# ISE 315: Engineering Statistics

*Lecture 20: Multiple Linear Regression (Part 1)*

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 20

Multiple Linear Regression — Part 1 (Chapter 7)

## Lecture 20 Outline

- Quick review: SLR key ideas (from Chapter 6)

## Lecture 20 Outline

- Quick review: SLR key ideas (from Chapter 6)
- Why multiple regressors? (§7.1)

## Lecture 20 Outline

- Quick review: SLR key ideas (from Chapter 6)
- Why multiple regressors? (§7.1)
- The matrix approach and least squares (§7.2)

## Lecture 20 Outline

- Quick review: SLR key ideas (from Chapter 6)
- Why multiple regressors? (§7.1)
- The matrix approach and least squares (§7.2)
- Estimating $\sigma^2$ and standard errors (§7.3)

## Lecture 20 Outline

- Quick review: SLR key ideas (from Chapter 6)
- Why multiple regressors? (§7.1)
- The matrix approach and least squares (§7.2)
- Estimating $\sigma^2$ and standard errors (§7.3)
- ANOVA for MLR and the $F$-test for significance (§7.4)

## Lecture 20 Outline

- Quick review: SLR key ideas (from Chapter 6)
- Why multiple regressors? (§7.1)
- The matrix approach and least squares (§7.2)
- Estimating $\sigma^2$ and standard errors (§7.3)
- ANOVA for MLR and the $F$-test for significance (§7.4)
- $R^2$ and $R^2_{\text{adj}}$ (§7.5)

## Lecture 20 Outline

- Quick review: SLR key ideas (from Chapter 6)
- Why multiple regressors? (§7.1)
- The matrix approach and least squares (§7.2)
- Estimating $\sigma^2$ and standard errors (§7.3)
- ANOVA for MLR and the $F$-test for significance (§7.4)
- $R^2$ and $R^2_{\text{adj}}$ (§7.5)
- Reading software output (§7.6)

# Quick Review

From SLR (Chapter 6) to MLR (Chapter 7)

## Chapter 6 Recap: SLR in One Slide

**Model:** $y = \beta_0 + \beta_1 x + \varepsilon$ (one predictor, (6.2))

## Chapter 6 Recap: SLR in One Slide

**Model:** $y = \beta_0 + \beta_1 x + \varepsilon$ (one predictor, (6.2))

**Estimation:** $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9), $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (6.10)

## Chapter 6 Recap: SLR in One Slide

**Model:** $y = \beta_0 + \beta_1 x + \varepsilon$ (one predictor, (6.2))

**Estimation:** $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9), $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (6.10)

**ANOVA:** $SS_T = SS_R + SS_E$ (6.19), $R^2 = SS_R/SS_T$ (6.21)

## Chapter 6 Recap: SLR in One Slide

**Model:** $y = \beta_0 + \beta_1 x + \varepsilon$ (one predictor, (6.2))

**Estimation:** $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9), $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (6.10)

**ANOVA:** $SS_T = SS_R + SS_E$ (6.19), $R^2 = SS_R/SS_T$ (6.21)

**Inference:** $T_0 = \hat{\beta}_1/\text{se}(\hat{\beta}_1) \sim t_{n-2}$ (6.18)

## Chapter 6 Recap: SLR in One Slide

**Model:** $y = \beta_0 + \beta_1 x + \varepsilon$ (one predictor, (6.2))

**Estimation:** $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9), $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (6.10)

**ANOVA:** $SS_T = SS_R + SS_E$ (6.19), $R^2 = SS_R/SS_T$ (6.21)

**Inference:** $T_0 = \hat{\beta}_1/\text{se}(\hat{\beta}_1) \sim t_{n-2}$ (6.18)

**The big limitation:** In practice, the response $y$ is rarely explained by a single variable. The delivery time of a truck route depends on distance, number of stops, load, and more. We need a model that can handle **multiple predictors**.

# Why Multiple Regressors?

Section 7.1

## Motivation: One Predictor Is Often Not Enough

**The benchmark score of a perception model** in an AI safety evaluation depends on many factors:

- Network size ($x_1$)
- Training data quality ($x_2$)
- Hours of adversarial fine-tuning ($x_3$)

# Motivation: One Predictor Is Often Not Enough

**The benchmark score of a perception model** in an AI safety evaluation depends on many factors:

- Network size ($x_1$)
- Training data quality ($x_2$)
- Hours of adversarial fine-tuning ($x_3$)

Using just one predictor (say, network size) might give a low $R^2$. Adding more predictors can explain more of the variability in $y$.

## Motivation: One Predictor Is Often Not Enough

**The benchmark score of a perception model** in an AI safety evaluation depends on many factors:

- Network size ($x_1$)
- Training data quality ($x_2$)
- Hours of adversarial fine-tuning ($x_3$)

Using just one predictor (say, network size) might give a low $R^2$. Adding more predictors can explain more of the variability in $y$.

**The multiple linear regression (MLR) model** extends SLR to $k$ predictors:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{7.1}$$

## Motivation: One Predictor Is Often Not Enough

**The benchmark score of a perception model** in an AI safety evaluation depends on many factors:

- Network size ($x_1$)
- Training data quality ($x_2$)
- Hours of adversarial fine-tuning ($x_3$)

Using just one predictor (say, network size) might give a low $R^2$. Adding more predictors can explain more of the variability in $y$.

**The multiple linear regression (MLR) model** extends SLR to $k$ predictors:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{7.1}$$

The model is still **linear in the parameters** $\beta_j$. This is what "linear" refers to — the relationship between $y$ and the $\beta$'s, not necessarily between $y$ and the $x$'s.

# The MLR Model: Assumptions

*Section 7.1*

For $n$ observations and $k$ regressors:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \qquad i = 1, 2, \ldots, n$$

# The MLR Model: Assumptions

*Section 7.1*

For $n$ observations and $k$ regressors:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \qquad i = 1, 2, \ldots, n$$

**Assumptions** (same spirit as SLR):

- $E(\varepsilon_i) = 0$ for all $i$
- $\mathrm{Var}(\varepsilon_i) = \sigma^2$ for all $i$ (constant variance)
- The errors $\varepsilon_i$ are uncorrelated
- (For inference) $\varepsilon_i \sim N(0, \sigma^2)$

# The MLR Model: Assumptions

*Section 7.1*

For $n$ observations and $k$ regressors:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \qquad i = 1, 2, \ldots, n$$

**Assumptions** (same spirit as SLR):

- $E(\varepsilon_i) = 0$ for all $i$
- $\mathrm{Var}(\varepsilon_i) = \sigma^2$ for all $i$ (constant variance)
- The errors $\varepsilon_i$ are uncorrelated
- (For inference) $\varepsilon_i \sim N(0, \sigma^2)$

**Key difference from SLR:** We now have $p = k + 1$ parameters to estimate $(\beta_0, \beta_1, \ldots, \beta_k)$, and we need $n > p$ observations.

# The MLR Model: Assumptions

*Section 7.1*

For $n$ observations and $k$ regressors:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i, \qquad i = 1, 2, \ldots, n$$

**Assumptions** (same spirit as SLR):

- $E(\varepsilon_i) = 0$ for all $i$
- $\mathrm{Var}(\varepsilon_i) = \sigma^2$ for all $i$ (constant variance)
- The errors $\varepsilon_i$ are uncorrelated
- (For inference) $\varepsilon_i \sim N(0, \sigma^2)$

**Key difference from SLR:** We now have $p = k + 1$ parameters to estimate $(\beta_0, \beta_1, \ldots, \beta_k)$, and we need $n > p$ observations.

**Interpretation of $\beta_j$:** the expected change in $y$ per unit change in $x_j$, *holding all other $x$'s fixed*.

# The Matrix Approach

Section 7.2

# Matrix Formulation of MLR

*Section 7.2*

Write all $n$ equations simultaneously as:

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times p}{\mathbf{X}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \tag{7.2}$$

Textbook §7.2: Chapter 7 ((7.2) onward)

## Matrix Formulation of MLR

*Section 7.2*

Write all $n$ equations simultaneously as:

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times p}{\mathbf{X}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \tag{7.2}$$

where:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

Textbook §7.2: Chapter 7 ((7.2) onward)

## Matrix Formulation of MLR

*Section 7.2*

Write all $n$ equations simultaneously as:

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times p}{\mathbf{X}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \tag{7.2}$$

where:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

The first column of $\mathbf{X}$ is all 1's — it corresponds to the intercept $\beta_0$.

