ISE 315: Engineering Statistics

Lecture 19: Simple Linear Regression (Part 3) — Prediction, R^2, and Model Adequacy

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 19

Simple Linear Regression — Part 3: Intervals for Prediction, R^2, Residual Analysis, and Cautions (Chapter 6, cont'd)

Lecture 19 Outline

- Quick review: slope inference and ANOVA (from Lecture 18)

Lecture 19 Outline

- Quick review: slope inference and ANOVA (from Lecture 18)
- Predicting at x_0: CI for the mean response (6.26) vs. PI for a new observation (6.27), via Procedure 6.3

Lecture 19 Outline

- Quick review: slope inference and ANOVA (from Lecture 18)
- Predicting at x_0: CI for the mean response (6.26) vs. PI for a new observation (6.27), via Procedure 6.3
- Coefficient of determination R^2 (6.21) and correlation r (6.22)

Lecture 19 Outline

- Quick review: slope inference and ANOVA (from Lecture 18)
- Predicting at x_0: CI for the mean response (6.26) vs. PI for a new observation (6.27), via Procedure 6.3
- Coefficient of determination R^2 (6.21) and correlation r (6.22)
- Residual analysis: Procedure 6.4 and standardized residuals (6.28)

Lecture 19 Outline

- Quick review: slope inference and ANOVA (from Lecture 18)
- Predicting at x_0: CI for the mean response (6.26) vs. PI for a new observation (6.27), via Procedure 6.3
- Coefficient of determination R^2 (6.21) and correlation r (6.22)
- Residual analysis: Procedure 6.4 and standardized residuals (6.28)
- Abuses of regression: extrapolation, causation, R^2 worship

Lecture 19 Outline

- Quick review: slope inference and ANOVA (from Lecture 18)
- Predicting at x_0: CI for the mean response (6.26) vs. PI for a new observation (6.27), via Procedure 6.3
- Coefficient of determination R^2 (6.21) and correlation r (6.22)
- Residual analysis: Procedure 6.4 and standardized residuals (6.28)
- Abuses of regression: extrapolation, causation, R^2 worship
- Exam-style practice problem (AI safety red-team data)

Quick Review

What We Did in Lecture 18

Lecture 18 Recap: The Fuel Model Passed Its Tests

The model (truck load x in tonnes, fuel y in L/100 km, $n = 8$):

$$\hat{y} = 16.04 + 1.124\,x, \qquad \hat{\sigma}^2 = 0.884, \qquad S_{xx} = 168, \quad \bar{x} = 13$$

Lecture 18 Recap: The Fuel Model Passed Its Tests

The model (truck load x in tonnes, fuel y in L/100 km, $n = 8$):

$$\hat{y} = 16.04 + 1.124\,x, \qquad \hat{\sigma}^2 = 0.884, \qquad S_{xx} = 168, \quad \bar{x} = 13$$

Slope test (6.18): $T_0 = 1.124/0.0725 = 15.50 > t_{0.025,\,6} = 2.447 \Rightarrow$ reject H_0: $\beta_1 = 0$.

Lecture 18 Recap: The Fuel Model Passed Its Tests

The model (truck load x in tonnes, fuel y in L/100 km, $n = 8$):

$$\hat{y} = 16.04 + 1.124\,x, \qquad \hat{\sigma}^2 = 0.884, \qquad S_{xx} = 168, \quad \bar{x} = 13$$

Slope test (6.18): $T_0 = 1.124/0.0725 = 15.50 > t_{0.025,\,6} = 2.447 \Rightarrow$ reject H_0: $\beta_1 = 0$.

ANOVA (6.19), (6.20): $SS_T = 217.48 = 212.18 + 5.30$, $F_0 = 240.0 = T_0^2$. Same verdict.

Lecture 18 Recap: The Fuel Model Passed Its Tests

The model (truck load x in tonnes, fuel y in L/100 km, $n = 8$):

$$\hat{y} = 16.04 + 1.124\,x, \qquad \hat{\sigma}^2 = 0.884, \qquad S_{xx} = 168, \quad \bar{x} = 13$$

Slope test (6.18): $T_0 = 1.124/0.0725 = 15.50 > t_{0.025,\,6} = 2.447 \Rightarrow$ reject H_0: $\beta_1 = 0$.

ANOVA (6.19), (6.20): $SS_T = 217.48 = 212.18 + 5.30$, $F_0 = 240.0 = T_0^2$. Same verdict.

95% CI for the slope (6.23): $(0.947,\ 1.301)$ L/100 km per tonne.

Lecture 18 Recap: The Fuel Model Passed Its Tests

The model (truck load x in tonnes, fuel y in L/100 km, $n = 8$):

$$\hat{y} = 16.04 + 1.124\,x, \qquad \hat{\sigma}^2 = 0.884, \qquad S_{xx} = 168, \quad \bar{x} = 13$$

Slope test (6.18): $T_0 = 1.124/0.0725 = 15.50 > t_{0.025,\,6} = 2.447 \Rightarrow$ reject H_0: $\beta_1 = 0$.

ANOVA (6.19), (6.20): $SS_T = 217.48 = 212.18 + 5.30$, $F_0 = 240.0 = T_0^2$. Same verdict.

95% CI for the slope (6.23): $(0.947,\ 1.301)$ L/100 km per tonne.

The relationship is real. **Today we put the line to work** — and then check whether we should trust it at all.

Confidence and Prediction Intervals

Two Different Questions at the Same x_0

Two Different Questions at $x_0 = 16$ Tonnes

Procedure 6.3

Question A (planning): across *all* future 16-tonne trips, what is the **average** fuel consumption? $\rightarrow$ CI for the mean response $E(Y \mid x_0)$, (6.26)

Two Different Questions at $x_0 = 16$ Tonnes

Procedure 6.3

Question A (planning): across *all* future 16-tonne trips, what is the **average** fuel consumption? $\rightarrow$ CI for the mean response $E(Y \mid x_0)$, (6.26)

Question B (dispatch): *tomorrow's specific* 16-tonne trip — what fuel will **that one trip** use? $\rightarrow$ PI for a new observation Y_{new}, (6.27)

Two Different Questions at $x_0 = 16$ Tonnes

Procedure 6.3

Question A (planning): across *all* future 16-tonne trips, what is the **average** fuel consumption? $\rightarrow$ CI for the mean response $E(Y \mid x_0)$, (6.26)

Question B (dispatch): *tomorrow's specific* 16-tonne trip — what fuel will **that one trip** use? $\rightarrow$ PI for a new observation Y_{new}, (6.27)

Procedure 6.3 handles both:

1. Check that x_0 is inside the observed range of x.
2. Compute the point estimate $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ (6.3).
3. Decide: mean response (CI) or single new observation (PI)?
4. Compute the interval — (6.26) or (6.27) — and interpret in context.

