

Chapter 10

Simple Linear Regression

We look at scatter diagrams, correlation and linear regression for *paired* data sets.

10.1 Introduction

A scatter diagram is a graph of paired data.

Exercise 10.1 (Scatter Plots)

See Lab 13: Scatter Plots

1. *Reading Ability Versus Level of Illumination.* Consider the following table and corresponding scatter plot¹ of the bivariate data of level of illumination versus ability to read.

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

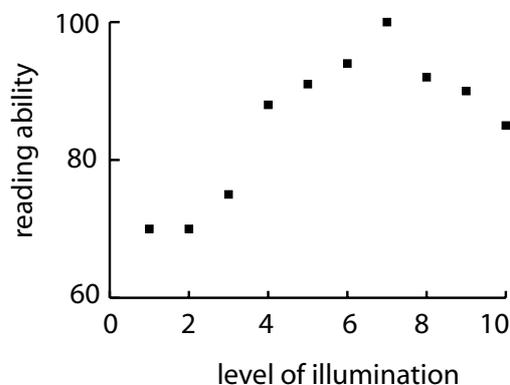


Figure 10.1 (Scatter Plot for Reading Ability Versus Level of Illumination)

¹Type (x, y) into L_1 and L_2 , then 2nd STAT PLOT ENTER ON, choose the scatter plot, make Xlist L_1 and Ylist L_2 , then ZOOM 9:ZoomStat. Also, GROUP the data into a file called READ.

- (a) There are (circle one) **10** / **20** / **30** data points.
- (b) One particular data point is (circle one) **(70, 75)** / **(75, 2)** / **(2, 70)**.
- (c) The data point (9,90) means (circle one)
- i. for a level of illumination of 9, the reading ability is 90.
 - ii. for a reading ability of 9, the level of illumination is 90.
- (d) Reading ability is (circle one) **positively** / **not** / **negatively** associated with level of illumination. As the level of illumination increases, the reading ability (circle one) **increases** / **decreases**.
- (e) **True** / **False** This association is *curved*—it is not linear. It is not linear because a straight line, called a *SD line*, cannot be drawn on this graph such that all the points of the scatter plot would fall on this line.
- (f) Even though it is not linear, a straight line could be drawn on the scatter plot such that the points are all “fairly close” to this straight line. Consequently, the scatter plot might be said to be (circle one) **weakly** / **moderately** / **strongly** linear.
- (g) Reading ability depends on the level of illumination and, certainly, the level of illumination does not depend on reading ability. For this reason, the variable “reading ability” is called the (circle one) **dependent** / **independent** variable and “level of illumination” is called the (circle one) **dependent** / **independent** variable.
- (h) **True** / **False** Sometimes it is not so obvious which is the dependent variable and which is the independent variable. For example, it is not immediately clear which is the independent variable and dependent variable for a scatter plot of husband’s IQ scores and wife’s IQ scores. If you were interested in knowing the husband’s IQ score, *given* the wife’s IQ score, say, then the wife’s IQ score would be the independent variable and the husband’s IQ score would be the dependent variable.
- (i) Both reading ability and level of illumination are examples of quantitative data. Scatter plots can only be drawn for quantitative data. They cannot, sensibly, be drawn for qualitative data because (circle none, one or more)
- i. qualitative data has no order
 - ii. the distance between qualitative data points is not meaningful
- (j) **True** / **False** Scatter plot data and experimental design data (where ANOVA procedures are applied) are very similar to one another. Both involve dependent and independent data. In both, the dependent (response) variable consists of interval data. However, whereas the *independent* (explanatory) variable in scatter plots always consists of a *range* of ratio data, the *independent* variable in experimental designs always consists of a *finite fixed* number of interval (and, possibly, even nominal or ordinal) data values.

- (k) (Review.) If we sampled at random another ten individuals, we would get (circle one) **the same** / **different** scatter plot of (reading ability, level of illumination) scores than those given in the diagram above. The data in the diagram above is an example of a (circle one) **sample** / **population**. This data is said to be (circle one) **observed** / **not observed**.

2. *Grain Yield Versus Distance Upslope From Water.*

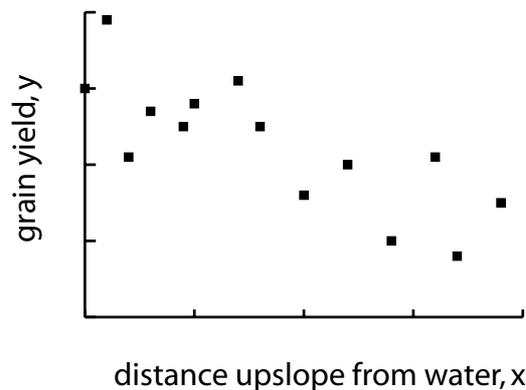


Figure 10.2 (Grain Yield Versus Distance Upslope From Water)

- (a) Describe this scatter plot.
- i. **positive** / **negative**
 - ii. **weak** / **moderate** / **strong**
 - iii. **linear** / **curved**
 - iv. or: **random scatter**
- (b) Does this graph make sense? (Circle none, one or more.)
- i. **No**, because this graph implies that distance up slope depends on grain yield, when, in fact, it should be the other way around.
 - ii. **Yes**, because this graph implies that distance up slope depends on grain yield, which makes sense.
 - iii. **No**, because this graph implies that grain yield is the independent variable.
 - iv. **Yes**, because this graph implies that the grain yield, y , is the dependent variable.
- (c) (Review.) In a *second* random sample of (distance, yield) data, the scatter plot of such points probably (circle one) **would** / **would not** be the same scatter plot as the first random sample of data points given above.

3. *More Scatter Plots*

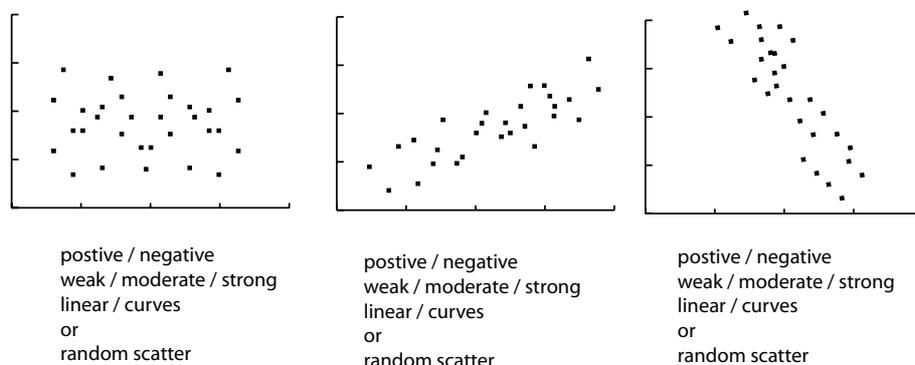


Figure 10.3 (More Scatter Plots)

- (a) Describe each scatter plot by circling the most appropriate descriptions given below each graph.
- (b) (Review) The data in all of the diagrams above are examples of (circle one) **samples / populations**. This data is said to be (circle one) **observed / not observed**.

