Practice Problems 1b

1. **Do <u>NOT</u> use a computer for this problem.**

An Anytown State University student wishes to examine the relationship between the monthly rent amount for an apartment (y) (in $) and the number of bedrooms (x). The data are as follows:

x	y
2	567
2	728
2	754
3	808
2	841
2	978
4	1,127
3	1,145
4	1,190
3	1,197
4	1,264
3	1,284
4	1,301
3	1,310
4	1,416
3	1,434

Consider the model

$$Y = \beta_0 + \beta_1 x + \varepsilon, \qquad i = 1, 2, \ldots, 16.$$

where ε's are i.i.d. $N(0, \sigma^2)$.

$$\sum x = 48, \qquad \sum y = 17{,}344,$$

$$\sum x^2 = 154, \qquad \sum y^2 = 19{,}883{,}146,$$

$$\sum x y = 54{,}462,$$

$$\sum (x - \bar{x})^2 = 10, \qquad \sum (y - \bar{y})^2 = 1{,}082{,}250,$$

$$\sum (x - \bar{x})(y - \bar{y}) = \sum (x - \bar{x})\, y = 2{,}430.$$

a) Find the equation of the least-squares regression line.

b) Perform the significance of the regression test at a 5% level of significance. Specify the null and the alternative hypotheses. Report the value of the test statistic, the critical value(s), and the decision.

c) Test $H_0: \beta_1 = 202$ vs. $H_1: \beta_1 > 202$ at a 5% level of significance. Report the value of the test statistic, the p-value, and the decision.

d) Construct a 95% confidence interval for the average monthly rent amount for an apartment that has 2 bedrooms.

e) What is the p-value of the test $H_0: \beta_0 = 0$ vs. $H_1: \beta_0 \neq 0$? (You may give a range.)

2. **Do NOT use a computer for this problem.**

An Anytown State University student wishes to examine the relationship between the monthly rent amount for an apartment (y) (in \$), the apartment size (x_1) (in thousands of sq. ft.), the number of bedrooms (x_2), and the distance between the apartment and the Student Union Building (x_3) (in miles). The data are as follows:

x_1	x_2	x_3	y
1.9	2	2.5	567
1.6	2	0.8	728
1.8	2	1.6	754
1.6	3	3.8	808
1.7	2	1.7	841
1.6	2	1.3	978
1.9	4	3.5	1,127
1.5	3	3.4	1,145
2.0	4	1.9	1,190
1.9	3	3.0	1,197
1.8	4	2.1	1,264
1.8	3	2.1	1,284
1.7	4	1.7	1,301
2.0	3	0.9	1,310
2.2	4	2.7	1,416
1.8	3	0.6	1,434

Consider the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon, \qquad i = 1, 2, \ldots, 16.$$

where ε's are i.i.d. $N(0, \sigma^2)$.

Then

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} 16 & 28.8 & 48 & 33.6 \\ 28.8 & 52.34 & 87.4 & 60.48 \\ 48 & 87.4 & 154 & 104.8 \\ 33.6 & 60.48 & 104.8 & 85.06 \end{bmatrix},$$

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \begin{bmatrix} 7.0483 & -3.942 & 0.171 & -0.192 \\ -3.942 & 2.58 & -0.29 & 0.08 \\ 0.171 & -0.29 & 0.145 & -0.04 \\ -0.192 & 0.08 & -0.04 & 0.08 \end{bmatrix},$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \begin{bmatrix} 17{,}344 \\ 31{,}514.2 \\ 54{,}462 \\ 36{,}092.4 \end{bmatrix}, \qquad \hat{\boldsymbol{\beta}} = \begin{bmatrix} ? \\ ? \\ ? \\ ? \end{bmatrix},$$

$\Sigma(y_i - \hat{y}_i)^2 = 360{,}000$,

and $\Sigma(y_i - \bar{y})^2 = 1{,}082{,}250$.

a) Obtain the least-squares estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$.