Textbook §7.2: Chapter 7 ((7.2) onward)

# Least Squares Estimation in Matrix Form

*Section 7.2*

We minimize sum squared residuals: $S(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2$.

# Least Squares Estimation in Matrix Form

*Section 7.2*

We minimize sum squared residuals: $S(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2$.

In matrix form: minimize $S = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

# Least Squares Estimation in Matrix Form

*Section 7.2*

We minimize sum squared residuals: $S(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2$.

In matrix form: minimize $S = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

Taking the derivative and setting it to zero gives the **normal equations**:

$$\mathbf{X}'\mathbf{X}\,\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{7.3}$$

# Least Squares Estimation in Matrix Form

*Section 7.2*

We minimize sum squared residuals: $S(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2$.

In matrix form: minimize $S = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

Taking the derivative and setting it to zero gives the **normal equations**:

$$\mathbf{X}'\mathbf{X}\,\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{7.3}$$

If $\mathbf{X}'\mathbf{X}$ is invertible, the solution is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{7.4}$$

# Least Squares Estimation in Matrix Form

*Section 7.2*

We minimize sum squared residuals: $S(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2$.

In matrix form: minimize $S = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$.

Taking the derivative and setting it to zero gives the **normal equations**:

$$\mathbf{X}'\mathbf{X}\,\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \tag{7.3}$$

If $\mathbf{X}'\mathbf{X}$ is invertible, the solution is:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{7.4}$$

This is the **one formula** that replaces the separate $\hat{\beta}_1 = S_{xy}/S_{xx}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ formulas from SLR. The fitting recipe is Procedure 7.1. On the exam, $(\mathbf{X}'\mathbf{X})^{-1}$ will be given.

# Worked Example: MLR with Two Regressors

*AI safety evaluation (Table 7.1 in the textbook)*

**Problem.** An AI safety team evaluates $n = 8$ candidate perception models on a fixed benchmark. Coded regressors: $x_1 = +1$ for a large network, $-1$ for small; $x_2 = +1$ for curated training data, $-1$ for uncurated. Response $y =$ benchmark score (points out of 100). Fit $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$.

| Run $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $x_{i1}$ (size) | $-1$ | $+1$ | $-1$ | $+1$ | $-1$ | $+1$ | $-1$ | $+1$ |
| $x_{i2}$ (data quality) | $-1$ | $-1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $+1$ |
| $y_i$ (score) | 56 | 68 | 74 | 82 | 58 | 66 | 72 | 84 |

## Worked Example: MLR with Two Regressors

*AI safety evaluation (Table 7.1 in the textbook)*

**Problem.** An AI safety team evaluates $n = 8$ candidate perception models on a fixed benchmark. Coded regressors: $x_1 = +1$ for a large network, $-1$ for small; $x_2 = +1$ for curated training data, $-1$ for uncurated. Response $y =$ benchmark score (points out of 100). Fit $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$.

| Run $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $x_{i1}$ (size) | $-1$ | $+1$ | $-1$ | $+1$ | $-1$ | $+1$ | $-1$ | $+1$ |
| $x_{i2}$ (data quality) | $-1$ | $-1$ | $+1$ | $+1$ | $-1$ | $-1$ | $+1$ | $+1$ |
| $y_i$ (score) | 56 | 68 | 74 | 82 | 58 | 66 | 72 | 84 |

We follow Procedure 7.1: build $\mathbf{X}$ and $\mathbf{y}$, form $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{y}$, invert, multiply.

# Worked Example (cont'd): $\mathbf{X}$, $\mathbf{y}$, and $\mathbf{X}'\mathbf{X}$

**Step 1: Build $\mathbf{X}$ ($8 \times 3$) and $\mathbf{y}$.**

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \\ 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 56 \\ 68 \\ 74 \\ 82 \\ 58 \\ 66 \\ 72 \\ 84 \end{bmatrix}$$

# Worked Example (cont'd): $\mathbf{X}$, $\mathbf{y}$, and $\mathbf{X}'\mathbf{X}$

**Step 1: Build $\mathbf{X}$ ($8 \times 3$) and $\mathbf{y}$.**

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \\ 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} 56 \\ 68 \\ 74 \\ 82 \\ 58 \\ 66 \\ 72 \\ 84 \end{bmatrix}$$

**Step 2: Compute $\mathbf{X}'\mathbf{X}$ and invert.** The $(1,1)$ entry is $n = 8$; each off-diagonal is a sum like $\sum x_{i1}$ or $\sum x_{i1}x_{i2}$, and the balanced $\pm 1$ coding makes all of them 0:

$$\begin{bmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{bmatrix} \qquad \begin{bmatrix} 0.125 & 0 & 0 \\ 0 & 0.125 & 0 \\ 0 & 0 & 0.125 \end{bmatrix}$$

# Worked Example (cont'd): Computing $\hat{\boldsymbol{\beta}}$

**Step 3: Compute $\mathbf{X}'\mathbf{y}$** — each entry is a sum of the $y_i$ weighted by one column of $\mathbf{X}$:

$$\begin{aligned}
\textstyle\sum_i y_i &= 56 + 68 + 74 + 82 + 58 + 66 + 72 + 84 = 560 \\
\textstyle\sum_i x_{i1} y_i &= -56 + 68 - 74 + 82 - 58 + 66 - 72 + 84 = 40 \\
\textstyle\sum_i x_{i2} y_i &= -56 - 68 + 74 + 82 - 58 - 66 + 72 + 84 = 64
\end{aligned}$$

so $\mathbf{X}'\mathbf{y} = (560,\ 40,\ 64)'$.

# Worked Example (cont'd): Computing $\hat{\boldsymbol{\beta}}$

**Step 3: Compute $\mathbf{X}'\mathbf{y}$** — each entry is a sum of the $y_i$ weighted by one column of $\mathbf{X}$:

$$\begin{aligned}\textstyle\sum_i y_i &= 56 + 68 + 74 + 82 + 58 + 66 + 72 + 84 = 560\\ \textstyle\sum_i x_{i1} y_i &= -56 + 68 - 74 + 82 - 58 + 66 - 72 + 84 = 40\\ \textstyle\sum_i x_{i2} y_i &= -56 - 68 + 74 + 82 - 58 - 66 + 72 + 84 = 64\end{aligned}$$

so $\mathbf{X}'\mathbf{y} = (560,\ 40,\ 64)'$.

**Step 4: Multiply**, using (7.4): $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0.125 & 0 & 0 \\ 0 & 0.125 & 0 \\ 0 & 0 & 0.125 \end{bmatrix} \begin{bmatrix} 560 \\ 40 \\ 64 \end{bmatrix} = \begin{bmatrix} 0.125 \times 560 \\ 0.125 \times 40 \\ 0.125 \times 64 \end{bmatrix} = \begin{bmatrix} 70 \\ 5 \\ 8 \end{bmatrix}$$

# Worked Example (cont'd): Computing $\hat{\boldsymbol{\beta}}$

**Step 3: Compute $\mathbf{X}'\mathbf{y}$** — each entry is a sum of the $y_i$ weighted by one column of $\mathbf{X}$:

$$\begin{aligned}\textstyle\sum_i y_i &= 56 + 68 + 74 + 82 + 58 + 66 + 72 + 84 = 560\\ \textstyle\sum_i x_{i1} y_i &= -56 + 68 - 74 + 82 - 58 + 66 - 72 + 84 = 40\\ \textstyle\sum_i x_{i2} y_i &= -56 - 68 + 74 + 82 - 58 - 66 + 72 + 84 = 64\end{aligned}$$

so $\mathbf{X}'\mathbf{y} = (560,\ 40,\ 64)'$.

**Step 4: Multiply**, using (7.4): $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$:

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0.125 & 0 & 0 \\ 0 & 0.125 & 0 \\ 0 & 0 & 0.125 \end{bmatrix} \begin{bmatrix} 560 \\ 40 \\ 64 \end{bmatrix} = \begin{bmatrix} 0.125 \times 560 \\ 0.125 \times 40 \\ 0.125 \times 64 \end{bmatrix} = \begin{bmatrix} 70 \\ 5 \\ 8 \end{bmatrix}$$

## Worked Example (cont'd): Interpretation and Residuals

**Interpretation:** Holding data quality fixed, going from a small to a large network ($x_1 = -1$ to $+1$, a two-unit change) raises the predicted score by $2 \times 5 = 10$ points. Holding size fixed, curated data raises it by $2 \times 8 = 16$ points. Data quality moves the score more than model size.

