12.2.17 No-fines concrete, made from a uniformly graded coarse aggregate and a cement-water paste, is beneficial in areas prone to excessive rainfall because of its excellent drainage properties. The article **"Pavement Thickness Design for No-Fines Concrete Parking Lots,"** ***J. of Trans. Engr.*, 1995: 476–484)** employed a least squares analysis in studying how y = porosity (%) is related to x = unit weight (pcf) in concrete specimens. Consider the following representative data:

x	99.0	101.1	102.7	103.0	105.4	107.0	108.7	110.8
y	28.8	27.9	27.0	25.2	22.8	21.5	20.9	19.6

x	112.1	112.4	113.6	113.8	115.1	115.4	120.0
y	17.1	18.9	16.0	16.7	13.0	13.6	10.8

Relevant summary quantities are $\sum x_i = 1640.1, \sum y_i = 299.8, \sum x_i^2 = 179{,}849.73, \sum x_i y_i = 32{,}308.59, \sum y_i^2 = 6430.06$.

(a) Obtain the equation of the estimated regression line. Then create a scatterplot of the data and graph the estimated line. Does it appear that the model relationship will explain a great deal of the observed variation in y?

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x, n = 15$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - (\sum x_i)^2/n} = \frac{32{,}308.59 - (1640.1)(299.8)/15}{179{,}849.73 - (1640.1)^2/15} = -0.905$$

$$\hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \frac{299.8 - (-0.905)(1640.1)}{15} = 118.91$$

Ans: $\boldsymbol{y = 118.91 - 0.905x}$

(b) Interpret the slope of the least square line.

The slope of the regression line is –0.905, which means that the porosity decreases by 0.905% per 1-pcf increase.

Ans: The porosity decreases by 0.905% per 1-pcf increase

(c) What happens if the estimated line is used to predict porosity when unit weight is 135? Why is this not a good idea?

$\hat{y} = 118.91 - 0.905(135) = -3.265$ (negative prediction)

It is not a good idea to use this prediction because the porosity cannot be smaller than zero.

Ans: –3.265, the porosity cannot be negative

(d) Calculate the residuals corresponding to the first two observations.

$\hat{y}_1 = 118.91 - 0.905(99) = 29.315, y_1 - \hat{y}_1 = 28.8 - 29.315 = -0.515$

$\hat{y}_2 = 118.91 - 0.905(101.1) = 27.4145, y_2 - \hat{y}_2 = 27.9 - 27.4145 = 0.4855$

Ans: –0.515, 0.4855

(e) Calculate and interpret a point estimate of σ.

$\hat{y}_3 = 118.91 - 0.905(102.7) = 25.9665, y_3 - \hat{y}_3 = 27 - 25.9665 = 1.0335$
$\hat{y}_4 = 118.91 - 0.905(103) = 25.695, y_4 - \hat{y}_4 = 25.2 - 25.695 = -0.495$
$\hat{y}_5 = 118.91 - 0.905(105.4) = 23.523, y_5 - \hat{y}_5 = 22.8 - 23.523 = -0.723$
$\hat{y}_6 = 118.91 - 0.905(107) = 22.075, y_6 - \hat{y}_6 = 21.5 - 22.075 = -0.575$
$\hat{y}_7 = 118.91 - 0.905(108.7) = 20.5365, y_7 - \hat{y}_7 = 20.9 - 20.5365 = 0.3635$
$\hat{y}_8 = 118.91 - 0.905(110.8) = 18.636, y_8 - \hat{y}_8 = 19.6 - 18.636 = 0.964$
$\hat{y}_9 = 118.91 - 0.905(112.1) = 17.4595, y_9 - \hat{y}_9 = 17.1 - 17.4595 = -0.3595$
$\hat{y}_{10} = 118.91 - 0.905(112.4) = 17.188, y_{10} - \hat{y}_{10} = 18.9 - 17.188 = 1.712$
$\hat{y}_{11} = 118.91 - 0.905(113.6) = 16.102, y_{11} - \hat{y}_{11} = 16 - 16.102 = -0.102$
$\hat{y}_{12} = 118.91 - 0.905(113.8) = 15.921, y_{12} - \hat{y}_{12} = 16.7 - 15.921 = 0.779$
$\hat{y}_{13} = 118.91 - 0.905(115.1) = 14.7445, y_{13} - \hat{y}_{13} = 13 - 14.7445 = -1.7445$
$\hat{y}_{14} = 118.91 - 0.905(115.4) = 14.473, y_{14} - \hat{y}_{14} = 13.6 - 14.473 = -0.873$
$\hat{y}_{15} = 118.91 - 0.905(120) = 10.31, y_{15} - \hat{y}_{15} = 10.8 - 10.31 = 0.49$

