40 + 0.110\,x_1$.

**Step 2: Read the slopes.** Each extra kilometer costs 0.060 h on the highway but 0.110 h in the city — the effect of distance depends on route type.

**Step 3: Is the interaction real?** $T_0 = 0.050/0.012 = 4.167 > t_{0.025,\,26} = 2.056$. **Reject** $H_0$: $\beta_{12} = 0$; the slopes differ.

# Polynomial Terms: Fitting Curves with a Linear Model

*Section 7.12*

If a residual plot shows curvature, add powers of the regressor:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \tag{7.20}$$

Setting $x_1 = x$ and $x_2 = x^2$ gives a standard two-regressor linear model — the fitted *curve* is not a line, but the model is linear in the $\beta$'s.

# Polynomial Terms: Fitting Curves with a Linear Model

*Section 7.12*

If a residual plot shows curvature, add powers of the regressor:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \tag{7.20}$$

Setting $x_1 = x$ and $x_2 = x^2$ gives a standard two-regressor linear model — the fitted *curve* is not a line, but the model is linear in the $\beta$'s.

**Example:** Warehouse picking error rate vs. picking speed is U-shaped: slow picking wastes attention, fast picking causes misses. A quadratic term captures this with the same least squares machinery.

# Polynomial Terms: Fitting Curves with a Linear Model

*Section 7.12*

If a residual plot shows curvature, add powers of the regressor:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \tag{7.20}$$

Setting $x_1 = x$ and $x_2 = x^2$ gives a standard two-regressor linear model — the fitted *curve* is not a line, but the model is linear in the $\beta$'s.

**Example:** Warehouse picking error rate vs. picking speed is U-shaped: slow picking wastes attention, fast picking causes misses. A quadratic term captures this with the same least squares machinery.

**Transformed responses**, same idea: fitting $\ln y = \beta_0 + \beta_1 x + \varepsilon$ is a linear regression with response $\ln y$.

# Polynomial Terms: Fitting Curves with a Linear Model

*Section 7.12*

If a residual plot shows curvature, add powers of the regressor:

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \tag{7.20}$$

Setting $x_1 = x$ and $x_2 = x^2$ gives a standard two-regressor linear model — the fitted *curve* is not a line, but the model is linear in the $\beta$'s.

**Example:** Warehouse picking error rate vs. picking speed is U-shaped: slow picking wastes attention, fast picking causes misses. A quadratic term captures this with the same least squares machinery.

**Transformed responses**, same idea: fitting $\ln y = \beta_0 + \beta_1 x + \varepsilon$ is a linear regression with response $\ln y$.

**Caution — back-transform as a whole:** From $\widehat{\ln y} = 2.0 + 0.50\,x$, the prediction at $x = 3$ is $\hat{y} = e^{2.0+0.50(3)} = e^{3.5} = 33.12$, **not** $e^{2.0} + e^{0.50}x$. The exponential of

# Multicollinearity & the VIF

Section 7.10

# Multicollinearity: When Regressors Move Together

*Section 7.10*

"The effect of $x_j$ holding the others fixed" quietly assumes the data contain variation in $x_j$ *while the others are fixed*. When two regressors move together — distance and travel time, model size and training cost, price and plan tier — the data contain almost no such variation.

# Multicollinearity: When Regressors Move Together

*Section 7.10*

"The effect of $x_j$ holding the others fixed" quietly assumes the data contain variation in $x_j$ *while the others are fixed*. When two regressors move together — distance and travel time, model size and training cost, price and plan tier — the data contain almost no such variation.

**What multicollinearity does NOT break:** the fitted surface as a whole, and predictions inside the data region.

# Multicollinearity: When Regressors Move Together

*Section 7.10*

"The effect of $x_j$ holding the others fixed" quietly assumes the data contain variation in $x_j$ *while the others are fixed*. When two regressors move together — distance and travel time, model size and training cost, price and plan tier — the data contain almost no such variation.

**What multicollinearity does NOT break:** the fitted surface as a whole, and predictions inside the data region.

**What it destroys — the individual coefficients:**

- small changes in the data swing $\hat{\beta}_j$ wildly
- standard errors balloon; signs can contradict engineering sense
- the overall $F$ can be highly significant while *every* individual $t$ is insignificant

# Multicollinearity: When Regressors Move Together

*Section 7.10*

"The effect of $x_j$ holding the others fixed" quietly assumes the data contain variation in $x_j$ *while the others are fixed*. When two regressors move together — distance and travel time, model size and training cost, price and plan tier — the data contain almost no such variation.