10.2 The Simple Linear Regression Model

We take a first look at fitting a line to paired data.

Exercise 10.2 (Linear Regression: A First Look) An *estimate* of the least-squares line is given by

$$\hat{\mu}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

where $\hat{\beta}_0$ is the y -intercept and $\hat{\beta}_1$ is the slope.

See Lab 13: Least Squares Line

1. *A First Look: Reading Ability Versus Level of Illumination.*

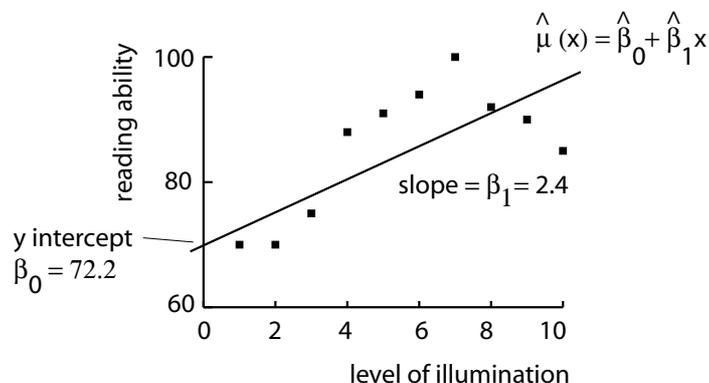


Figure 10.4 (Least-Squares Line for Reading Ability Versus Level of Illumination)

- (a) The *least-squares line* (or linear regression line²) is given by the equation, $\hat{\mu}(x) = 72.2 + 2.418x$. The *y-intercept* of this line, $\hat{\beta}_0$, is (circle one) **72.2** / **2.418**. The *slope*, $\hat{\beta}_1$, of this line is (circle one) **72.2** / **2.418**.
- (b) **True** / **False** The least-squares line $\hat{\mu}(x) = 72.2 + 2.418x$ could also be written as $\hat{\mu}(x) = 2.418x + 72.2$.
- (c) **True** / **False** The *y-intercept*, $\hat{\beta}_0 = 72.2$, is interpreted to mean the average reading ability is 72.2, if the illumination level is zero.
- (d) **True** / **False** The slope, $\hat{\beta}_1 = 2.418$, is interpreted to mean, on average, reading ability increases 2.418 units for an increase of *one* unit of illumination. The slope of the least squares line is usually much more important information to the statistician than the *y-intercept*.
- (e) The *estimated* value of the *response* (or dependent variable) reading ability, $\hat{\mu}(x)$, for the *explanatory* (or independent variable) illumination at the given level $x = 6.5$, is $\hat{\mu}(6.5) = 72.2 + 2.418(6.5) \approx 87.9$. Consequently, the predicted reading ability, $\hat{\mu}(x)$, at level $x = 5.5$ is (circle one) **83.9** / **85.5** / **87.6**.
- (f) Draw a *vertical* line which passes through 6.5 on the “level of illumination” axis. Now draw an *horizontal* line which passes through the point where the solid regression line and the previously drawn vertical line intersect. This horizontal line should intersect the “reading ability” line at about (circle one) **87.9** / **96.1** / **120.6**. Drawing the lines in this way corresponds to using the regression line to predict $\hat{\mu}(x) = 87.9$ for $x = 6.5$.
- (g) *Point of Averages*. The average of x is $\bar{x} = \frac{1+2+\dots+10}{10} = 5.5$ and the average of y is $\bar{y} = \frac{70+70+\dots+85}{10} \approx$ (circle one) (circle one) **83.9** / **85.5** / **87.6**. The point of averages, $(\bar{x}, \bar{y}) = (5.5, 85.5)$ is (circle one) **on** / **not on** the regression line. (Does $85.5 = 72.2 + 2.418(5.5)$?)
- (h) At level $x = 7$, $\hat{\mu}(x)$ is equal to (circle one) **87.9** / **89.1** / **120.6**. In this case, the estimated (or fitted) value, $\hat{\mu}(x)$, can be compared to the *observed* value, $y = 100$; in fact, the difference between the two is called the *error* (or residual) and is given by $e = y - \hat{\mu}(x) =$ (circle one) **9.2** / **10.9** / **12.6**.
- (i) If we were to draw the residual for $x = 7$ on the scatter plot, we would (circle one)
- draw line parallel to the regression line
 - draw a line connecting the point (1,70) to the point (2,70)
 - draw a line connecting the point (1,70) to (10,85)
 - draw a *vertical* line connecting (7,100) to the regression line

²Type (x, y) into L_1 , L_2 and $72.2 + 2.418X$ into $Y=$, then 2nd STAT PLOT, choose the scatter plot and make Xlist L_1 , Ylist L_2 and then ZOOM 9:ZoomStat.

- v. draw a *horizontal* line connecting (7,100) to the regression line
- (j) There are (circle one) **1** / **5** / **10** residuals on the scatter plot.
- (k) At level $x = 20$, $\hat{\mu}(x) \approx 72.2 + 2.418x = 72.2 + 2.418(20) \approx 120.6$. In this case, since $x = 20$ is way outside the range of data, (1,10), the predicted value, $\hat{\mu}(x) = 120.6$, is most likely a (circle one) **poor** / **good** estimate of reading ability.
- (l) (Review.) If we sampled at random another ten individuals, we would get (circle one) **the same** / **different** scatter plot of (reading ability, level of illumination) scores than those given in the diagram above. The data in the diagram above is a example of a (circle one) **sample** / **population**.
- (m) (Review.) If we sampled at random another ten individuals, we would get (circle one) **the same** / **different** *least-squares* line as given in the diagram above.
- (n) (Review.) Since the scatter diagram of data is a *sample*, this must mean that anything calculated from this data, such as the slope, $\hat{\beta}_1$, and *y*-intercept, $\hat{\beta}_0$, and fitted value, $\hat{\mu}(x) = \hat{\beta}_0 + \hat{\beta}_1x$, must all be (circle one) **statistics** / **parameters**.