$$\begin{aligned} SSE = \sum (y_i - \hat{y}_i)^2 &= (-0.515)^2 + (0.4855)^2 + (1.0335)^2 + (-0.495)^2 + (-0.723)^2 \\ &+ (-0.575)^2 + (0.3635)^2 + (0.964)^2 + (-0.3595)^2 + (1.712)^2 + (-0.102)^2 \\ &+ (0.779)^2 + (-1.7445)^2 + (-0.873)^2 + (0.49)^2 = 11.452 \end{aligned}$$

$\hat{\sigma} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{11.452}{15-2}} = 0.939$. The average error made on predictions is 0.9380%

Ans: $\hat{\boldsymbol{\sigma}} = \mathbf{0.939}$

(f) What proportion of observed variation in porosity can be attributed to the approximate linear relationship between unit weight and porosity?

$$SST = S_{yy} = \sum y_i^2 - \left(\sum y_i\right)^2 / n = 6430.06 - 299.8^2/15 = 438.057$$

$$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{11.452}{438.057} = 0.974$$

Conclusion: 97.40% of the observed variation in porosity can be attributed to the approximate linear relationship between unit weight and porosity.

Ans: 0.974

12.2.20 The bond behavior of reinforcing bars is an important determinant of strength and stability. The article **"Experimental Study on the Bond Behavior of Reinforcing Bars Embedded in Concrete Subjected to Lateral Pressure" (*J. of Materials in Civil Engr.*, 2012: 125–133)** reported the results of one experiment in which varying levels of lateral pressure were applied to 21 concrete cube specimens, each with an embedded 16-mm plain steel round bar, and the corresponding bond capacity was determined. Due to differing concrete tube strengths (f_{cu}, in MPa), the applied lateral pressure was equivalent to a fixed proportion of the specimen's f_{cu} (0, .1f_{cu}, …, 6f_{cu}). Also, since bond strength can be heavily influenced by the specimen's f_{cu}, bond capacity was expressed as the ratio of bond strength (MPa) to $\sqrt{f_{cu}}$.

Pressure	0	0	0	.1	.1	.1	.2
Ratio	0.123	0.100	0.101	0.172	0.133	0.107	0.217
Pressure	.2	.2	.3	.3	.3	.4	.4
Ratio	0.172	0.151	0.263	0.227	0.252	0.310	0.365
Pressure	.4	.5	.5	.5	.6	.6	.6
Ratio	0.239	0.365	0.319	0.312	0.394	0.386	0.320

(a) Does a scatterplot of the data support the use of the simple linear regression model?
Ans: Ignored per instruction
(b) Use the accompanying Minitab output to give point estimates of the slope and intercept of the population regression line.

```
The regression equation is
Ratio = 0.101 + 0.461 Pressure

Predictor      Coef   SE Coef      T      P
Constant    0.10121   0.01308   7.74  0.000
Pressure    0.46071   0.03627  12.70  0.000

S=0.0332397  R-Sq=89.5%  R-Sq(adj)=88.9%

Analysis of Variance
Source          DF       SS       MS      F      P
Regression       1  0.17830  0.17830 161.37  0.000
Residual Error  19  0.02099  0.00110
Total           20  0.19929
```

Ans: $\textbf{slope} = \hat{\boldsymbol{\beta}}_1 = \mathbf{0.46071}, \boldsymbol{y}\textbf{– intercept} = \hat{\boldsymbol{\beta}}_0 = \mathbf{0.10121}$

$$\hat{\boldsymbol{y}} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1 \boldsymbol{x} = \mathbf{0.10121} + \mathbf{0.46071}\boldsymbol{x}$$

(c) Calculate a point estimate of the true average bond capacity when lateral pressure is .45f_{cu}.
$\hat{y} = 0.10121 + 0.46071(0.45) = 0.309$
Bond strength = Ratio × $\sqrt{f_{cu}} = 0.309\sqrt{0.45} = 0.207$
The point estimate of the average bond strength is 0.2069 MPa when lateral pressure is 0.45 f_{cu}.

