ISE 315: Engineering Statistics

Lecture 17: Simple Linear Regression — A Full Worked Fit

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 17

Simple Linear Regression, Continued: A Full Worked Fit (Chapter 6)

Recap: The Model and the Estimators

From Lecture 15

The simple linear regression model (6.2):

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$$

Recap: The Model and the Estimators

From Lecture 15

The simple linear regression model (6.2):

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$$

Least squares building blocks and estimators:

Quantity	Formula	Book eq.
Spread of x	$S_{xx} = \sum x_i^2 - (\sum x_i)^2/n$	(6.7)
Co-movement	$S_{xy} = \sum x_i y_i - (\sum x_i)(\sum y_i)/n$	(6.8)
Slope	$\hat{\beta}_1 = S_{xy}/S_{xx}$	(6.9)
Intercept	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$	(6.10)
[illegible]	$\hat{\sigma}^2 = SS_E/(n-2)$	(6.14)

Recap: The Model and the Estimators

From Lecture 15

The simple linear regression model (6.2):

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n$$

Least squares building blocks and estimators:

Quantity	**Formula**	**Book eq.**
Spread of x	$S_{xx} = \sum x_i^2 - (\sum x_i)^2/n$	(6.7)
Co-movement	$S_{xy} = \sum x_i y_i - (\sum x_i)(\sum y_i)/n$	(6.8)
Slope	$\hat{\beta}_1 = S_{xy}/S_{xx}$	(6.9)
Intercept	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$	(6.10)
[illegible]	$\hat{\sigma}^2 = SS_E/(n-2)$	(6.14)

Recap: The AI Verification Fit

Quick Warm-Up

In Lecture 15 we fit the AI safety verification data (x = testing hours, y = failures per 1000 inspections, $n = 8$):

$$\hat{y} = 54.87 - 0.8452\,x, \qquad \hat{\sigma}^2 = 1.621$$

Recap: The AI Verification Fit

Quick Warm-Up

In Lecture 15 we fit the AI safety verification data (x = testing hours, y = failures per 1000 inspections, $n = 8$):

$$\hat{y} = 54.87 - 0.8452\,x, \qquad \hat{\sigma}^2 = 1.621$$

Quick check: the team can afford $x = 32$ hours of verification. What failure rate does the model predict?

Recap: The AI Verification Fit

Quick Warm-Up

In Lecture 15 we fit the AI safety verification data (x = testing hours, y = failures per 1000 inspections, $n = 8$):

$$\hat{y} = 54.87 - 0.8452\,x, \qquad \hat{\sigma}^2 = 1.621$$

Quick check: the team can afford $x = 32$ hours of verification. What failure rate does the model predict?

$$\hat{y} = 54.87 - 0.8452(32) = 54.87 - 27.05 = 27.82 \text{ failures per 1000}$$

Recap: The AI Verification Fit

Quick Warm-Up

In Lecture 15 we fit the AI safety verification data (x = testing hours, y = failures per 1000 inspections, $n = 8$):

$$\hat{y} = 54.87 - 0.8452\,x, \qquad \hat{\sigma}^2 = 1.621$$

Quick check: the team can afford $x = 32$ hours of verification. What failure rate does the model predict?

$$\hat{y} = 54.87 - 0.8452(32) = 54.87 - 27.05 = 27.82 \text{ failures per 1000}$$

$x = 32$ is **inside** the observed range $10 \le x \le 45$, so this is interpolation — a

Lecture 17 Outline

- The fitting recipe: Procedure 6.1, end to end

Lecture 17 Outline

- The fitting recipe: Procedure 6.1, end to end
- Worked example: truck load vs. fuel consumption

Lecture 17 Outline

- The fitting recipe: Procedure 6.1, end to end
- Worked example: truck load vs. fuel consumption
- Every step with book numbers: S_{xx}, S_{xy}, $\hat{\beta}_1$, $\hat{\beta}_0$

Lecture 17 Outline

- The fitting recipe: Procedure 6.1, end to end
- Worked example: truck load vs. fuel consumption
- Every step with book numbers: S_{xx}, S_{xy}, $\hat{\beta}_1$, $\hat{\beta}_0$
- Residuals: computing and checking them ((6.4))

Lecture 17 Outline

- The fitting recipe: Procedure 6.1, end to end
- Worked example: truck load vs. fuel consumption
- Every step with book numbers: S_{xx}, S_{xy}, $\hat{\beta}_1$, $\hat{\beta}_0$
- Residuals: computing and checking them ((6.4))
- SS_E two ways, then $\hat{\sigma}^2$ ((6.11), (6.13), (6.14))