2. A First Look: Distance Upslope From Water Versus Grain Yield

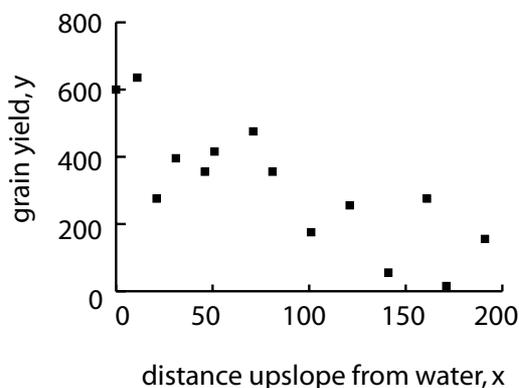


Figure 10.5 (Distance Upslope From Water Versus Grain Yield)

- (a) The least-squares line is given by the equation, $\hat{\mu}(x) = 550 - 1.5x$. The *y*-intercept of this line, $\hat{\beta}_0$, is (circle one) **550** / **-1.5**. The *slope*, $\hat{\beta}_1$, of this line is (circle one) **550** / **-1.5**. Draw this line on the scatter diagram above.
- (b) **True** / **False**. The slope, $\hat{\beta}_1 = -1.5$, is interpreted to mean, on average, grain yield decreases 1.5 tons for a one foot increase in distance from water.
- (c) The predicted grain yield, $\hat{\mu}(x)$, for $x = 100$ feet from water is $\hat{\mu}(x) = 550 - 1.5x = 550 - 1.5(100) =$ (circle one) **200** / **300** / **400**.

- (d) Draw a *vertical* line which passes through 100 on the “distance upslope” axis. Now draw an *horizontal* line which passes through the point where the solid regression line and the previously drawn vertical line intersect. This horizontal line should intersect the “grain yield” line at about (circle one) **200 / 300 / 400**.
- (e) At level $x = 150$, it is *observed* that $y = 400$. At $x = 150$, the *predicted* $\hat{\mu}(x)$ is equal to (circle one) $\hat{\mu}(x) = 550 - 1.5x = 550 - 1.5(150) =$ (circle one) **200 / 325 / 450** and so the residual at $x = 150$ is $y - \hat{\mu}(x) =$ (circle one) **-75 / -25 / 75**.
- (f) (Review.) If we gathered another random sample of data, we would get (circle one) **the same / different** *least-squares line* as given in the diagram above. The slope, $\hat{\beta}_1$, and y -intercept, $\hat{\beta}_0$, and predicted values, $\hat{\mu}(x) = \hat{\beta}_0 + \hat{\beta}_1x$, are all (circle one) **statistics / parameters**.

10.3 Estimating Parameters

We look at how to calculate the slope and intercept parameters in a simple linear regression, using the method of least squares.

Exercise 10.3 (Calculating the Least-Squares Line)

1. *Calculating the Least-Squares Line By Calculator.*

(a) *Reading Ability Versus Level of Illumination.*

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

Verify that the least-squares line is $\hat{\mu}(x) = 72.2 + 2.418x$.

(First, type the (x, y) values of the given data set into the (L_1, L_2) (or UNGROUP it) lists. Then type STAT CALC 8: LinReg(a + bx) ENTER.)

(b) *Circumference Versus Heights of Trees*³.

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

The least-squares line is (circle one)

$$\hat{\mu}(x) = 29.44 + 4.70x / \hat{\mu}(x) = 4.70 + 29.44x / \hat{\mu}(x) = 2.94 + 47.0x.$$

(c) *Student Population Versus Annual Pizza Sales*⁴.

³GROUP as TREE.

⁴GROUP as PIZZA.

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

The least-squares line is (circle one)

$$\hat{\mu}(x) = 60 + 5x \quad / \quad \hat{\mu}(x) = 5 + 60x \quad / \quad \hat{\mu}(x) = 65 + 65x.$$

The scatter plot and graph of this data are given below.

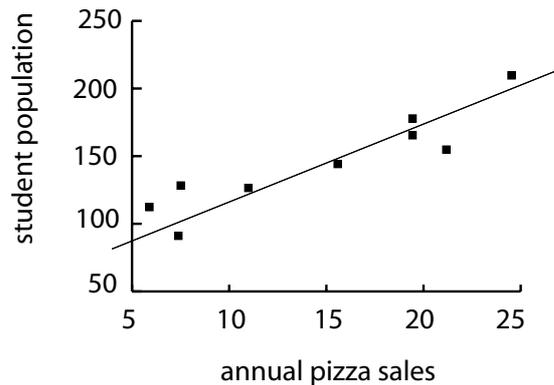


Figure 10.6 (Student Population Versus Annual Pizza Sales)

(Paste the linear regression $y = 60 + 5x$ onto the scatter plot by typing Y= VARS 5:Statistics ENTER EQ 1:RegEQ ENTER, then, as before, ZOOM 9:ZoomStat ENTER.)

2. *Calculating the Least-Squares Line By Formula.* The formula for the least-squares line is

$$\hat{\mu}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

where

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_x}$$

and

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

where \bar{x} is the average of the x values and \bar{y} is the average of the y values,

$$SS_{xy} = \sum_{i=1}^n x_i y_i - \left[\sum_{i=1}^n x_i \sum_{i=1}^n y_i / n \right]$$

and

$$SS_x = \sum_{i=1}^n x_i^2 - \left[\left(\sum_{i=1}^n x_i \right)^2 / n \right]$$

and n is the number of points in the scatter plot.