2. (continued)

b) Perform the significance of the regression test at a 5% level of significance. Specify the null and the alternative hypotheses. Report the value of the test statistic, the critical value(s), and the decision.

c) Construct a 95% prediction interval for the monthly rent amount for an apartment that has 1.8 thousand sq. ft., 2 bedrooms, and is 2.3 miles away from the Student Union Building.

d) Test $H_0: \beta_3 = 0$ vs. $H_1: \beta_3 \neq 0$ at a 5% level of significance. Report the value of the test statistic, the critical value(s), and the decision. Give a range for the p-value.

e) Test $H_0: \beta_2 = \beta_3 = 0$ vs. $H_1:$ (either $\beta_2 \neq 0$, or $\beta_3 \neq 0$) at a 5% level of significance. Report the value of the test statistic, the critical value(s), and the decision.

"Hint": You do have Σx_1, Σy, Σx_1^2, $\Sigma x_1 y$.

3. **<u>Do</u> use a computer for this problem.**

An Anytown State University student wishes to examine the relationship between the monthly rent amount for an apartment (y) (in \$) and the number of bedrooms (x). The data are stored in `420Pr02.csv`.

Consider the model

$$Y = \beta_0 + \beta_1 x + \varepsilon, \qquad i = 1, 2, \ldots, 16.$$

where ε's are i.i.d. $N(0, \sigma^2)$.

a) Find the equation of the least-squares regression line.

b) Perform the significance of the regression test at a 5% level of significance. On your printout, circle or highlight the values of the test statistic and the p-value for this test.

d) Construct a 95% confidence interval for the average monthly rent amount for an apartment that has 2 bedrooms.

e) What is the p-value of the test $H_0: \beta_0 = 0$ vs. $H_1: \beta_0 \neq 0$?

x	y
2	567
2	728
2	754
3	808
2	841
2	978
4	1,127
3	1,145
4	1,190
3	1,197
4	1,264
3	1,284
4	1,301
3	1,310
4	1,416
3	1,434

4. **Do use a computer for this problem.**

An Anytown State University student wishes to examine the relationship between the monthly rent amount for an apartment (y) (in $), the apartment size ($x_1$) (in thousands of sq. ft.), the number of bedrooms (x_2), and the distance between the apartment and the Student Union Building (x_3) (in miles). The data are stored in `420Pr02.csv`.

x_1	x_2	x_3	y
1.9	2	2.5	567
1.6	2	0.8	728
1.8	2	1.6	754
1.6	3	3.8	808
1.7	2	1.7	841
1.6	2	1.3	978
1.9	4	3.5	1,127
1.5	3	3.4	1,145
2.0	4	1.9	1,190
1.9	3	3.0	1,197
1.8	4	2.1	1,264
1.8	3	2.1	1,284
1.7	4	1.7	1,301
2.0	3	0.9	1,310
2.2	4	2.7	1,416
1.8	3	0.6	1,434

Consider the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon, \qquad i = 1, 2, \ldots, 16.$$

where ε's are i.i.d. $N(0, \sigma^2)$.

a) Obtain the least-squares estimates $\hat\beta_0$, $\hat\beta_1$, $\hat\beta_2$, and $\hat\beta_3$.

b) Perform the significance of the regression test at a 5% level of significance. On your printout, circle or highlight the values of the test statistic and the p-value for this test.

c) Construct a 95% prediction interval for the monthly rent amount for an apartment that has 1.8 thousand sq. ft., 2 bedrooms, and is 2.3 miles away from the Student Union Building.

d) Test $H_0: \beta_3 = 0$ vs. $H_1: \beta_3 \neq 0$ at a 5% level of significance. On your printout, circle or highlight the values of the test statistic and the p-value for this test.

e) Test $H_0: \beta_2 = \beta_3 = 0$ vs. $H_1:$ (either $\beta_2 \neq 0$, or $\beta_3 \neq 0$) at a 5% level of significance. On your printout, circle or highlight the values of the test statistic and the p-value for this test.

5. Suppose the distance between the apartment and the Student Union Building (X) (in miles) and the monthly rent amount for a 4-bedroom apartment (Y) (in \$) in Anytown have a bivariate normal distribution with

$$\mu_X = 2, \qquad \sigma_X = 0.6, \qquad \mu_Y = 1{,}210, \qquad \sigma_Y = 200, \qquad \rho = -0.28.$$

What is the probability that the monthly rent amount is under \$1,250 for a 4-bedroom apartment located 2.6 miles away from with the Student Union Building?

