13.1.12 (a) Could a linear regression result in residuals 23, –27, 5, 17, –8, 9, and 15? Why or why not?

The sum of all residuals should be equal to zero. Because 23 –27 + 5 + 17 – 8 + 9 + 15 = 34 $\neq$ 0, a linear regression cannot result in such residuals.

(b) Could a linear regression result in residuals 23, –27, 5, 17, –8, –12, and 2 corresponding to *x* values 3, –4, 8, 12, –14, –20, and 25? Why or why not? [*Hint*: See Exercise 10.]

For every *i* = 1, 2, …, 7, the x_i and e_i have the same time, resulting a sum of all residuals greater than zero. Thus, a linear regression cannot result in such *x*-values and corresponding residuals.

Ans: (a) not possible; (b) not possible

13.2.22 In each of the following cases, decide whether the given function is intrinsically linear. If so, identify x' and y', and then explain how a random error term ϵ can be introduced to yield an intrinsically linear probabilistic model.

(a) $y = 1/(\alpha + \beta x)$

$$y = 1/(\alpha + \beta x) \rightarrow \frac{1}{y} = \alpha + \beta x, \therefore \boldsymbol{x' = x; y' = \frac{1}{y}; \frac{1}{y} = \alpha + \beta x + \epsilon}$$

(b) $y = 1/\left(1 + e^{\alpha + \beta x}\right)$

$$y = 1/\left(1 + e^{\alpha + \beta x}\right) \rightarrow \frac{1}{y} = 1 + 1 + e^{\alpha + \beta x} \rightarrow \frac{1}{y} - 1 = e^{\alpha + \beta x} \rightarrow \ln\left(\frac{1}{y} - 1\right) = \alpha + \beta x$$

$$\therefore \boldsymbol{x' = x;\ y' = \ln\left(\frac{1}{y} - 1\right); \ln\left(\frac{1}{y} - 1\right) = \alpha + \beta x + \epsilon}$$

(c) $y = e^{e^{\alpha + \beta x}}$ (a Gompertz curve)

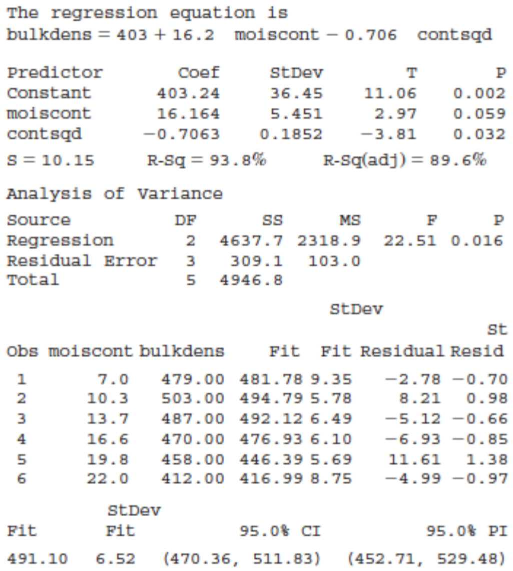
$$y = e^{e^{\alpha + \beta x}} \rightarrow \ln y = e^{\alpha + \beta x} \rightarrow \ln(\ln y) = \alpha + \beta x$$

$$\therefore \boldsymbol{x' = x; y' = \ln(\ln y); \ln(\ln y) = \alpha + \beta x + \epsilon}$$

(d) $y = \alpha + \beta e^{\lambda x}$

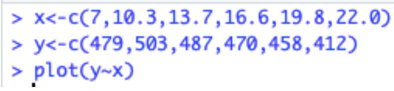
Ans: The given function is not intrinsically linear

13.3.26 The article **"Physical Properties of Cumin Seed"** (***J. of Agric. Engr. Res.***, **1996: 93–98)** considered a quadratic regression of y = bulk density on x = moisture content. Data from a graph in the article follows, along with Minitab output from the quadratic fit.