**What multicollinearity does NOT break:** the fitted surface as a whole, and predictions inside the data region.

**What it destroys — the individual coefficients:**

- small changes in the data swing $\hat{\beta}_j$ wildly
- standard errors balloon; signs can contradict engineering sense
- the overall $F$ can be highly significant while *every* individual $t$ is insignificant

**Where the instability lives:** $\mathrm{Var}(\hat{\beta}_j) = \sigma^2 C_{jj}$ (7.7). Nearly collinear columns

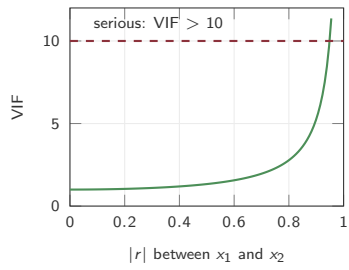
# The Variance Inflation Factor

*Figure 7.3 in the textbook*

Regress $x_j$ on all the *other* regressors; call the resulting coefficient of determination $R_j^2$. Then:

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \qquad (7.17)$$

$\text{VIF}_j$ is exactly the factor by which $\text{Var}(\hat{\beta}_j)$ is inflated relative to a design where $x_j$ is unrelated to the rest.



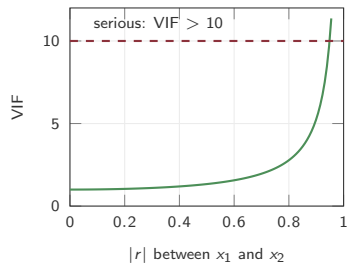
# The Variance Inflation Factor

*Figure 7.3 in the textbook*

Regress $x_j$ on all the *other* regressors; call the resulting coefficient of determination $R_j^2$. Then:

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \qquad (7.17)$$

$\text{VIF}_j$ is exactly the factor by which $\text{Var}(\hat{\beta}_j)$ is inflated relative to a design where $x_j$ is unrelated to the rest.



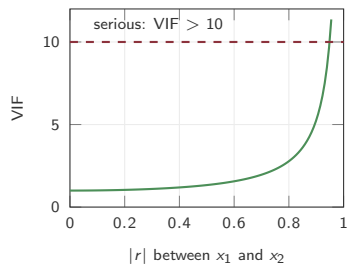

**Benchmarks:** orthogonal regressor $\Rightarrow R_j^2 = 0$, $\text{VIF} = 1$ (no inflation). As $R_j^2 \to 1$, VIF explodes.

# The Variance Inflation Factor

*Figure 7.3 in the textbook*

Regress $x_j$ on all the *other* regressors; call the resulting coefficient of determination $R_j^2$. Then:

$$\text{VIF}_j = \frac{1}{1 - R_j^2} \qquad (7.17)$$




$\text{VIF}_j$ is exactly the factor by which $\text{Var}(\hat{\beta}_j)$ is inflated relative to a design where $x_j$ is unrelated to the rest.

**Benchmarks:** orthogonal regressor $\Rightarrow R_j^2 = 0$, $\text{VIF} = 1$ (no inflation). As $R_j^2 \to 1$, VIF explodes.

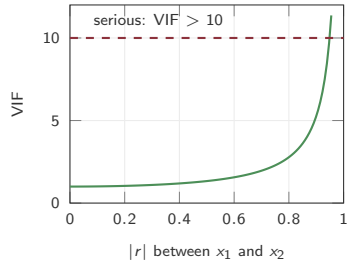
**Working rules:** $\text{VIF} > 5$ deserves attention; $\text{VIF} > 10$ signals serious multicollinearity.

# The Variance Inflation Factor

*Figure 7.3 in the textbook*

Regress $x_j$ on all the *other* regressors; call the resulting coefficient of determination $R_j^2$. Then:

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2} \qquad (7.17)$$




$\mathrm{VIF}_j$ is exactly the factor by which $\mathrm{Var}(\hat{\beta}_j)$ is inflated relative to a design where $x_j$ is unrelated to the rest.

**Benchmarks:** orthogonal regressor $\Rightarrow R_j^2 = 0$, $\mathrm{VIF} = 1$ (no inflation). As $R_j^2 \to 1$, VIF explodes.

