# Chapter 14

# Inference on the Least-Squares Regression Model and Multiple Regression

## 14.1 Testing the Significance of the Least-Squares Regression Model

Test and CI for slope, $\beta_1$, of regression model $y_i = \beta_1 x_i + \beta_0 + \epsilon_i$, is

$$t_0 = \frac{b_1 - \beta_1}{\frac{s_e}{\sqrt{\sum (x_i - \bar{x})^2}}} = \frac{b_1 - \beta_1}{s_{b_1}}, \qquad b_1 \pm t_{\frac{\alpha}{2}} \left( \frac{s_e}{\sqrt{\sum (x_i - \bar{x})^2}} \right),$$

where $\mu_{y|x} = \beta_1 x + \beta_0$ and where standard error of estimate, $s_e$, is

$$s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$$

where $\hat{y}_i = b_1 x_i + b_0$ and where points are sampled at random and residuals, $\epsilon_i$, are normal with constant variance and where $t_{\frac{\alpha}{2}}$ has $n-2$ degrees of freedom.

**Exercise 14.1 (Testing Significance of Least-Squares Regression Model)**

1. *Scatterplot, least-squares line, residuals review: height vs circumference of trees.*

| circumference, $x$ | 2.1 | 1.7 | 1.1 | 1.5 | 2.7 |
|---|---|---|---|---|---|
| height, $y$ | 40 | 37 | 35 | 36 | 42 |

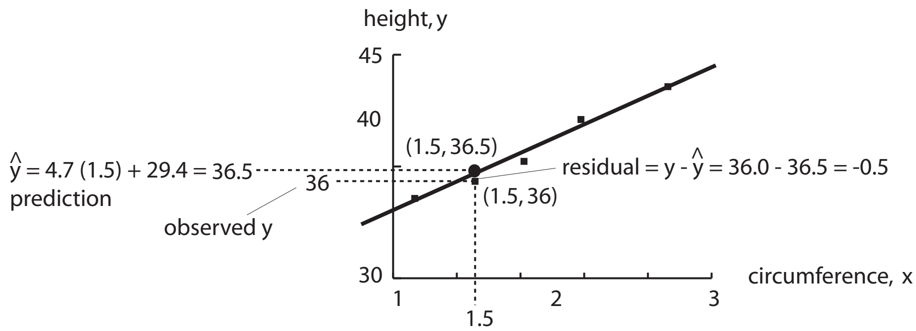


Figure 14.1 (Scatterplot, least-squares line, residuals)

(a) *Scatter diagram*

- Turn off all STAT PLOTS and Y = plots, then type $(x, y)$ values into $(L_1, L_2)$ lists.
- Display scatter diagram by pressing
  - 2nd STAT PLOT ENTER
  - On ENTER
  - Type: scatter plot figure first row, far left ENTER
  - Xlist: L1 (for $x$ values) ENTER
  - Ylist: L2 (for $y$ values) ENTER
  - Mark: (choose any one of the three)

  then hit ZOOM 9:ZoomStat. TRACE key used to view various $(x, y)$ points.

(b) *Calculating least-squares regression line.*

$\boldsymbol{\hat{y} = 2.438x + 4.704}$
$\boldsymbol{\hat{y} = 4.704x + 29.438}$
$\boldsymbol{\hat{y} = 5.944x + 47.04}$.

Type data into $L_1$, $L_2$; then STAT CALC LinReg$(ax + b)$ $L_1$, $L_2$.

(c) *Slope and y-intercept of least-squares regression line,* $\hat{y} = 4.704x + 29.438$.

*Slope* is $b_1 =$ (circle one) **4.704 / 29.438**.
Slope, $b_1 = 4.704$, means, on average, height increases 4.704 feet for an increase of *one* foot of circumference.

The *y–intercept* is $b_0 =$ (circle one) **4.704 / 29.438**.

If we sampled at random another five trees, (choose one)
**same / different** *slope, y-intercept* would most likely occur. This implies $b_0, b_1$ are **statistics / parameters** used to estimate $\beta_0, \beta_1$.

Least–squares $\hat{y} = 4.704x + 29.438$ estimates model $\mu_{y|x} = \beta_1 x + \beta_0$.