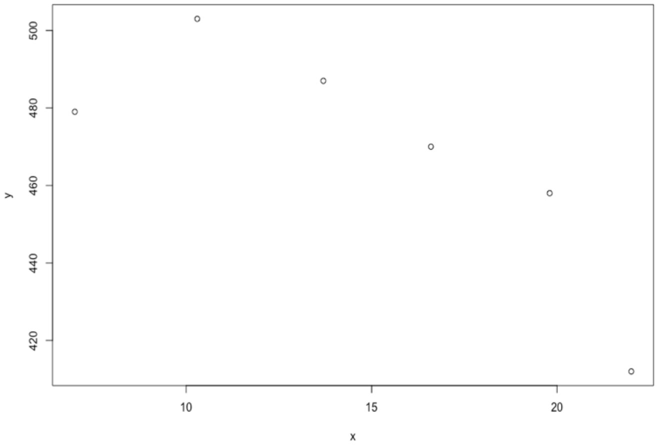


```
The regression equation is
bulkdens = 403 + 16.2  moiscont − 0.706  contsqd

Predictor        Coef      StDev         T       P
Constant       403.24      36.45     11.06   0.002
moiscont       16.164      5.451      2.97   0.059
contsqd       −0.7063     0.1852     −3.81   0.032
S = 10.15       R-Sq = 93.8%       R-Sq(adj) = 89.6%

Analysis of Variance
Source          DF      SS      MS       F      P
Regression       2  4637.7  2318.9   22.51  0.016
Residual Error   3   309.1   103.0
Total            5  4946.8
                                  StDev
                                                St
Obs moiscont bulkdens     Fit   Fit Residual Resid
 1       7.0   479.00  481.78  9.35    −2.78 −0.70
 2      10.3   503.00  494.79  5.78     8.21  0.98
 3      13.7   487.00  492.12  6.49    −5.12 −0.66
 4      16.6   470.00  476.93  6.10    −6.93 −0.85
 5      19.8   458.00  446.39  5.69    11.61  1.38
 6      22.0   412.00  416.99  8.75    −4.99 −0.97

          StDev
Fit       Fit          95.0% CI            95.0% PI
491.10    6.52   (470.36, 511.83)   (452.71, 529.48)
```

(a) Does a scatterplot of the data appear consistent with the quadratic regression model?



```
> x<-c(7,10.3,13.7,16.6,19.8,22.0)
> y<-c(479,503,487,470,458,412)
> plot(y~x)
```



The scatterplot suggest that the data is consistent with the quadratic regression model.

Ans: Yes

(b) What proportion of observed variation in density can be attributed to the model relationship?

$r^2 = 0.938$, so 93.8% of observed variation in density can be attributed to the model relationship.

Ans: 93.8%

(c) Calculate a 95% CI for true average density when moisture content is 13.7.

$x^* = 13.7, \hat{y} = 492.12, s_{\hat{y}} = 6.49, t_{\alpha/2,n-(k+1)} = t_{0.025,(6-(2+1))} = t_{0.025,3} = 3.182$

95% CI $= \hat{y} \pm t_{\alpha/2,n-(k+1)} \cdot s_{\hat{y}} = 492.12 \pm 3.182 \cdot 6.49 = (\mathbf{471.466}, \mathbf{512.774})$

(d) The last line of output is from a request for estimation and prediction information when moisture content is 14. Calculate a 99% PI for density when moisture content is 14.

$$x^* = 14, \hat{y} = 491.10, \sqrt{s^2 + s_{\hat{y}}^2} = \frac{529.48 - 491.10}{3.182} = 12.06, t_{\alpha/2,n-(k+1)} = t_{0.005,3} = 5.841$$

$$99\% \text{ PI} = \hat{y} \pm t_{\alpha/2,n-(k+1)} \cdot \sqrt{s^2 + s_{\hat{y}}^2} = 491.10 \pm 5.841 \cdot 12.06 = (\mathbf{420.65}, \mathbf{561.55})$$

(e) Does the quadratic predictor appear to provide useful information? Test the appropriate hypotheses at significance level .05.

$H_0: \beta_2 = 0, H_a: \beta_2 \neq 0, t = -3.81, p-$ value $= 0.032 < 0.05 = \alpha$ (reject H_0)

Ans: The quadratic predictor provides useful information

13.3.28 The viscosity (y) of an oil was measured by a cone and plate viscometer at six different cone speeds (x). It was assumed that a quadratic regression model was appropriate, and the estimated regression function resulting from the $n = 6$ observations was

$$y = -113.0937 + 3.3684x - .01780x^2$$

(a) Estimate $\mu_{Y \cdot 75}$, the expected viscosity when speed is 75 rpm.

$\mu_{Y \cdot 75} = -113.0937 + 3.3684 \cdot 75 - .01780(75)^2 = \mathbf{39.4113}$

(b) What viscosity would you predict for a cone speed of 60 rpm?