Lecture 17 Outline

- The fitting recipe: Procedure 6.1, end to end
- Worked example: truck load vs. fuel consumption
- Every step with book numbers: S_{xx}, S_{xy}, $\hat{\beta}_1$, $\hat{\beta}_0$
- Residuals: computing and checking them ((6.4))
- SS_E two ways, then $\hat{\sigma}^2$ ((6.11), (6.13), (6.14))
- Exam-style practice problem (startup pricing experiment)

Procedure 6.1: The Road Map for Today

1. Compute the five sums: $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum x_i y_i$, $\sum y_i^2$.

Procedure 6.1: The Road Map for Today

1. Compute the five sums: $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum x_i y_i$, $\sum y_i^2$.
2. Compute $\bar{x}$, $\bar{y}$, then S_{xx} (6.7) and S_{xy} (6.8).

Procedure 6.1: The Road Map for Today

1. Compute the five sums: $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum x_i y_i$, $\sum y_i^2$.
2. Compute $\bar{x}$, $\bar{y}$, then S_{xx} (6.7) and S_{xy} (6.8).
3. Slope: $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9).

Procedure 6.1: The Road Map for Today

1. Compute the five sums: $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum x_i y_i$, $\sum y_i^2$.
2. Compute $\bar{x}$, $\bar{y}$, then S_{xx} (6.7) and S_{xy} (6.8).
3. Slope: $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9).
4. Intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (6.10).

Procedure 6.1: The Road Map for Today

1. Compute the five sums: $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum x_i y_i$, $\sum y_i^2$.
2. Compute $\bar{x}$, $\bar{y}$, then S_{xx} (6.7) and S_{xy} (6.8).
3. Slope: $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9).
4. Intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (6.10).
5. Write and plot the fitted line (6.3); interpret the slope.

Procedure 6.1: The Road Map for Today

1. Compute the five sums: $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum x_i y_i$, $\sum y_i^2$.
2. Compute $\bar{x}$, $\bar{y}$, then S_{xx} (6.7) and S_{xy} (6.8).
3. Slope: $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9).
4. Intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (6.10).
5. Write and plot the fitted line (6.3); interpret the slope.
6. Compute residuals $e_i = y_i - \hat{y}_i$ (6.4); check $\sum e_i = 0$.

Procedure 6.1: The Road Map for Today

1. Compute the five sums: $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum x_i y_i$, $\sum y_i^2$.
2. Compute $\bar{x}$, $\bar{y}$, then S_{xx} (6.7) and S_{xy} (6.8).
3. Slope: $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9).
4. Intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (6.10).
5. Write and plot the fitted line (6.3); interpret the slope.
6. Compute residuals $e_i = y_i - \hat{y}_i$ (6.4); check $\sum e_i = 0$.
7. Compute SS_E (6.11) — and verify with the shortcut (6.13).

Procedure 6.1: The Road Map for Today

1. Compute the five sums: $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum x_i y_i$, $\sum y_i^2$.
2. Compute $\bar{x}$, $\bar{y}$, then S_{xx} (6.7) and S_{xy} (6.8).
3. Slope: $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9).
4. Intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$ (6.10).
5. Write and plot the fitted line (6.3); interpret the slope.
6. Compute residuals $e_i = y_i - \hat{y}_i$ (6.4); check $\sum e_i = 0$.
7. Compute SS_E (6.11) — and verify with the shortcut (6.13).
8. Estimate the error variance: $\hat{\sigma}^2 = SS_E/(n-2)$ (6.14).

Procedure 6.1: The Road Map for Today

1. Compute the five sums: $\sum x_i$, $\sum y_i$, $\sum x_i^2$, $\sum x_i y_i$, $\sum y_i^2$.
2. Compute $\bar{x}$, $\bar{y}$, then S_{xx} (6.7) and S_{xy} (6.8).
3. Slope: $\hat{\beta}_1 = S_{xy}/S_{xx}$ (6.9).
4. Intercept: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (6.10).
5. Write and plot the fitted line (6.3); interpret the slope.
6. Compute residuals $e_i = y_i - \hat{y}_i$ (6.4); check $\sum e_i = 0$.
7. Compute SS_E (6.11) — and verify with the shortcut (6.13).
8. Estimate the error variance: $\hat{\sigma}^2 = SS_E/(n-2)$ (6.14).

Eight steps, every one on the formula sheet. Let's do all of them once, carefully.