# Worked Example (cont'd): Interpretation and Residuals

**Interpretation:** Holding data quality fixed, going from a small to a large network ($x_1 = -1$ to $+1$, a two-unit change) raises the predicted score by $2 \times 5 = 10$ points. Holding size fixed, curated data raises it by $2 \times 8 = 16$ points. Data quality moves the score more than model size.

**Step 5: Fitted values and residuals:** $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$, $e_i = y_i - \hat{y}_i$.

| Run $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $y_i$ | 56 | 68 | 74 | 82 | 58 | 66 | 72 | 84 |
| $\hat{y}_i = 70 + 5x_1 + 8x_2$ | 57 | 67 | 73 | 83 | 57 | 67 | 73 | 83 |
| $e_i$ | $-1$ | $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $+1$ |

## Worked Example (cont'd): Interpretation and Residuals

**Interpretation:** Holding data quality fixed, going from a small to a large network ($x_1 = -1$ to $+1$, a two-unit change) raises the predicted score by $2 \times 5 = 10$ points. Holding size fixed, curated data raises it by $2 \times 8 = 16$ points. Data quality moves the score more than model size.

**Step 5: Fitted values and residuals:** $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$, $e_i = y_i - \hat{y}_i$.

| Run $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $y_i$ | 56 | 68 | 74 | 82 | 58 | 66 | 72 | 84 |
| $\hat{y}_i = 70 + 5x_1 + 8x_2$ | 57 | 67 | 73 | 83 | 57 | 67 | 73 | 83 |
| $e_i$ | $-1$ | $+1$ | $+1$ | $-1$ | $+1$ | $-1$ | $-1$ | $+1$ |

The residuals sum to zero. This is always true when the model contains an intercept — use it as a free arithmetic check.

# Estimating $\sigma^2$ and Standard Errors

Section 7.3

## Estimating the Error Variance $\sigma^2$

*Section 7.3*

Fitting $p = k + 1$ parameters uses up $p$ degrees of freedom, so:

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} = MS_E, \qquad SS_E = \sum_{i=1}^{n} e_i^2 = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \tag{7.5}$$

The matrix shortcut avoids computing residuals one by one.

# Estimating the Error Variance $\sigma^2$

*Section 7.3*

Fitting $p = k + 1$ parameters uses up $p$ degrees of freedom, so:

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} = MS_E, \qquad SS_E = \sum_{i=1}^{n} e_i^2 = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \tag{7.5}$$

The matrix shortcut avoids computing residuals one by one.

**Evaluation example, Step 1:** $\mathbf{y}'\mathbf{y} = 56^2 + 68^2 + \cdots + 84^2 = 39{,}920.$

# Estimating the Error Variance $\sigma^2$

*Section 7.3*

Fitting $p = k + 1$ parameters uses up $p$ degrees of freedom, so:

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} = MS_E, \qquad SS_E = \sum_{i=1}^{n} e_i^2 = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \tag{7.5}$$

The matrix shortcut avoids computing residuals one by one.

**Evaluation example, Step 1:** $\mathbf{y}'\mathbf{y} = 56^2 + 68^2 + \cdots + 84^2 = 39{,}920.$

**Step 2:** $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = [70 \;\; 5 \;\; 8] \begin{bmatrix} 560 \\ 40 \\ 64 \end{bmatrix} = 39{,}200 + 200 + 512 = 39{,}912.$

# Estimating the Error Variance $\sigma^2$

*Section 7.3*

Fitting $p = k + 1$ parameters uses up $p$ degrees of freedom, so:

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} = MS_E, \qquad SS_E = \sum_{i=1}^{n} e_i^2 = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \tag{7.5}$$

The matrix shortcut avoids computing residuals one by one.

**Evaluation example, Step 1:** $\mathbf{y}'\mathbf{y} = 56^2 + 68^2 + \cdots + 84^2 = 39{,}920.$

**Step 2:** $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = [70 \;\; 5 \;\; 8] \begin{bmatrix} 560 \\ 40 \\ 64 \end{bmatrix} = 39{,}200 + 200 + 512 = 39{,}912.$

**Step 3:** $SS_E = 39{,}920 - 39{,}912 = 8.$ (Check: eight residuals of $\pm 1$ give $\sum e_i^2 = 8$. ✓)

# Estimating the Error Variance $\sigma^2$

*Section 7.3*

Fitting $p = k + 1$ parameters uses up $p$ degrees of freedom, so:

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} = MS_E, \qquad SS_E = \sum_{i=1}^{n} e_i^2 = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} \tag{7.5}$$

The matrix shortcut avoids computing residuals one by one.

**Evaluation example, Step 1:** $\mathbf{y}'\mathbf{y} = 56^2 + 68^2 + \cdots + 84^2 = 39{,}920$.

**Step 2:** $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = [70 \;\; 5 \;\; 8] \begin{bmatrix} 560 \\ 40 \\ 64 \end{bmatrix} = 39{,}200 + 200 + 512 = 39{,}912.$

**Step 3:** $SS_E = 39{,}920 - 39{,}912 = 8.$ (Check: eight residuals of $\pm 1$ give $\sum e_i^2 = 8$. ✓)

**Step 4:** With $n = 8$, $p = 3$: $\hat{\sigma}^2 = \dfrac{8}{8-3} = \dfrac{8}{5} = \boxed{1.600}$

## Standard Errors of Regression Coefficients

*Section 7.3*

Let $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is:

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\mathbf{C} \tag{7.6}$$

# Standard Errors of Regression Coefficients

*Section 7.3*

Let $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is:

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\mathbf{C} \tag{7.6}$$

The **variance** of the $j$-th coefficient is the $j$-th diagonal entry:

$$\mathrm{Var}(\hat{\beta}_j) = \sigma^2 C_{jj}, \qquad \mathrm{se}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}} = \sqrt{MS_E \cdot C_{jj}} \tag{7.7}$$

# Standard Errors of Regression Coefficients

*Section 7.3*

Let $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is:

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\mathbf{C} \tag{7.6}$$

The **variance** of the $j$-th coefficient is the $j$-th diagonal entry:

$$\mathrm{Var}(\hat{\beta}_j) = \sigma^2 C_{jj}, \qquad \mathrm{se}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}} = \sqrt{MS_E \cdot C_{jj}} \tag{7.7}$$

**Evaluation example:** $\mathbf{C} = \mathrm{diag}(0.125,\ 0.125,\ 0.125)$ and $\hat{\sigma}^2 = 1.600$, so all three coefficients share the same standard error:

$$\mathrm{se}(\hat{\beta}_j) = \sqrt{1.600 \times 0.125} = \sqrt{0.2000} = 0.4472$$

# Standard Errors of Regression Coefficients

*Section 7.3*

Let $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$. The variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ is:

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2\mathbf{C} \tag{7.6}$$

The **variance** of the $j$-th coefficient is the $j$-th diagonal entry:

$$\mathrm{Var}(\hat{\beta}_j) = \sigma^2 C_{jj}, \qquad \mathrm{se}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 C_{jj}} = \sqrt{MS_E \cdot C_{jj}} \tag{7.7}$$

**Evaluation example:** $\mathbf{C} = \mathrm{diag}(0.125,\ 0.125,\ 0.125)$ and $\hat{\sigma}^2 = 1.600$, so all three coefficients share the same standard error:

$$\mathrm{se}(\hat{\beta}_j) = \sqrt{1.600 \times 0.125} = \sqrt{0.2000} = 0.4472$$

# ANOVA for MLR

Testing Overall Model Significance (Section 7.4)

# ANOVA Decomposition in MLR

*Section 7.4*

The same decomposition as SLR, with different degrees of freedom:

$$\underbrace{SS_T}_{n-1} = \underbrace{SS_R}_{k} + \underbrace{SS_E}_{n-k-1} \tag{7.8}$$

where $k$ is the number of regressors and $p = k + 1$ is the total number of parameters.

Textbook §7.4: Chapter 7 ((7.8), (7.9))

# ANOVA Decomposition in MLR

*Section 7.4*

The same decomposition as SLR, with different degrees of freedom:

$$\underbrace{SS_T}_{n-1} = \underbrace{SS_R}_{k} + \underbrace{SS_E}_{n-k-1} \tag{7.8}$$

where $k$ is the number of regressors and $p = k + 1$ is the total number of parameters.