**Working rules:** $\mathrm{VIF} > 5$ deserves attention; $\mathrm{VIF} > 10$ signals serious multicollinearity.

For *two* regressors, $R_j^2$ is just the squared sample correlation between them — quick

## Example: Correlated Capability Metrics

*AI safety evaluation suite, $k = 2$*

---

**Setup:** A risk-score regression uses two capability metrics of the evaluated models: $x_1$ = parameter count and $x_2$ = training compute. Their sample correlation is $r = 0.95$.

## Example: Correlated Capability Metrics

*AI safety evaluation suite, $k = 2$*

---

**Setup:** A risk-score regression uses two capability metrics of the evaluated models: $x_1$ = parameter count and $x_2$ = training compute. Their sample correlation is $r = 0.95$.

**Step 1:** $R_1^2 = r^2 = 0.95^2 = 0.9025$.

## Example: Correlated Capability Metrics

*AI safety evaluation suite, $k = 2$*

---

**Setup:** A risk-score regression uses two capability metrics of the evaluated models: $x_1$ = parameter count and $x_2$ = training compute. Their sample correlation is $r = 0.95$.

**Step 1:** $R_1^2 = r^2 = 0.95^2 = 0.9025$.

**Step 2:** By (7.17):

$$\text{VIF} = \frac{1}{1 - 0.9025} = \frac{1}{0.0975} = 10.26$$

## Example: Correlated Capability Metrics

*AI safety evaluation suite, $k = 2$*

---

**Setup:** A risk-score regression uses two capability metrics of the evaluated models: $x_1$ = parameter count and $x_2$ = training compute. Their sample correlation is $r = 0.95$.

**Step 1:** $R_1^2 = r^2 = 0.95^2 = 0.9025$.

**Step 2:** By (7.17):

$$\text{VIF} = \frac{1}{1 - 0.9025} = \frac{1}{0.0975} = 10.26$$

**Step 3: Interpret.** VIF $> 10$: serious. Each coefficient's variance is about ten times what a clean design would give; each standard error is inflated by $\sqrt{10.26} = 3.203$. Effects that would be clearly significant become statistically invisible.

## Example: Correlated Capability Metrics

*AI safety evaluation suite, $k = 2$*

---

**Setup:** A risk-score regression uses two capability metrics of the evaluated models: $x_1$ = parameter count and $x_2$ = training compute. Their sample correlation is $r = 0.95$.

**Step 1:** $R_1^2 = r^2 = 0.95^2 = 0.9025$.

**Step 2:** By (7.17):

$$\text{VIF} = \frac{1}{1 - 0.9025} = \frac{1}{0.0975} = 10.26$$

**Step 3: Interpret.** VIF $> 10$: serious. Each coefficient's variance is about ten times what a clean design would give; each standard error is inflated by $\sqrt{10.26} = 3.203$. Effects that would be clearly significant become statistically invisible.

**Safety-report consequence:** Ranking "which capability drives the risk score" from

## Practice + What to Do About Multicollinearity

**Practice:** A dwell-time model uses $x_1$ = route distance and $x_2$ = scheduled drive time, with $r = 0.98$. Compute the VIF and classify it.

## Practice + What to Do About Multicollinearity

**Practice:** A dwell-time model uses $x_1$ = route distance and $x_2$ = scheduled drive time, with $r = 0.98$. Compute the VIF and classify it.

**Step 1:** $R_1^2 = 0.98^2 = 0.9604$.

## Practice + What to Do About Multicollinearity

**Practice:** A dwell-time model uses $x_1$ = route distance and $x_2$ = scheduled drive time, with $r = 0.98$. Compute the VIF and classify it.

**Step 1:** $R_1^2 = 0.98^2 = 0.9604$.

**Step 2:** $\text{VIF} = 1/(1 - 0.9604) = 1/0.0396 = 25.25 \;\Rightarrow$ far beyond 10: **serious**.

## Practice + What to Do About Multicollinearity

**Practice:** A dwell-time model uses $x_1$ = route distance and $x_2$ = scheduled drive time, with $r = 0.98$. Compute the VIF and classify it.

**Step 1:** $R_1^2 = 0.98^2 = 0.9604$.
**Step 2:** $\text{VIF} = 1/(1 - 0.9604) = 1/0.0396 = 25.25 \;\Rightarrow$ far beyond 10: **serious**.