Overlay least-squares on scatter diagram:

- Create scatterplot.
- Calculate least-squares line.
- Press Y =
- VARS, down to 5:Statistics ENTER, over to EQ ENTER
- GRAPH
- Press TRACE , then press UP or DOWN to trace along observed values or least-squares.
- On least-squares regression, type $x$ value ENTER for a $y$ value.

(d) *Residuals, standard error of estimate, $s_e$.*

| circumference, $x$ | 2.1 | 1.7 | 1.1 | 1.5 | 2.7 | total |
|---|---|---|---|---|---|---|
| observed height, $y$ | 40 | 37 | 35 | 36 | 42 | 190 |
| predicted height, $\hat{y}$ | 39.3 | 37.4 | 34.6 | 36.5 | 42.1 | 190 |
| residual, $y - \hat{y}$ | 0.7 | -0.4 | 0.4 | -0.5 | -0.1 | 0 |
| residual$^2$, $(y - \hat{y})^2$ | 0.5 | 0.2 | 0.2 | 0.2 | 0.0 | 1.1 |

Type data into $L_1$, $L_2$ then STAT CALC LinReg($ax + b$) $L_1$, $L_2$.

Define $L_3 = 4.704L_1 + 29.438$. Define $L_4 = L_2 - L_3$. Define $L_5 = L_4^2$.

Total residuals$^2$, $\sum(y - \hat{y})^2 \approx 1.1$, measures how close points are to least-squares line. Standard error of estimate, $s_e$, measures "average" distance observed data is from least–squares line,

$$s_e = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} \approx \sqrt{\frac{1.1}{5-2}} \approx$$

(circle one) **0.40** / **0.60** / **0.80**.

(e) Standard error of estimate is related to least–squares line in much same way standard deviation is related to (circle one) **average** / **variance**.

2. *Calculating standard error of estimate, $s_e$, using TI–84+.*

(a) *Height versus circumference of trees.*

| circumference, $x$ | 2.1 | 1.7 | 1.1 | 1.5 | 2.7 |
|---|---|---|---|---|---|
| height, $y$ | 40 | 37 | 35 | 36 | 42 |

$s_e \approx$ (circle one) **0.40** / **0.60** / **0.80**.

(Type $x$, $y$ into $L_1$, $L_2$. Then STAT TESTS LinRegTTest... $L_1$ $L_2$ $1 \neq 0$ Calculate ENTER.

Arrow down to read $s = 0.60....$)

(b) *Reading ability versus brightness.*

| brightness, $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ability to read, $y$ | 70 | 70 | 75 | 88 | 91 | 94 | 100 | 92 | 90 | 85 |

$s_e \approx$ (circle one) **6.45** / **7.03** / **7.83**.

(Type $x$, $y$ into $L_3$, $L_4$. Then STAT TESTS LinRegTTest... $L_3$ $L_4$ $1 \neq 0$ Calculate ENTER.)

(c) *Grain yield versus distance from water.*

| dist, x | 0 | 10 | 20 | 30 | 45 | 50 | 70 | 80 | 100 | 120 | 140 | 160 | 170 | 190 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| yield, y | 500 | 590 | 410 | 470 | 450 | 480 | 510 | 450 | 360 | 400 | 300 | 410 | 280 | 350 |

$s_e \approx$ (circle one) **21.2** / **43.8** / **54.8**.

(Type $x$, $y$ into $L_5$, $L_6$. Then STAT TESTS LinRegTTest... $L_3$ $L_4$ $1 \neq 0$ Calculate ENTER.)

3. *Inference for slope, $\beta_1$, of linear regression: reading ability.*

| illumination, $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ability to read, $y$ | 70 | 70 | 75 | 88 | 91 | 94 | 100 | 92 | 90 | 85 |

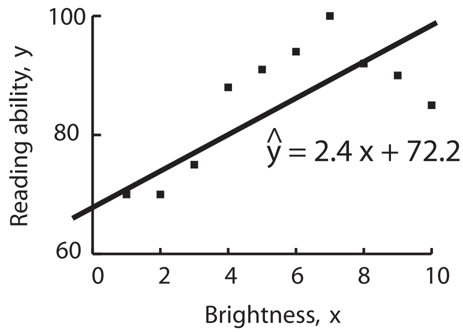


Figure 4.2 (Scatter diagram and least-squares, reading ability vs brightness)

Based on $n = 10$ data points, we find sample slope $b_1 \approx 2.418$. Test if population slope, $\beta_1$, is *positive* at a level of significance of 5%. Also, calculate a 95% CI.