Ans: 0.207

(d) What is a point estimate of the error standard deviation σ, and how would you interpret it?

$$SSE = \sum (y_i - \hat{y}_i)^2 = 0.02099 \text{ (from Minitab output)}, \hat{\sigma} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{0.02099}{21-2}} = 0.0332$$

The average error made on predictions of the ratio of the bond strength to $\sqrt{f_{cu}}$ is 0.0332 MPa to $\sqrt{f_{cu}}$.

Ans: $\hat{\boldsymbol{\sigma}} = \mathbf{0.0332}$

(e) What is the value of total variation, and what proportion of it can be explained by the model relationship?

$$SST = S_{yy} = \sum y_i^2 - \left(\sum y_i\right)^2/n = 0.19929 \text{ (from Minitab output)}$$

$$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{0.02099}{0.19929} = 0.895 \text{ (matches the Minitab output)}$$

Conclusion: 89.50% of the total variation can be explained by the model relationship.

Ans: $\boldsymbol{r^2 = 0.895}$

12.2.29 Consider the following three data sets, in which the variables of interest are x = commuting distance and y = commuting time. Based on a scatterplot and the values of s and r^2, in which situation would simple linear regression be most (least) effective, and why?

Data Set	1		2		3	
	x	y	x	y	x	y
	15	42	5	16	5	8
	16	35	10	32	10	16
	17	45	15	44	15	22
	18	42	20	45	20	23
	19	49	25	63	25	31
	20	46	50	115	50	60
S_{xx}	17.50		1270.8333		1270.8333	
S_{xy}	29.50		2722.5		1431.6667	
$\hat{\beta}_1$	1.685714		2.142295		1.126557	
$\hat{\beta}_0$	13.666672		7.868852		3.196729	
SST	114.83		5897.5		1627.33	
SSE	65.10		65.10		14.48	

$$\hat{\sigma}_1 = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{65.10}{6-2}} = 4.03, \hat{\sigma}_2 = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{65.10}{6-2}} = 4.03, \hat{\sigma}_3 = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{14.48}{6-2}} = 1.90$$

$$r_1^2 = 1 - \frac{65.10}{114.83} = 0.433, r_2^2 = 1 - \frac{65.10}{5897.5} = 0.989, r_3^2 = 1 - \frac{14.48}{1627.33} = 0.991$$

The larger r^2 is, the better; and the smaller s is, the better. Data set 3 has the largest r^2 and the smallest s; whereas Data set 1 has the smallest r^2 and the largest s.

Ans: most effective: set 3, least effective: set 1

12.3.35 How does lateral acceleration—side forces experienced in turns that are largely under driver control—affect nausea as perceived by bus passengers? The article **"Motion Sickness in Public Road Transport: The Effect of Driver, Route, and Vehicle" (*Ergonomics,* 1999: 1646–1664)** reported data on x = motion sickness does (calculated in accordance with a British standard for evaluating similar motion at sea) and y = reported nausea (%). Relevant summary quantities are $n = 17, \sum x_i = 222.1, \sum y_i = 193, \sum x_i^2 = 3056.69, \sum x_i\, y_i = 2759.6, \sum y_i^2 = 2975$. Values of dose in in the sample ranged from 6.0 to 17.6.

(a) Assuming that the simple linear regression model is valid for relating these two variables (this is supported by the raw data), calculate and interpret an estimate of the slope parameter that conveys information about the precision and reliability of estimation.