6. An Anytown State University student wishes to examine the relationship between the monthly rent amount for an apartment (y) (in \$) and the apartment size (x) (in thousands of sq. ft.). The data are as follows:

x	y
1.9	567
1.6	728
1.8	754
1.6	808
1.7	841
1.6	978
1.9	1,127
1.5	1,145
2.0	1,190
1.9	1,197
1.8	1,264
1.8	1,284
1.7	1,301
2.0	1,310
2.2	1,416
1.8	1,434

Assume that (X, Y) have a bivariate normal distribution.

$$\sum x = 28.8, \qquad \sum y = 17{,}344,$$

$$\sum x^2 = 52.34, \qquad \sum y^2 = 19{,}883{,}146,$$

$$\sum x\,y = 31{,}514.2,$$

$$\sum (x - \overline{x})^2 = 0.5, \qquad \sum (y - \overline{y})^2 = 1{,}082{,}250,$$

$$\sum (x - \overline{x})(y - \overline{y}) = \sum (x - \overline{x})\,y = 295.$$

a) Find the sample correlation coefficient r.

b) Find the p-value of the test $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$.

Use $T = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ and either EXCEL TDIST or R `pt`.

c) Find the (approximate) p-value of the test $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$. Use $W = \dfrac{1}{2}\ln\left(\dfrac{1+r}{1-r}\right)$.

d) Find the (approximate) p-value of the test $H_0 : \rho = 0.20$ vs. $H_1 : \rho > 0.20$.

e) Construct a 90% confidence interval for ρ.

$100(1-\alpha)\,\%$ confidence interval for ρ:

$$\left(\frac{e^a - 1}{e^a + 1}, \frac{e^b - 1}{e^b + 1}\right), \quad \text{where} \quad a = \ln\frac{1+r}{1-r} - \frac{2\,z_{\alpha/2}}{\sqrt{n-3}}, \quad b = \ln\frac{1+r}{1-r} + \frac{2\,z_{\alpha/2}}{\sqrt{n-3}}.$$

Answers:

1. **Do <u>NOT</u> use a computer for this problem.**

An Anytown State University student wishes to examine the relationship between the monthly rent amount for an apartment (y) (in $) and the number of bedrooms (x). The data are as follows:

x	y
2	567
2	728
2	754
3	808
2	841
2	978
4	1,127
3	1,145
4	1,190
3	1,197
4	1,264
3	1,284
4	1,301
3	1,310
4	1,416
3	1,434

Consider the model

$$Y = \beta_0 + \beta_1 x + \varepsilon, \qquad i = 1, 2, \ldots, 16.$$

where ε's are i.i.d. $N(0, \sigma^2)$.

$$\sum x = 48, \qquad \sum y = 17{,}344,$$

$$\sum x^2 = 154, \qquad \sum y^2 = 19{,}883{,}146,$$

$$\sum x\,y = 54{,}462,$$

$$\sum (x - \bar{x})^2 = 10, \qquad \sum (y - \bar{y})^2 = 1{,}082{,}250,$$

$$\sum (x - \bar{x})(y - \bar{y}) = \sum (x - \bar{x})\,y = 2{,}430.$$

a) Find the equation of the least-squares regression line.

$$\bar{x} = \frac{\sum x}{n} = \frac{48}{16} = 3. \qquad \bar{y} = \frac{\sum y}{n} = \frac{17{,}344}{16} = 1{,}084.$$

SXX = 10, SXY = 2,430. $\hat{\beta}_1 = \frac{\text{SXY}}{\text{SXX}} = \frac{2{,}430}{10} = \mathbf{243}.$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} = 1{,}084 - 243 \cdot 3 = \mathbf{355}.$

Least-squares regression line: $\boldsymbol{\hat{y} = 355 + 243\,x}.$

b) Perform the significance of the regression test at a 5% level of significance. Specify the null and the alternative hypotheses. Report the value of the test statistic, the critical value(s), and the decision.

$H_0 : \boldsymbol{\beta_1 = 0}$ vs $H_1 : \boldsymbol{\beta_1 \neq 0}$.

$$\text{SSRegression} = \hat{\beta}_1^{\,2} \times \text{SXX} = 243^2 \times 10 = 590{,}490.$$

$$\text{RSS} = \text{SYY} - \text{SSRegression} = 1{,}082{,}250 - 590{,}490 = 491{,}760.$$

Source	*SS*	*df*	*MS*	*F*
Regression	590,490	$2 - 1 = 1$	590,490	**16.81**
Residuals	491,760	$n - 2 = 14$	35,125.7143	
Total	1,082,250	$n - 1 = 15$		

Critical Value: $F_{0.05}(1, 14) = \mathbf{4.60}$.