(a) *Check assumptions.*

| brightness, x | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ability to read, y | 70 | 70 | 75 | 88 | 91 | 94 | 100 | 92 | 90 | 85 |
| predicted, $\hat{y}$ | 74.6 | 77.0 | 79.5 | 81.9 | 84.3 | 86.7 | 89.1 | 91.5 | 94.0 | 96.4 |
| residual, $y - \hat{y}$ | -4.6 | -7.0 | -4.5 | 6.1 | 6.7 | 7.3 | 10.9 | 0.5 | -4.0 | -8.6 |

Type data into $L_3$, $L_4$ then STAT CALC LinReg($ax + b$) $L_3$, $L_4$.

Define $L_5 = 2.4L_1 + 72.2$. Define $L_6 = L_5 - L_4$ for residuals.

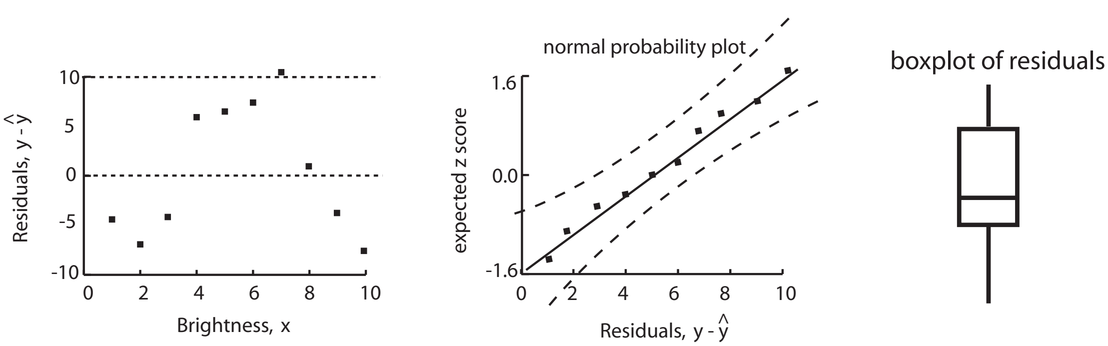


Figure 4.3 (Diagnostics of residuals, reading ability vs brightness)

i. *Pattern?*
According to either scatter diagram or residual plot,
there **is a** / **is no** pattern: points are curved, not linear.

ii. *Constant variance?*
According to residual plot, residuals vary -10 and 10 over entire range of brightness; that is, data variance is **constant** / **variable**.

iii. *Residuals normal?*
Normal probability plot indicates residuals
**normal** / **not normal** because data within dotted bounds.

iv. *Outliers?*
Boxplot indicates **outliers** / **no outliers**.

(b) *Hypothesis test, right–sided.*

i. *Statement.* Choose one.
A. $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 < 0$
B. $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 > 0$
C. $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$

ii. *Test.*

| brightness, $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $(x - \bar{x})^2 = (x - 5.5)^2$ | 20.25 | 12.25 | 6.25 | 2.25 | 0.25 | 0.25 | 2.25 | 6.25 | 12.25 | 20.25 |

(Type $x$, $y$ into $L_3$, $L_4$. Define $L_5 = (L_3 - 5.5)^2$ ENTER. STAT CALC $L_5$, read $\sum x = 82.5$.)
Chance $b_1 = 2.418$ or more, *if* $\beta_1 = 0$, is

$$\text{p–value} = P(b_1 \geq 2.418) = P\left(\frac{b_1 - \beta_1}{\frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}}} \geq \frac{2.418 - 0}{\frac{7.827}{\sqrt{82.5}}}\right) \approx P\left(t \geq 2.81\right) \approx$$