Worked Example: Truck Load vs. Fuel Consumption

Supply Chain Data

A fleet analyst wants to predict fuel consumption from payload so she can quote delivery costs. For $n = 8$ line-haul trips she records the load x (tonnes) and the fuel consumption y (L/100 km):

Worked Example: Truck Load vs. Fuel Consumption

Supply Chain Data

A fleet analyst wants to predict fuel consumption from payload so she can quote delivery costs. For $n = 8$ line-haul trips she records the load x (tonnes) and the fuel consumption y (L/100 km):

x (tonnes)	6	8	10	12	14	16	18	20
y (L/100 km)	23.0	24.2	28.1	28.6	32.9	33.4	37.2	37.8

Worked Example: Truck Load vs. Fuel Consumption

Supply Chain Data

A fleet analyst wants to predict fuel consumption from payload so she can quote delivery costs. For $n = 8$ line-haul trips she records the load x (tonnes) and the fuel consumption y (L/100 km):

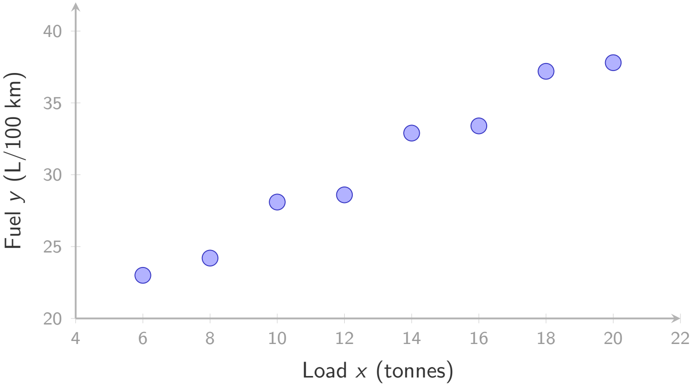
x (tonnes)	6	8	10	12	14	16	18	20
y (L/100 km)	23.0	24.2	28.1	28.6	32.9	33.4	37.2	37.8

Questions we will answer:

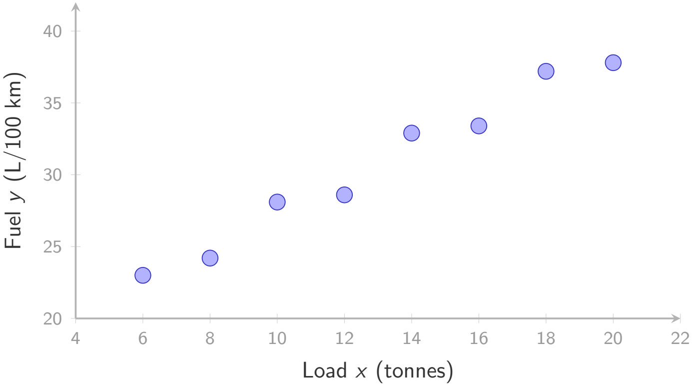
- How much extra fuel does each additional tonne cost?
- How much scatter remains around the fitted line?

Textbook: Chapter 6, §6.3–§6.4 ((6.7)–(6.14))

Step 0: Always Plot First



Step 0: Always Plot First



The scatter is clearly **linear and increasing**, with modest noise. A straight-line model (6.2) is reasonable. Now we fit it.

Step 1: The Five Sums

Everything in Procedure 6.1 is built from five sums. With $n = 8$:

Step 1: The Five Sums

Everything in Procedure 6.1 is built from five sums. With $n = 8$:

$$\sum x_i = 6 + 8 + \cdots + 20 = 104 \qquad \sum y_i = 23.0 + 24.2 + \cdots + 37.8 = 245.2$$

Step 1: The Five Sums

Everything in Procedure 6.1 is built from five sums. With $n = 8$:

$$\sum x_i = 6 + 8 + \cdots + 20 = 104$$

$$\sum y_i = 23.0 + 24.2 + \cdots + 37.8 = 245.2$$

$$\sum x_i^2 = 36 + 64 + \cdots + 400 = 1520$$

$$\sum x_i y_i = 138 + 193.6 + \cdots + 756 = 3376.4$$

Step 1: The Five Sums

Everything in Procedure 6.1 is built from five sums. With $n = 8$:

$$\sum x_i = 6 + 8 + \cdots + 20 = 104$$

$$\sum y_i = 23.0 + 24.2 + \cdots + 37.8 = 245.2$$

$$\sum x_i^2 = 36+64+\cdots+400 = 1520$$

$$\sum x_i y_i = 138+193.6+\cdots+756 = 3376.4$$

$$\sum y_i^2 = 529 + 585.64 + \cdots + 1428.84 = 7732.86$$

Step 1: The Five Sums

Everything in Procedure 6.1 is built from five sums. With $n = 8$:

$$\sum x_i = 6 + 8 + \cdots + 20 = 104$$

$$\sum y_i = 23.0 + 24.2 + \cdots + 37.8 = 245.2$$

$$\sum x_i^2 = 36 + 64 + \cdots + 400 = 1520$$

$$\sum x_i y_i = 138 + 193.6 + \cdots + 756 = 3376.4$$

$$\sum y_i^2 = 529 + 585.64 + \cdots + 1428.84 = 7732.86$$

Step 2a: The Spread of x

Using (6.7)