Decision: **Reject $\mathbf{H_0}$** at $\alpha = 0.05$.

OR

$$s^2 = \frac{\text{RSS}}{n-2} = \frac{491{,}760}{14} \approx 35{,}125.7143.$$

Test Statistic:

$$T = \frac{\hat{\beta}_1 - \beta_{10}}{s \big/ \sqrt{\text{SXX}}} = \frac{243 - 0}{\sqrt{35{,}125.7143} \big/ \sqrt{10}} \approx \mathbf{4.1}.$$

Critical Values: $\pm\, t_{0.025}(14) = \pm\, \mathbf{2.145}$.

Decision: **Reject $\mathbf{H_0}$** at $\alpha = 0.05$.

c) Test $H_0 : \beta_1 = 202$ vs. $H_1 : \beta_1 > 202$ at a 5% level of significance. Report the value of the test statistic, the p-value, and the decision.

Test Statistic:

$$T = \frac{\hat{\beta}_1 - \beta_{10}}{s \big/ \sqrt{SXX}} = \frac{243 - 202}{\sqrt{35{,}125.7143} \big/ \sqrt{10}} \approx \mathbf{0.692}.$$

Rejection Region:

Reject H_0 if $T > t_{0.05}(14 \text{ df}) = 1.761$.

Do NOT Reject $\mathbf{H_0}$ at $\alpha = 0.05$.

OR

P-value $\approx$ **0.25** $> 0.05 = \alpha$.

Do NOT Reject $\mathbf{H_0}$ at $\alpha = 0.05$.

d) Construct a 95% confidence interval for the average monthly rent amount for an apartment that has 2 bedrooms.

Confidence interval for $\mu(x)$: $\hat{y} \pm t_{\alpha/2} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{SXX}}$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 355 + 243 \cdot 2 = 841.$

$n - 2 = 14$ degrees of freedom. $t_{0.025}(14) = 2.145.$

$841 \pm 2.145 \cdot \sqrt{35{,}125.7143} \cdot \sqrt{\frac{1}{16} + \frac{(2-3)^2}{10}}$ **841 ± 162**

(679 , 1,003)

e) What is the p-value of the test $H_0 : \beta_0 = 0$ vs. $H_1 : \beta_0 \neq 0$?
(You may give a range.)

$$T = \frac{\hat{\beta}_0 - \beta_{00}}{s_e \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{SXX}}} = \frac{355 - 0}{\sqrt{35{,}125.7143} \cdot \sqrt{\frac{1}{16} + \frac{3^2}{10}}} = \mathbf{1.9307}.$$

$n - 2 = 14$ degrees of freedom.

$t_{0.05}(14) = 1.761 < T < t_{0.025}(14) = 2.145$.

$0.025 < \text{one tail} < 0.05$.

Two – tailed. P-value = two tails.

0.05 < p-value < 0.10.

2. **Do <u>NOT</u> use a computer for this problem.**

An Anytown State University student wishes to examine the relationship between the monthly rent amount for an apartment (y) (in \$), the apartment size (x_1) (in thousands of sq. ft.), the number of bedrooms (x_2), and the distance between the apartment and the Student Union Building (x_3) (in miles). The data are as follows:

x_1	x_2	x_3	y
1.9	2	2.5	567
1.6	2	0.8	728
1.8	2	1.6	754
1.6	3	3.8	808
1.7	2	1.7	841
1.6	2	1.3	978
1.9	4	3.5	1,127
1.5	3	3.4	1,145
2.0	4	1.9	1,190
1.9	3	3.0	1,197
1.8	4	2.1	1,264
1.8	3	2.1	1,284
1.7	4	1.7	1,301
2.0	3	0.9	1,310
2.2	4	2.7	1,416
1.8	3	0.6	1,434

Consider the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon, \qquad i = 1, 2, \dots, 16.$$

where ε's are i.i.d. $N(0, \sigma^2)$.