**0.002** / **0.011** / **0.058** (with $n - 2 = 10 - 2 = 8$ df)
(2nd DISTR tcdf(2.81,E99,8)
OR: Type $x$, $y$ into $L_3$, $L_4$. Then STAT TESTS LinRegTTest... $L_3$ $L_4$ 1 > 0 Calculate ENTER)
Level of significance $\alpha =$ (choose one) **0.01** / **0.05** / **0.10**.

iii. *Conclusion.*
Since p–value $= 0.011 < \alpha = 0.050$,
**do not reject** / **reject** null $H_0 : \beta_1 = 0$.
Data indicates population slope
**smaller than** / **equals** / **greater than** zero (0).
In other words, reading ability
**is** / **is not** positively associated with brightness.

(c) *Confidence interval for $\beta_1$.*
A 95% confidence interval for $\beta_1$ is

$$b_1 \pm t_{\frac{\alpha}{2}} \frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}} = 2.42 \pm 2.306\left(\frac{7.827}{\sqrt{82.5}}\right) \approx$$

(circle one) **$2.42 \pm 0.99$** / **$2.42 \pm 1.99$** / **$2.42 \pm 2.99$**
For $t_{\frac{\alpha}{2}}$, use PRGM INVT 8 0.975; for $b_1$ and $s_e$, use STAT TESTS E:LinRegTTest;
for rest, use STAT CALC 1-Var Stats $L_3$
OR type $x$, $y$ into $L_3$, $L_4$ then STAT TESTS LinRegTTInt... $L_3$ $L_4$ 1 0.95 Calculate ENTER.

4. *Inference for slope, $\beta_1$, of linear regression: elastic band.*
As elastic band stretched, width of band decreases. Based on $n = 5$ data points, $b_1 \approx -4.704$. Test if population slope, $\beta_1$, is *less* than zero at level of significance of 5%. Also, calculate 95% CI for $\beta_1$.

| | | | | | |
|---|---|---|---|---|---|
| band width, $x$ | -2.1 | -1.7 | -1.1 | -1.5 | -2.7 |
| stretch length, $y$ | 40 | 37 | 35 | 36 | 42 |

Convert height vs circumference data set to this data set: Define $L_1 = -1 \times L_1$.

(a) *Hypothesis test, left–sided.*

i. *Statement.* Choose one.
   A. $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 < 0$
   B. $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 > 0$
   C. $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$

ii. *Test.*
Chance $b_1 = -4.704$ or less, if $\beta_1 = 0$, is

$$\text{p–value} = P(b_1 \leq -4.704) = P\left(\frac{b_1 - \beta_1}{\frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}}} \leq \frac{-4.704 - 0}{\frac{0.59718}{\sqrt{1.488}}}\right) \approx P(t \leq -9.61) \approx$$

**0.0012 / 0.0023 / 0.0058**

(2nd DISTR tcdf(-E99,-9.61,3)

OR: Type $x$, $y$ into $L_1$, $L_2$. Then STAT TESTS LinRegTTest... $L_1$ $L_2$ 1 $< 0$ Calculate ENTER)

Level of significance $\alpha =$ (choose one) **0.01 / 0.05 / 0.10**.

iii. *Conclusion.*
Since p–value $= 0.0012 < \alpha = 0.0500$,
**do not reject / reject** null $H_0 : \beta_1 = 0$.
Data indicates population slope
**smaller than / equals / greater than** zero (0).
In other words, stretch length is
**negatively / positively** associated with band width.

(b) *Confidence Interval For $\beta_1$.*
A 95% confidence interval for $\beta_1$ is

$$b_1 \pm t_{\frac{\alpha}{2}} \frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}} = -4.704 \pm 3.18 \left(\frac{0.59718}{\sqrt{1.488}}\right) \approx$$

(circle one) **$-4.704 \pm 0.743$ / $-4.704 \pm 0.843$ / $-4.704 \pm 1.557$**

For $t_{\frac{\alpha}{2}}$, use PRGM INVT 3 0.975; for $b_1$ and $s_e$, use STAT TESTS E:LinRegTTest;

for rest, use STAT CALC 1-Var Stats $L_1$

OR type $x$, $y$ into $L_1$, $L_2$ then STAT TESTS LinRegTTInt... $L_1$ $L_2$ 1 0.95 Calculate ENTER.