- (a) If $SS_{xy} = 23$, $SS_x = 122$, $\bar{x} = 12.2$ and $\bar{y} = 16.2$, then
 $\hat{\beta}_1 = \frac{SS_{xy}}{SS_x} = \frac{23}{122} =$ (circle one) **0.19 / 0.34 / 0.56**
 and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 16.2 - 0.19(12.2) =$ (circle one) **12.34 / 13.89 / 14.67**
 and so $\hat{\mu}(x) = \hat{\beta}_0 + \hat{\beta}_1 x =$ (circle one)
 $\hat{\mu}(x) = 11.89 + 0.19x / \hat{\mu}(x) = 12.89 + 0.19x / \hat{\mu}(x) = 13.89 + 0.19x$.
- (b) If $SS_{xy} = 13$, $SS_x = 122$, $\bar{x} = 10.2$ and $\bar{y} = 16.2$, then
 $\hat{\beta}_1 = \frac{SS_{xy}}{SS_x} = \frac{13}{122} =$ (circle one) **0.11 / 0.34 / 0.56**
 and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 16.2 - 0.11(10.2) =$ (circle one) **12.34 / 13.89 / 15.08**
 and so $\hat{\mu}(x) = \hat{\beta}_0 + \hat{\beta}_1 x =$ (circle one)
 $\hat{\mu}(x) = 11.89 + 0.19x / \hat{\mu}(x) = 12.89 + 0.19x / \hat{\mu}(x) = 15.08 + 0.11x$.
- (c) If $\sum_{i=1}^n x_i = -13$, $\sum_{i=1}^n y_i = 12$, $\sum_{i=1}^n x_i^2 = 160$, $\sum_{i=1}^n x_i y_i = -345$, and $n = 5$, then
 $SS_{xy} = \sum_{i=1}^n x_i y_i - [\sum_{i=1}^n x_i \sum_{i=1}^n y_i / n] = -345 - [(-13)(12)] =$
 (circle one) **-189 / -234 / -345**
 $SS_x = \sum_{i=1}^n x_i^2 - [(\sum_{i=1}^n x_i)^2 / n] = 160 - [(-13)^2 / 5] =$
 (circle one) **110.2 / 126.2 / 231.3**
 and so $\hat{\beta}_1 = \frac{SS_{xy}}{SS_x} = \frac{-189}{126.2} =$ (circle one) **-0.19 / -1.34 / -1.50**
 and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{12}{5} - (-1.50)\frac{-13}{5} =$ (circle one) **-1.5 / -2.3 / -4.6**
 and so $\hat{\mu}(x) = \hat{\beta}_0 + \hat{\beta}_1 x =$
 (circle one) $\hat{\mu}(x) = -1.5 + 1.5x / \hat{\mu}(x) = -1.5 - 1.5x / \hat{\mu}(x) = 3 + 1.5x$.
- (d) *Circumference Versus Heights of Trees.*

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

(Type data into L_1 and L_2 and then use STAT CALC 2-Var Stats.)

Since $n =$ (circle one) **5 / 10 / 15**

$\sum_{i=1}^n x_i =$ (circle one) **9.1 / 10.3 / 11.4**

$\sum_{i=1}^n y_i =$ (circle one) **89 / 134 / 190**

$\sum_{i=1}^n x_i^2 =$ (circle one) **18.05 / 19.05 / 20.05**

$\sum_{i=1}^n x_i y_i =$ (circle one) **352.8 / 634.1 / 745.4**

then $SS_{xy} = \sum_{i=1}^n x_i y_i - [\sum_{i=1}^n x_i \sum_{i=1}^n y_i / n] = 352.8 - [(9.1)(190)/5] =$

(circle one) **5 / 6 / 7**

$SS_x = \sum_{i=1}^n x_i^2 - [(\sum_{i=1}^n x_i)^2 / n] = 18.05 - [(9.1)^2 / 5] =$

(circle one) **1.321 / 1.488 / 2.233**

and so $\hat{\beta}_1 = \frac{SS_{xy}}{SS_x} = \frac{7}{1.488} =$ (circle one) **3.704 / 4.704 / 5.704**

and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{190}{5} - (4.704)\frac{9.1}{5} =$ (circle one) **28.44 / 29.44 / 30.44**

and so the least-squares line is

(circle one) $\hat{\mu}(x) = 29.438 + 4.704x$ / $\hat{\mu}(x) = 4.704 + 29.438x$ /
 $\hat{\mu}(x) = 2.944 + 47.04x$.

10.4 Inferences About Expected and Predicted Responses

We look at statistical tests, confidence intervals and prediction intervals for the parameters $\mu(x)$, β_0 and β_1 which appear in the regression line

$$\mu(x) = \beta_0 + \beta_1 x$$

In general, the test statistic is given by

$$t_{\hat{\theta}} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}}$$

and the confidence interval is given by

$$\hat{\theta} \pm t(n-2, \alpha/2) \hat{\sigma}_{\hat{\theta}}$$

where

$$\begin{aligned} \hat{\theta} &= c_0 \hat{\beta}_0 + c_1 \hat{\beta}_1 \\ \hat{\sigma}_{\hat{\theta}}^2 &= \hat{\sigma}^2 \{s_{00} c_0^2 + 2s_{01} c_0 c_1 + s_{11} c_1^2\} \end{aligned}$$

where

$$\begin{aligned} s_{00} &= \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right] \\ s_{11} &= \frac{1}{S_{xx}} \\ s_{01} &= -\frac{\bar{x}}{S_{xx}} \\ \hat{\sigma}^2 &= \frac{S_{yy} - \hat{\beta}_1^2 S_{xx}}{n-2} \end{aligned}$$

and where

$$\begin{aligned} \hat{\beta}_1^2 &= \frac{S_{xy}}{S_{xx}} \\ \hat{\beta}_0^2 &= \bar{y} - \hat{\beta}_1^2 \bar{x} \\ S_{xx} &= \sum_{i=1}^n x_i^2 - [(\sum_{i=1}^n x_i)^2 / n] \end{aligned}$$

$$SS_{xy} = \sum_{i=1}^n x_i y_i - \left[\sum_{i=1}^n x_i \sum_{i=1}^n y_i / n \right]$$

$$S_{yy} = \sum_{i=1}^n y_i^2 - \left[\left(\sum_{i=1}^n y_i \right)^2 / n \right]$$

Exercise 10.4 (Inference In Linear Regression)

1. *Test of β_0 : Reading Ability Versus Illumination.* Consider the following table of the reading ability versus level of illumination data.

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

Test⁵ if the y -intercept, β_0 (the average reading ability when the illumination level is zero (0)), is different from 74 at $\alpha = 0.05$. Also calculate a 95% CI for β_0 .

(Type data in L_1, L_2 , then use PRGM REGINF ENTER {1,0,0,0} ENTER 74 ENTER 0.05 ENTER)

- (a) *Test Statistic Versus Critical Value.*

- i. *Statement.* The statement of the test is (check none, one or more):

- A. $H_0 : \beta_0 = 74$ versus $H_1 : \beta_0 > 74$.
- B. $H_0 : \beta_0 = 74$ versus $H_1 : \beta_0 < 74$.
- C. $H_0 : \beta_0 = 74$ versus $H_1 : \beta_0 \neq 74$.

- ii. *Test.* In this case,

$\hat{\theta} = \hat{\beta}_0 =$ (circle one) **72.2 / 82.2 / 92.2**

(This is ESTM in PGRM REGINF.)

and

$\hat{\sigma}_\theta^2 =$ (circle one) **25.35 / 26.35 / 28.59**

(This is the *square* of SE in PGRM REGINF.)

and so the test statistic is

$$t_{\hat{\theta}} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}} = \frac{72.2 - 74}{\sqrt{28.59}} =$$

(circle one) **-0.33 / -0.45 / -0.55.**

(TOBS in PGRM REGINF)

The critical value at $\alpha = 0.05$, with $n - 2 = 10 - 2 = 8$ degrees of freedom, is (circle one) **-2.31 / -3.89 / -4.26**

(TCRT in PGRM REGINF or use PRGM INVT ENTER 8 ENTER 0.025 ENTER)

⁵Does it really make sense to test β_0 in this case? Is the y -intercept within the range of data?

iii. *Conclusion.* Since the test statistic, -0.33 , is larger than the critical value, -2.31 , we (circle one) **accept** / **reject** the null hypothesis that $\beta_0 = 74$.