**Remedies, in rough order of preference:**

1. **Design:** collect data where the regressors vary independently — a designed experiment (like the coded $\pm 1$ evaluation runs) has VIF $= 1$ by construction
2. **Drop** one of the redundant regressors
3. **Combine** them into one meaningful index (e.g., average speed $=$ distance/time)
4. **Accept** the entanglement: use the model for prediction only, and refuse to interpret individual coefficients

# Building a Model Step by Step

Section 7.13

## Building and Vetting an MLR Model

*Procedure 7.5*

1. **Choose candidates from knowledge, not from the data** — regressors the physics, process, or business logic says should matter, including indicators, interactions, polynomials.

## Building and Vetting an MLR Model

*Procedure 7.5*

1. **Choose candidates from knowledge, not from the data** — regressors the physics, process, or business logic says should matter, including indicators, interactions, polynomials.
2. **Fit the full candidate model** (7.4) and run the overall $F$-test (Procedure 7.2). Not significant? Stop and rethink the candidate list.

## Building and Vetting an MLR Model

*Procedure 7.5*

1. **Choose candidates from knowledge, not from the data** — regressors the physics, process, or business logic says should matter, including indicators, interactions, polynomials.
2. **Fit the full candidate model** (7.4) and run the overall $F$-test (Procedure 7.2). Not significant? Stop and rethink the candidate list.
3. **Check for multicollinearity**: compute VIFs (7.17); resolve any VIF $> 10$ before interpreting coefficients.

## Building and Vetting an MLR Model

*Procedure 7.5*

1. **Choose candidates from knowledge, not from the data** — regressors the physics, process, or business logic says should matter, including indicators, interactions, polynomials.
2. **Fit the full candidate model** (7.4) and run the overall $F$-test (Procedure 7.2). Not significant? Stop and rethink the candidate list.
3. **Check for multicollinearity**: compute VIFs (7.17); resolve any VIF $> 10$ before interpreting coefficients.
4. **Assess terms one at a time**: $t$-tests (7.12); doubtful *groups* via the partial $F$-test. Remove non-contributors **one at a time, refitting after each removal**.

## Building and Vetting an MLR Model

*Procedure 7.5*

1. **Choose candidates from knowledge, not from the data** — regressors the physics, process, or business logic says should matter, including indicators, interactions, polynomials.
2. **Fit the full candidate model** (7.4) and run the overall $F$-test (Procedure 7.2). Not significant? Stop and rethink the candidate list.
3. **Check for multicollinearity**: compute VIFs (7.17); resolve any VIF $> 10$ before interpreting coefficients.
4. **Assess terms one at a time**: $t$-tests (7.12); doubtful *groups* via the partial $F$-test. Remove non-contributors **one at a time, refitting after each removal**.
5. **Compare competing models with $R^2_{\text{adj}}$** (7.11), never with plain $R^2$.

## Building and Vetting an MLR Model

*Procedure 7.5*

1. **Choose candidates from knowledge, not from the data** — regressors the physics, process, or business logic says should matter, including indicators, interactions, polynomials.
2. **Fit the full candidate model** (7.4) and run the overall $F$-test (Procedure 7.2). Not significant? Stop and rethink the candidate list.
3. **Check for multicollinearity**: compute VIFs (7.17); resolve any VIF $> 10$ before interpreting coefficients.
4. **Assess terms one at a time**: $t$-tests (7.12); doubtful *groups* via the partial $F$-test. Remove non-contributors **one at a time, refitting after each removal**.
5. **Compare competing models with $R^2_{\text{adj}}$** (7.11), never with plain $R^2$.
6. **Check the residuals** of the surviving model; patterns send you back to step

## Building and Vetting an MLR Model

*Procedure 7.5*

1. **Choose candidates from knowledge, not from the data** — regressors the physics, process, or business logic says should matter, including indicators, interactions, polynomials.
2. **Fit the full candidate model** (7.4) and run the overall $F$-test (Procedure 7.2). Not significant? Stop and rethink the candidate list.
3. **Check for multicollinearity**: compute VIFs (7.17); resolve any VIF $> 10$ before interpreting coefficients.
4. **Assess terms one at a time**: $t$-tests (7.12); doubtful *groups* via the partial $F$-test. Remove non-contributors **one at a time, refitting after each removal**.
5. **Compare competing models with $R^2_{\text{adj}}$** (7.11), never with plain $R^2$.
6. **Check the residuals** of the surviving model; patterns send you back to step

## Why Remove Regressors One at a Time?

The $t$-tests are **marginal**: removing one regressor changes every other regressor's estimate, standard error, and $t$-value.