5. *Inference for slope, $\beta_1$, of linear regression: pizza.*
Consider data on sales ($y$, in \$1000s) of pizzas versus student population ($x$, in 1000s). Based on $n = 10$ data points, $b_1 = 5$. Test if population slope, $\beta_1$, is *different* than zero at a level of significance of 5%. Also, calculate 95% CI.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| number students, $x$ | 2 | 6 | 8 | 8 | 12 | 16 | 20 | 20 | 22 | 26 |
| pizza sales, $y$ | 58 | 105 | 88 | 118 | 117 | 137 | 157 | 169 | 149 | 202 |

(a) *Hypothesis test, two–sided.*

i. *Statement.* Choose one.

A. $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 < 0$

B. $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 > 0$

C. $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$

ii. *Test.*

$$\text{p–value} = 2 \times P(b_1 \geq 5) = 2 \times P\left(\frac{b_1 - \beta_1}{\frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}}} \geq \frac{5 - 0}{\frac{13.83}{\sqrt{568}}}\right) \approx 2 \times P\left(t \geq 8.62\right) \approx$$

**0.00** / **0.01** / **0.06** (with $n - 2 = 10 - 2 = 8$ df)

(2nd DISTR tcdf(8.62,E99,8) times two (2)

OR: Type $x$, $y$ into $L_5$, $L_6$. Then STAT TESTS LinRegTTest... $L_5$ $L_6$ 1 $\neq$ 0 Calculate ENTER)

Level of significance $\alpha =$ (choose one) **0.01** / **0.05** / **0.10**.

iii. *Conclusion.*

Since p–value $= 0.00 < \alpha = 0.05$,
**do not reject** / **reject** null $H_0 : \beta_1 = 0$.
Data indicates population slope
**smaller than** / **equals** / **does not equal** zero (0).
In other words, sales
**is** / **is not** associated with student number.

(b) *Confidence interval for $\beta_1$.*
A 95% confidence interval for $\beta_1$ is

$$b_1 \pm t_{\frac{\alpha}{2}} \frac{s_e}{\sqrt{\sum(x_i - \bar{x})^2}} = 5 \pm 2.31\left(\frac{13.829}{\sqrt{568}}\right) =$$

(circle one) $\mathbf{5 \pm 0.74}$ / $\mathbf{5 \pm 1.3}$ / $\mathbf{5 \pm 2.2}$

For $t_{\frac{\alpha}{2}}$, use PRGM INVT 8 0.975; for $b_1$ and $s_e$, use STAT TESTS E:LinRegTTest;
for rest, use STAT CALC 1-Var Stats $L_5$
OR type $x$, $y$ into $L_1$, $L_2$ then STAT TESTS LinRegTTInt... $L_1$ $L_2$ 1 0.95 Calculate ENTER.

## 14.2 Confidence and Prediction Intervals

Confidence interval[1] (CI) for *mean* response and prediction interval (PI) for *individual* response of regression model $y_i = \beta_1 x_i + \beta_0 + \epsilon_i$ are given, respectively, as

$$\hat{y} \pm t_{\frac{\alpha}{2}} \cdot s_e \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} \qquad \hat{y} \pm t_{\frac{\alpha}{2}} \cdot s_e \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}},$$

for given (fixed) $x^*$ where points are sampled at random and residuals, $\epsilon_i$, are normal with constant variance and where $t_{\frac{\alpha}{2}}$ has $n - 2$ degrees of freedom.

**Exercise 14.2 (Confidence and Prediction Intervals)**

1. *CI and PI: reading versus illumination.*

| brightness, $x$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ability to read, $y$ | 70 | 70 | 75 | 88 | 91 | 94 | 100 | 92 | 90 | 85 |

Calculate 95% CI and 95% PI for $\hat{y}$ at $x^* = 3.5$ and at $x^* = 6.5$.