**Computing sums of squares in matrix form:**

$$SS_T = \mathbf{y}'\mathbf{y} - n\bar{y}^2, \qquad SS_R = \hat{\beta}'\mathbf{X}'\mathbf{y} - n\bar{y}^2, \qquad SS_E = \mathbf{y}'\mathbf{y} - \hat{\beta}'\mathbf{X}'\mathbf{y}$$

Textbook §7.4: Chapter 7 ((7.8), (7.9))

# ANOVA Decomposition in MLR

*Section 7.4*

The same decomposition as SLR, with different degrees of freedom:

$$\underbrace{SS_T}_{n-1} = \underbrace{SS_R}_{k} + \underbrace{SS_E}_{n-k-1} \tag{7.8}$$

where $k$ is the number of regressors and $p = k + 1$ is the total number of parameters.

**Computing sums of squares in matrix form:**

$$SS_T = \mathbf{y}'\mathbf{y} - n\bar{y}^2, \qquad SS_R = \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{y}^2, \qquad SS_E = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y}$$

**Mean squares:** $MS_R = SS_R/k$ and $MS_E = SS_E/(n-k-1) = \hat{\sigma}^2$ by (7.5). Note: the denominator of $MS_E$ is $n - p = n - k - 1$, not $n - 2$.

Textbook §7.4: Chapter 7 ((7.8), (7.9))

# The ANOVA Table for MLR

*Table 7.2 in the textbook*

| **Source** | **df** | **SS** | **MS** | $F$ |
|---|---|---|---|---|
| Regression | $k$ | $SS_R$ | $MS_R = SS_R/k$ | $F_0 = MS_R/MS_E$ |
| Residual | $n-k-1$ | $SS_E$ | $MS_E = SS_E/(n-k-1)$ | |
| Total | $n-1$ | $SS_T$ | | |

# The ANOVA Table for MLR

*Table 7.2 in the textbook*

| **Source** | **df** | **SS** | **MS** | $F$ |
|---|---|---|---|---|
| Regression | $k$ | $SS_R$ | $MS_R = SS_R/k$ | $F_0 = MS_R/MS_E$ |
| Residual | $n-k-1$ | $SS_E$ | $MS_E = SS_E/(n-k-1)$ | |
| Total | $n-1$ | $SS_T$ | | |

**Compare with SLR:** In SLR, $k = 1$, so regression df $= 1$ and residual df $= n-2$. Everything else is the same structure.

## The ANOVA Table for MLR

*Table 7.2 in the textbook*

| **Source** | **df** | **SS** | **MS** | $F$ |
|---|---|---|---|---|
| Regression | $k$ | $SS_R$ | $MS_R = SS_R/k$ | $F_0 = MS_R/MS_E$ |
| Residual | $n-k-1$ | $SS_E$ | $MS_E = SS_E/(n-k-1)$ | |
| Total | $n-1$ | $SS_T$ | | |

**Compare with SLR:** In SLR, $k = 1$, so regression df $= 1$ and residual df $= n-2$. Everything else is the same structure.

**Exam tip:** Always verify that the degrees of freedom add up: $k+(n-k-1) = n-1$. If they do not, you have made an error.

# Worked Example (cont'd): ANOVA Table

*Table 7.3 in the textbook*

---

$n = 8$, $k = 2$, $\bar{y} = 560/8 = 70$. From before: $\mathbf{y}'\mathbf{y} = 39{,}920$, $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 39{,}912$.

# Worked Example (cont'd): ANOVA Table

*Table 7.3 in the textbook*

$n = 8$, $k = 2$, $\bar{y} = 560/8 = 70$. From before: $\mathbf{y}'\mathbf{y} = 39{,}920$, $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 39{,}912$.

$$SS_T = 39{,}920 - 8(70^2) = 39{,}920 - 39{,}200 = 720$$
$$SS_R = 39{,}912 - 39{,}200 = 712$$
$$SS_E = 720 - 712 = 8$$

# Worked Example (cont'd): ANOVA Table

*Table 7.3 in the textbook*

$n = 8$, $k = 2$, $\bar{y} = 560/8 = 70$. From before: $\mathbf{y}'\mathbf{y} = 39{,}920$, $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 39{,}912$.

$$SS_T = 39{,}920 - 8(70^2) = 39{,}920 - 39{,}200 = 720$$
$$SS_R = 39{,}912 - 39{,}200 = 712$$
$$SS_E = 720 - 712 = 8$$

| **Source** | **df** | **SS** | **MS** | $F_0$ |
|---|---|---|---|---|
| Regression | 2 | 712 | 356.0 | 222.5 |
| Residual | 5 | 8 | 1.6 | |
| Total | 7 | 720 | | |

# Worked Example (cont'd): ANOVA Table

*Table 7.3 in the textbook*

$n = 8$, $k = 2$, $\bar{y} = 560/8 = 70$. From before: $\mathbf{y}'\mathbf{y} = 39{,}920$, $\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} = 39{,}912$.

$$SS_T = 39{,}920 - 8(70^2) = 39{,}920 - 39{,}200 = 720$$
$$SS_R = 39{,}912 - 39{,}200 = 712$$
$$SS_E = 720 - 712 = 8$$

| **Source** | **df** | **SS** | **MS** | $F_0$ |
|---|---|---|---|---|
| Regression | 2 | 712 | 356.0 | 222.5 |
| Residual | 5 | 8 | 1.6 | |
| Total | 7 | 720 | | |

## $F$-Test for Significance of Regression

*Procedure 7.2*

---

**Question:** Does the model with $k$ regressors explain a significant amount of variability?

# $F$-Test for Significance of Regression

*Procedure 7.2*

---

**Question:** Does the model with $k$ regressors explain a significant amount of variability?

**Hypotheses** (Procedure 7.2, step 1):

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \qquad \text{vs.} \qquad H_1 : \beta_j \neq 0 \text{ for at least one } j$$

Note: $\beta_0$ is **not** included in the null hypothesis.

## $F$-Test for Significance of Regression

*Procedure 7.2*

**Question:** Does the model with $k$ regressors explain a significant amount of variability?

**Hypotheses** (Procedure 7.2, step 1):

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \qquad \text{vs.} \qquad H_1 : \beta_j \neq 0 \text{ for at least one } j$$

Note: $\beta_0$ is **not** included in the null hypothesis.

**Test statistic:**

$$F_0 = \frac{MS_R}{MS_E} = \frac{SS_R/k}{SS_E/(n-k-1)} \sim F_{k,\, n-k-1} \tag{7.9}$$

## $F$-Test for Significance of Regression

*Procedure 7.2*

**Question:** Does the model with $k$ regressors explain a significant amount of variability?

**Hypotheses** (Procedure 7.2, step 1):

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \qquad \text{vs.} \qquad H_1 : \beta_j \neq 0 \text{ for at least one } j$$

Note: $\beta_0$ is **not** included in the null hypothesis.

**Test statistic:**

$$F_0 = \frac{MS_R}{MS_E} = \frac{SS_R/k}{SS_E/(n-k-1)} \sim F_{k,\, n-k-1} \tag{7.9}$$

**Decision:** Reject $H_0$ if $F_0 > f_{\alpha,\, k,\, n-k-1}$.

## $F$-Test for Significance of Regression

*Procedure 7.2*

**Question:** Does the model with $k$ regressors explain a significant amount of variability?

**Hypotheses** (Procedure 7.2, step 1):

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0 \qquad \text{vs.} \qquad H_1 : \beta_j \neq 0 \text{ for at least one } j$$

Note: $\beta_0$ is **not** included in the null hypothesis.

**Test statistic:**

$$F_0 = \frac{MS_R}{MS_E} = \frac{SS_R/k}{SS_E/(n-k-1)} \sim F_{k,\, n-k-1} \tag{7.9}$$

**Decision:** Reject $H_0$ if $F_0 > f_{\alpha,\, k,\, n-k-1}$.

**Interpretation:** Rejecting $H_0$ means *at least one* regressor has a significant linear

## Worked Example (cont'd): Is the Evaluation Model Significant?

$\alpha = 0.05$

---

**Step 1: Hypotheses.** $H_0 : \beta_1 = \beta_2 = 0$ vs. $H_1$ : at least one $\beta_j \neq 0$; $\alpha = 0.05$.

## Worked Example (cont'd): Is the Evaluation Model Significant?

$\alpha = 0.05$

---

**Step 1: Hypotheses.** $H_0 : \beta_1 = \beta_2 = 0$ vs. $H_1$ : at least one $\beta_j \neq 0$; $\alpha = 0.05$.

**Step 2: Test statistic**, from the ANOVA table and (7.9):

$$F_0 = \frac{MS_R}{MS_E} = \frac{356.0}{1.6} = 222.5$$

## Worked Example (cont'd): Is the Evaluation Model Significant?

$\alpha = 0.05$

---

**Step 1: Hypotheses.** $H_0 : \beta_1 = \beta_2 = 0$ vs. $H_1$ : at least one $\beta_j \neq 0$; $\alpha = 0.05$.

**Step 2: Test statistic**, from the ANOVA table and (7.9):

$$F_0 = \frac{MS_R}{MS_E} = \frac{356.0}{1.6} = 222.5$$

**Step 3: Critical value.** With 2 and 5 degrees of freedom, $f_{0.05,\,2,\,5} = 5.79$ (from the $F$-table excerpt, Table 7.4).