Two Different Questions at $x_0 = 16$ Tonnes

Procedure 6.3

Question A (planning): across *all* future 16-tonne trips, what is the **average** fuel consumption? $\rightarrow$ CI for the mean response $E(Y \mid x_0)$, (6.26)

Question B (dispatch): *tomorrow's specific* 16-tonne trip — what fuel will **that one trip** use? $\rightarrow$ PI for a new observation Y_{new}, (6.27)

Procedure 6.3 handles both:

1. Check that x_0 is inside the observed range of x.
2. Compute the point estimate $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$ (6.3).
3. Decide: mean response (CI) or single new observation (PI)?
4. Compute the interval — (6.26) or (6.27) — and interpret in context.

Textbook: §6.11 ((6.26) onward)

CI for the Mean Response

Using (6.26)

A $100(1-\alpha)\%$ confidence interval on $E(Y \mid x_0)$ — equation (6.26):

$$\hat{y}_0 \pm t_{\alpha/2,\, n-2} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \qquad (6.26)$$

CI for the Mean Response

Using (6.26)

A $100(1-\alpha)\%$ confidence interval on $E(Y \mid x_0)$ — equation (6.26):

$$\hat{y}_0 \pm t_{\alpha/2,\, n-2} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \tag{6.26}$$

Read the standard error: it has a $1/n$ term (uncertainty in the line's height) plus a $(x_0 - \bar{x})^2 / S_{xx}$ term (uncertainty in the line's tilt).

CI for the Mean Response

Using (6.26)

A $100(1-\alpha)\%$ confidence interval on $E(Y \mid x_0)$ — equation (6.26):

$$\hat{y}_0 \pm t_{\alpha/2,\, n-2} \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \tag{6.26}$$

Read the standard error: it has a $1/n$ term (uncertainty in the line's height) plus a $(x_0 - \bar{x})^2 / S_{xx}$ term (uncertainty in the line's tilt).

Consequence: the interval is **narrowest at** $x_0 = \bar{x}$ and widens as x_0 moves toward the edges of the data. The line is best pinned down at its center of mass.

PI for a New Observation

Using (6.27)

A $100(1-\alpha)\%$ prediction interval for Y_{new} at x_0 — equation (6.27):

$$\hat{y}_0 \pm t_{\alpha/2,\, n-2} \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \qquad (6.27)$$

PI for a New Observation

Using (6.27)

A $100(1-\alpha)\%$ prediction interval for Y_{new} at x_0 — equation (6.27):

$$\hat{y}_0 \pm t_{\alpha/2,\,n-2}\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)} \tag{6.27}$$

The extra "$1+$" is the variance of one new observation scattering around the true line. It does **not** shrink as n grows — one trip is always noisy, no matter how well we know the line.

PI for a New Observation

Using (6.27)

A $100(1-\alpha)\%$ prediction interval for Y_{new} at x_0 — equation (6.27):

$$\hat{y}_0 \pm t_{\alpha/2,\,n-2}\sqrt{\hat{\sigma}^2\left(1+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)} \qquad (6.27)$$

The extra "$1+$" is the variance of one new observation scattering around the true line. It does **not** shrink as n grows — one trip is always noisy, no matter how well we know the line.

Consequence: a PI is **always wider** than the CI at the same x_0 and the same confidence level.

CI vs. PI at a Glance

Table 6.8 in the textbook

	CI (6.26)	**PI (6.27)**
Target	$E(Y \mid x_0)$: the mean	Y_{new}: one new observation
Extra "1+" term	no	yes
Width	narrower	wider
Shrinks as $n \to \infty$?	yes, toward zero width	no, floor of $\pm z\,\sigma$ remains
Typical use	planning, budgeting averages	quoting one specific job

CI vs. PI at a Glance

Table 6.8 in the textbook

	CI (6.26)	**PI (6.27)**
Target	$E(Y \mid x_0)$: the mean	Y_{new}: one new observation
Extra "1+" term	no	yes
Width	narrower	wider
Shrinks as $n \to \infty$?	yes, toward zero width	no, floor of $\pm z\,\sigma$ remains
Typical use	planning, budgeting averages	quoting one specific job

Read the question carefully. "Estimate the *mean* fuel use at 16 tonnes" $\rightarrow$ CI. "Predict the fuel use of the *next* 16-tonne trip" $\rightarrow$ PI. The single word *mean* vs. *next* decides the formula.

Worked Example: 95% CI for Mean Fuel Use at $x_0 = 16$

Step 1: $x_0 = 16$ is inside the observed range 6–20 tonnes. ✓

Worked Example: 95% CI for Mean Fuel Use at $x_0 = 16$

Step 1: $x_0 = 16$ is inside the observed range 6–20 tonnes. ✓

Step 2 — point estimate (6.3): $\hat{y}_0 = 16.04 + 1.124(16) = 16.04 + 17.98 = 34.02$.

Worked Example: 95% CI for Mean Fuel Use at $x_0 = 16$

Step 1: $x_0 = 16$ is inside the observed range 6–20 tonnes. ✓

Step 2 — point estimate (6.3): $\hat{y}_0 = 16.04 + 1.124(16) = 16.04 + 17.98 = 34.02$.

Step 3 — standard error (no "1+"; this is a CI) (6.26):

$$se = \sqrt{0.884\left(\frac{1}{8} + \frac{(16-13)^2}{168}\right)} = \sqrt{0.884\,(0.125 + 0.0536)} = \sqrt{0.1579} = 0.397$$

Worked Example: 95% CI for Mean Fuel Use at $x_0 = 16$

Step 1: $x_0 = 16$ is inside the observed range 6–20 tonnes. ✓

Step 2 — point estimate (6.3): $\hat{y}_0 = 16.04 + 1.124(16) = 16.04 + 17.98 = 34.02$.