(a) *Confidence interval (CI) and prediction interval (PI) at* $x^* = 3.5$

i. *Calculate* $\hat{y}$ *at* $x^* = 3.5$.
Since least–squares line is $\hat{y} = 2.418x + 72.2$, at $x^* = 3.5$,
$\hat{y} = 2.418(3.5) + 72.2 =$ (circle one) **78.9** / **80.7** / **88.9**.
(Type $x$ data in $L_1$ and $y$ data into $L_2$. Then STAT CALC LinReg($ax + b$) ENTER)

ii. *Critical value,* $t_{\frac{\alpha}{2}}$.
$95\% = (1 - \alpha) \cdot 100\% = (1 - 0.05) \cdot 100\%$, $n - 2 = 10 - 2 = 8$ df
$t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025} =$ (circle one) (circle one) **2.31** / **2.58** / **2.78**.
(PRGM ENTER INVT ENTER 8 ENTER 0.975 ENTER; yes, 0.975–why?)

iii. *Standard error of estimate,* $s_e$.
Standard error of estimate, $s_e =$ (circle one) **7.33** / **7.53** / **7.84**.
(STAT TESTS LinRegTTest... $L_1$ $L_2$ $1 \neq 0$ Calculate ENTER)

iv. *Sum of squares of* $x$.
$\sum(x_i - \bar{x})^2 =$ (circle one) **82.5** / **108.5** / **122.5**.
(Type $x$, $y$ into $L_1$, $L_2$. Define $L_3 = (L_1 - 5.5)^2$ ENTER. STAT CALC $L_3$, read $\sum x = 82.5$.)

v. *CI of mean response at* $x^* = 3.5$.
Also, $n = 10$, $\bar{x} = 5.5$, and so

$$\hat{y} \pm t_{\frac{\alpha}{2}} s_e \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 80.7 \pm (2.31)(7.84)\sqrt{\frac{1}{10} + \frac{(3.5 - 5.5)^2}{82.5}} =$$

[1] The calculator does *not* calculate either confidence interval or prediction interval using a preset menu. Both are calculated a piece at a time.

(circle one) **80.7 ± 7.0** / **80.7 ± 8.2** / **80.7 ± 9.2**.
Length of prediction interval at $x^* = 3.5$ is
$L =$ (choose one) **7.0** / **14.0** / **21.0**.
(Remember, length is twice width, that 7.0 is being added and subtracted to 80.7.)

vi. *PI of individual response at* $x^* = 3.5$.
Also, $n = 10$, $\bar{x} = 5.5$, and so

$$\hat{y} \pm t_{\frac{\alpha}{2}} s_e \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 80.7 \pm (2.31)(7.84)\sqrt{1 + \frac{1}{10} + \frac{(3.5 - 5.5)^2}{82.5}} =$$

(circle one) **80.7 ± 19.4** / **80.7 ± 21.3** / **80.7 ± 22.3**.
Length of confidence interval at $x^* = 3.5$ is
$L =$ (choose one) **19.4** / **21.3** / **38.8**.
(Remember, length is twice width, that 19.4 is being added and subtracted to 80.7.)

vii. *Comparing CI and PI lengths at* $x^* = 3.5$.
Length of CI of mean response, $L = 14.0$, is
**smaller than** / **same as** / **larger than**
length of PI of individual response, $L = 38.8$.
More certain of average reading ability than individual reading ability.

(b) *Confidence interval (CI) and prediction interval (PI) at* $x^* = 6.5$

i. *Calculate* $\hat{y}$ *at* $x^* = 6.5$.
Since least–squares line is $\hat{y} = 72.2 + 2.418x$, at $x^* = 6.5$,
$\hat{y} = 72.2 + 2.418(6.5) =$ (circle one) **86.9** / **87.9** / **88.9**.

ii. *Critical value,* $t_{\frac{\alpha}{2}}$.
$95\% = (1 - \alpha) \cdot 100\% = (1 - 0.05) \cdot 100\%$, $n - 2 = 10 - 2 = 8$ df
$t_{\frac{\alpha}{2}} = t_{\frac{0.05}{2}} = t_{0.025} =$ (circle one) (circle one) **2.31** / **2.58** / **2.78**.
(PRGM ENTER INVT ENTER 8 ENTER 0.975 ENTER; yes, 0.975–why?)