## Why Remove Regressors One at a Time?

The $t$-tests are **marginal**: removing one regressor changes every other regressor's estimate, standard error, and $t$-value.

**Deleting all insignificant regressors in one sweep is wrong.** Two regressors can be individually insignificant only because each is standing in the other's shadow (collinearity again). Removing both at once throws away real signal.

## Why Remove Regressors One at a Time?

The $t$-tests are **marginal**: removing one regressor changes every other regressor's estimate, standard error, and $t$-value.

**Deleting all insignificant regressors in one sweep is wrong.** Two regressors can be individually insignificant only because each is standing in the other's shadow (collinearity again). Removing both at once throws away real signal.

**The safe loop:** remove the weakest term, refit, look again. Repeat until every remaining term earns its place.

## Example: Does Dropping Tickets Improve the Churn Model?

*Steps 4–5 of Procedure 7.5*

---

**Full model** ($k = 3$: price, onboarding, tickets; $n = 30$): $R^2 = 0.7420$.

$$R^2_{\text{adj}} = 1 - (1 - 0.7420)\,\frac{29}{26} = 1 - 0.2580 \times 1.1154 = 0.7122$$

## Example: Does Dropping Tickets Improve the Churn Model?

*Steps 4–5 of Procedure 7.5*

---

**Full model** ($k = 3$: price, onboarding, tickets; $n = 30$): $R^2 = 0.7420$.

$$R^2_{\text{adj}} = 1 - (1 - 0.7420)\,\frac{29}{26} = 1 - 0.2580 \times 1.1154 = 0.7122$$

**Step 4:** The tickets coefficient was the clear non-contributor ($T_0 = 0.50$). Remove it and refit.

## Example: Does Dropping Tickets Improve the Churn Model?

*Steps 4–5 of Procedure 7.5*

---

**Full model** ($k = 3$: price, onboarding, tickets; $n = 30$): $R^2 = 0.7420$.

$$R^2_{\text{adj}} = 1 - (1 - 0.7420)\,\frac{29}{26} = 1 - 0.2580 \times 1.1154 = 0.7122$$

**Step 4:** The tickets coefficient was the clear non-contributor ($T_0 = 0.50$). Remove it and refit.

**Reduced model** ($k = 2$): $R^2 = 0.7395$ — lower, as it must be. But:

$$R^2_{\text{adj}} = 1 - (1 - 0.7395)\,\frac{29}{27} = 1 - 0.2605 \times 1.0741 = 0.7202$$

which is **higher**.

## Example: Does Dropping Tickets Improve the Churn Model?

*Steps 4–5 of Procedure 7.5*

---

**Full model** ($k = 3$: price, onboarding, tickets; $n = 30$): $R^2 = 0.7420$.

$$R^2_{\text{adj}} = 1 - (1 - 0.7420)\,\frac{29}{26} = 1 - 0.2580 \times 1.1154 = 0.7122$$

**Step 4:** The tickets coefficient was the clear non-contributor ($T_0 = 0.50$). Remove it and refit.

**Reduced model** ($k = 2$): $R^2 = 0.7395$ — lower, as it must be. But:

$$R^2_{\text{adj}} = 1 - (1 - 0.7395)\,\frac{29}{27} = 1 - 0.2605 \times 1.0741 = 0.7202$$

which is **higher**.

**Step 5 verdict:** The smaller model explains essentially the same variability while

## Common Mistakes to Avoid

- **Wrong degrees of freedom.** In MLR: df $= n - k - 1$, **not** $n - 2$ (that is only for SLR with $k = 1$).

## Common Mistakes to Avoid

- **Wrong degrees of freedom.** In MLR: df $= n - k - 1$, **not** $n - 2$ (that is only for SLR with $k = 1$).
- **Mixing up overall $F$ and individual $t$.** "Is the model significant?" $\neq$ "Is $x_j$ significant?" With collinear regressors, the $F$ can be significant while every $t$ is not.

## Common Mistakes to Avoid

- **Wrong degrees of freedom.** In MLR: df $= n - k - 1$, **not** $n - 2$ (that is only for SLR with $k = 1$).
- **Mixing up overall $F$ and individual $t$.** "Is the model significant?" $\neq$ "Is $x_j$ significant?" With collinear regressors, the $F$ can be significant while every $t$ is not.
- **Wrong subscript on $\beta$.** If the question asks about "number of stops" and that is $x_2$, the test is on $\beta_2$, not $\beta_1$.