(b) *P-Value Versus Level of Significance.*

i. *Statement.* The statement of the test is (check none, one or more):

A. $H_0 : \beta_0 = 74$ versus $H_1 : \beta_0 > 74$.

B. $H_0 : \beta_0 = 74$ versus $H_1 : \beta_0 < 74$.

C. $H_0 : \beta_0 = 74$ versus $H_1 : \beta_0 \neq 74$.

ii. *Test.* Since the test statistic is $t_{\hat{\theta}} = -0.33$, with $n - 2 = 10 - 2 = 8$ degrees of freedom, and this is a two-sided test,

$$\text{p-value} = P(t \leq -0.33) + P(t \geq 0.33) =$$

(circle one) **0.56** / **0.75** / **0.86**

(PVAL in PGRM REGINF or use 2nd DISTR $\text{tcdf}(-E99, -0.33, 8)$.)

The level of significance is 0.05.

iii. *Conclusion.* Since the p-value, 0.75, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the y -intercept is $\beta_0 = 74$.

(c) *Confidence Interval of β_0 .* A 95% CI is given by

$$\begin{aligned} \hat{\theta} \pm t(n-2, \alpha/2) \hat{\sigma}_{\hat{\theta}} &= \hat{\beta}_0 \pm t(n-2, \alpha/2) \sqrt{\hat{\sigma}_{\hat{\theta}}^2} \\ &= 72.2 \pm t(10-2, 0.05/2) \sqrt{28.59} = \end{aligned}$$

(circle one) **(59.87, 84.53)** / **(60.87, 85.53)** / **(61.87, 86.53)**

(CI in PGRM REGINF)

2. *Test of β_1 : Reading Ability Versus Illumination.* Consider the following table of the reading ability versus level of illumination data.

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

Test if the slope, β_1 (the change in reading ability for unit increase in illumination level), is different from zero (0) at $\alpha = 0.05$. Calculate a 95% CI for β_1 .

(Use⁶ PRGM REGINF where {B0,B1,MN,PI} = {0,1,0,0} ENTER 0 ENTER 0.05 ENTER.)

(a) *Test Statistic Versus Critical Value.*

⁶In addition to this program, there is a pre-set program, STAT TESTS E:LinRegTTest, which also tests $\beta_1 = 0$.

i. *Statement.* The statement of the test is (check none, one or more):

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.* In this case,

$\hat{\theta} = \hat{\beta}_1 =$ (circle one) **1.34** / **2.42** / **3.45**

(ESTM in PRGM REGINF)

and

$\hat{\sigma}_{\hat{\theta}}^2 =$ (circle one) **0.47** / **0.57** / **0.74** (*square* of SE in PGRM REGINF)

and so the test statistic is

$$t_{\hat{\theta}} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}} = \frac{2.42 - 0}{\sqrt{0.74}} =$$

(circle one) **2.81** / **3.45** / **3.55**.

(TOBS in PGRM REGINF)

The critical value at $\alpha = 0.05$, with $n - 2 = 10 - 2 = 8$ degrees of freedom, is

(circle one) **2.31** / **3.89** / **4.26**

(TCRT in REGINF or use PRGM INVT ENTER 8 ENTER 0.025 ENTER)

iii. *Conclusion.* Since the test statistic, 2.81, is larger than the critical value, 2.31, we (circle one) **accept** / **reject** the null hypothesis that $\beta_1 = 0$.

(b) *P-Value Versus Level of Significance.*

i. *Statement.* The statement of the test is (check none, one or more):

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.* Since the test statistic is $t_{\hat{\theta}} = 2.81$, with $n - 2 = 10 - 2 = 8$ degrees of freedom, and this is a two-sided test,

$$\text{p-value} = P(t \leq -2.81) + P(t \geq 2.81) =$$

(circle one) **0.01** / **0.02** / **0.03**

(PVAL in PGRM REGINF or use 2nd DISTR *tcdf*(2.81,E99,8).)

The level of significance is 0.05.

iii. *Conclusion.* Since the p-value, 0.02, is smaller than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the slope is $\beta_1 = 0$.

(c) *Confidence Interval of β_1* . A 95% CI is given by

$$\begin{aligned}\hat{\theta} \pm t(n-2, \alpha/2)\hat{\sigma}_{\hat{\theta}} &= \hat{\beta}_1 \pm t(n-2, \alpha/2)\sqrt{\hat{\sigma}_{\hat{\theta}}^2} \\ &= 2.42 \pm t(10-2, 0.05/2)\sqrt{(0.74)} =\end{aligned}$$

(circle one) **(0.43, 4.41)** / **(1.33, 5.41)** / **(2.33, 6.41)**
(CI in PGRM REGINF)

3. *Test of $\mu(x)$: Reading Ability Versus Illumination*. Consider the following table of the reading ability versus level of illumination data.

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

Test if the expected response, $\mu(x)$ is different from 75 at $x = 2.4$ and at $\alpha = 0.05$. Also calculate a 95% confidence interval as well as 95% prediction interval.

(Use PRGM REGINF where {B0,B1,MN,PI} = {0,0,1,1}.)

(a) *Test Statistic Versus Critical Value*.

i. *Statement*. The statement of the test is (check none, one or more):

A. $H_0 : \mu(2.4) = 75$ versus $H_1 : \mu(2.4) > 75$.

B. $H_0 : \mu(2.4) = 75$ versus $H_1 : \mu(2.4) < 75$.

C. $H_0 : \mu(2.4) = 75$ versus $H_1 : \mu(2.4) \neq 75$.

ii. *Test*. In this case,

$\hat{\theta} = \hat{\mu}(x) =$ (circle one) **77.0** / **77.5** / **78.0**

(ESTM in PGRM REGINF)

and $\hat{\sigma}_{\hat{\theta}}^2 =$ (circle one) **1.642** / **2.642** / **13.26**

(square of SE in PGRM REGINF)

and so the test statistic is

$$t_{\hat{\theta}} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}} = \frac{78 - 75}{\sqrt{13.26}} =$$

(circle one) **0.41** / **0.67** / **0.83**.