## Worked Example (cont'd): Is the Evaluation Model Significant?

$\alpha = 0.05$

---

**Step 1: Hypotheses.** $H_0 : \beta_1 = \beta_2 = 0$ vs. $H_1$ : at least one $\beta_j \neq 0$; $\alpha = 0.05$.

**Step 2: Test statistic**, from the ANOVA table and (7.9):

$$F_0 = \frac{MS_R}{MS_E} = \frac{356.0}{1.6} = 222.5$$

**Step 3: Critical value.** With 2 and 5 degrees of freedom, $f_{0.05,2,5} = 5.79$ (from the $F$-table excerpt, Table 7.4).

**Step 4: Decision.** $222.5 > 5.79$, so we **reject** $H_0$.

## Worked Example (cont'd): Is the Evaluation Model Significant?

$\alpha = 0.05$

---

**Step 1: Hypotheses.** $H_0 : \beta_1 = \beta_2 = 0$ vs. $H_1$ : at least one $\beta_j \neq 0$; $\alpha = 0.05$.

**Step 2: Test statistic**, from the ANOVA table and (7.9):

$$F_0 = \frac{MS_R}{MS_E} = \frac{356.0}{1.6} = 222.5$$

**Step 3: Critical value.** With 2 and 5 degrees of freedom, $f_{0.05,2,5} = 5.79$ (from the $F$-table excerpt, Table 7.4).

**Step 4: Decision.** $222.5 > 5.79$, so we **reject** $H_0$.

**Step 5: Conclusion in context.** At least one of network size and data quality has a real linear relationship with the benchmark score. The correct wording is "the regression is significant" — not "every regressor is significant."

## Important Distinction: Overall $F$ vs. Individual $t$

**The overall $F$-test** asks: "Is the model *as a whole* useful?"

- $H_0$: *all* slope coefficients are zero simultaneously
- A significant $F$ means at least one $\beta_j \neq 0$

## Important Distinction: Overall $F$ vs. Individual $t$

**The overall $F$-test** asks: "Is the model *as a whole* useful?"

- $H_0$: *all* slope coefficients are zero simultaneously
- A significant $F$ means at least one $\beta_j \neq 0$

**Individual $t$-tests** (next lecture, (7.12)) ask: "Does *this particular* $x_j$ contribute, given the other regressors are in the model?"

- $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$ for a specific $j$

## Important Distinction: Overall $F$ vs. Individual $t$

**The overall $F$-test** asks: "Is the model *as a whole* useful?"

- $H_0$: *all* slope coefficients are zero simultaneously
- A significant $F$ means at least one $\beta_j \neq 0$

**Individual $t$-tests** (next lecture, (7.12)) ask: "Does *this particular* $x_j$ contribute, given the other regressors are in the model?"

- $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$ for a specific $j$

**On the exam:** Read the question carefully. If it says

- "significance of the regression model," use the $F$-test.
- "significance of the contribution of $x_j$," use the $t$-test on $\beta_j$.

# Example: Reading an ANOVA Table from Software Output

*Delivery routes (Table 7.5)*

**Delivery time model** (hours): $\hat{y} = 3.147 + 0.093\,x_1 + 0.633\,x_2 - 0.302\,x_3$

$x_1$ = distance (km), $x_2$ = number of stops, $x_3$ = load (tonnes).

| **Source** | **df** | **SS** | **MS** | $F_0$ |
|---|---|---|---|---|
| Regression | 3 | 83.343 | 27.781 | 134.3 |
| Residual | 21 | 4.343 | 0.2068 | |
| Total | 24 | 87.686 | | |

# Example: Reading an ANOVA Table from Software Output

*Delivery routes (Table 7.5)*

**Delivery time model** (hours): $\hat{y} = 3.147 + 0.093\,x_1 + 0.633\,x_2 - 0.302\,x_3$
$x_1$ = distance (km), $x_2$ = number of stops, $x_3$ = load (tonnes).

| Source | df | SS | MS | $F_0$ |
|---|---|---|---|---|
| Regression | 3 | 83.343 | 27.781 | 134.3 |
| Residual | 21 | 4.343 | 0.2068 | |
| Total | 24 | 87.686 | | |

**How many observations?** $n - 1 = 24$, so $n = 25$.
**How many regressors?** $k = 3$ (Distance, Stops, Load).
**How many parameters?** $p = k + 1 = 4$; residual df $= 25 - 4 = 21$. ✓

# Example: Reading an ANOVA Table from Software Output

*Delivery routes (Table 7.5)*

**Delivery time model** (hours): $\hat{y} = 3.147 + 0.093\,x_1 + 0.633\,x_2 - 0.302\,x_3$
$x_1$ = distance (km), $x_2$ = number of stops, $x_3$ = load (tonnes).

| Source | df | SS | MS | $F_0$ |
|---|---|---|---|---|
| Regression | 3 | 83.343 | 27.781 | 134.3 |
| Residual | 21 | 4.343 | 0.2068 | |
| Total | 24 | 87.686 | | |

**How many observations?** $n - 1 = 24$, so $n = 25$.
**How many regressors?** $k = 3$ (Distance, Stops, Load).
**How many parameters?** $p = k + 1 = 4$; residual df $= 25 - 4 = 21$. ✓

**Is the model significant at $\alpha = 0.05$?**
$F_0 = 134.3$. Critical value: $f_{0.05,\,3,\,21} = 3.07$.

# $R^2$ and $R^2_{\text{adj}}$

How Well Does the Model Fit? (Section 7.5)

## Coefficient of Determination $R^2$

*Section 7.5*

Same definition as SLR:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{7.10}$$

It measures the proportion of variability in $y$ explained by the model.

## Coefficient of Determination $R^2$

*Section 7.5*

Same definition as SLR:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{7.10}$$

It measures the proportion of variability in $y$ explained by the model.

**Evaluation example:** $R^2 = 712/720 = 0.9889$. The two coded regressors explain 98.89% of the score variability.

## Coefficient of Determination $R^2$

*Section 7.5*

Same definition as SLR:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{7.10}$$

It measures the proportion of variability in $y$ explained by the model.

**Evaluation example:** $R^2 = 712/720 = 0.9889$. The two coded regressors explain 98.89% of the score variability.

**Delivery example:** $R^2 = 83.343/87.686 = 0.9505$, so 95.05% of the variability in delivery time is explained by the three predictors.

## Coefficient of Determination $R^2$

*Section 7.5*

Same definition as SLR:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{7.10}$$

It measures the proportion of variability in $y$ explained by the model.

**Evaluation example:** $R^2 = 712/720 = 0.9889$. The two coded regressors explain 98.89% of the score variability.

**Delivery example:** $R^2 = 83.343/87.686 = 0.9505$, so 95.05% of the variability in delivery time is explained by the three predictors.

**The problem with $R^2$:** Adding *any* regressor to the model will **never decrease** $R^2$, even if the regressor is pure noise. So $R^2$ can be artificially inflated by including useless variables.

## Adjusted Coefficient of Determination $R^2_{\text{adj}}$

To penalize unnecessary regressors, we charge the model for each parameter it uses:

$$R^2_{\text{adj}} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} = 1 - \frac{MS_E}{SS_T/(n-1)} \tag{7.11}$$

where $p = k + 1$ is the number of parameters.

## Adjusted Coefficient of Determination $R^2_{\text{adj}}$

To penalize unnecessary regressors, we charge the model for each parameter it uses:

$$R^2_{\text{adj}} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} = 1 - \frac{MS_E}{SS_T/(n-1)} \tag{7.11}$$

where $p = k + 1$ is the number of parameters.

Unlike $R^2$, the adjusted version **can decrease** when a useless regressor is added: $SS_E$ drops a little, but the residual df drop by a whole unit, so $MS_E$ typically rises.

## Adjusted Coefficient of Determination $R^2_{\text{adj}}$

To penalize unnecessary regressors, we charge the model for each parameter it uses:

$$R^2_{\text{adj}} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} = 1 - \frac{MS_E}{SS_T/(n-1)} \tag{7.11}$$

where $p = k + 1$ is the number of parameters.