Then

$$\mathbf{X}^T\mathbf{X} = \begin{bmatrix} 16 & 28.8 & 48 & 33.6 \\ 28.8 & 52.34 & 87.4 & 60.48 \\ 48 & 87.4 & 154 & 104.8 \\ 33.6 & 60.48 & 104.8 & 85.06 \end{bmatrix},$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 7.0483 & -3.942 & 0.171 & -0.192 \\ -3.942 & 2.58 & -0.29 & 0.08 \\ 0.171 & -0.29 & 0.145 & -0.04 \\ -0.192 & 0.08 & -0.04 & 0.08 \end{bmatrix},$$

$$\mathbf{X}^T\mathbf{Y} = \begin{bmatrix} 17{,}344 \\ 31{,}514.2 \\ 54{,}462 \\ 36{,}092.4 \end{bmatrix}, \qquad \hat{\boldsymbol{\beta}} = \begin{bmatrix} ? \\ ? \\ ? \\ ? \end{bmatrix},$$

$\sum(y_i - \hat{y}_i)^2 = 360{,}000,$

and $\sum(y_i - \bar{y})^2 = 1{,}082{,}250.$

a) Obtain the least-squares estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} = \begin{bmatrix} 7.0483 & -3.942 & 0.171 & -0.192 \\ -3.942 & 2.58 & -0.29 & 0.08 \\ 0.171 & -0.29 & 0.145 & -0.04 \\ -0.192 & 0.08 & -0.04 & 0.08 \end{bmatrix} \begin{bmatrix} 17{,}344 \\ 31{,}514.2 \\ 54{,}462 \\ 36{,}092.4 \end{bmatrix} = \begin{bmatrix} \mathbf{400} \\ \mathbf{30} \\ \mathbf{280} \\ \mathbf{-100} \end{bmatrix}.$$

b) Perform the significance of the regression test at a 5% level of significance. Specify the null and the alternative hypotheses. Report the value of the test statistic, the critical value(s), and the decision.

$H_0: \boldsymbol{\beta_1 = \beta_2 = \beta_3 = 0}$ vs H_1: at least one of $\beta_1, \beta_2, \beta_3$ is not zero.

$n = 16$, $p = (\#\text{ of } \beta\text{'s}) = (\#\text{ of columns of matrix } \mathbf{X}) = 4$.

Source	*SS*	*df*	*MS*	*F*
Regression	722,250	$p - 1 = 3$	240,750	**8.025**
Residuals	360,000	$n - p = 12$	30,000	
Total	1,082,250	$n - 1 = 15$		

Critical Value: $F_{0.05}(3, 12) = \mathbf{3.49}$.

Decision: **Reject $\mathbf{H_0}$** at $\alpha = 0.05$.

c) Construct a 95% prediction interval for the monthly rent amount for an apartment that has 1.8 thousand sq. ft., 2 bedrooms, and is 2.3 miles away from the Student Union Building.

$$X_0^T = [\ 1 \quad 1.8 \quad 2 \quad 2.3\ ]$$

$$\hat{Y}_0 = 1 \times 400 + 1.8 \times 30 + 2 \times 280 + 2.3 \times (-100) = 784.$$

$$X_0^T \mathbf{C} X_0 = [1 \quad 1.8 \quad 2 \quad 2.3] \cdot \begin{bmatrix} 7.0483 & -3.942 & 0.171 & -0.192 \\ -3.942 & 2.58 & -0.29 & 0.08 \\ 0.171 & -0.29 & 0.145 & -0.04 \\ -0.192 & 0.08 & -0.04 & 0.08 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1.8 \\ 2 \\ 2.3 \end{bmatrix}$$

$$= [-0.1469 \quad 0.306 \quad -0.153 \quad 0.056] \cdot \begin{bmatrix} 1 \\ 1.8 \\ 2 \\ 2.3 \end{bmatrix} = 0.2267.$$

$$\hat{\text{Var}}(\hat{Y}_0) = [1 + X_0^T \mathbf{C} X_0]\, s^2 = (1 + 0.2267) \times 30{,}000 = 36{,}801.$$

$$t_{0.025}(12) = 2.179.$$

$784 \pm 2.179 \times \sqrt{36{,}801}$ $\qquad$ $\mathbf{784 \pm 418}$ $\qquad$ $\mathbf{(366, 1{,}202)}$

d) Test $H_0: \beta_3 = 0$ vs. $H_1: \beta_3 \neq 0$ at a 5% level of significance. Report the value of the test statistic, the critical value(s), and the decision. Give a range for the p-value.

$$\hat{\text{Var}}(\hat{\beta}_3) = C_{33} \times s^2 = 0.08 \times 30{,}000 = 2{,}400.$$