Step 3 — standard error (no "1+"; this is a CI) (6.26):

$$se = \sqrt{0.884\left(\frac{1}{8} + \frac{(16-13)^2}{168}\right)} = \sqrt{0.884\,(0.125 + 0.0536)} = \sqrt{0.1579} = 0.397$$

Step 4 — interval, with $t_{0.025,\,6} = 2.447$:

$$34.02 \pm 2.447 \times 0.397 = 34.02 \pm 0.97 \quad \Longrightarrow \quad \boxed{(33.05,\ 34.99)}$$

Worked Example: 95% CI for Mean Fuel Use at $x_0 = 16$

Step 1: $x_0 = 16$ is inside the observed range 6–20 tonnes. ✓

Step 2 — point estimate (6.3): $\hat{y}_0 = 16.04 + 1.124(16) = 16.04 + 17.98 = 34.02$.

Step 3 — standard error (no "1+"; this is a CI) (6.26):

$$se = \sqrt{0.884\left(\frac{1}{8} + \frac{(16-13)^2}{168}\right)} = \sqrt{0.884\,(0.125 + 0.0536)} = \sqrt{0.1579} = 0.397$$

Step 4 — interval, with $t_{0.025,\,6} = 2.447$:

$$34.02 \pm 2.447 \times 0.397 = 34.02 \pm 0.97 \quad \Longrightarrow \quad \boxed{(33.05,\ 34.99)}$$

We are 95% confident the *average* fuel use across all 16-tonne trips is between 33.1 and 35.0 L/100 km — tight enough for the analyst's cost model.

Worked Example: 95% PI for Tomorrow's 16-Tonne Trip

Standard error — same pieces, plus the "$1+$" (6.27):

$$se = \sqrt{0.884\,(1 + 0.125 + 0.0536)} = \sqrt{0.884 \times 1.179} = \sqrt{1.042} = 1.021$$

Worked Example: 95% PI for Tomorrow's 16-Tonne Trip

Standard error — same pieces, plus the "1 +" (6.27):

$$se = \sqrt{0.884\,(1 + 0.125 + 0.0536)} = \sqrt{0.884 \times 1.179} = \sqrt{1.042} = 1.021$$

Interval:

$$34.02 \pm 2.447 \times 1.021 = 34.02 \pm 2.50 \quad \Longrightarrow \quad \boxed{(31.52,\ 36.52)}$$

Worked Example: 95% PI for Tomorrow's 16-Tonne Trip

Standard error — same pieces, plus the "$1+$" (6.27):

$$se = \sqrt{0.884\,(1 + 0.125 + 0.0536)} = \sqrt{0.884 \times 1.179} = \sqrt{1.042} = 1.021$$

Interval:

$$34.02 \pm 2.447 \times 1.021 = 34.02 \pm 2.50 \quad \Longrightarrow \quad \boxed{(31.52,\ 36.52)}$$

Compare at the same $x_0 = 16$:

	Interval	**Width**
CI for the mean	(33.05, 34.99)	1.94
PI for one trip	(31.52, 36.52)	5.00

Worked Example: 95% PI for Tomorrow's 16-Tonne Trip

Standard error — same pieces, plus the "$1+$" (6.27):

$$se = \sqrt{0.884\,(1 + 0.125 + 0.0536)} = \sqrt{0.884 \times 1.179} = \sqrt{1.042} = 1.021$$

Interval:

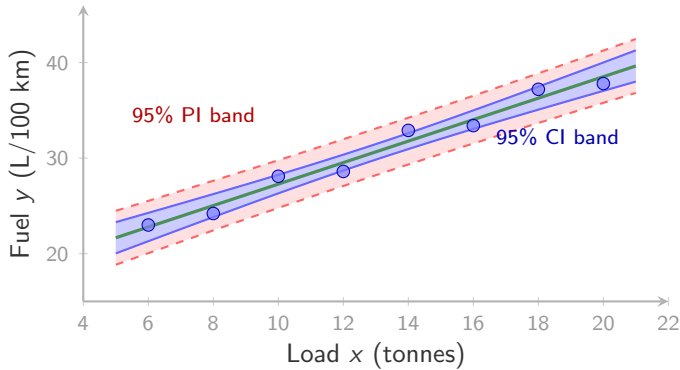
$$34.02 \pm 2.447 \times 1.021 = 34.02 \pm 2.50 \quad \Longrightarrow \quad \boxed{(31.52,\ 36.52)}$$

Compare at the same $x_0 = 16$:

	Interval	**Width**
CI for the mean	(33.05, 34.99)	1.94
PI for one trip	(31.52, 36.52)	5.00

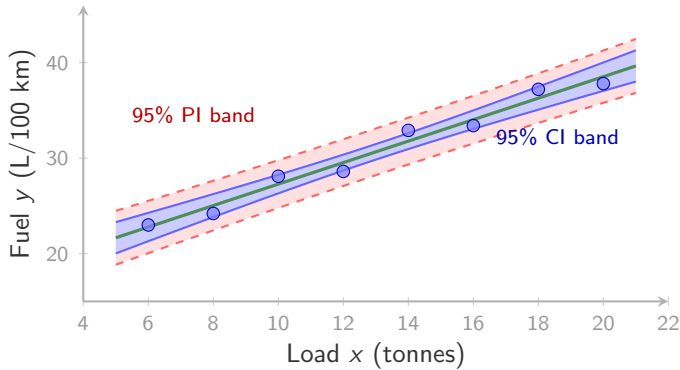
The Full Picture: CI and PI Bands Along the Line

Compare Figure 6.7 in the textbook



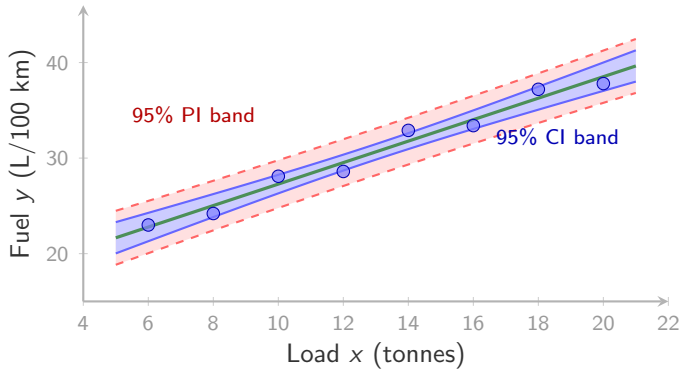
The Full Picture: CI and PI Bands Along the Line

Compare Figure 6.7 in the textbook



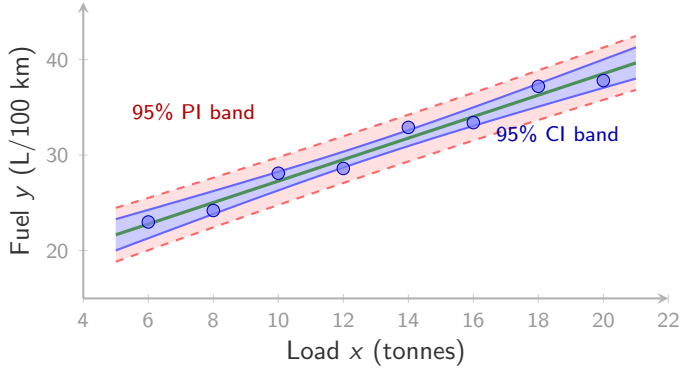
The Full Picture: CI and PI Bands Along the Line

Compare Figure 6.7 in the textbook



The Full Picture: CI and PI Bands Along the Line

Compare Figure 6.7 in the textbook



Both bands are narrowest at $\bar{x} = 13$ and flare toward the edges; the PI band stays wider everywhere. Beyond $x = 20$ there is no data at all — the bands warn you, then extrapolation ignores the warning.