(TOBS in REGINF)

The critical value at $\alpha = 0.05$, with $n - 2 = 10 - 2 = 8$ degrees of freedom, is

(circle one) **2.31** / **3.89** / **4.26**

(TCRT in REGINF)

iii. *Conclusion*. Since the test statistic, 0.83, is smaller than the critical value, 2.31, we (circle one) **accept** / **reject** the null hypothesis that $\mu(2.4) = 75$.

(b) *P-Value Versus Level of Significance.*

i. *Statement.* The statement of the test is (check none, one or more):

A. $H_0 : \mu(2.4) = 75$ versus $H_1 : \mu(2.4) > 75$.

B. $H_0 : \mu(2.4) = 75$ versus $H_1 : \mu(2.4) < 75$.

C. $H_0 : \mu(2.4) = 75$ versus $H_1 : \mu(2.4) \neq 75$.

ii. *Test.* Since the test statistic is $t_{\hat{\theta}} = 3.64$, with $n - 2 = 10 - 2 = 8$ degrees of freedom, and this is a two-sided test,

$$\text{p-value} = P(t \leq -3.64) + P(t \geq 3.64) =$$

(circle one) **0.21** / **0.32** / **0.43**

(PVAL in PRGM REGINF or use 2nd DISTR $t\text{cdf}(3.64, E99, 8)$.)

The level of significance is 0.05.

iii. *Conclusion.* Since the p-value, 0.43, is larger than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the expected response is $\mu(2.4) = 75$.

(c) *Confidence Interval of $\mu(x)$.* A 95% CI at $x = 2.4$ is given by

$$\hat{\theta} \pm t(n - 2, \alpha/2)\hat{\sigma}_{\hat{\theta}} = 78 \pm t(10 - 2, 0.05/2)(3.642) =$$

(circle one) **(49.61, 86.4)** / **(59.61, 96.4)** / **(69.61, 86.4)**

(CI in PGRM REGINF)

(d) *Prediction Interval of $\mu(x)$.* A 95% PI at $x = 2.4$ for the average of $m = 2$ future values is given by

$$\hat{\theta} \pm t(n - 2, \alpha/2)\hat{\sigma}_{\hat{\theta}_f} = 78 \pm t(10 - 2, 0.05/2)(6.63) =$$

(circle one) **(42.73, 93.28)** / **(58.2, 97.91)** / **(62.73, 93.28)**

(PI in PGRM REGINF)

4. *Confidence Band and Prediction Band: Reading Ability Versus Illumination.*

Consider the following table of the reading ability versus level of illumination data.

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

Calculate and plot the confidence band and prediction band for the expected response, $\mu(x)$, at $\alpha = 0.05$.

(Use PRGM REGINF where $\{B0, B1, MN, PI\} = \{0, 0, 1, 1\}$.)

(a) Previously, we found that the confidence interval for $x = 2.4$ and $\alpha = 0.05$ to be (69.61, 86.4); we can also find the prediction interval to be (58.1, 97.91) when $m = 1$. Complete the following table of CIs and PIs for various values of x .

x	Confidence Interval	Prediction Interval
1.5	(66.04, 85.61)	(55.3, 96.36)
2.4	(69.61, 86.4)	(58.1, 97.91)
3.0	————	(59.88, 99.03)
5.5	(79.79, 91.21)	(66.57, 104.43)
7.5	(83.38, 97.29)	(70.99, 109.68)
9.5	(85.39, 104.96)	(74.64, 115.7)

- (b) The confidence band and prediction band are constructed by connecting all of the confidence intervals and predictions intervals for the various values of x , as shown in the following diagram⁷.

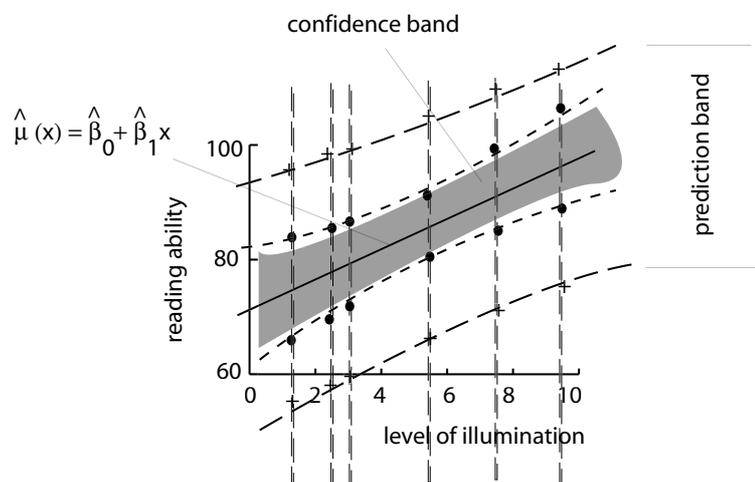


Figure 10.7 (Confidence Band and Prediction Band)

The confidence band is (circle one) **wider** / **narrower** than the prediction band.

5. *Making An ANOVA-like Problem Look Like A Regression Problem: Three Drugs.* The responses⁸ of fifteen different patients are measured for one drug at three dosage levels (in mg).

10 mg	5.90	5.92	5.91	5.89	5.88
20 mg	5.51	5.50	5.50	5.49	5.50
30 mg	5.01	5.00	4.99	4.98	5.02

- (a) The table of data above can be presented in a way that more clearly shows this problem to be a regression problem (circle one):
- i. Arrangement A.

⁷The definition of confidence bands given here is not the same as it is other references; for example, Neter et al., where $\hat{\theta} \pm W\hat{\sigma}$ is used instead of $\hat{\theta} \pm t\hat{\sigma}$, as it is here.

⁸GROUP as DRGREG.

Section 4. Inferences About Expected and Predicted Responses (ATTENDANCE 13)257

dosage, x	10	20	30	10	20	30	...	10	20	30
response, y	5.90	5.92	5.91	5.89	5.88	5.51	...	4.99	4.98	5.02

ii. Arrangement B.

dosage, x	10	10	10	10	10	20	...	30	30	30
response, y	5.90	5.92	5.91	5.89	5.88	5.51	...	4.99	4.98	5.02

(b) A scatter plot indicates this data is
 (circle one) **positive linear** / **negative linear**.
 (Type x and y into L_1 and L_2 respectively, type 2nd STAT PLOT, turn one of the stat-plots on, choose the scatter plot, make Xlist L_1 and Ylist L_2 , then ZOOM 9:ZoomStat ENTER.)

(c) The linear regression line is (circle one)

$$\hat{\mu}(x) = 0.045 - 6.37x /$$

$$\hat{\mu}(x) = 6.37 - 0.045x /$$

$$\hat{\mu}(x) = -0.045 - 6.37x$$

(Use STAT CALC 8:LinReg(a+bx).)