## Common Mistakes to Avoid

- **Wrong degrees of freedom.** In MLR: df $= n - k - 1$, **not** $n - 2$ (that is only for SLR with $k = 1$).
- **Mixing up overall $F$ and individual $t$.** "Is the model significant?" $\neq$ "Is $x_j$ significant?" With collinear regressors, the $F$ can be significant while every $t$ is not.
- **Wrong subscript on $\beta$.** If the question asks about "number of stops" and that is $x_2$, the test is on $\beta_2$, not $\beta_1$.
- **Using $m$ indicators for $m$ categories.** Use $m - 1$; the extra one makes $\mathbf{X}'\mathbf{X}$ singular.

## Common Mistakes to Avoid

- **Wrong degrees of freedom.** In MLR: df $= n - k - 1$, **not** $n - 2$ (that is only for SLR with $k = 1$).
- **Mixing up overall $F$ and individual $t$.** "Is the model significant?" $\neq$ "Is $x_j$ significant?" With collinear regressors, the $F$ can be significant while every $t$ is not.
- **Wrong subscript on $\beta$.** If the question asks about "number of stops" and that is $x_2$, the test is on $\beta_2$, not $\beta_1$.
- **Using $m$ indicators for $m$ categories.** Use $m - 1$; the extra one makes $\mathbf{X}'\mathbf{X}$ singular.
- **Comparing models with $R^2$.** It never decreases when you add regressors; use $R^2_{\text{adj}}$ (7.11).

## Common Mistakes to Avoid

- **Wrong degrees of freedom.** In MLR: df $= n - k - 1$, **not** $n - 2$ (that is only for SLR with $k = 1$).
- **Mixing up overall $F$ and individual $t$.** "Is the model significant?" $\neq$ "Is $x_j$ significant?" With collinear regressors, the $F$ can be significant while every $t$ is not.
- **Wrong subscript on $\beta$.** If the question asks about "number of stops" and that is $x_2$, the test is on $\beta_2$, not $\beta_1$.
- **Using $m$ indicators for $m$ categories.** Use $m - 1$; the extra one makes $\mathbf{X}'\mathbf{X}$ singular.
- **Comparing models with $R^2$.** It never decreases when you add regressors; use $R^2_{\text{adj}}$ (7.11).
- **Interpreting coefficients under high VIF.** Check (7.17) first; VIF $> 10$

## Lecture 21 Summary

- **Individual $t$-test** (7.12): $T_0 = (\hat{\beta}_j - \beta_{j,0})/\text{se}(\hat{\beta}_j) \sim t_{n-p}$. A marginal test: "given the other regressors are in the model."

## Lecture 21 Summary

- **Individual $t$-test** (7.12): $T_0 = (\hat{\beta}_j - \beta_{j,0})/\text{se}(\hat{\beta}_j) \sim t_{n-p}$. A marginal test: "given the other regressors are in the model."
- **CI for** $\beta_j$ (7.13): $\hat{\beta}_j \pm t_{\alpha/2,\, n-p}\, \text{se}(\hat{\beta}_j)$. Use the duality to test claims: reject any $\beta_{j,0}$ outside the interval.

## Lecture 21 Summary

- **Individual $t$-test** (7.12): $T_0 = (\hat{\beta}_j - \beta_{j,0})/\text{se}(\hat{\beta}_j) \sim t_{n-p}$. A marginal test: "given the other regressors are in the model."
- **CI for $\beta_j$** (7.13): $\hat{\beta}_j \pm t_{\alpha/2,\, n-p}\, \text{se}(\hat{\beta}_j)$. Use the duality to test claims: reject any $\beta_{j,0}$ outside the interval.
- **Indicator variables** (7.18): $z \in \{0, 1\}$ gives two parallel lines; $\beta_2$ is the gap. $m$ categories need $m - 1$ indicators.