R^2 and Correlation

How Much Does the Line Explain?

Coefficient of Determination R^2

Using (6.21)

Definition — equation (6.21):

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{6.21}$$

Coefficient of Determination R^2

Using (6.21)

Definition — equation (6.21):

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{6.21}$$

R^2 is the **proportion of the total variability** in y explained by the linear relationship with x. It lives between 0 and 1.

Coefficient of Determination R^2

Using (6.21)

Definition — equation (6.21):

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{6.21}$$

R^2 is the **proportion of the total variability** in y explained by the linear relationship with x. It lives between 0 and 1.

- $R^2 = 1$: all points exactly on the line ($SS_E = 0$)
- $R^2 = 0$: the line explains nothing ($SS_R = 0$)
- In between: part explained, part left in the residuals

Coefficient of Determination R^2

Using (6.21)

Definition — equation (6.21):

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{6.21}$$

R^2 is the **proportion of the total variability** in y explained by the linear relationship with x. It lives between 0 and 1.

- $R^2 = 1$: all points exactly on the line ($SS_E = 0$)
- $R^2 = 0$: the line explains nothing ($SS_R = 0$)
- In between: part explained, part left in the residuals

Caution now, details later: a high R^2 does **not** certify that the model is adequate, and it does not prove causation. The residual checks in the second half of today

Worked Example: R^2 for the Fuel Model

From Lecture 18's ANOVA table: $SS_R = 212.18$, $SS_E = 5.30$, $SS_T = 217.48$.

Worked Example: R^2 for the Fuel Model

From Lecture 18's ANOVA table: $SS_R = 212.18$, $SS_E = 5.30$, $SS_T = 217.48$.

Apply (6.21):

$$R^2 = \frac{SS_R}{SS_T} = \frac{212.18}{217.48} = 0.9756$$

Worked Example: R^2 for the Fuel Model

From Lecture 18's ANOVA table: $SS_R = 212.18$, $SS_E = 5.30$, $SS_T = 217.48$.

Apply (6.21):

$$R^2 = \frac{SS_R}{SS_T} = \frac{212.18}{217.48} = 0.9756$$

Check with the other form: $1 - SS_E/SS_T = 1 - 5.30/217.48 = 1 - 0.0244 = 0.9756$ ✓

Worked Example: R^2 for the Fuel Model

From Lecture 18's ANOVA table: $SS_R = 212.18$, $SS_E = 5.30$, $SS_T = 217.48$.

Apply (6.21):

$$R^2 = \frac{SS_R}{SS_T} = \frac{212.18}{217.48} = 0.9756$$

Check with the other form: $1 - SS_E/SS_T = 1 - 5.30/217.48 = 1 - 0.0244 = 0.9756$ ✓

Interpretation: about 97.6% of the trip-to-trip variability in fuel consumption is explained by payload. The remaining 2.4% — driver behavior, wind, traffic — is the scatter that $\hat{\sigma}^2 = 0.884$ measures.

The Sample Correlation Coefficient r

Using (6.22)

Definition — equation (6.22):

$$r = \frac{S_{xy}}{\sqrt{S_{xx}\, S_{yy}}} \tag{6.22}$$

r measures the strength *and direction* of the linear association: $-1 \leq r \leq 1$.

The Sample Correlation Coefficient r

Using (6.22)

Definition — equation (6.22):

$$r = \frac{S_{xy}}{\sqrt{S_{xx}\, S_{yy}}} \tag{6.22}$$

r measures the strength *and direction* of the linear association: $-1 \leq r \leq 1$.

Two facts to remember:

- In simple linear regression, $r^2 = R^2$.
- The sign of r always matches the sign of $\hat{\beta}_1$ (both inherit it from S_{xy}).

The Sample Correlation Coefficient r

Using (6.22)

Definition — equation (6.22):

$$r = \frac{S_{xy}}{\sqrt{S_{xx}\, S_{yy}}} \tag{6.22}$$

r measures the strength *and direction* of the linear association: $-1 \leq r \leq 1$.

Two facts to remember:

- In simple linear regression, $r^2 = R^2$.
- The sign of r always matches the sign of $\hat{\beta}_1$ (both inherit it from S_{xy}).

Fuel data:

$$r = \frac{188.8}{\sqrt{168 \times 217.48}} = \frac{188.8}{\sqrt{36536.6}} = \frac{188.8}{191.1} = 0.9878$$

The Sample Correlation Coefficient r

Using (6.22)

Definition — equation (6.22):

$$r = \frac{S_{xy}}{\sqrt{S_{xx}\, S_{yy}}} \tag{6.22}$$

r measures the strength *and direction* of the linear association: $-1 \leq r \leq 1$.
Two facts to remember:

- In simple linear regression, $r^2 = R^2$.
- The sign of r always matches the sign of $\hat{\beta}_1$ (both inherit it from S_{xy}).

Fuel data:

$$r = \frac{188.8}{\sqrt{168 \times 217.48}} = \frac{188.8}{\sqrt{36536.6}} = \frac{188.8}{191.1} = 0.9878$$

Quick Check: R^2 for the Startup Pricing Experiment

From Lecture 17's practice problem

Recall: discount x (%) vs. conversion y (%), $n = 5$, $\hat{y} = 2.55 + 0.13\,x$, with $S_{xx} = 250$, $S_{xy} = 32.5$, $S_{yy} = 4.30$. **Compute R^2 and r.**

Quick Check: R^2 for the Startup Pricing Experiment

From Lecture 17's practice problem

Recall: discount x (%) vs. conversion y (%), $n = 5$, $\hat{y} = 2.55 + 0.13\,x$, with $S_{xx} = 250$, $S_{xy} = 32.5$, $S_{yy} = 4.30$. **Compute R^2 and r.**

Sums of squares: $SS_T = S_{yy} = 4.30$, $SS_R = \hat{\beta}_1 S_{xy} = 0.13 \times 32.5 = 4.225$.