iii. *Standard error of estimate,* $s_e$.
Standard error of estimate, $s_e =$ (circle one) **7.33** / **7.52** / **7.84**.
(STAT TESTS LinRegTTest... $L_1$ $L_2$ $1 \neq 0$ Calculate ENTER)

iv. *Sum of squares of* $x$.
$\sum(x_i - \bar{x})^2 =$ (circle one) **82.5** / **108.5** / **122.5**.
(Type $x$, $y$ into $L_1$, $L_2$. Define $L_3 = (L_1 - 5.5)^2$ ENTER. STAT CALC $L_3$, read $\sum x = 82.5$.)

v. *CI of mean response at* $x^* = 6.5$.
Also, $n = 10$, $\bar{x} = 5.5$, and so

$$\hat{y} \pm t_{\frac{\alpha}{2}} s_e \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 87.9 \pm (2.31)(7.84)\sqrt{\frac{1}{10} + \frac{(6.5 - 5.5)^2}{82.5}} =$$

(circle one) **87.9 ± 6.1** / **87.9 ± 7.1** / **87.9 ± 8.1**.
Length of CI at $x^* = 6.5$ is
$L =$ (choose one) **6.1** / **12.2** / **18.3**.

vi. *PI at* $x^* = 6.5$.
Also, $n = 10$, $\bar{x} = 5.5$, and so

$$\hat{y} \pm t_{\frac{\alpha}{2}} s_e \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 87.9 \pm (2.31)(7.84)\sqrt{1 + \frac{1}{10} + \frac{(3.5 - 5.5)^2}{82.5}} =$$

(circle one) **87.9 ± 19.1** / **87.9 ± 20.3** / **87.9 ± 22.3**.
Length of PI at $x^* = 6.5$ is
$L =$ (choose one) **19.1** / **20.3** / **38.2**.

(c) *Related questions.*

i. *Comparing predicted* $\hat{y}$*s at* $x^* = 3.5$ *and at* $x^* = 6.5$.
The $\hat{y} = 80.7$ at $x^* = 3.5$ is
**smaller than** / **same as** / **larger than**
the $\hat{y} = 87.9$ at $x^* = 6.5$
Predicted reading ability improves for increased brightness.

ii. *Comparing CI lengths at* $x^* = 3.5$ *and at* $x^* = 6.5$.
Length of CI at $x^* = 3.5$, $L = 14.0$
**smaller than** / **same as** / **larger than**
length of CI at $x^* = 6.5$, $L = 12.2$.
Uncertainty in mean reading ability greater at $x^* = 3.5$ than $x^* = 6.5$.

iii. *Comparing PI lengths at* $x^* = 3.5$ *and at* $x^* = 6.5$.
Length of CI at $x^* = 3.5$, $L = 38.8$
**smaller than** / **same as** / **larger than**
length of CI at $x^* = 6.5$, $L = 38.2$.
Uncertainty in individual ability greater at $x^* = 3.5$ than $x^* = 6.5$.

iv. *Confidence (prediction) band, from confidence (prediction) intervals.*
**True** / **False**
CIs (PIs) change for different $x^*$ and, together, create a confidence (prediction) *band* of intervals.

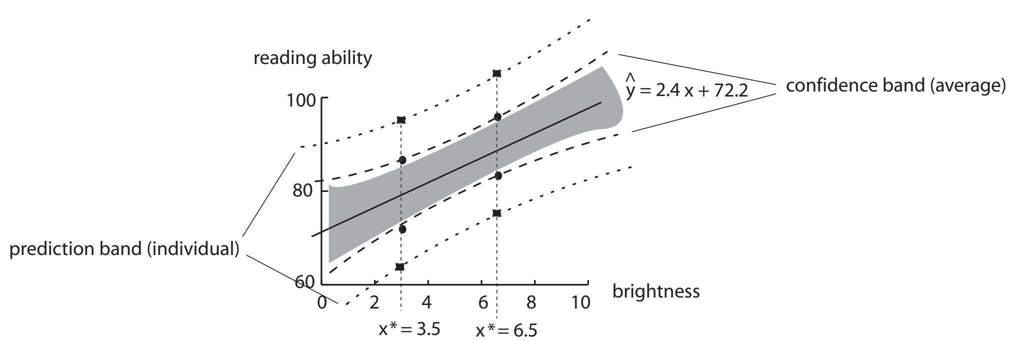


Figure 4.4 (CI, PI, confidence and prediction bands)

v. **True** / **False**
Confidence (prediction) *band* narrowest at point of averages $(\bar{x}, \bar{y})$.