## Lecture 21 Summary

- **Individual $t$-test** (7.12): $T_0 = (\hat{\beta}_j - \beta_{j,0})/\mathrm{se}(\hat{\beta}_j) \sim t_{n-p}$. A marginal test: "given the other regressors are in the model."
- **CI for** $\beta_j$ (7.13): $\hat{\beta}_j \pm t_{\alpha/2,\,n-p}\,\mathrm{se}(\hat{\beta}_j)$. Use the duality to test claims: reject any $\beta_{j,0}$ outside the interval.
- **Indicator variables** (7.18): $z \in \{0, 1\}$ gives two parallel lines; $\beta_2$ is the gap. $m$ categories need $m - 1$ indicators.
- **Interactions** (7.19): the product $x_1 x_2$ lets the slope of $x_1$ depend on $x_2$: slope $= \beta_1 + \beta_{12} x_2$. Polynomials (7.20) fit curves, still linear in the $\beta$'s.

## Lecture 21 Summary

- **Individual $t$-test** (7.12): $T_0 = (\hat{\beta}_j - \beta_{j,0})/\text{se}(\hat{\beta}_j) \sim t_{n-p}$. A marginal test: "given the other regressors are in the model."
- **CI for** $\beta_j$ (7.13): $\hat{\beta}_j \pm t_{\alpha/2,\, n-p}\, \text{se}(\hat{\beta}_j)$. Use the duality to test claims: reject any $\beta_{j,0}$ outside the interval.
- **Indicator variables** (7.18): $z \in \{0, 1\}$ gives two parallel lines; $\beta_2$ is the gap. $m$ categories need $m - 1$ indicators.
- **Interactions** (7.19): the product $x_1 x_2$ lets the slope of $x_1$ depend on $x_2$: slope $= \beta_1 + \beta_{12} x_2$. Polynomials (7.20) fit curves, still linear in the $\beta$'s.
- **Multicollinearity**: $\text{VIF}_j = 1/(1 - R_j^2)$ (7.17); $> 5$ attention, $> 10$ serious. It inflates $\text{se}(\hat{\beta}_j)$, not the predictions.

## Lecture 21 Summary

- **Individual $t$-test** (7.12): $T_0 = (\hat{\beta}_j - \beta_{j,0})/\text{se}(\hat{\beta}_j) \sim t_{n-p}$. A marginal test: "given the other regressors are in the model."
- **CI for** $\beta_j$ (7.13): $\hat{\beta}_j \pm t_{\alpha/2,\, n-p}\, \text{se}(\hat{\beta}_j)$. Use the duality to test claims: reject any $\beta_{j,0}$ outside the interval.
- **Indicator variables** (7.18): $z \in \{0, 1\}$ gives two parallel lines; $\beta_2$ is the gap. $m$ categories need $m - 1$ indicators.
- **Interactions** (7.19): the product $x_1 x_2$ lets the slope of $x_1$ depend on $x_2$: slope $= \beta_1 + \beta_{12} x_2$. Polynomials (7.20) fit curves, still linear in the $\beta$'s.
- **Multicollinearity**: $\text{VIF}_j = 1/(1 - R_j^2)$ (7.17); $> 5$ attention, $> 10$ serious. It inflates $\text{se}(\hat{\beta}_j)$, not the predictions.
- **Model building** (Procedure 7.5): candidates from knowledge $\rightarrow$ overall $F$ $\rightarrow$ VIFs $\rightarrow$ $t$-tests one at a time $\rightarrow$ compare with $R^2_{\text{adj}}$ $\rightarrow$ residuals $\rightarrow$ report.

## Announcements

- **Major Exam 2 is coming up!** Chapters 5, 6, and 7.

## Announcements

- **Major Exam 2 is coming up!** Chapters 5, 6, and 7.
- **Study strategy:**
  - Review Lectures 15–21 slides and worked examples
  - Redo the practice handouts by hand, and the homework problems
  - Familiarize yourself with the formula sheet
  - For each problem, identify the question type first, then pick the formula

## Announcements

- **Major Exam 2 is coming up!** Chapters 5, 6, and 7.
- **Study strategy:**
  - Review Lectures 15–21 slides and worked examples
  - Redo the practice handouts by hand, and the homework problems
  - Familiarize yourself with the formula sheet
  - For each problem, identify the question type first, then pick the formula
- **Optional practice session**: Sunday, 8–10 PM online (Teams link will be provided)

## Announcements

- **Major Exam 2 is coming up!** Chapters 5, 6, and 7.
- **Study strategy:**
  - Review Lectures 15–21 slides and worked examples
  - Redo the practice handouts by hand, and the homework problems
  - Familiarize yourself with the formula sheet
  - For each problem, identify the question type first, then pick the formula
- **Optional practice session**: Sunday, 8–10 PM online (Teams link will be provided)
- **Quiz on Chapter 7** — submit before class.