Quick Check: R^2 for the Startup Pricing Experiment

From Lecture 17's practice problem

Recall: discount x (%) vs. conversion y (%), $n = 5$, $\hat{y} = 2.55 + 0.13\,x$, with $S_{xx} = 250$, $S_{xy} = 32.5$, $S_{yy} = 4.30$. **Compute R^2 and r.**

Sums of squares: $SS_T = S_{yy} = 4.30$, $SS_R = \hat{\beta}_1 S_{xy} = 0.13 \times 32.5 = 4.225$.

Apply (6.21): $$R^2 = \frac{4.225}{4.30} = 0.9826$$

Quick Check: R^2 for the Startup Pricing Experiment

From Lecture 17's practice problem

Recall: discount x (%) vs. conversion y (%), $n = 5$, $\hat{y} = 2.55 + 0.13\,x$, with $S_{xx} = 250$, $S_{xy} = 32.5$, $S_{yy} = 4.30$. **Compute R^2 and r.**

Sums of squares: $SS_T = S_{yy} = 4.30$, $SS_R = \hat{\beta}_1 S_{xy} = 0.13 \times 32.5 = 4.225$.

Apply (6.21): $$R^2 = \frac{4.225}{4.30} = 0.9826$$

Apply (6.22): $$r = \frac{32.5}{\sqrt{250 \times 4.30}} = \frac{32.5}{32.79} = 0.9913, \quad \text{and } r^2 = 0.9826 \checkmark$$

Quick Check: R^2 for the Startup Pricing Experiment

From Lecture 17's practice problem

Recall: discount x (%) vs. conversion y (%), $n = 5$, $\hat{y} = 2.55 + 0.13\,x$, with $S_{xx} = 250$, $S_{xy} = 32.5$, $S_{yy} = 4.30$. **Compute R^2 and r.**

Sums of squares: $SS_T = S_{yy} = 4.30, \quad SS_R = \hat{\beta}_1 S_{xy} = 0.13 \times 32.5 = 4.225.$

Apply (6.21): $R^2 = \dfrac{4.225}{4.30} = 0.9826$

Apply (6.22): $r = \dfrac{32.5}{\sqrt{250 \times 4.30}} = \dfrac{32.5}{32.79} = 0.9913, \quad$ and $r^2 = 0.9826$ ✓

About 98% of the variation in conversion across the five pilot groups is explained by the discount level. With $n = 5$, though, that impressive number rests on very little data — another reason the intervals from Lecture 18 matter.

Residual Analysis

Is the Linear Model Adequate?

Why Residual Analysis?

Procedure 6.4

A significant F-test and a high R^2 say the line *explains variability*. They do **not** say the model *assumptions hold*: errors with mean zero, constant variance σ^2, uncorrelated, and (for our tests and intervals) normally distributed.

Why Residual Analysis?

Procedure 6.4

A significant F-test and a high R^2 say the line *explains variability*. They do **not** say the model *assumptions hold*: errors with mean zero, constant variance σ^2, uncorrelated, and (for our tests and intervals) normally distributed.

Procedure 6.4 — the adequacy checklist:

1. Compute the residuals $e_i = y_i - \hat{y}_i$ (6.4).
2. Plot e_i vs. $\hat{y}_i$: look for random scatter, no curve, no funnel.
3. Check normality (normal probability plot, or a histogram for large n).
4. Standardize: $d_i = e_i / \sqrt{\hat{\sigma}^2}$ (6.28); flag $|d_i| > 2$ as potential outliers.

Why Residual Analysis?

Procedure 6.4

A significant F-test and a high R^2 say the line *explains variability*. They do **not** say the model *assumptions hold*: errors with mean zero, constant variance σ^2, uncorrelated, and (for our tests and intervals) normally distributed.

Procedure 6.4 — the adequacy checklist:

1. Compute the residuals $e_i = y_i - \hat{y}_i$ (6.4).
2. Plot e_i vs. $\hat{y}_i$: look for random scatter, no curve, no funnel.
3. Check normality (normal probability plot, or a histogram for large n).
4. Standardize: $d_i = e_i / \sqrt{\hat{\sigma}^2}$ (6.28); flag $|d_i| > 2$ as potential outliers.

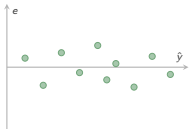
If the residuals look like structureless noise, the model has earned its intervals. If they show a pattern, the model is missing something.

Textbook: §6.12 ((6.28))

Residual Patterns: One Good, Two Problematic

Compare Figure 6.8 in the textbook

Random scatter

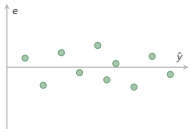


Assumptions look fine.

Residual Patterns: One Good, Two Problematic

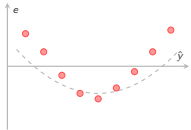
Compare Figure 6.8 in the textbook

Random scatter



Assumptions look fine.

Curved (U-shape)

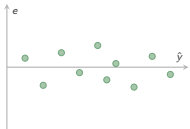


Relationship is nonlinear.

Residual Patterns: One Good, Two Problematic

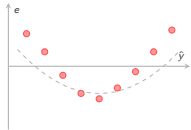
Compare Figure 6.8 in the textbook

Random scatter



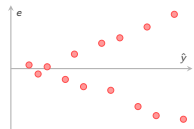
Assumptions look fine.

Curved (U-shape)



Relationship is nonlinear.

Funnel (fan)

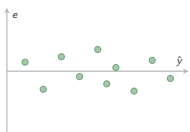


Variance is not constant.

Residual Patterns: One Good, Two Problematic

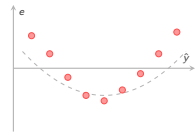
Compare Figure 6.8 in the textbook

Random scatter



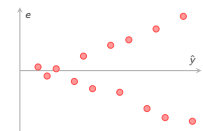
Assumptions look fine.

Curved (U-shape)



Relationship is nonlinear.

Funnel (fan)



Variance is not constant.

A curve means the straight-line form is wrong; a funnel means σ^2 changes with the response. Either way, the CI and PI formulas from today stop being trustworthy — fix the model first (Chapter 7 adds the tools).