Write the formula first — equation (6.7) in the textbook:

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n} \tag{6.7}$$

Step 2a: The Spread of x

Using (6.7)

Write the formula first — equation (6.7) in the textbook:

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n} \tag{6.7}$$

Substitute the sums from Step 1:

$$S_{xx} = 1520 - \frac{(104)^2}{8} = 1520 - \frac{10816}{8}$$

Step 2a: The Spread of x

Using (6.7)

Write the formula first — equation (6.7) in the textbook:

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum x_i)^2}{n} \tag{6.7}$$

Substitute the sums from Step 1:

$$S_{xx} = 1520 - \frac{(104)^2}{8} = 1520 - \frac{10816}{8}$$

$$S_{xx} = 1520 - 1352 = 168$$

Step 2a: The Spread of x

Using (6.7)

Write the formula first — equation (6.7) in the textbook:

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum x_i\right)^2}{n} \tag{6.7}$$

Substitute the sums from Step 1:

$$S_{xx} = 1520 - \frac{(104)^2}{8} = 1520 - \frac{10816}{8}$$

$$S_{xx} = 1520 - 1352 = 168$$

Step 2b: The Co-Movement of x and y

Using (6.8)

Again, formula first — equation (6.8):

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n} \tag{6.8}$$

Step 2b: The Co-Movement of x and y

Using (6.8)

Again, formula first — equation (6.8):

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n} \tag{6.8}$$

Substitute:

$$S_{xy} = 3376.4 - \frac{(104)(245.2)}{8} = 3376.4 - \frac{25500.8}{8}$$

Step 2b: The Co-Movement of x and y

Using (6.8)

Again, formula first — equation (6.8):

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n} \tag{6.8}$$

Substitute:

$$S_{xy} = 3376.4 - \frac{(104)(245.2)}{8} = 3376.4 - \frac{25500.8}{8}$$

$$S_{xy} = 3376.4 - 3187.6 = 188.8$$

Step 2b: The Co-Movement of x and y

Using (6.8)

Again, formula first — equation (6.8):

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{\left(\sum x_i\right)\left(\sum y_i\right)}{n} \tag{6.8}$$

Substitute:

$$S_{xy} = 3376.4 - \frac{(104)(245.2)}{8} = 3376.4 - \frac{25500.8}{8}$$

$$S_{xy} = 3376.4 - 3187.6 = 188.8$$

$S_{xy} > 0$: heavier loads move *with* higher fuel consumption. The slope will be positive — exactly what the scatter plot showed.

Step 3: The Slope

Using (6.9)

The least squares slope — equation (6.9):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \tag{6.9}$$

Step 3: The Slope

Using (6.9)

The least squares slope — equation (6.9):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \tag{6.9}$$

Substitute:

$$\hat{\beta}_1 = \frac{188.8}{168} = 1.124$$

Step 3: The Slope

Using (6.9)

The least squares slope — equation (6.9):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \tag{6.9}$$

Substitute:

$$\hat{\beta}_1 = \frac{188.8}{168} = 1.124$$

Interpretation (say it in context): each additional tonne of payload increases expected fuel consumption by about **1.12 L per 100 km**.

Step 3: The Slope

Using (6.9)

The least squares slope — equation (6.9):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \tag{6.9}$$

Substitute:

$$\hat{\beta}_1 = \frac{188.8}{168} = 1.124$$

Interpretation (say it in context): each additional tonne of payload increases expected fuel consumption by about **1.12 L per 100 km**.

For a cost estimate: on the 400-km Dammam–Riyadh run, one extra tonne means roughly $4 \times 1.124 \approx 4.5$ extra litres of diesel per trip.

Step 4: The Intercept and the Fitted Line

Using (6.10) and (6.3)

The intercept — equation (6.10):

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{6.10}$$

Step 4: The Intercept and the Fitted Line

Using (6.10) and (6.3)

The intercept — equation (6.10):

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{6.10}$$

Substitute ($\bar{x} = 13$, $\bar{y} = 30.65$, $\hat{\beta}_1 = 1.124$):

$$\hat{\beta}_0 = 30.65 - (1.124)(13) = 30.65 - 14.61 = 16.04$$

Step 4: The Intercept and the Fitted Line

Using (6.10) and (6.3)

The intercept — equation (6.10):

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{6.10}$$

Substitute ($\bar{x} = 13$, $\bar{y} = 30.65$, $\hat{\beta}_1 = 1.124$):

$$\hat{\beta}_0 = 30.65 - (1.124)(13) = 30.65 - 14.61 = 16.04$$

Step 5: write the fitted regression line (6.3):

$$\boxed{\hat{y} = 16.04 + 1.124\,x}$$

Step 4: The Intercept and the Fitted Line

Using (6.10) and (6.3)