2. *CI and PI: elastic band.*

| band width, $x$ | -2.1 | -1.7 | -1.1 | -1.5 | -2.7 |
|---|---|---|---|---|---|
| stretch length, $y$ | 40 | 37 | 35 | 36 | 42 |

Calculate a 92% CI and 92% PI for $\hat{y}$ at $x^* = -1.8$.

(a) *Calculate least–squares line, $\hat{y} = b_1 x + b_0$.*
$\boldsymbol{\hat{y} = 29.44 - 4.70x}$ / $\boldsymbol{\hat{y} = 29.44 - 4.98x}$ / $\boldsymbol{\hat{y} = 29.44 - 5.89x}$.
(Type $x$, $y$ into $L_1$, $L_2$; then use STAT CALC LinReg($ax + b$).)

(b) *Predicted value, $\hat{y}$.*
Since least–squares line is $\hat{y} = -4.70x + 29.44$, at $x = -1.8$,
$\hat{y} = -4.70(-1.8) + 29.44 =$ (circle one) **26.9** / **37.9** / **48.9**.
(For very accurate answer, use Y =; then use VARS 5:Statistics EQ ENTER; then 2nd QUIT; then VARS Y–VARS ENTER $(-1.8)$.)

(c) *Critical value, $t_{\frac{\alpha}{2}}$.*
$92\% = (1-\alpha)\cdot 100\% = (1-0.08)\cdot 100\%$, $n - 2 = 5 - 2 = 3$ df
$t_{\frac{\alpha}{2}} = t_{\frac{0.08}{2}} = t_{0.04} =$ (circle one) (circle one) **2.61** / **2.88** / **2.98**.
(PRGM ENTER INVT ENTER 3 ENTER 0.96 ENTER; yes, 0.96–why?)

(d) *Standard error of estimate, $s_e$.*
Standard error of estimate, $s_e =$ (circle one) **0.50** / **0.60** / **0.70**.
(Use STAT TESTS E:LinRegTTest...)

(e) *Sum of squares of $x$.*
$\sum(x_i - \bar{x})^2 =$ (circle one) **0.567** / **0.978** / **1.488**.
(Type $x$, $y$ into $L_1$, $L_2$. Define $L_3 = (L_1 - (-1.82))^2$ ENTER. STAT CALC $L_3$, read $\sum x$.)

(f) *CI of mean stretch length at $x^* = -1.8$.*
Also, $n = 5$, $\bar{x} = -1.82$, and so

$$\hat{y} \pm t_{\frac{\alpha}{2}} s_e \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 37.9 \pm (2.61)(0.60)\sqrt{\frac{1}{5} + \frac{(-1.8-(-1.82))^2}{1.488}} =$$

(circle one) $\mathbf{37.9 \pm 0.3}$ / $\mathbf{37.9 \pm 0.5}$ / $\mathbf{37.9 \pm 0.7}$.

(g) *PI of individual stretch length at $x^* = -1.82$.*
Also, $n = 5$, $\bar{x} = -1.82$, and so

$$\hat{y} \pm t_{\frac{\alpha}{2}} s_e \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum(x_i - \bar{x})^2}} = 37.9 \pm (2.61)(0.60)\sqrt{1 + \frac{1}{5} + \frac{(-1.8-(-1.82))^2}{1.488}} =$$

(circle one) $\mathbf{37.9 \pm 1.72}$ / $\mathbf{37.9 \pm 2.34}$ / $\mathbf{37.9 \pm 3.49}$.

## 14.3 Multiple Regression

Not covered.