Standardized Residuals for the Fuel Data

Using (6.28)

Scale each residual by $\hat{\sigma} = \sqrt{0.884} = 0.940$ — equation (6.28):

$$d_i = \frac{e_i}{\sqrt{\hat{\sigma}^2}} \tag{6.28}$$

If the model is right, about 95% of the d_i should land in $[-2, 2]$.

Standardized Residuals for the Fuel Data

Using (6.28)

Scale each residual by $\hat{\sigma} = \sqrt{0.884} = 0.940$ — equation (6.28):

$$d_i = \frac{e_i}{\sqrt{\hat{\sigma}^2}} \tag{6.28}$$

If the model is right, about 95% of the d_i should land in $[-2, 2]$.

i	1	2	3	4	5	6	7	8
e_i	$+0.22$	-0.83	$+0.82$	-0.93	$+1.13$	-0.62	$+0.93$	-0.72
d_i	$+0.23$	-0.88	$+0.87$	-0.99	$+1.20$	-0.66	$+0.99$	-0.77

Standardized Residuals for the Fuel Data

Using (6.28)

Scale each residual by $\hat{\sigma} = \sqrt{0.884} = 0.940$ — equation (6.28):

$$d_i = \frac{e_i}{\sqrt{\hat{\sigma}^2}} \tag{6.28}$$

If the model is right, about 95% of the d_i should land in $[-2, 2]$.

i	1	2	3	4	5	6	7	8
e_i	$+0.22$	-0.83	$+0.82$	-0.93	$+1.13$	-0.62	$+0.93$	-0.72
d_i	$+0.23$	-0.88	$+0.87$	-0.99	$+1.20$	-0.66	$+0.99$	-0.77

All $|d_i| < 2$: **no outliers**. The signs alternate irregularly with no drift and no funnel, so Procedure 6.4 passes.

Standardized Residuals for the Fuel Data

Using (6.28)

Scale each residual by $\hat{\sigma} = \sqrt{0.884} = 0.940$ — equation (6.28):

$$d_i = \frac{e_i}{\sqrt{\hat{\sigma}^2}} \tag{6.28}$$

If the model is right, about 95% of the d_i should land in $[-2, 2]$.

i	1	2	3	4	5	6	7	8
e_i	$+0.22$	-0.83	$+0.82$	-0.93	$+1.13$	-0.62	$+0.93$	-0.72
d_i	$+0.23$	-0.88	$+0.87$	-0.99	$+1.20$	-0.66	$+0.99$	-0.77

All $|d_i| < 2$: **no outliers**. The signs alternate irregularly with no drift and no funnel, so Procedure 6.4 passes.

Verdict: the fuel model has now survived the slope test, the F-test, and the residual

Three Ways Regression Gets Abused

1. Extrapolation. The model is only supported inside the observed x range. A startup that fits conversions on 5–25% discounts and then quotes the line at a 60% discount is reading tea leaves — next slide.

Three Ways Regression Gets Abused

1. Extrapolation. The model is only supported inside the observed x range. A startup that fits conversions on 5–25% discounts and then quotes the line at a 60% discount is reading tea leaves — next slide.

2. Correlation $\neq$ causation. Heavier trucks used more fuel, but load was not randomly assigned: heavier trips may also run different routes or older tractors. Regression on observational data supports *prediction*, not *intervention*. Causal claims need a designed experiment (Chapters 8–9).

Three Ways Regression Gets Abused

1. Extrapolation. The model is only supported inside the observed x range. A startup that fits conversions on 5–25% discounts and then quotes the line at a 60% discount is reading tea leaves — next slide.

2. Correlation $\neq$ causation. Heavier trucks used more fuel, but load was not randomly assigned: heavier trips may also run different routes or older tractors. Regression on observational data supports *prediction*, not *intervention*. Causal claims need a designed experiment (Chapters 8–9).

3. R^2 worship. An AI team's eval-score model can post $R^2 = 0.98$ and still fail: a curved residual plot means its prediction intervals are wrong even though its fit looks superb. High R^2 + failed Procedure 6.4 = a model you should not deploy.

Three Ways Regression Gets Abused

1. Extrapolation. The model is only supported inside the observed x range. A startup that fits conversions on 5–25% discounts and then quotes the line at a 60% discount is reading tea leaves — next slide.

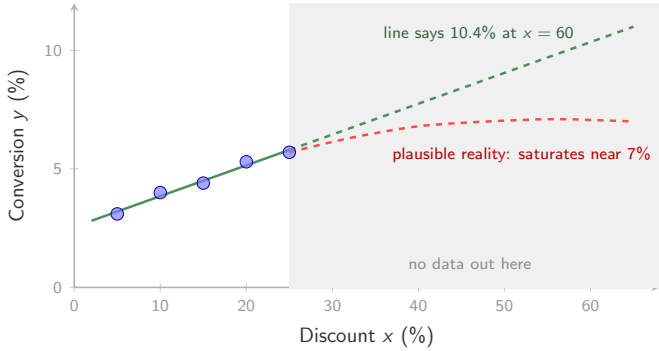
2. Correlation $\neq$ causation. Heavier trucks used more fuel, but load was not randomly assigned: heavier trips may also run different routes or older tractors. Regression on observational data supports *prediction*, not *intervention*. Causal claims need a designed experiment (Chapters 8–9).

3. R^2 worship. An AI team's eval-score model can post $R^2 = 0.98$ and still fail: a curved residual plot means its prediction intervals are wrong even though its fit looks superb. High R^2 + failed Procedure 6.4 = a model you should not deploy.

Textbook: §6.14

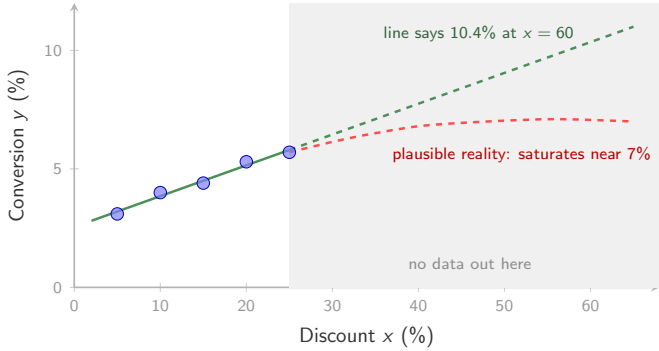
Extrapolation: The Startup Discount Trap