The intercept — equation (6.10):

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{6.10}$$

Substitute ($\bar{x} = 13$, $\bar{y} = 30.65$, $\hat{\beta}_1 = 1.124$):

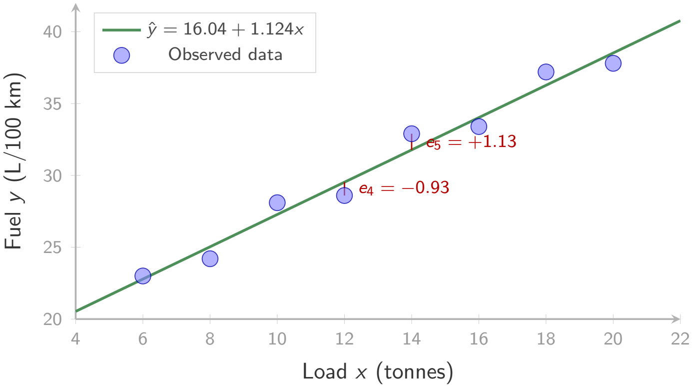
$$\hat{\beta}_0 = 30.65 - (1.124)(13) = 30.65 - 14.61 = 16.04$$

Step 5: write the fitted regression line (6.3):

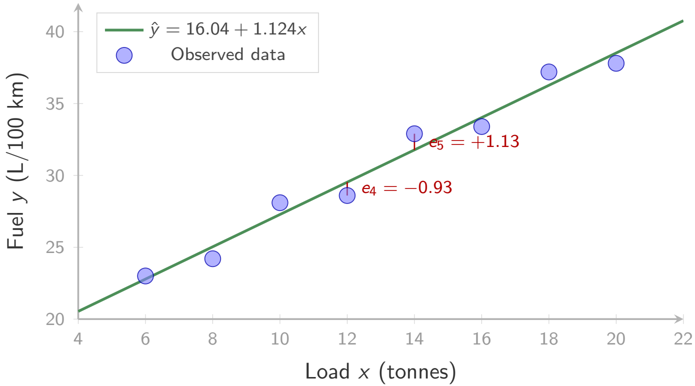
$$\boxed{\hat{y} = 16.04 + 1.124\,x}$$

The intercept (≈ 16 L/100 km) is the model's estimate for an *empty* truck. Note

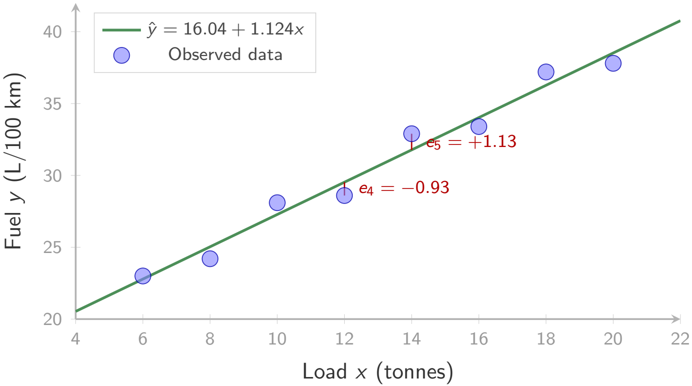
The Fitted Line Over the Data



The Fitted Line Over the Data



The Fitted Line Over the Data



The vertical gaps between points and line are the **residuals** (6.4). We compute all eight next.

Step 6: The Residuals

Using (6.4): $e_i = y_i - \hat{y}_i$

For each trip, compute the fitted value $\hat{y}_i = 16.04 + 1.124\, x_i$, then the residual:

Step 6: The Residuals

Using (6.4): $e_i = y_i - \hat{y}_i$

For each trip, compute the fitted value $\hat{y}_i = 16.04 + 1.124\, x_i$, then the residual:

i	x_i	y_i	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$
1	6	23.0	22.78	$+0.22$
2	8	24.2	25.03	-0.83
3	10	28.1	27.28	$+0.82$
4	12	28.6	29.53	-0.93
5	14	32.9	31.77	$+1.13$
6	16	33.4	34.02	-0.62
7	18	37.2	36.27	$+0.93$
8	20	37.8	38.52	-0.72

Step 6: The Residuals

Using (6.4): $e_i = y_i - \hat{y}_i$

For each trip, compute the fitted value $\hat{y}_i = 16.04 + 1.124\, x_i$, then the residual:

i	x_i	y_i	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$
1	6	23.0	22.78	$+0.22$
2	8	24.2	25.03	-0.83
3	10	28.1	27.28	$+0.82$
4	12	28.6	29.53	-0.93
5	14	32.9	31.77	$+1.13$
6	16	33.4	34.02	-0.62
7	18	37.2	36.27	$+0.93$
8	20	37.8	38.52	-0.72