6. *Test of β_1 : Three Drugs.* The responses of fifteen different patients are measured for one drug at three dosage levels (in mg).

10 mg	5.90	5.92	5.91	5.89	5.88
20 mg	5.51	5.50	5.50	5.49	5.50
30 mg	5.01	5.00	4.99	4.98	5.02

Test if the slope, β_1 (the change in patient responses for unit increase in drug dosage), is different from zero (0) at $\alpha = 0.05$. Also calculate a 95% CI for β_1 . (Use PRGM REGINF where $\{B0, B1, MN, PI\} = \{0, 1, 0, 0\}$.)

(a) *Test Statistic Versus Critical Value.*

i. *Statement.* The statement of the test is (check none, one or more):

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.* To test β_1 , notice

$$\hat{\theta} = \hat{\beta}_1 = -0.045 \text{ (ESTM)}$$

$$\hat{\sigma}_{\hat{\theta}}^2 = 0.001^2 \text{ (SE)}^2$$

and so the test statistic

$$t_{\hat{\theta}} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}} = \frac{-0.045 - 0}{0.001} =$$

(circle one) **-45.67 / -49.99 / -58.55** (TOBS).

The lower critical value at $\alpha = 0.05$, with $n - 2 = 15 - 2 = 13$ degrees of freedom, is

(circle one) **-2.16 / -3.89 / -4.26** (TCRT).

- iii. *Conclusion.* Since the test statistic, -49.99, is smaller than the critical value, -2.16, we (circle one) **accept / reject** the null hypothesis that $\beta_1 = 0$.

(b) *P-Value Versus Level of Significance.*

- i. *Statement.* The statement of the test is (check 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.* To test β_1 , notice the p-value is (circle one) **0 / 0.01 / 0.02** (PVAL).

- iii. *Conclusion.* Since the p-value, 0, is smaller than the level of significance, 0.05, we (circle one) **accept / reject** the null hypothesis that the slope is $\beta_1 = 0$.

(c) *Confidence Interval of $\hat{\beta}_1$.* A 95% CI is given by

(circle one) **(-0.05, -0.04) / (-0.06, -0.05) / (-0.07, -0.06)**

(CI in PGRM REGINF)

7. *Test of $\mu(x)$: Three Drugs.* The responses of fifteen different patients are measured for one drug at three dosage levels (in mg).

10 mg	5.90	5.92	5.91	5.89	5.88
20 mg	5.51	5.50	5.50	5.49	5.50
30 mg	5.01	5.00	4.99	4.98	5.02

Test if the expected response, $\mu(x)$ is different from 5.2 at $x = 15$ and at $\alpha = 0.05$ and calculate a prediction interval with $m = 1$.

(Use PRGM REGINF where $\{B0, B1, MN, PI\} = \{0, 0, 1, 1\}$.)

(a) *Test Statistic Versus Critical Value.*

- i. *Statement.* The statement of the test is (circle one):

A. $H_0 : \mu(15) = 5.2$ versus $H_1 : \mu(15) > 5.2$.

B. $H_0 : \mu(15) = 5.2$ versus $H_1 : \mu(15) < 5.2$.

C. $H_0 : \mu(15) = 5.2$ versus $H_1 : \mu(15) \neq 5.2$.

- ii. *Test.* To test $\mu(x)$, notice

$$\begin{aligned}\hat{\theta} &= \hat{\mu}(x) = 5.692 \text{ (ESTM)} \\ \hat{\sigma}_{\hat{\theta}}^2 &= 0.01^2 \text{ (SE)}^2\end{aligned}$$

and so the test statistic is

$$t_{\hat{\theta}} = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}} = \frac{5.692 - 5.2}{0.01} =$$

(circle one) **57.05 / 67.05 / 77.05** (TOBS).

The upper critical value at $\alpha = 0.05$, with $n - 2 = 15 - 2 = 13$ degrees of freedom, is

(circle one) **2.16 / 3.89 / 4.26** (TCRT).

- iii. *Conclusion.* Since the test statistic, 57.05, is larger than the critical value, 2.16, we (circle one) **accept / reject** the null hypothesis that $\mu_{15} = 5.2$.

(b) *P-Value Versus Level of Significance.*

- i. *Statement.* The statement of the test is (circle one):

A. $H_0 : \mu(15) = 5.2$ versus $H_1 : \mu(15) > 5.2$.

B. $H_0 : \mu(15) = 5.2$ versus $H_1 : \mu(15) < 5.2$.

C. $H_0 : \mu(15) = 5.2$ versus $H_1 : \mu(15) \neq 5.2$.

- ii. *Test.* The p-value is

$$\text{p-value} = P(t \leq -57.05) + P(t \geq 57.05) =$$

(circle one) **0 / 0.32 / 0.43** (PVAL)

The level of significance is 0.05.

- iii. *Conclusion.* Since the p-value, 0, is smaller than the level of significance, 0.05, we (circle one) **accept / reject** the null hypothesis that the expected response is $\mu_{15} = 5.2$.

(c) *Confidence Interval of $\mu(x)$.* A 95% CI at $x = 15$ is given by

$$\hat{\theta} \pm t(n - 2, \alpha/2)\hat{\sigma}_{\hat{\theta}} = 5.692 \pm t(13 - 2, 0.05/2)(0.01) =$$

(circle one) **(5.67, 5.71) / (6.67, 6.71) / (7.67, 7.71)**

(CI in PGRM REGINF)

(d) *Prediction Interval of $\mu(x)$.* A 95% PI at $x = 15$ for $m = 1$ future values is given by

$$\hat{\theta} \pm t(n - 2, \alpha/2)\hat{\sigma}_{\hat{\theta}_f} = 5.692 \pm t(13 - 2, 0.05/2)(0.03) =$$

(circle one) **(5.63, 5.76) / (6.67, 6.71) / (7.67, 7.71)**

Notice how the PI is wider than the CI.

10.5 Simple Linear Regression In An ANOVA Setting

We test the slope, β_1 , of linear regression using the analysis of variance technique. We also look at the coefficient of determination. The ANOVA table is

Source	Degrees of Freedom	Sum Of Squares	Mean Squares
Regression (β_1)	1	$SS[R] = \hat{\beta}_1^2 S_{xx}$	MS[R]
Error	n - 2	$SS[E] = S_{yy} - \hat{\beta}_1^2 S_{xx}$	MS[E]
Total	n - 1	$SS[TOT] = S_{yy}$	

where the observed F to test β_1 is

$$F = \frac{MS[R]}{MS[E]}$$

and the observed *coefficient of determination* is

$$R^2 = \frac{SS[R]}{SS[T]}$$

Exercise 10.5 (Linear Regression Using ANOVA)

See Lab 13: Linear Regression Using ANOVA

(Use PRGM REGANOVA in the following exercises⁹.)