$S_{xy} = 2759.6 - (222.1)(193)/17 = 238.112, S_{xx} = 3056.69 - (222.1)^2/17 = 155.019$

$S_{yy} = 2975 - (193)^2/17 = 784.482, \hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}} = \dfrac{238.112}{155.019} = 1.536$

$\hat{\beta}_0 = \dfrac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \dfrac{193 - (1.536)(222.1)}{17} = -8.714$

$MSE = \dfrac{SSE}{n-2} = \dfrac{S_{yy} - \hat{\beta} S_{xy}}{n-2} = \dfrac{784.482 - 1.536(238.112)}{17-2} = 27.916$

$s = \sqrt{MSE} = \sqrt{27.916} = 5.284, s_{\hat{\beta}_1} = \dfrac{s}{\sqrt{S_{xx}}} = \dfrac{5.284}{\sqrt{155.019}} = 0.424,$

$t_{0.05/2,17-2} = t_{0.025,15} = 2.131$ (from Table A. 5)

95% CI: $1.536 \pm 2.131(0.424) = 1.536 \pm 0.904 = (0.632, 2.44)$

Conclusion: We are 95% confident that the slope of the population regression line lies between 0.632 and 2.44.

Ans: (0.632, 2.44) is a 95% CI

(b) Does it appear that there is a useful linear relationship between these two variables? Test appropriate hypotheses using $\alpha = .01$.

$H_0: \beta_1 = 0; H_a: \beta_1 \neq 0, t = \dfrac{1.536}{0.424} = 3.623$

Table A.5 shows that $t_{0.005,15} = 2.947, t_{0.001,15} = 3.733$, so 0.001 < p-value < 0.005. Since the p-value is less than the significance level $\alpha = .01$, we reject the null hypothesis.

Conclusion: There is sufficient evidence to support the claim that there is a useful linear relationship between two variables.

Ans: Yes, *t* = 3.623, *P*-value < 0.005

(c) Would it be sensible to use the simple linear regression model as a basis for prediction % nausea when dose = 5.0? Explain your reasoning.

It is not sensible to use the simple linear regression model as a basis for predicting the % of nausea when dose = 5.0 because 5.0 does lie in the given range of doses from 6.0 to 17.6 (extrapolation).

Ans: No

12.4.47 The simple linear regression model provides a very good fit to the data on rainfall and runoff volume given in Exercise 16 of Section 12.2. The equation of the least squares line is $y = -1.128 + .82697x, r^2 = .975$, and $s = 5.24$.

(a) Use the fact that $s_{\hat{Y}} = 1.44$ when rainfall volume is 40 m^3 to predict runoff in a way that conveys information about reliability and precision. Does the resulting interval suggest that precise information about the value of runoff for this future observation is available? Explain your reasoning.

$t_{\alpha/2,n-2} = t_{0.025,13} = 2.160$ (from Table A. 5), $\hat{y}_{40} = -1.128 + 0.82697 \cdot 40 = 31.9508$

PI: $\hat{y}_{40} \pm t_{0.025,13}\sqrt{s^2 + s_{\hat{Y}}^2} = 31.9508 \pm 2.16\sqrt{5.24^2 + 1.44^2} = (20.213, 43.689)$

Conclusion: The interval is wide comparing to the value, and the information is not precise.

Ans: 95% PI is (20.231, 43.689), no

(b) Calculate a PI for runoff when rainfall is 50 using the same prediction level as in part (a). What can be said about the simultaneous prediction level for the two intervals you have calculated?

$\hat{y}_{50} = -1.128 + 0.82697 \cdot 50 = 40.2205$

$$\sum x_i = 5 + 12 + 14 + 17 + 23 + 30 + 40 + 47 + 55 + 67 + 72 + 81 + 96 + 112 + 127$$
$$= 798, \bar{x} = 798/15 = 53.2$$

$$\sum x_i^2 = 5^2 + 12^2 + 14^2 + 17^2 + 23^2 + 30^2 + 40^2 + 47^2 + 55^2 + 67^2 + 72^2 + 81^2 + 96^2$$
$$+ 112^2 + 127^2 = 63{,}040$$

$$S_{xx} = \sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2 = 63{,}040 + \frac{798^2}{15} = 20{,}586.4$$

$$s_{\hat{Y}_{50}} = s\sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} = 5.24\sqrt{\frac{1}{15} + \frac{(50 - 53.2)^2}{20586.4}} = 1.358$$

PI: $\hat{y}_{50} \pm t_{0.025,13}\sqrt{s^2 + s_{\hat{Y}}^2} = 40.2205 \pm 2.16\sqrt{5.24^2 + 1.358^2} = (28.528, 51.913)$

For the simultaneous prediction level, $100(1 - 2 \cdot 0.05)\% = 90\%$.