Built-in check: $\sum e_i = 0.22 - 0.83 + 0.82 - 0.93 + 1.13 - 0.62 + 0.93 - 0.72 = 0.00$ ✓

Step 7: The Error Sum of Squares, Two Ways

(6.11) vs. the Shortcut (6.13)

Way 1 — direct, from the residual table (6.11):

$$SS_E = \sum_{i=1}^{8} e_i^2 = 0.22^2 + 0.83^2 + \cdots + 0.72^2 = 5.30$$

Step 7: The Error Sum of Squares, Two Ways

(6.11) vs. the Shortcut (6.13)

Way 1 — direct, from the residual table (6.11):

$$SS_E = \sum_{i=1}^{8} e_i^2 = 0.22^2 + 0.83^2 + \cdots + 0.72^2 = 5.30$$

Way 2 — shortcut, no residuals needed. First S_{yy} from (6.12):

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 7732.86 - \frac{(245.2)^2}{8} = 7732.86 - 7515.38 = 217.48$$

Step 7: The Error Sum of Squares, Two Ways

(6.11) vs. the Shortcut (6.13)

Way 1 — direct, from the residual table (6.11):

$$SS_E = \sum_{i=1}^{8} e_i^2 = 0.22^2 + 0.83^2 + \cdots + 0.72^2 = 5.30$$

Way 2 — shortcut, no residuals needed. First S_{yy} from (6.12):

$$S_{yy} = \sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n} = 7732.86 - \frac{(245.2)^2}{8} = 7732.86 - 7515.38 = 217.48$$

Then (6.13):

$$SS_E = S_{yy} - \hat{\beta}_1 \, S_{xy} = 217.48 - (1.124)(188.8) = 217.48 - 212.18 = 5.30$$

Step 7: The Error Sum of Squares, Two Ways

(6.11) vs. the Shortcut (6.13)

Way 1 — direct, from the residual table (6.11):

$$SS_E = \sum_{i=1}^{8} e_i^2 = 0.22^2 + 0.83^2 + \cdots + 0.72^2 = 5.30$$

Way 2 — shortcut, no residuals needed. First S_{yy} from (6.12):

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 7732.86 - \frac{(245.2)^2}{8} = 7732.86 - 7515.38 = 217.48$$

Then (6.13):

$$SS_E = S_{yy} - \hat{\beta}_1 S_{xy} = 217.48 - (1.124)(188.8) = 217.48 - 212.18 = 5.30$$

Step 8: Estimating the Error Variance

Using (6.14)

Formula first — equation (6.14):

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} \tag{6.14}$$

Step 8: Estimating the Error Variance

Using (6.14)

Formula first — equation (6.14):

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} \tag{6.14}$$

Substitute ($n = 8$, so $n - 2 = 6$ degrees of freedom):

$$\hat{\sigma}^2 = \frac{5.30}{6} = 0.884 \qquad \Rightarrow \qquad \hat{\sigma} = \sqrt{0.884} = 0.940 \text{ L}/100 \text{ km}$$

Step 8: Estimating the Error Variance

Using (6.14)

Formula first — equation (6.14):

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} \tag{6.14}$$

Substitute ($n = 8$, so $n - 2 = 6$ degrees of freedom):

$$\hat{\sigma}^2 = \frac{5.30}{6} = 0.884 \qquad \Rightarrow \qquad \hat{\sigma} = \sqrt{0.884} = 0.940 \text{ L}/100 \text{ km}$$

Interpretation: after accounting for load, trip-to-trip fuel consumption still scatters by about ± 0.94 L/100 km around the line — driver behavior, wind, traffic.

Step 8: Estimating the Error Variance

Using (6.14)

Formula first — equation (6.14):

$$\hat{\sigma}^2 = \frac{SS_E}{n-2} \tag{6.14}$$

Substitute ($n = 8$, so $n - 2 = 6$ degrees of freedom):

$$\hat{\sigma}^2 = \frac{5.30}{6} = 0.884 \qquad \Rightarrow \qquad \hat{\sigma} = \sqrt{0.884} = 0.940 \text{ L}/100 \text{ km}$$

Interpretation: after accounting for load, trip-to-trip fuel consumption still scatters by about ± 0.94 L/100 km around the line — driver behavior, wind, traffic.