Compare Figure 6.10 in the textbook



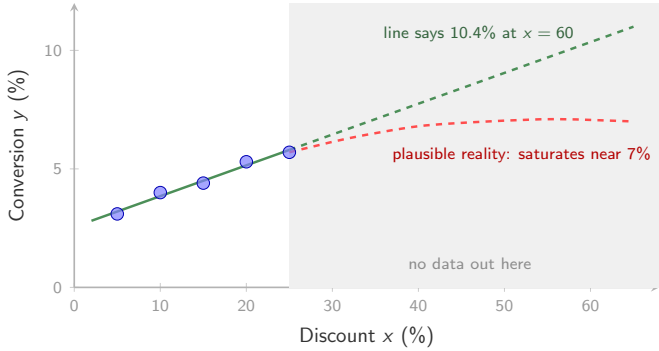
Extrapolation: The Startup Discount Trap

Compare Figure 6.10 in the textbook



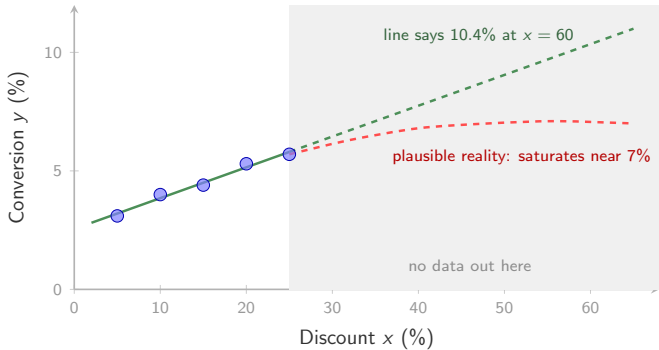
Extrapolation: The Startup Discount Trap

Compare Figure 6.10 in the textbook



Extrapolation: The Startup Discount Trap

Compare Figure 6.10 in the textbook



At $x = 60$ the fitted line promises a 10.4% conversion, but conversions can saturate — and a 60% discount gives away most of the revenue in SAR per signup. Inside 5–25% the model is evidence; outside, it is only a guess.

Practice Problem [5 pts]

Exam-Style: AI Safety Red-Team Data, Continued

Problem statement. Continue with Lecture 18's red-team data (x = red-team hours, y = critical defects found, $n = 6$, hours ranging 4–14):

$$\hat{y} = 2.0 + 1.5\,x, \quad \bar{x} = 9, \quad S_{xx} = 70, \quad S_{xy} = 105, \quad S_{yy} = 159.42, \quad \hat{\sigma}^2 = 0.48$$

Practice Problem [5 pts]

Exam-Style: AI Safety Red-Team Data, Continued

Problem statement. Continue with Lecture 18's red-team data (x = red-team hours, y = critical defects found, $n = 6$, hours ranging 4–14):

$$\hat{y} = 2.0 + 1.5\,x, \quad \bar{x} = 9, \quad S_{xx} = 70, \quad S_{xy} = 105, \quad S_{yy} = 159.42, \quad \hat{\sigma}^2 = 0.48$$

The team will budget **12 red-team hours** for the next release. Use $t_{0.025,\,4} = 2.776$.

Practice Problem [5 pts]

Exam-Style: AI Safety Red-Team Data, Continued

Problem statement. Continue with Lecture 18's red-team data (x = red-team hours, y = critical defects found, $n = 6$, hours ranging 4–14):

$$\hat{y} = 2.0 + 1.5\,x, \quad \bar{x} = 9, \quad S_{xx} = 70, \quad S_{xy} = 105, \quad S_{yy} = 159.42, \quad \hat{\sigma}^2 = 0.48$$

The team will budget **12 red-team hours** for the next release. Use $t_{0.025,\,4} = 2.776$.

(a) [1 pt] Predict the number of critical defects at $x_0 = 12$; confirm this is not extrapolation.
(b) [1.5 pts] Give a 95% CI for the *mean* number of defects at 12 hours (6.26).
(c) [1.5 pts] Give a 95% PI for the defects found in *the next release* at 12 hours (6.27).
(d) [1 pt] Compute R^2 (6.21) and r (6.22). Which interval — (b) or (c) — should

Practice Problem — Solution

Parts (a) and (b)

(a) $x_0 = 12$ lies inside the observed range 4–14 hours ✓, so Procedure 6.3 applies:

$$\hat{y}_0 = 2.0 + 1.5(12) = 2.0 + 18.0 = \boxed{20.0 \text{ defects}}$$

Practice Problem — Solution

Parts (a) and (b)

(a) $x_0 = 12$ lies inside the observed range 4–14 hours ✓, so Procedure 6.3 applies:

$$\hat{y}_0 = 2.0 + 1.5(12) = 2.0 + 18.0 = \boxed{20.0 \text{ defects}}$$

(b) Standard error for the *mean* response (6.26) (no "1+"):

$$se = \sqrt{0.48\left(\frac{1}{6} + \frac{(12-9)^2}{70}\right)} = \sqrt{0.48\,(0.1667 + 0.1286)} = \sqrt{0.1417} = 0.376$$

Practice Problem — Solution

Parts (a) and (b)

(a) $x_0 = 12$ lies inside the observed range 4–14 hours ✓, so Procedure 6.3 applies:

$$\hat{y}_0 = 2.0 + 1.5(12) = 2.0 + 18.0 = \boxed{20.0 \text{ defects}}$$

(b) Standard error for the *mean* response (6.26) (no "1+"):

$$se = \sqrt{0.48\left(\frac{1}{6} + \frac{(12-9)^2}{70}\right)} = \sqrt{0.48\,(0.1667 + 0.1286)} = \sqrt{0.1417} = 0.376$$

Interval:

$$20.0 \pm 2.776 \times 0.376 = 20.0 \pm 1.05 \quad \Longrightarrow \quad \boxed{(18.95,\ 21.05)}$$

Practice Problem — Solution

Parts (a) and (b)

(a) $x_0 = 12$ lies inside the observed range 4–14 hours ✓, so Procedure 6.3 applies:

$$\hat{y}_0 = 2.0 + 1.5(12) = 2.0 + 18.0 = \boxed{20.0 \text{ defects}}$$

(b) Standard error for the *mean* response (6.26) (no "1+"):

$$se = \sqrt{0.48\left(\frac{1}{6} + \frac{(12-9)^2}{70}\right)} = \sqrt{0.48\,(0.1667 + 0.1286)} = \sqrt{0.1417} = 0.376$$

Interval:

$$20.0 \pm 2.776 \times 0.376 = 20.0 \pm 1.05 \quad \Longrightarrow \quad \boxed{(18.95,\ 21.05)}$$

Practice Problem — Solution (Continued)

Parts (c) and (d)