Test Statistic: $\quad t = \dfrac{-100 - 0}{\sqrt{2{,}400}} \approx \mathbf{-2.041}.$

Critical Values: $\quad \pm t_{0.025}(12) = \mathbf{\pm 2.179}.$

Decision: **Do NOT Reject $\mathbf{H_0}$** at $\alpha = 0.05$.

e) Test $H_0 : \beta_2 = \beta_3 = 0$ vs. $H_1 : (\text{either } \beta_2 \neq 0, \text{ or } \beta_3 \neq 0)$ at a 5% level of significance. Report the value of the test statistic, the critical value(s), and the decision.

Null model: $Y = \beta_0 + \beta_1 x_1 + \varepsilon$ – Simple Linear Regression

$q = (\text{\# of } \beta\text{'s}) = 2.$

$\sum x_1 = 28.8,\qquad \sum y = 17{,}344,\qquad \sum x_1^2 = 52.34,\qquad \sum x_1 y = 31{,}514.2.$

$$S X_1 X_1 = 52.34 - \frac{1}{16}(28.8)^2 = 0.5.$$

$$S X_1 Y = 31{,}514.2 - \frac{1}{16}(28.8)(17{,}344) = 295.$$

$$\hat{\beta}_1 = \frac{SX_1Y}{SX_1X_1} = \frac{295}{0.5} = 590.$$

$$SSRegr_{Null} = \hat{\beta}_1^2 \, SX_1X_1 = 590^2 \times 0.5 = 174{,}050.$$

$$RSS_{Null} = SYY - SSRegr_{Null} = 1{,}082{,}250 - 174{,}050 = 908{,}200.$$

	SS	*DF*	*MS*	*F*
Diff.	$RSS_{Null} - RSS_{Full}$	$p - q$	…	…
Full	RSS_{Full}	$n - p$	…	
Null	RSS_{Null}	$n - q$		

	SS	*DF*	*MS*	*F*	
Diff.	548,200	2	274,100	**9.136667**	← Test Statistic
Full	360,000	12	30,000		
Null	908,200	14			

Critical Value: $F_{0.05}(2, 12) = \mathbf{3.89}$.

Decision: **Reject H_0** at $\alpha = 0.05$.

3. **Do use a computer for this problem.**

An Anytown State University student wishes to examine the relationship between the monthly rent amount for an apartment (y) (in $) and the number of bedrooms (x). The data are stored in `420Pr02.csv`.

Consider the model

$$Y = \beta_0 + \beta_1 x + \varepsilon, \qquad i = 1, 2, \dots, 16.$$

where ε's are i.i.d. $N(0, \sigma^2)$.

a) Find the equation of the least-squares regression line.

b) Perform the significance of the regression test at a 5% level of significance. On your printout, circle or highlight the values of the test statistic and the p-value for this test.

d) Construct a 95% confidence interval for the average monthly rent amount for an apartment that has 2 bedrooms.

e) What is the p-value of the test $H_0: \beta_0 = 0$ vs. $H_1: \beta_0 \neq 0$?

x	y
2	567
2	728
2	754
3	808
2	841
2	978
4	1,127
3	1,145
4	1,190
3	1,197
4	1,264
3	1,284
4	1,301
3	1,310
4	1,416
3	1,434

```
x <- c(2,2,2,3,2,2,4,3,4,3,4,3,4,3,4,3)
y <- c(567,728,754,808,841,978,1127,
    1145,1190,1197,1264,1284,1301,1310,1416,1434)> attach(Pr02)
```

```
fit <- lm(y ~ x)
summary(fit)

Call:
lm(formula = y ~ x2)

Residuals:
   Min     1Q Median     3Q    Max
  -276   -119    -13    119    350

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   355.00     183.87   1.931  0.07403 .
x             243.00      59.27   4.100  0.00108 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 187.4 on 14 degrees of freedom
Multiple R-squared:  0.5456,    Adjusted R-squared:  0.5132
F-statistic: 16.81 on 1 and 14 DF,  p-value: 0.001082

new <- data.frame( x=2 )
predict.lm(fit, new, interval=c("confidence"), level=0.95)
  fit      lwr     upr
1 841 678.9596 1003.04
```