The complete answer: $\hat{y} = 16.04 + 1.124\,x$ with $\hat{\sigma}^2 = 0.884$. These three numbers ($\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\sigma}^2$) are the full output of Procedure 6.1 — and the inputs to the

Practice Problem [5 pts]

Exam-Style: Startup Pricing Experiment

Problem statement. Your startup tests how the **discount level** (**x**, %) offered at checkout affects the **signup conversion rate** (**y**, %). Five discount levels were each shown to a large pilot group:

Practice Problem [5 pts]

Exam-Style: Startup Pricing Experiment

Problem statement. Your startup tests how the **discount level** (**x**, %) offered at checkout affects the **signup conversion rate** (**y**, %). Five discount levels were each shown to a large pilot group:

Observation	**1**	**2**	**3**	**4**	**5**
x (discount, %)	5	10	15	20	25
y (conversion, %)	3.1	4.0	4.4	5.3	5.7

Practice Problem [5 pts]

Exam-Style: Startup Pricing Experiment

Problem statement. Your startup tests how the **discount level** (**x**, %) offered at checkout affects the **signup conversion rate** (**y**, %). Five discount levels were each shown to a large pilot group:

Observation	**1**	**2**	**3**	**4**	**5**
x (discount, %)	5	10	15	20	25
y (conversion, %)	3.1	4.0	4.4	5.3	5.7

Summary statistics:

$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$
75	22.5	1375	105.55	370

Practice Problem [5 pts]

Exam-Style: Startup Pricing Experiment

Problem statement. Your startup tests how the **discount level** (**x**, %) offered at checkout affects the **signup conversion rate** (**y**, %). Five discount levels were each shown to a large pilot group:

Observation	**1**	**2**	**3**	**4**	**5**
x (discount, %)	5	10	15	20	25
y (conversion, %)	3.1	4.0	4.4	5.3	5.7

Summary statistics:

$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$
75	22.5	1375	105.55	370

Practice Problem — Solution

Part (a): Fit the Model

Compute S_{xx} (6.7) and S_{xy} (6.8):

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 1375 - \frac{(75)^2}{5} = 1375 - 1125 = 250$$

$$S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} = 370 - \frac{(75)(22.5)}{5} = 370 - 337.5 = 32.5$$

Practice Problem — Solution

Part (a): Fit the Model

Compute S_{xx} (6.7) and S_{xy} (6.8):

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 1375 - \frac{(75)^2}{5} = 1375 - 1125 = 250$$

$$S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} = 370 - \frac{(75)(22.5)}{5} = 370 - 337.5 = 32.5$$

Slope (6.9):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{32.5}{250} = 0.13$$

Practice Problem — Solution

Part (a): Fit the Model

Compute S_{xx} (6.7) and S_{xy} (6.8):

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 1375 - \frac{(75)^2}{5} = 1375 - 1125 = 250$$

$$S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} = 370 - \frac{(75)(22.5)}{5} = 370 - 337.5 = 32.5$$

Slope (6.9):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{32.5}{250} = 0.13$$

Intercept (6.10): $\bar{x} = 75/5 = 15, \quad \bar{y} = 22.5/5 = 4.5.$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 4.5 - 0.13(15) = 4.5 - 1.95 = 2.55$$

Practice Problem — Solution

Part (a): Fit the Model

Compute S_{xx} (6.7) and S_{xy} (6.8):

$$S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 1375 - \frac{(75)^2}{5} = 1375 - 1125 = 250$$

$$S_{xy} = \sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{n} = 370 - \frac{(75)(22.5)}{5} = 370 - 337.5 = 32.5$$

Slope (6.9):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{32.5}{250} = 0.13$$

Intercept (6.10): $\bar{x} = 75/5 = 15, \quad \bar{y} = 22.5/5 = 4.5.$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 4.5 - 0.13(15) = 4.5 - 1.95 = 2.55$$

Practice Problem — Solution (Continued)

Parts (b), (c), and (d)

(b) Estimate conversion at $x = 20$:

$$\hat{y} = 2.55 + 0.13(20) = 2.55 + 2.60 = \boxed{5.15\%}$$

Practice Problem — Solution (Continued)

Parts (b), (c), and (d)

(b) Estimate conversion at $x = 20$:

$$\hat{y} = 2.55 + 0.13(20) = 2.55 + 2.60 = \boxed{5.15\%}$$

(c) Residual (6.4) for the observation at $x = 20$ (where $y = 5.3$):

$$e = y - \hat{y} = 5.3 - 5.15 = \boxed{+0.15}$$

Since $e > 0$ (actual $>$ predicted), the model **underestimates** the true value there.

Practice Problem — Solution (Continued)

Parts (b), (c), and (d)

(b) Estimate conversion at $x = 20$:

$$\hat{y} = 2.55 + 0.13(20) = 2.55 + 2.60 = \boxed{5.15\%}$$

(c) Residual (6.4) for the observation at $x = 20$ (where $y = 5.3$):

$$e = y - \hat{y} = 5.3 - 5.15 = \boxed{+0.15}$$

Since $e > 0$ (actual $>$ predicted), the model **underestimates** the true value there.