(c) Standard error for *one new release* (6.27) (with the "$1+$"):

$$se = \sqrt{0.48\,(1 + 0.1667 + 0.1286)} = \sqrt{0.6217} = 0.788$$

$$20.0 \pm 2.776 \times 0.788 = 20.0 \pm 2.19 \quad \Longrightarrow \quad \boxed{(17.81,\ 22.19)}$$

Practice Problem — Solution (Continued)

Parts (c) and (d)

(c) Standard error for *one new release* (6.27) (with the "$1+$"):

$$se = \sqrt{0.48\,(1 + 0.1667 + 0.1286)} = \sqrt{0.6217} = 0.788$$

$$20.0 \pm 2.776 \times 0.788 = 20.0 \pm 2.19 \quad \Longrightarrow \quad \boxed{(17.81,\ 22.19)}$$

(d) $SS_R = 1.5 \times 105 = 157.5$, $SS_T = 159.42$, so by (6.21) and (6.22):

$$R^2 = \frac{157.5}{159.42} = 0.9880, \qquad r = \frac{105}{\sqrt{70 \times 159.42}} = \frac{105}{105.6} = 0.9940$$

Check: $r^2 = 0.9880 = R^2$ ✓

Practice Problem — Solution (Continued)

Parts (c) and (d)

(c) Standard error for *one new release* (6.27) (with the "$1+$"):

$$se = \sqrt{0.48\,(1 + 0.1667 + 0.1286)} = \sqrt{0.6217} = 0.788$$

$$20.0 \pm 2.776 \times 0.788 = 20.0 \pm 2.19 \quad \Longrightarrow \quad \boxed{(17.81,\ 22.19)}$$

(d) $SS_R = 1.5 \times 105 = 157.5$, $SS_T = 159.42$, so by (6.21) and (6.22):

$$R^2 = \frac{157.5}{159.42} = 0.9880, \qquad r = \frac{105}{\sqrt{70 \times 159.42}} = \frac{105}{105.6} = 0.9940$$

Check: $r^2 = 0.9880 = R^2$ ✓

Which interval? The release manager cares about **the next single release**, so quote the **PI** (17.8, 22.2). The CI describes a long-run average that no single

Lecture 19 Summary

- Procedure 6.3: check the range, compute $\hat{y}_0$, then pick the interval. CI for the mean (6.26); PI for one new observation (6.27) — the "1+" makes the PI wider, always.

Lecture 19 Summary

- Procedure 6.3: check the range, compute $\hat{y}_0$, then pick the interval. CI for the mean (6.26); PI for one new observation (6.27) — the "1+" makes the PI wider, always.
- Both intervals are narrowest at $\bar{x}$ and flare toward the edges of the data.

Lecture 19 Summary

- Procedure 6.3: check the range, compute $\hat{y}_0$, then pick the interval. CI for the mean (6.26); PI for one new observation (6.27) — the "1+" makes the PI wider, always.
- Both intervals are narrowest at $\bar{x}$ and flare toward the edges of the data.
- $R^2 = SS_R/SS_T$ (6.21) is the fraction of variability explained; $r = S_{xy}/\sqrt{S_{xx}S_{yy}}$ (6.22) adds the direction, and $r^2 = R^2$.

Lecture 19 Summary

- Procedure 6.3: check the range, compute $\hat{y}_0$, then pick the interval. CI for the mean (6.26); PI for one new observation (6.27) — the "1+" makes the PI wider, always.
- Both intervals are narrowest at $\bar{x}$ and flare toward the edges of the data.
- $R^2 = SS_R/SS_T$ (6.21) is the fraction of variability explained; $r = S_{xy}/\sqrt{S_{xx}S_{yy}}$ (6.22) adds the direction, and $r^2 = R^2$.
- Procedure 6.4: plot e_i vs. $\hat{y}_i$, check normality, flag $|d_i| > 2$ using (6.28). Random scatter = adequate model; curve or funnel = fix the model first.

Lecture 19 Summary

- Procedure 6.3: check the range, compute $\hat{y}_0$, then pick the interval. CI for the mean (6.26); PI for one new observation (6.27) — the "1+" makes the PI wider, always.
- Both intervals are narrowest at $\bar{x}$ and flare toward the edges of the data.
- $R^2 = SS_R/SS_T$ (6.21) is the fraction of variability explained; $r = S_{xy}/\sqrt{S_{xx}S_{yy}}$ (6.22) adds the direction, and $r^2 = R^2$.
- Procedure 6.4: plot e_i vs. $\hat{y}_i$, check normality, flag $|d_i| > 2$ using (6.28). Random scatter = adequate model; curve or funnel = fix the model first.
- Abuses to avoid: extrapolation outside the x range, causal claims from observational fits, and trusting R^2 over the residual checks.

Lecture 19 Summary

- Procedure 6.3: check the range, compute $\hat{y}_0$, then pick the interval. CI for the mean (6.26); PI for one new observation (6.27) — the "1+" makes the PI wider, always.
- Both intervals are narrowest at $\bar{x}$ and flare toward the edges of the data.
- $R^2 = SS_R/SS_T$ (6.21) is the fraction of variability explained; $r = S_{xy}/\sqrt{S_{xx}S_{yy}}$ (6.22) adds the direction, and $r^2 = R^2$.
- Procedure 6.4: plot e_i vs. $\hat{y}_i$, check normality, flag $|d_i| > 2$ using (6.28). Random scatter = adequate model; curve or funnel = fix the model first.
- Abuses to avoid: extrapolation outside the x range, causal claims from observational fits, and trusting R^2 over the residual checks.
- **Next lecture:** multiple linear regression — more than one predictor at a time.

Reminders

- **Quiz on CI vs. PI** is due before class — it hinges on spotting *mean* vs. *next observation*

Reminders

- **Quiz on CI vs. PI** is due before class — it hinges on spotting *mean* vs. *next observation*
- Homework 6 (regression) is due next Thursday on Blackboard

Reminders

- **Quiz on CI vs. PI** is due before class — it hinges on spotting *mean* vs. *next observation*
- Homework 6 (regression) is due next Thursday on Blackboard
- Office hours: Tuesdays 9–10 AM, Room 22-219 or Zoom

Reminders

- **Quiz on CI vs. PI** is due before class — it hinges on spotting *mean* vs. *next observation*
- Homework 6 (regression) is due next Thursday on Blackboard
- Office hours: Tuesdays 9–10 AM, Room 22-219 or Zoom
- Read §7.1 and §7.2 to prepare for multiple regression