$\mu_{Y \cdot 60} = -113.0937 + 3.3684 \cdot 60 - .01780(60)^2 = \mathbf{24.9303}$

(c) If $\sum y_i^2 = 8386.43, \sum y_i = 210.70, \sum x_i y_i = 17{,}002.00$, and $\sum x_i^2 y_i = 1{,}419{,}780$, compute SSE$[= \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i - \hat{\beta}_2 \sum x_i^2 y_i]$ and s.

$8386.43 - (-113.0937) \cdot 210.70 - 3.3684 \cdot 17{,}002 - (-0.0178) \cdot 1{,}419{,}780 = \mathbf{217.82}$

$$s = \sqrt{\frac{\text{SSE}}{n - (k+1)}} = \sqrt{\frac{217.82}{6 - (2+1)}} = \mathbf{8.52}$$

(d) From part (c), SST $= 8386.43 - (210.70)^2/6 = 987.35$. Using SSE computed in part (c), what is the computed value of R^2?

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}} = \frac{217.82}{987.35} = \mathbf{0.7794}$$

(e) If the estimated standard deviation of $\hat{\beta}_2$ is $s_{\hat{\beta}_2} = .00226$, test $H_0: \beta_2 = 0$ versus $H_0: \beta_2 \neq 0$ at level .01, and interpret the result.

$$t = \frac{-0.01780 - 0}{0.00226} = -7.88, |t| = 7.88 > 5.841, \therefore p\text{-value} < 0.01 \rightarrow \text{reject } H_0$$

Ans: The quadratic predictor provides useful information

13.4.39 Let y = sales at a fast-food outlet (1000s of %), x_1 = number of competing outlets within a 1-mile radius, x_2 = population within a 1-mile radius (1000s of people), and x_3 be an indicator variable that equals 1 if the outlet has a drive-up window and 0 otherwise. Suppose that the true regression model is

$$Y = 10.00 - 1.2x_1 + 6.8x_2 + 15.3x_3 + \epsilon$$

(a) What is the mean value of sales when the number of competing outlets is 2, there are 8000 people within a 1-mile radius, and the outlet has a drive-up window?

$x_1 = 2, x_2 = 8, x_3 = 1, \hat{y}_{2,8,1} = 10.00 - 1.2(2) + 6.8(8) + 15.3(1) = 77.3$

Ans: $77,300

(b) What is the mean value of sales for an outlet without a drive-up window that has three competing outlets and 5000 people within a 1-mile radius?

$x_1 = 3, x_2 = 5, x_3 = 0, \hat{y}_{3,5,0} = 10.00 - 1.2(3) + 6.8(5) + 15.3(0) = 40.4$

Ans: $40,400

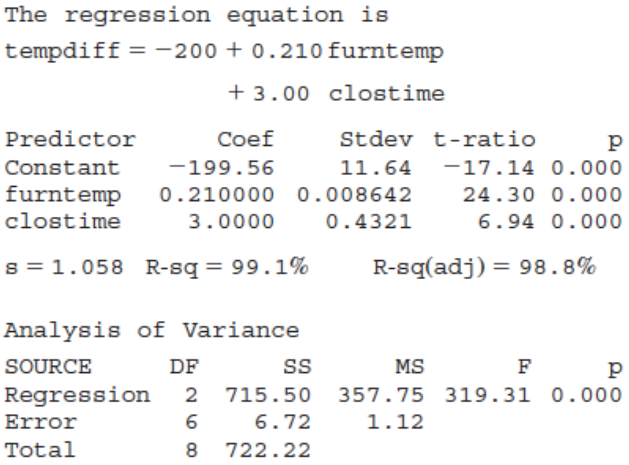
(c) Interpret β_3.

Ans: When both the number of competing outlets within a mile and the population within a 1-mile radius (1000s of people) are held constant, the average change in sales *y* will increase by $15,300 when an outlet has a drive-up window.