(d) Expected change for 8 additional discount points:

$$\Delta\hat{y} = \hat{\beta}_1 \times 8 = 0.13 \times 8 = \boxed{1.04 \text{ percentage points}}$$

Interpretation: the slope $\hat{\beta}_1 = 0.13$ means each extra point of discount is expected

Lecture 17 Summary

- Procedure 6.1 runs: sums $\to S_{xx}$ (6.7), S_{xy} (6.8) $\to \hat{\beta}_1$ (6.9) $\to \hat{\beta}_0$ (6.10) $\to$ fitted line (6.3).

Lecture 17 Summary

- Procedure 6.1 runs: sums $\to S_{xx}$ (6.7), S_{xy} (6.8) $\to \hat{\beta}_1$ (6.9) $\to \hat{\beta}_0$ (6.10) $\to$ fitted line (6.3).
- **Residuals** $e_i = y_i - \hat{y}_i$ (6.4) always sum to zero — use this as a check.

Lecture 17 Summary

- Procedure 6.1 runs: sums $\rightarrow S_{xx}$ (6.7), S_{xy} (6.8) $\rightarrow \hat{\beta}_1$ (6.9) $\rightarrow \hat{\beta}_0$ (6.10) $\rightarrow$ fitted line (6.3).
- **Residuals** $e_i = y_i - \hat{y}_i$ (6.4) always sum to zero — use this as a check.
- SS_E two ways: direct (6.11) or shortcut $S_{yy} - \hat{\beta}_1 S_{xy}$ (6.13); they must agree.

Lecture 17 Summary

- Procedure 6.1 runs: sums $\to S_{xx}$ (6.7), S_{xy} (6.8) $\to \hat{\beta}_1$ (6.9) $\to \hat{\beta}_0$ (6.10) $\to$ fitted line (6.3).
- **Residuals** $e_i = y_i - \hat{y}_i$ (6.4) always sum to zero — use this as a check.
- SS_E two ways: direct (6.11) or shortcut $S_{yy} - \hat{\beta}_1 S_{xy}$ (6.13); they must agree.
- $\hat{\sigma}^2 = SS_E/(n-2)$ (6.14) measures the remaining scatter around the line.

Lecture 17 Summary

- Procedure 6.1 runs: sums $\rightarrow S_{xx}$ (6.7), S_{xy} (6.8) $\rightarrow \hat{\beta}_1$ (6.9) $\rightarrow \hat{\beta}_0$ (6.10) $\rightarrow$ fitted line (6.3).
- **Residuals** $e_i = y_i - \hat{y}_i$ (6.4) always sum to zero — use this as a check.
- SS_E two ways: direct (6.11) or shortcut $S_{yy} - \hat{\beta}_1 S_{xy}$ (6.13); they must agree.
- $\hat{\sigma}^2 = SS_E/(n-2)$ (6.14) measures the remaining scatter around the line.
- Always interpret the slope **in context**, and never extrapolate far outside the observed x range.

Lecture 17 Summary

- Procedure 6.1 runs: sums $\rightarrow S_{xx}$ (6.7), S_{xy} (6.8) $\rightarrow \hat{\beta}_1$ (6.9) $\rightarrow \hat{\beta}_0$ (6.10) $\rightarrow$ fitted line (6.3).
- **Residuals** $e_i = y_i - \hat{y}_i$ (6.4) always sum to zero — use this as a check.
- SS_E two ways: direct (6.11) or shortcut $S_{yy} - \hat{\beta}_1 S_{xy}$ (6.13); they must agree.
- $\hat{\sigma}^2 = SS_E/(n-2)$ (6.14) measures the remaining scatter around the line.
- Always interpret the slope **in context**, and never extrapolate far outside the observed x range.
- **Next lecture:** tests and confidence intervals on β_1 and β_0, and the regression ANOVA table.

Reminders

- **Quiz on Chapter 6 basics** is due before class — it covers today's eight steps

Reminders

- **Quiz on Chapter 6 basics** is due before class — it covers today's eight steps
- Homework 6 (regression fitting) is posted on Blackboard

Reminders

- **Quiz on Chapter 6 basics** is due before class — it covers today's eight steps
- Homework 6 (regression fitting) is posted on Blackboard
- Office hours: Tuesdays 9–10 AM, Room 22-219 or Zoom.

Reminders

- **Quiz on Chapter 6 basics** is due before class — it covers today's eight steps
- Homework 6 (regression fitting) is posted on Blackboard
- Office hours: Tuesdays 9–10 AM, Room 22-219 or Zoom.
- Read §6.7 and §6.8 to prepare for the next lecture.