Unlike $R^2$, the adjusted version **can decrease** when a useless regressor is added: $SS_E$ drops a little, but the residual df drop by a whole unit, so $MS_E$ typically rises.

**Evaluation example:**

$$R^2_{\text{adj}} = 1 - \frac{8/5}{720/7} = 1 - \frac{1.600}{102.86} = 1 - 0.01556 = 0.9844$$

## Adjusted Coefficient of Determination $R^2_{\text{adj}}$

To penalize unnecessary regressors, we charge the model for each parameter it uses:

$$R^2_{\text{adj}} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} = 1 - \frac{MS_E}{SS_T/(n-1)} \tag{7.11}$$

where $p = k + 1$ is the number of parameters.

Unlike $R^2$, the adjusted version **can decrease** when a useless regressor is added: $SS_E$ drops a little, but the residual df drop by a whole unit, so $MS_E$ typically rises.

**Evaluation example:**

$$R^2_{\text{adj}} = 1 - \frac{8/5}{720/7} = 1 - \frac{1.600}{102.86} = 1 - 0.01556 = 0.9844$$

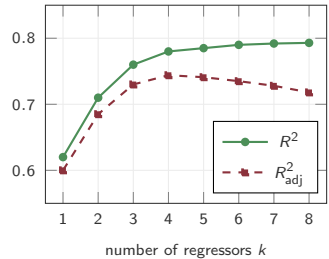
**Delivery example:**

$$R^2_{\text{adj}} = 1 - \frac{4.343/21}{54.750/24} = 1 - \frac{0.2068}{2.2813} = 1 - 0.0566 = 0.9434$$

# When to Use $R^2$ vs. $R^2_{\text{adj}}$

*Figure 7.2 in the textbook*

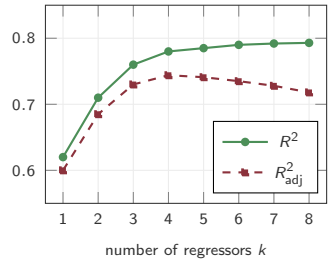
| | $R^2$ | $R^2_{\text{adj}}$ |
|---|---|---|
| Never decreases with more $x$'s? | Yes | No |
| Can it decrease? | No | Yes |
| Compare models, different $k$? | No | Yes |
| "Proportion explained"? | Yes | Not exactly |



# When to Use $R^2$ vs. $R^2_{\text{adj}}$

*Figure 7.2 in the textbook*

| | $R^2$ | $R^2_{\text{adj}}$ |
|---|---|---|
| Never decreases with more $x$'s? | Yes | No |
| Can it decrease? | No | Yes |
| Compare models, different $k$? | No | Yes |
| "Proportion explained"? | Yes | Not exactly |

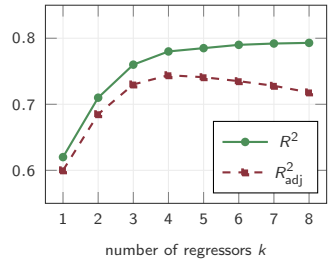



As regressors are added one at a time, $R^2$ never falls, but $R^2_{\text{adj}}$ peaks (here at $k = 4$) and then declines: the later regressors do not earn the degrees of freedom they cost.

# When to Use $R^2$ vs. $R^2_{\text{adj}}$

*Figure 7.2 in the textbook*

| | $R^2$ | $R^2_{\text{adj}}$ |
|---|---|---|
| Never decreases with more $x$'s? | Yes | No |
| Can it decrease? | No | Yes |
| Compare models, different $k$? | No | Yes |
| "Proportion explained"? | Yes | Not exactly |




As regressors are added one at a time, $R^2$ never falls, but $R^2_{\text{adj}}$ peaks (here at $k = 4$) and then declines: the later regressors do not earn the degrees of freedom they cost.

**Rule of thumb:** Use $R^2$ to report how well the current model fits. Use $R^2_{\text{adj}}$ (7.11) to compare models with different numbers of regressors.

# SLR vs. MLR: What Changes, What Stays the Same

| **Concept** | **SLR (Ch. 6)** | **MLR (Ch. 7)** |
|---|---|---|
| Model | $y = \beta_0 + \beta_1 x + \varepsilon$ (6.2) | $y = \beta_0 + \sum \beta_j x_j + \varepsilon$ (7.1) |
| Parameters | $p = 2$ | $p = k + 1$ |
| Estimation | $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9) | $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (7.4) |
| Residual df | $n - 2$ | $n - k - 1$ |
| $\hat{\sigma}^2$ | $SS_E/(n-2)$ (6.14) | $SS_E/(n-k-1)$ (7.5) |
| $F$-test df | $F_{1,\,n-2}$ (6.20) | $F_{k,\,n-k-1}$ (7.9) |
| $R^2$ | $SS_R/SS_T$ (6.21) | $SS_R/SS_T$ (7.10) |
| $R^2_{\text{adj}}$ | Rarely used | Important (7.11) |

# SLR vs. MLR: What Changes, What Stays the Same

| Concept | SLR (Ch. 6) | MLR (Ch. 7) |
|---|---|---|
| Model | $y = \beta_0 + \beta_1 x + \varepsilon$ (6.2) | $y = \beta_0 + \sum \beta_j x_j + \varepsilon$ (7.1) |
| Parameters | $p = 2$ | $p = k + 1$ |
| Estimation | $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9) | $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ (7.4) |
| Residual df | $n - 2$ | $n - k - 1$ |
| $\hat{\sigma}^2$ | $SS_E/(n-2)$ (6.14) | $SS_E/(n-k-1)$ (7.5) |
| $F$-test df | $F_{1,\,n-2}$ (6.20) | $F_{k,\,n-k-1}$ (7.9) |
| $R^2$ | $SS_R/SS_T$ (6.21) | $SS_R/SS_T$ (7.10) |
| $R^2_{\text{adj}}$ | Rarely used | Important (7.11) |

**The key takeaway:** The structure (ANOVA, $F$-test, residuals) is identical. Only the degrees of freedom and estimation formulas change.

# Prediction and Residuals in MLR

**Prediction** at a new point $\mathbf{x}_0 = (1, x_{01}, x_{02}, \ldots, x_{0k})'$:

$$\hat{y}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \cdots + \hat{\beta}_k x_{0k}$$

## Prediction and Residuals in MLR

**Prediction** at a new point $\mathbf{x}_0 = (1, x_{01}, x_{02}, \ldots, x_{0k})'$:

$$\hat{y}_0 = \mathbf{x}_0' \hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \cdots + \hat{\beta}_k x_{0k}$$

**Practice:** Using the delivery model $\hat{y} = 3.147 + 0.093\,x_1 + 0.633\,x_2 - 0.302\,x_3$, predict the delivery time for Distance $= 30$ km, Stops $= 3$, Load $= 2$ tonnes.

# Prediction and Residuals in MLR

**Prediction** at a new point $\mathbf{x}_0 = (1, x_{01}, x_{02}, \ldots, x_{0k})'$:

$$\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \cdots + \hat{\beta}_k x_{0k}$$

**Practice:** Using the delivery model $\hat{y} = 3.147 + 0.093\,x_1 + 0.633\,x_2 - 0.302\,x_3$, predict the delivery time for Distance $= 30$ km, Stops $= 3$, Load $= 2$ tonnes.

$$\hat{y}_0 = 3.147 + 0.093(30) + 0.633(3) - 0.302(2) = 3.147 + 2.790 + 1.899 - 0.604$$

$$= \boxed{7.232 \text{ hours}}$$

## Prediction and Residuals in MLR

**Prediction** at a new point $\mathbf{x}_0 = (1, x_{01}, x_{02}, \ldots, x_{0k})'$:

$$\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}} = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \hat{\beta}_2 x_{02} + \cdots + \hat{\beta}_k x_{0k}$$

**Practice:** Using the delivery model $\hat{y} = 3.147 + 0.093\,x_1 + 0.633\,x_2 - 0.302\,x_3$, predict the delivery time for Distance $= 30$ km, Stops $= 3$, Load $= 2$ tonnes.