1. *Linear Regression Using ANOVA: Reading Ability Versus Illumination.* Consider the following table of the reading ability versus level of illumination data.

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

Test if the slope β_1 (the change in reading ability for unit increase in illumination level) is different from zero (0) at $\alpha = 0.05$; use an ANOVA procedure to do this.

(a) *Test Statistic Versus Critical Value.*

- i. *Statement.* The statement of the test is (check none, one or more):

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

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

⁹In addition to this program, there is also a pre-set program, STAT TESTS E:LinRegTTest, which also tests $\beta_1 = 0$

Section 5. Simple Linear Regression In An ANOVA Setting (ATTENDANCE 13)261

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

- ii. *Test.* The degrees of freedom for the regression are (always) 1 and for the error are $n - 2 = 10 - 2 = 8$ and so the ANOVA table is

Source	Degrees of Freedom	Sum Of Squares	Mean Squares
Regression (β_1)	1	482.4	482.4
Error	8	490.1	61.3
Total	9	972.5	

and so the test statistic is

$$F \text{ test statistic} = \frac{482.4}{61.3} =$$

(circle one) **6.88** / **7.88** / **8.88**.

The critical value at $\alpha = 0.05$, with 1 and $n - 2 = 10 - 2 = 8$ degrees of freedom, is

(circle one) **5.32** / **6.32** / **7.32**

(Use PRGM INVF ENTER 1 ENTER 8 ENTER 0.95 ENTER)

- iii. *Conclusion.* Since the test statistic, 7.88, is larger than the critical value, 5.32, we (circle one) **accept** / **reject** the null hypothesis that the slope β_1 is zero.

(b) *P-Value Versus Level of Significance.*

- i. *Statement.* The statement of the test is (check none, one or more):

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.* Since the test statistic is $F = 7.88$, the p-value, with 1 and $n - 2 = 10 - 2 = 8$ degrees of freedom, is given by

$$\text{p-value} = P(F \geq 7.88)$$

which equals (circle one) **0.00** / **0.022** / **0.043**.

(Use PRGM REGANOVA or use 2nd DISTR 9:Fcdf(7.88,E99,1,8).)

The level of significance is 0.05.

- iii. *Conclusion.* Since the p-value, 0.022, is smaller than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the slope β_1 is zero.

(c) *ANOVA Versus Linear Regression.* The conclusions reached using the F -distribution with the ANOVA procedure here is (circle one) **the same as** / **different from** the conclusions reached previously using the t -distribution.

2. *Linear Regression Using ANOVA: Three Drugs.* The responses of fifteen different patients are measured for one drug at three dosage levels (in mg).

10 mg	5.90	5.92	5.91	5.89	5.88
20 mg	5.51	5.50	5.50	5.49	5.50
30 mg	5.01	5.00	4.99	4.98	5.02

Test if the slope, β_1 (the change in patient responses for unit increase in drug dosage), is different from zero (0) at $\alpha = 0.05$; use the ANOVA procedure to do this.

(a) *Test Statistic Versus Critical Value.*

- i. *Statement.* The statement of the test is (check none, one or more):

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.* The degrees of freedom for the regression are (always) 1 and for the error are $n - 2 = 15 - 2 = 13$ and the ANOVA table is

Source	Degrees of Freedom	Sum Of Squares	Mean Squares
Regression (β_1)	1	2.025	2.025
Error	13	0.0105	0.00081
Total	14	2.0355	

and so the test statistic is

$$F \text{ test statistic} = \frac{2.025}{0.00081} =$$

(circle one) **2499.2 / 2599.2 / 2699.2.**

The upper critical value at $\alpha = 0.05$, with 1 and $n - 2 = 15 - 2 = 13$ degrees of freedom, is

(circle one) **4.32 / 4.67 / 5.67**

(Use PRGM INV F ENTER 1 ENTER 13 ENTER 0.95 ENTER)

- iii. *Conclusion.* Since the test statistic, 2499.2, is larger than the critical value, 4.67, we (circle one) **accept / reject** the null hypothesis that the slope β_1 is zero.

(b) *P-Value Versus Level of Significance.*

- i. *Statement.* The statement of the test is (check none, one or more):

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.* Since the test statistic is $F = 2499.2$, the p-value, with 1 and $n - 2 = 13 - 2 = 13$ degrees of freedom, is given by

$$\text{p-value} = P(F \geq 2499.2)$$

which equals (circle one) **0.00** / **0.022** / **0.043**.

(Use 2nd DISTR 9:Fcdf(2499.2,E99,1,13).)

The level of significance is 0.05.

- iii. *Conclusion.* Since the p-value, 0.0, is smaller than the level of significance, 0.05, we (circle one) **accept** / **reject** the null hypothesis that the slope β_1 is zero.

Exercise 10.6 (Coefficient of Determination) The observed coefficient of determination, R^2 , is given by

$$R^2 = \frac{SS[R]}{SS[T]}$$

1. *Coefficient of Determination: Reading Ability Versus Illumination.* Consider the following table of the reading ability versus level of illumination data.

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

Since the ANOVA table is given by

Source	Degrees of Freedom	Sum Of Squares	Mean Squares
Regression (β_1)	1	482.4	482.4
Error	8	490.1	61.3
Total	9	972.5	

the coefficient of determination, $R^2 = \frac{SS[R]}{SS[T]} = \frac{482.4}{972.5}$, is given by

(circle one) **0.49** / **0.70** / **0.99**.

2. *Coefficient of Determination: Three Drugs.* The responses of fifteen different patients are measured for one drug at three dosage levels (in mg).

10 mg	5.90	5.92	5.91	5.89	5.88
20 mg	5.51	5.50	5.50	5.49	5.50
30 mg	5.01	5.00	4.99	4.98	5.02

Since the ANOVA table is given by

Source	Degrees of Freedom	Sum Of Squares	Mean Squares
Regression (β_1)	1	2.025	2.025
Error	13	0.0105	0.00081
Total	14	2.0355	

the coefficient of determination, $R^2 = \frac{SS[R]}{SS[T]}$, is given by

(circle one) **0.49** / **0.70** / **0.99**.

3. The coefficient of determination measures what percentage of the variation in the observed response variable can be explained by the linear regression. For example, for the reading ability problem, (circle one) **49%** / **99%** of the variation can be explained by $\hat{\mu}(x) = 72.2 + 2.4x$. If the points of the scatter plot are close to the linear regression line, R^2 is large.