Conclusion: The prediction level is at least 90%

Ans: (28.528, 51.913), at least 90%

12.4.48 The catch basin in a storm-sewer system is the interface between surface runoff and the sewer. The catch-basin insert is a device for retrofitting catch basins to improve pollutant-removal properties. The article **"An Evaluation of the Urban Stormwater Pollutant Removal Efficiency of Catch Basin Inserts" (*Water Envir. Res.*, 2005: 500–510)** reported on tests of various inserts under controlled conditions for which inflow is close to what can be expected in the field. Consider the following data, read from a graph in the article, for one particular type of insert on x = amount filtered (1000s of liters) and y = % total suspended solids removed. Summary quantities are

$$\sum x_i = 1251, \sum x_i^2 = 199{,}365, \sum y_i = 250.6, \sum y_i^2 = 9249.36, \sum x_i\, y_i = 21{,}904.4$$

x	23	45	68	91	114	136	159	182	205	228
y	53.3	26.9	54.8	33.8	29.9	8.2	17.2	12.2	3.2	11.1

(b) Obtain the equation of the least squares line.

$$S_{xy} = \sum x_i y_i - \sum x_i \sum y_i/n = 21{,}904.4 - (1251)(250.6)/10 = -9445.66$$

$$S_{xx} = \sum x_i^2 - \left(\sum x_i\right)^2/n = 199{,}365 - (1251)^2/10 = 42864.9$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{-9445.66}{42864.9} = -0.220359$$

$$\hat{\beta}_0 = \frac{\sum y_i - \hat{\beta}_1 \sum x_i}{n} = \frac{250.6 - (-0.220359)(1251)}{10} = 52.6269, \hat{y} = 52.6269 - 0.220359x$$

Ans: $\boldsymbol{\hat{y} = 52.6269 - 0.220359x}$

(c) What proportion of observed variation in % removed can be attributed to the model relationship?

$$SSE = \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i\, y_i = 9249.36 - 52.6269(250.6) + 0.220359(21{,}904.4)$$
$$= 887.8905$$

$$SST = \sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2 = 9249.36 - \frac{250.6^2}{10} = 2969.324$$

$$r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{887.8905}{2969.324} = 0.7$$

Conclusion: 70% of observed variation can be attributed to the approximated linear relationship.

Ans: $\boldsymbol{r^2 = 0.7}$

(d) Does the simple linear regression model specify a useful relationship? Carry out an appropriate test of hypotheses using a significance level of .05.

$H_0: \beta_1 = 0; H_a: \beta_1 \neq 0,$

$$\hat{\sigma} = s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{887.8905}{10-2}} = 10.535, s_{\hat{\beta}_1} = \frac{s}{\sqrt{S_{xx}}} = \frac{10.535}{\sqrt{42864.9}} = 0.05088$$

$$t = \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} = \frac{-0.220359}{0.05088} = -4.33, t_{0.025,8} = 2.306 < |t|$$

Conclusion: Reject hypothesis.

Ans: The regression is significant

(e) Is there strong evidence for concluding that there is at least a 2% decrease in true average suspended solid removal associated with a 10,000-liter increase in the amount filtered? Test appropriate hypotheses using $\alpha = .05$.

$H_0: \beta_{10} = -0.2; H_a: \beta_{10} < -0.2,$

$$t = \frac{\hat{\beta}_1 - \beta_{10}}{s_{\hat{\beta}_1}} = \frac{-0.220359 - (-0.2)}{0.05088} = -0.4, t_{0.025,8} = 2.306 > |t|$$

Conclusion: Do not reject null hypothesis. The regression slope is not less than –0.2

Ans: Do not reject null hypothesis

12.4.49 You are told that a 95% CI for expected lead content when traffic flow is 15, based on a sample of *n* = 10 observations, is (462.1, 597.7). Calculate a CI with confidence level 99% for expected lead content when traffic flow is 15.

$$\hat{y} = \frac{462.1 + 597.7}{2} = 529.9, t_{0.025,8} = 2.306, t_{0.005,8} = 3.355 \text{ (from Table A. 5)}$$

$$s_{\hat{Y}} = \frac{597.7 - 529.9}{2.306} = 29.402, 99\% \text{ CI}: 529.9 \pm 3.355 \cdot 29.402 = (431.3, 628.5)$$

Ans: (431.3, 628.5)