$$\hat{y}_0 = 3.147 + 0.093(30) + 0.633(3) - 0.302(2) = 3.147 + 2.790 + 1.899 - 0.604$$

$$= \boxed{7.232 \text{ hours}}$$

**Residual:** If the actual delivery time was 7.0 hours, then $e = 7.0 - 7.232 = -0.232$

# Reading Regression Software Output

*Coefficient table (Table 7.6)*

On the exam, you will often see a table like this (delivery model again):

| **Variable** | **Coef** | **SE** | ***t*-Stat** | ***p*-value** |
|---|---|---|---|---|
| Intercept | 3.147 | 0.448 | 7.03 | $< 0.001$ |
| Distance ($x_1$) | 0.093 | 0.0146 | 6.37 | $< 0.001$ |
| Stops ($x_2$) | 0.633 | 0.132 | 4.80 | $< 0.001$ |
| Load ($x_3$) | $-0.302$ | 0.146 | $-2.07$ | 0.051 |

# Reading Regression Software Output

*Coefficient table (Table 7.6)*

On the exam, you will often see a table like this (delivery model again):

| Variable | Coef | SE | $t$-Stat | $p$-value |
|---|---|---|---|---|
| Intercept | 3.147 | 0.448 | 7.03 | $< 0.001$ |
| Distance ($x_1$) | 0.093 | 0.0146 | 6.37 | $< 0.001$ |
| Stops ($x_2$) | 0.633 | 0.132 | 4.80 | $< 0.001$ |
| Load ($x_3$) | $-0.302$ | 0.146 | $-2.07$ | 0.051 |

**What each column tells you:**

- **Coef** $= \hat{\beta}_j$ from (7.4)
- **SE** $= \text{se}(\hat{\beta}_j) = \sqrt{MS_E \cdot C_{jj}}$ (7.7), where $C_{jj}$ is from $(\mathbf{X}'\mathbf{X})^{-1}$
- $t$-**Stat** = Coef / SE (tests $H_0 : \beta_j = 0$; next lecture, (7.12))
- $p$-**value** = probability of a $t_{n-p}$ value at least this extreme under $H_0$

# Reading Regression Software Output

*Coefficient table (Table 7.6)*

On the exam, you will often see a table like this (delivery model again):

| Variable | Coef | SE | $t$-Stat | $p$-value |
|---|---|---|---|---|
| Intercept | 3.147 | 0.448 | 7.03 | $< 0.001$ |
| Distance ($x_1$) | 0.093 | 0.0146 | 6.37 | $< 0.001$ |
| Stops ($x_2$) | 0.633 | 0.132 | 4.80 | $< 0.001$ |
| Load ($x_3$) | $-0.302$ | 0.146 | $-2.07$ | 0.051 |

**What each column tells you:**

- **Coef** $= \hat{\beta}_j$ from (7.4)
- **SE** $= \text{se}(\hat{\beta}_j) = \sqrt{MS_E \cdot C_{jj}}$ (7.7), where $C_{jj}$ is from $(\mathbf{X}'\mathbf{X})^{-1}$
- $t$-**Stat** = Coef / SE (tests $H_0 : \beta_j = 0$; next lecture, (7.12))
- $p$-**value** = probability of a $t_{n-p}$ value at least this extreme under $H_0$

# Practice: Filling in a Software Output Table

*Startup pricing experiment*

**Given:** Your startup models weekly signups $y$ against subscription price $x$ (SAR): $\hat{y} = 102.33 - 4.015\,x$ (SLR, $n =$?)

| Source | df | SS | MS | $F$ |
|---|---|---|---|---|
| Regression | 1 | 2933.516 | ? | ? |
| Residual | 11 | ? | ? | |
| Total | ? | 3319.21 | | |

# Practice: Filling in a Software Output Table

*Startup pricing experiment*

**Given:** Your startup models weekly signups $y$ against subscription price $x$ (SAR): $\hat{y} = 102.33 - 4.015\,x$ (SLR, $n =$?)

| Source | df | SS | MS | $F$ |
|---|---|---|---|---|
| Regression | 1 | 2933.516 | ? | ? |
| Residual | 11 | ? | ? | |
| Total | ? | 3319.21 | | |

**Step 1:** $n =$? Residual df $= n - 2 = 11 \Rightarrow n = 13$. Total df $= n - 1 = 12$.

# Practice: Filling in a Software Output Table

*Startup pricing experiment*

**Given:** Your startup models weekly signups $y$ against subscription price $x$ (SAR): $\hat{y} = 102.33 - 4.015\,x$ (SLR, $n =$?)

| Source | df | SS | MS | $F$ |
|---|---|---|---|---|
| Regression | 1 | 2933.516 | ? | ? |
| Residual | 11 | ? | ? | |
| Total | ? | 3319.21 | | |

**Step 1:** $n =$? Residual df $= n - 2 = 11 \Rightarrow n = 13$. Total df $= n - 1 = 12$.
**Step 2:** $SS_E = SS_T - SS_R = 3319.21 - 2933.516 = 385.694$.

# Practice: Filling in a Software Output Table

*Startup pricing experiment*

**Given:** Your startup models weekly signups $y$ against subscription price $x$ (SAR): $\hat{y} = 102.33 - 4.015\,x$ (SLR, $n =$?)

| Source | df | SS | MS | $F$ |
|---|---|---|---|---|
| Regression | 1 | 2933.516 | ? | ? |
| Residual | 11 | ? | ? | |
| Total | ? | 3319.21 | | |

**Step 1:** $n =$? Residual df $= n - 2 = 11 \Rightarrow n = 13$. Total df $= n - 1 = 12$.
**Step 2:** $SS_E = SS_T - SS_R = 3319.21 - 2933.516 = 385.694$.
**Step 3:** $MS_R = 2933.516/1 = 2933.516$. $MS_E = 385.694/11 = 35.06$.

# Practice: Filling in a Software Output Table

*Startup pricing experiment*

**Given:** Your startup models weekly signups $y$ against subscription price $x$ (SAR): $\hat{y} = 102.33 - 4.015\,x$ (SLR, $n =$?)

| Source | df | SS | MS | $F$ |
|---|---|---|---|---|
| Regression | 1 | 2933.516 | ? | ? |
| Residual | 11 | ? | ? | |
| Total | ? | 3319.21 | | |

**Step 1:** $n =$? Residual df $= n - 2 = 11 \Rightarrow n = 13$. Total df $= n - 1 = 12$.
**Step 2:** $SS_E = SS_T - SS_R = 3319.21 - 2933.516 = 385.694$.
**Step 3:** $MS_R = 2933.516/1 = 2933.516$. $MS_E = 385.694/11 = 35.06$.
**Step 4:** $F_0 = 2933.516/35.06 = 83.66$. Also: $\hat{\sigma}^2 = MS_E = 35.06$.

# Practice: Filling in a Software Output Table

*Startup pricing experiment*

**Given:** Your startup models weekly signups $y$ against subscription price $x$ (SAR): $\hat{y} = 102.33 - 4.015\,x$ (SLR, $n =$?)

| Source | df | SS | MS | $F$ |
|---|---|---|---|---|
| Regression | 1 | 2933.516 | ? | ? |
| Residual | 11 | ? | ? | |
| Total | ? | 3319.21 | | |

**Step 1:** $n =$? Residual df $= n - 2 = 11 \Rightarrow n = 13$. Total df $= n - 1 = 12$.
**Step 2:** $SS_E = SS_T - SS_R = 3319.21 - 2933.516 = 385.694$.
**Step 3:** $MS_R = 2933.516/1 = 2933.516$. $MS_E = 385.694/11 = 35.06$.
**Step 4:** $F_0 = 2933.516/35.06 = 83.66$. Also: $\hat{\sigma}^2 = MS_E = 35.06$.

**Exam takeaway:** If you know the structure of the ANOVA table, you can fill in

## Practice: Extracting Key Quantities from Output

From the same pricing experiment ($\hat{y} = 102.33 - 4.015\,x$, $n = 13$, $MS_E = 35.06$):

## Practice: Extracting Key Quantities from Output

From the same pricing experiment ($\hat{y} = 102.33 - 4.015\,x$, $n = 13$, $MS_E = 35.06$):

**What is $R^2$?** $R^2 = SS_R/SS_T = 2933.516/3319.21 = 0.8838$.
So 88.38% of variability in signups is explained by price, and **11.62% is not explained**.

## Practice: Extracting Key Quantities from Output

From the same pricing experiment ($\hat{y} = 102.33 - 4.015\,x$, $n = 13$, $MS_E = 35.06$):

**What is $R^2$?** $R^2 = SS_R/SS_T = 2933.516/3319.21 = 0.8838$.
So 88.38% of variability in signups is explained by price, and **11.62% is not explained**.