4. **Do use a computer for this problem.**

An Anytown State University student wishes to examine the relationship between the monthly rent amount for an apartment (y) (in \$), the apartment size (x_1) (in thousands of sq. ft.), the number of bedrooms (x_2), and the distance between the apartment and the Student Union Building (x_3) (in miles). The data are stored in `420Pr02.csv`.

x_1	x_2	x_3	y
1.9	2	2.5	567
1.6	2	0.8	728
1.8	2	1.6	754
1.6	3	3.8	808
1.7	2	1.7	841
1.6	2	1.3	978
1.9	4	3.5	1,127
1.5	3	3.4	1,145
2.0	4	1.9	1,190
1.9	3	3.0	1,197
1.8	4	2.1	1,264
1.8	3	2.1	1,284
1.7	4	1.7	1,301
2.0	3	0.9	1,310
2.2	4	2.7	1,416
1.8	3	0.6	1,434

Consider the model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon, \qquad i = 1, 2, \ldots, 16.$$

where ε's are i.i.d. $N(0, \sigma^2)$.

a) Obtain the least-squares estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$.

b) Perform the significance of the regression test at a 5% level of significance. On your printout, circle or highlight the values of the test statistic and the p-value for this test.

c) Construct a 95% prediction interval for the monthly rent amount for an apartment that has 1.8 thousand sq. ft., 2 bedrooms, and is 2.3 miles away from the Student Union Building.

d) Test $H_0: \beta_3 = 0$ vs. $H_1: \beta_3 \neq 0$ at a 5% level of significance. On your printout, circle or highlight the values of the test statistic and the p-value for this test.

e) Test $H_0: \beta_2 = \beta_3 = 0$ vs. $H_1:$ (either $\beta_2 \neq 0$, or $\beta_3 \neq 0$) at a 5% level of significance. On your printout, circle or highlight the values of the test statistic and the p-value for this test.

```
fit <- lm(y ~ x1 + x2 + x3)

summary(fit)

Call:
lm(formula = y ~ x1 + x2 + x3)

Residuals:
   Min     1Q Median     3Q    Max 
  -200   -100    -50    125    200 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept)   400.00     459.84   0.870  0.40143   
x1             30.00     278.21   0.108  0.91591   
x2            280.00      65.95   4.245  0.00114 **
x3           -100.00      48.99  -2.041  0.06385 . 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 173.2 on 12 degrees of freedom
Multiple R-squared: 0.6674,     Adjusted R-squared: 0.5842 
F-statistic: 8.025 on 3 and 12 DF,  p-value: 0.003358


new <- data.frame( x1 = 1.8, x2 = 2, x3 = 2.3 )
predict.lm(fit, new, interval=c("prediction"), level=0.95)
   fit       lwr      upr
1  784  366.0256 1201.974


anova( lm(y ~ x1), fit )
Analysis of Variance Table

Model 1: y ~ x1
Model 2: y ~ x1 + x2 + x3
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)   
1     14  908200                                 
2     12  360000  2    548200 9.1367  0.003879 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. Suppose the distance between the apartment and the Student Union Building (X) (in miles) and the monthly rent amount for a 4-bedroom apartment (Y) (in \$) in Anytown have a bivariate normal distribution with

$$\mu_X = 2, \qquad \sigma_X = 0.6, \qquad \mu_Y = 1{,}210, \qquad \sigma_Y = 200, \qquad \rho = -0.28.$$

What is the probability that the monthly rent amount is under \$1,250 for a 4-bedroom apartment located 2.6 miles away from with the Student Union Building?

Need $P(\,Y < 1{,}250 \mid X = 2.6\,)$.

Given $X = 2.6$, Y has Normal distribution

with mean $1{,}210 + (-0.28) \cdot \dfrac{200}{0.6} \cdot (\,2.6 - 2\,) = 1{,}154$

and variance $\left(1 - (-0.28)^2\right) \cdot 200^2 = 36{,}864 \qquad$ (standard deviation 192).

$$P(\,Y < 1{,}250 \mid X = 2.6\,) \;=\; P\left(Z < \frac{1{,}250 - 1{,}154}{192}\right) \;=\; P(\,Z < 0.50\,) \;=\; \mathbf{0.6915}.$$

2. An Anytown State University student wishes to examine the relationship between the monthly rent amount for an apartment (y) (in $) and the apartment size (x) (in thousands of sq. ft.). The data are as follows:

Assume that (X, Y) have a bivariate normal distribution.

x	y
1.9	567
1.6	728
1.8	754
1.6	808
1.7	841
1.6	978
1.9	1,127
1.5	1,145
2.0	1,190
1.9	1,197
1.8	1,264
1.8	1,284
1.7	1,301
2.0	1,310
2.2	1,416
1.8	1,434

$$\sum x = 28.8, \qquad \sum y = 17{,}344,$$

$$\sum x^2 = 52.34, \qquad \sum y^2 = 19{,}883{,}146,$$

$$\sum x\,y = 31{,}514.2,$$

$$\sum (x - \bar{x})^2 = 0.5, \qquad \sum (y - \bar{y})^2 = 1{,}082{,}250,$$

$$\sum (x - \bar{x})(y - \bar{y}) = \sum (x - \bar{x})\,y = 295.$$

a) Find the sample correlation coefficient r.

$$r = \frac{295}{\sqrt{0.5}\,\sqrt{1{,}082{,}250}} \approx \mathbf{0.4010}.$$