13.4.42 An investigation of a die-casting process resulted in the accompanying data on x_1 = furnace temperature, x_2 = die close time, and y = temperature difference on the die surface **("A Multiple-Objective Decision-Making Approach for Assessing Simultaneous Improvement in Die Life and Casting Quality in a Die Casting Process," *Quality Engineering*, 1994: 371–383)**.

x_1	1250	1300	1350	1250	1300	1250	1300	1350	1350
x_2	6	7	6	7	6	8	8	7	8
y	80	95	101	85	92	87	96	106	108

Minitab output from fitting the multiple regression model with predictors x_1 and x_2 is given here.



```
The regression equation is
tempdiff = −200 + 0.210 furntemp
              + 3.00 clostime

Predictor      Coef     Stdev  t-ratio     p
Constant    −199.56     11.64   −17.14 0.000
furntemp   0.210000  0.008642    24.30 0.000
clostime     3.0000    0.4321     6.94 0.000

s = 1.058  R-sq = 99.1%     R-sq(adj) = 98.8%

Analysis of Variance
SOURCE      DF      SS      MS       F      p
Regression   2  715.50  357.75  319.31  0.000
Error        6    6.72    1.12
Total        8  722.22
```

(a) Carry out the model utility test.

$H_0: \beta_1 = \beta_2 = 0;\ H_a$: at least one $\beta_i \neq 0, F = 319.31, p-$ value $= 0.000$. Reject null hypothesis.

Ans: The linear model provides useful information

(b) Calculate and interpret a 95% confidence interval for β_2, the population regression coefficient of x_2.

$\hat{\beta}_2 = 3.000, s_{\hat{\beta}_2} = 0.4321, t_{\alpha/2,n-(k+1)} = t_{0.025,9-(2+1)} = t_{0.025,6} = 2.447,$

95% CI $= \hat{\beta}_2 \pm t_{\alpha/2,n-(k+1)} \cdot s_{\hat{\beta}_2} = 3.000 \pm 2.447 \cdot 0.4321 = (1.943, 4.057)$

Interpretation: When furnace temperature is constant, the average change in temperature difference on the die surface is between 1.943 and 4.057 degrees.

Ans: (1.943, 4.057)

(c) When $x_1 = 1300$ and $x_2 = 7$, the estimated standard deviation of $\hat{Y}$ is $s_{\hat{Y}} = .353$. Calculate a 95% confidence interval for true average temperature difference when furnace temperature is 1300 and die close time is 7.

$\hat{y}_{1300,7} = -200 + 0.210 \cdot 1300 + 3.00 \cdot 7 = 94.44$, 95% CI $= 94.44 \pm 2.447 \cdot 0.353$
$= (93.58, 95.30)$

Ans: (93.58, 95.30)

(d) Calculate a 95% prediction interval for the temperature difference resulting from a single experimental run with a furnace temperature of 1300 and a die close time of 7.

$s = 1.058, s_{\hat{y}} = 0.353$, 95% PI $= 94.44 \pm 2.447\sqrt{1.058^2 + 0.353^2} = (91.71, 97.17)$

Ans: (91.71, 97.17)

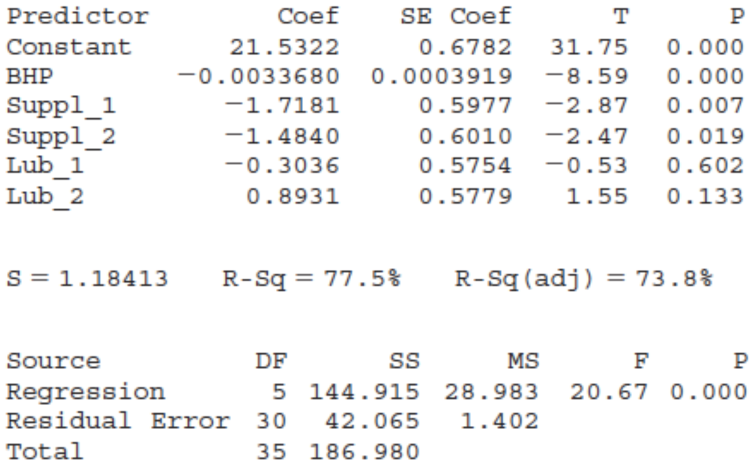
13.4.54 The use of high-strength steels (HSS) rather than aluminum and magnesium alloys in automotive body structures reduces vehicle weight. However, HSS use is still problematic because of difficulties with limited formability, increased springback, difficulties in joining, and reduced die life. The article **"Experimental Investigation of Springback Variation in Forming of High Strength Steels" (*J. of Manuf. Sci. and Engr.*, 2008: 1–9)** included data on y = springback from the wall opening angle and x_1 = blank holder pressure. Three different material suppliers and three different lubrication regimens (no lubrication, lubricant #1, and lubricant #2) were also utilized.