**$t$-statistic for the intercept** (given $\text{se}(\hat{\beta}_0) = 3.484$): $t_0 = 102.33/3.484 = 29.37$.

## Practice: Extracting Key Quantities from Output

From the same pricing experiment ($\hat{y} = 102.33 - 4.015\,x$, $n = 13$, $MS_E = 35.06$):

**What is $R^2$?** $R^2 = SS_R/SS_T = 2933.516/3319.21 = 0.8838$.
So 88.38% of variability in signups is explained by price, and **11.62% is not explained**.

**$t$-statistic for the intercept** (given $\text{se}(\hat{\beta}_0) = 3.484$): $t_0 = 102.33/3.484 = 29.37$.

**$t$-statistic for the slope** (given $\text{se}(\hat{\beta}_1) = 0.439$): $t_0 = -4.015/0.439 = -9.146$.

## Practice: Extracting Key Quantities from Output

From the same pricing experiment ($\hat{y} = 102.33 - 4.015\,x$, $n = 13$, $MS_E = 35.06$):

**What is $R^2$?** $R^2 = SS_R/SS_T = 2933.516/3319.21 = 0.8838$.
So 88.38% of variability in signups is explained by price, and **11.62% is not explained**.

**$t$-statistic for the intercept** (given $\text{se}(\hat{\beta}_0) = 3.484$): $t_0 = 102.33/3.484 = 29.37$.

**$t$-statistic for the slope** (given $\text{se}(\hat{\beta}_1) = 0.439$): $t_0 = -4.015/0.439 = -9.146$.

Notice: $T_0^2 = (-9.146)^2 = 83.65 \approx F_0 = 83.66$. This is the same SLR relationship we saw in Chapter 6: in SLR, $t^2 = F$ because there is only one regressor. In MLR with $k > 1$, this relationship no longer holds.

## Lecture 20 Summary

- **MLR model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, extending SLR to $k$ predictors. Matrix form (7.2): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.

## Lecture 20 Summary

- **MLR model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, extending SLR to $k$ predictors. Matrix form (7.2): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
- **Matrix estimation** (7.4): $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Build $\mathbf{X}$ (with a column of 1's), compute $\mathbf{X}'\mathbf{y}$, multiply by the given inverse (Procedure 7.1).

## Lecture 20 Summary

- **MLR model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, extending SLR to $k$ predictors. Matrix form (7.2): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
- **Matrix estimation** (7.4): $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Build $\mathbf{X}$ (with a column of 1's), compute $\mathbf{X}'\mathbf{y}$, multiply by the given inverse (Procedure 7.1).
- **Error variance** (7.5): $\hat{\sigma}^2 = SS_E/(n-p) = MS_E$; standard errors (7.7): $\text{se}(\hat{\beta}_j) = \sqrt{MS_E \cdot C_{jj}}$.

## Lecture 20 Summary

- **MLR model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, extending SLR to $k$ predictors. Matrix form (7.2): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
- **Matrix estimation** (7.4): $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Build $\mathbf{X}$ (with a column of 1's), compute $\mathbf{X}'\mathbf{y}$, multiply by the given inverse (Procedure 7.1).
- **Error variance** (7.5): $\hat{\sigma}^2 = SS_E/(n-p) = MS_E$; standard errors (7.7): $\text{se}(\hat{\beta}_j) = \sqrt{MS_E \cdot C_{jj}}$.
- **ANOVA** (7.8): same $SS_T = SS_R + SS_E$ structure. Regression df $= k$, residual df $= n - k - 1$.

## Lecture 20 Summary

- **MLR model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, extending SLR to $k$ predictors. Matrix form (7.2): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
- **Matrix estimation** (7.4): $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Build $\mathbf{X}$ (with a column of 1's), compute $\mathbf{X}'\mathbf{y}$, multiply by the given inverse (Procedure 7.1).
- **Error variance** (7.5): $\hat{\sigma}^2 = SS_E/(n-p) = MS_E$; standard errors (7.7): $\text{se}(\hat{\beta}_j) = \sqrt{MS_E \cdot C_{jj}}$.
- **ANOVA** (7.8): same $SS_T = SS_R + SS_E$ structure. Regression df $= k$, residual df $= n - k - 1$.
- **$F$-test** (7.9): $F_0 = MS_R/MS_E \sim F_{k,\,n-k-1}$ tests overall model significance (Procedure 7.2).
  $H_0 : \beta_1 = \cdots = \beta_k = 0$ vs. $H_1$ : at least one $\beta_j \neq 0$.

## Lecture 20 Summary

- **MLR model** (7.1): $y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$, extending SLR to $k$ predictors. Matrix form (7.2): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$.
- **Matrix estimation** (7.4): $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Build $\mathbf{X}$ (with a column of 1's), compute $\mathbf{X}'\mathbf{y}$, multiply by the given inverse (Procedure 7.1).
- **Error variance** (7.5): $\hat{\sigma}^2 = SS_E/(n-p) = MS_E$; standard errors (7.7): $\text{se}(\hat{\beta}_j) = \sqrt{MS_E \cdot C_{jj}}$.
- **ANOVA** (7.8): same $SS_T = SS_R + SS_E$ structure. Regression df $= k$, residual df $= n - k - 1$.
- **$F$-test** (7.9): $F_0 = MS_R/MS_E \sim F_{k,\,n-k-1}$ tests overall model significance (Procedure 7.2).
  $H_0 : \beta_1 = \cdots = \beta_k = 0$ vs. $H_1$ : at least one $\beta_j \neq 0$.
- $R^2_{\textbf{adj}}$ (7.11): $1 - \frac{SS_E/(n-p)}{SS_T/(n-1)}$. Unlike $R^2$ (7.10), it penalizes unnecessary

## Coming Up: Lecture 21 (MLR Part 2)

**Next lecture will cover:**

- Hypothesis tests on individual coefficients $\beta_j$ ($t$-tests, (7.12))
- Confidence intervals on $\beta_j$ (7.13)
- Indicator variables and interaction terms
- Multicollinearity and the variance inflation factor (7.17)
- Reading software output for decision-making, and a model-building strategy (Procedure 7.5)

## Coming Up: Lecture 21 (MLR Part 2)

**Next lecture will cover:**

- Hypothesis tests on individual coefficients $\beta_j$ ($t$-tests, (7.12))
- Confidence intervals on $\beta_j$ (7.13)
- Indicator variables and interaction terms
- Multicollinearity and the variance inflation factor (7.17)
- Reading software output for decision-making, and a model-building strategy (Procedure 7.5)

**Make sure you are comfortable** with today's material first — the ANOVA table and the matrix estimation are the foundation.

# Major Exam 2 Preview

*Chapters 5, 6, and 7*

**Exam structure** (same as Major 1):

| Type | Count | Time | Points |
|---|---|---|---|
| Regular written | 3 | $\sim$17 min each | 17 pts each |
| Short written | 3 | $\sim$9 min each | 9 pts each |
| MCQ (conceptual) | 8 | $\sim$2.75 min each | 2.75 pts each |

# Major Exam 2 Preview

*Chapters 5, 6, and 7*

**Exam structure** (same as Major 1):

| Type | Count | Time | Points |
|---|---|---|---|
| Regular written | 3 | ∼17 min each | 17 pts each |
| Short written | 3 | ∼9 min each | 9 pts each |
| MCQ (conceptual) | 8 | ∼2.75 min each | 2.75 pts each |

**My assigned topics:**

- Hypothesis tests and confidence intervals on regression coefficients
- ANOVA for regression
- Multiple regression: model fitting and interpretation

# Major Exam 2 Preview

*Chapters 5, 6, and 7*

**Exam structure** (same as Major 1):

| Type | Count | Time | Points |
|---|---|---|---|
| Regular written | 3 | ∼17 min each | 17 pts each |
| Short written | 3 | ∼9 min each | 9 pts each |
| MCQ (conceptual) | 8 | ∼2.75 min each | 2.75 pts each |

**My assigned topics:**

- Hypothesis tests and confidence intervals on regression coefficients
- ANOVA for regression
- Multiple regression: model fitting and interpretation

## Announcements

- **Quiz on Chapter 7 basics** — submit before class.

## Announcements

- **Quiz on Chapter 7 basics** — submit before class.
- **Next lecture: MLR Part 2**
  - CI/HT on individual coefficients, indicators, interactions, VIF
  - Bring: calculator and formula sheet (if provided)

## Announcements

- **Quiz on Chapter 7 basics** — submit before class.
- **Next lecture: MLR Part 2**
  - CI/HT on individual coefficients, indicators, interactions, VIF
  - Bring: calculator and formula sheet (if provided)
- **Office hours**: Tuesdays 9–10 AM, Room 22-219 (or by appointment via calendar)