```
cor(x1,y)
[1] 0.4010266
```

b) Find the p-value of the test $\mathrm{H}_0 : \rho = 0$ vs. $\mathrm{H}_1 : \rho \neq 0$.

Use $\mathrm{T} = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ and either EXCEL TDIST or R `pt`.

Test Statistic: $$\mathrm{T} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.4010 \cdot \sqrt{16-2}}{\sqrt{1-0.4010^2}} \approx \mathbf{1.638}.$$

```
2*(1-pt(1.638,16-2))
[1] 0.123693
```

=TDIST(1.638,16-2,2) **0.123693**

c) Find the (approximate) p-value of the test $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$.

Use $W = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right)$.

$$W = \frac{1}{2} \ln \frac{1+r}{1-r} = \frac{1}{2} \cdot \ln\left(\frac{1+0.4010}{1-0.4010}\right) \approx 0.42484.$$

Under H_0, $\mu_W = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} = \frac{1}{2} \cdot \ln\left(\frac{1+0}{1-0}\right) = 0,$

$$\sigma_W^2 = \frac{1}{n-3} = \frac{1}{13}.$$

Test Statistic: $Z = \frac{W - \mu_W}{\sigma_W} = \frac{0.42484 - 0}{\sqrt{1/13}} \approx \mathbf{1.53}.$

P-value = two tails = $2 \times P(Z > 1.53) = 2 \ 0.0630 = \mathbf{0.1260}$.

d) Find the (approximate) p-value of the test $H_0 : \rho = 0.20$ vs. $H_1 : \rho > 0.20$.

Under H_0, $\mu_W = \frac{1}{2} \ln \frac{1+\rho}{1-\rho} = \frac{1}{2} \cdot \ln\left(\frac{1+0.20}{1-0.20}\right) \approx 0.20273,$

$$\sigma_W^2 = \frac{1}{n-3} = \frac{1}{13}.$$

Test Statistic: $Z = \frac{W - \mu_W}{\sigma_W} = \frac{0.42484 - 0.20273}{\sqrt{1/13}} \approx \mathbf{0.80}.$

P-value = right tail = $P(Z > 0.80) = \mathbf{0.2119}$.

e) Construct a 90% confidence interval for ρ.

$100\,(1-\alpha)\,\%$ confidence interval for ρ:

$$\left(\frac{e^{a}-1}{e^{a}+1}\ ,\ \frac{e^{b}-1}{e^{b}+1} \right), \quad \text{where} \quad a = \ln\frac{1+r}{1-r} - \frac{2\,z_{\alpha/2}}{\sqrt{n-3}}, \quad b = \ln\frac{1+r}{1-r} + \frac{2\,z_{\alpha/2}}{\sqrt{n-3}}.$$

$$a = \ln\left(\frac{1+0.4010}{1-0.4010}\right) - \frac{2\times1.645}{\sqrt{16-3}} \approx -0.06274.$$

$$b = \ln\left(\frac{1+0.4010}{1-0.4010}\right) + \frac{2\times1.645}{\sqrt{16-3}} \approx 1.76223.$$

$$\left(\frac{e^{a}-1}{e^{a}+1}\ ,\ \frac{e^{b}-1}{e^{b}+1} \right) \approx \mathbf{(-0.03136,\ 0.70698)}.$$