(a) What predictors would you use in a model to incorporate supplier and lubrication information in addition to BHP?

$$\text{Suppl}_1 = \begin{cases} 1 & \text{material from supplier 1} \\ 0 & \text{otherwise} \end{cases} \qquad \text{Suppl}_2 = \begin{cases} 1 & \text{material from supplier 2} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Lub}_1 = \begin{cases} 1 & \text{lubricant 1} \\ 0 & \text{otherwise} \end{cases} \qquad \text{Lub}_2 = \begin{cases} 1 & \text{lubricant 2} \\ 0 & \text{otherwise} \end{cases}$$

(b) The accompanying Minitab output resulted from fitting the model of (a) (the article's authors also used Minitab; amusingly, they employed a significance level of .06 in various tests of hypotheses). Does there appear to be a useful relationship between the response variable and at least one of the predictors? Carry out a formal test of hypotheses.



```
Predictor          Coef    SE Coef       T      P
Constant        21.5322     0.6782   31.75  0.000
BHP          −0.0033680  0.0003919   −8.59  0.000
Suppl_1         −1.7181     0.5977   −2.87  0.007
Suppl_2         −1.4840     0.6010   −2.47  0.019
Lub_1           −0.3036     0.5754   −0.53  0.602
Lub_2            0.8931     0.5779    1.55  0.133


S = 1.18413   R-Sq = 77.5%   R-Sq(adj) = 73.8%


Source          DF       SS      MS      F      P
Regression       5  144.915  28.983  20.67  0.000
Residual Error  30   42.065   1.402
Total           35  186.980
```

$H_0\!: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0;\ H_a\!:$ at least one $\beta_i \neq 0, F = 20.67, p-$ value $= 0.000$. Reject null hypothesis.

Ans: Useful relationship between the response variable and at least one of the predictors

(c) When BHP is 1000, material is from supplier 1, and no lubrication is used, $s_{\hat{Y}} = .524$. Calculate a 95% PI for the springback that would result from making an additional observation under these conditions.

$$\hat{y}_{1000,1,0,0,0} = 21.5322 - 0.0033680 \cdot 1000 - 1.1781 \cdot 1 - 1.4840 \cdot 0 - 0.3036 \cdot 0 + 0.3931 \cdot 0 = 16.4461$$

$t_{0.025,36-(5+1)} = t_{0.025,30} = 2.042, s = 1.18413, s_{\hat{Y}} = .524$

95% PI $= 16.4461 \pm 2.042\sqrt{1.18413^2 + 0.524^2} = (13.802, 19.090)$

Ans: (13.802, 19.090)

(d) From the output, it appears that lubrication regimen may not be providing useful information. A regression with the corresponding predictors removed resulted in SSE = 48.426. What is the coefficient of multiple determination for this model, and what would you conclude about the importance of the lubrication regimen?

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}} = 1 - \frac{48.426}{186.980} = 0.741$$

Based on just coefficient of multiple determination, one should not make conclusion about importance of new model since the coefficients of determination are very close for the two models (0.775 versus 0.741). Use test statistic instead.

$H_0 \colon \beta_l = \beta_{l+1} = \cdots = \beta_k = 0;\ H_a \colon \text{at least one } \beta_i \neq 0\ (i = l, l+1, \ldots, 9)$

$$f = \frac{(SSE_l - SSE_k)/(k-l)}{SSE_k/[n-(k+1)]} = \frac{(48.426 - 42.065)/(5-3)}{42.065/30} = 2.27, F_{2,30} = 3.32 > 2.27$$

Conclusion: Do not reject null hypothesis

Ans: lubrication regiment does not provide useful information

(e) A model with predictors for BHP, supplier, and lubrication regimen, as well as predictors for interactions between BHP and both supplier and lubrication regiment, resulted in SSE = 28.216 and R^2 = .849. Does this model appear to improve on the model with just BHP and predictors for supplier?

$H_0 \colon \beta_l = \beta_{l+1} = \cdots = \beta_k = 0;\ H_a \colon \text{at least one } \beta_i \neq 0\ (i = l, l+1, \ldots, 9)$

$$f = \frac{(SSE_l - SSE_k)/(k-l)}{SSE_k/[n-(k+1)]} = \frac{(48.426 - 28.216)/(9-3)}{28.216/(36-10)} = 3.10, F_{6,26} = 2.47 > 3.10$$

Conclusion: reject null hypothesis

Ans: The new model shows improvement at given significance level