HW 10 Solution

Chapter 10

4. 
\[ x_\cdot = I J x_\cdot = 32(5.19) = 166.08, \text{ so } SST = 911.01 - (166.08)^2/32 = 49.95. \]
\[ SSTr = 8[(4.39 - 5.19)^2 + \ldots + (6.36 - 5.19)^2] = 20.38, \text{ so } SSE = SST - SSTr = 49.95 - 20.38 = 29.57. \]
Then \[ f = \frac{20.38/3}{29.57/28} = 6.43. \]
Since \[ 6.43 \geq F_{0.05,3,28} = 2.95, \text{ } H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \text{ is rejected at level 0.05.} \]
There are differences between at least two average flight times for the four treatments.

7.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>3</td>
<td>75,081.72</td>
<td>25,027.24</td>
<td>1.70</td>
</tr>
<tr>
<td>Error</td>
<td>16</td>
<td>235,419.04</td>
<td>14,713.69</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>310,500.76</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The hypothesis \[ H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 \text{ vs } H_a : \text{ at least two } \mu_i \text{'s are unequal.} \]
\[ 1.70 < F_{10,3,16} = 2.46, \text{ so the p-value } > .10, \text{ and we fail to reject } H_0. \]

9.

The summary quantities are \[ x_1 = 34.3, \ x_2 = 39.6, \ x_3 = 33.0, \ x_4 = 41.9, \ x_\cdot = 148.8, \]
\[ \sum \sum x_{ij}^2 = 946.68, \text{ so } (x_\cdot)^2/24 = (148.8)^2/24 = 922.56, \text{ } SST = 946.68 - 922.56 = 24.12, \]
\[ SSTr = \frac{(34.3)^2 + \ldots + (41.9)^2}{6} - 922.56 = 8.98, \text{ } SSE = 24.12 - 8.98 = 15.14 \]

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatments</td>
<td>3</td>
<td>8.98</td>
<td>2.99</td>
<td>3.95</td>
</tr>
<tr>
<td>Error</td>
<td>20</td>
<td>15.14</td>
<td>.757</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>24.12</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Since \[ 3.10 = F_{0.05,3,20} < 3.95 < 4.94 = F_{0.01,3,20}, \text{ } 0.01 < p-value < 0.05 \text{ and } H_0 \text{ is rejected at level 5\%} \]

14.

\[ I = 4, J = 8, \text{ so } Q_{0.05,4,28} = 3.87, \text{ } l = 3.87\sqrt{\frac{100}{8}} = 1.41 \]
\[ \mu_1 - \mu_2: -0.132 \pm 1.41 \]
\mu_1 - \mu_3: -1.10 \pm 1.41 \\
\mu_1 - \mu_4: -1.97 \pm 1.41 \\
\mu_2 - \mu_3: -0.97 \pm 1.41 \\
\mu_2 - \mu_4: -1.84 \pm 1.41 \\
\mu_3 - \mu_4: -0.87 \pm 1.41 \\

Treatment 4 appears to differ significantly from both 1 and 2, but there are no other significant differences.

Chapter 12

7.

a. \mu_{Y:2500} = 1800 + 1.3(2500) = 5050 \\
b. expected change = slope = \beta_1=1.3 \\
c. expected change = 100\beta_1=130 \\
d. expected change = -100 \beta_1=-130 \\

8.

a. \mu_{Y:2000} = 1800 + 1.3(2000) = 4400, and \sigma = 350, so \\
\quad P(Y > 5000) = P(Z > \frac{5000 - 4000}{350}) = P(Z > 1.71) = .0436 \\
b. Now \ E(Y) = 5050, so P(Y > 5000) = P(Z > .14) = .4443. \\
c. E(Y_2 - Y_1) = E(Y_2) - E(Y_1) = 5050 - 4400 = 650, and V(Y_2 - Y_1) = V(Y_2) + V(Y_1) = \\
\quad (350)^2 + (350)^2 = 245,000, so the standard deviation of Y_2 - Y_1 is 494.97. Thus \\
\quad P(Y_2 - Y_1 > 1000) = P(Z > \frac{1000 - 650}{494.97}) = P(Z > .71) = .2389 \\
d. The standard deviation of Y_2 - Y_1 = 494.97 (from c), and E(Y_2 - Y_1) = 1800 + 1.3x_2 - \\
\quad (1800 + 1.3x_1) = 1.3(x_2 - x_1). Thus \\
\quad P(Y_2 > Y_1) = P(Y_2 - Y_1 > 0) = P(Z > \frac{-1.3(x_2 - x_1)}{494.97}) = .95
This implies that $-1.645 = \frac{-1.3(x_2-x_1)}{494.97}$, so $x_2 - x_1 = 626.33$.

12.

a. $S_{xx} = 39095 - \frac{(517)^2}{14} = 20002.929$, $S_{xy} = 25825 - \frac{(517)(346)}{14} = 13047.714$;

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{13047.714}{20002.929} = .652$$

$$\hat{\beta}_0 = \frac{\sum y - \hat{\beta}_1 \sum x}{n} = \frac{346 - (6523)(517)}{14} = .6261$$

So the equation of the least square regression line is $y = .626 + .652x$.

b. $\hat{y}(35) = .626 + .652(35) = 23.456$. The residual is $y - \hat{y} = 21 - 23.456 = -2.456$.

c. $S_{yy} = 17454 - \frac{(346)^2}{14} = 8902.857$, and

$$SSE = \sum y_i^2 - \hat{\beta}_0 \sum y_i - \hat{\beta}_1 \sum x_i y_i = 17454 - 0.6261 \times 346 - .6523 \times 25825 = 391.7219$$

$$\hat{\sigma} = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{391.7219}{12}} = 5.713$$

d. $SST = S_{yy} = 8902.875$, $r^2 = 1 - \frac{SSE}{SST} = 1 - 391.722/8902.875 = .956$

e. After the largest observations are dropped, similarly, we have $\hat{\beta}_1 = .56445$, $\hat{\beta}_0 = 2.2891$, the least square line is $y = 2.2891 + .56445x$. And $r^2 = .6879$. (Could you explain why $r^2$ decrease so much after the two largest points were deleted?)

16.

a. The scatterplot shows a strong linear relationship between rainfall volume and runoff volume, thus it supports the use of the simple linear regression model.

b. $\bar{x} = 52.200$, $\bar{y} = 42.867$, $S_{xx} = 63040 - \frac{(798)^2}{15} = 20586.4$, $S_{yy} = 41999 - \frac{(643)^2}{15} = 14435.7$, and $S_{xy} = 51232 - \frac{(798)(643)}{15} = 17024.4$

$$\hat{\gamma}_0 = \frac{S_{xy}}{S_{xx}} = \frac{17024.4}{20586.4} = .82697$$

and

$$\hat{\gamma}_0 = 42.867 - (.82697)53.2 = -1.1278$$
c. $\mu_{y,50} = -1.1278 + (.82697)(50) = 40.2207$

d. $SSE = S_{yy} - \hat{\eta}_i S_{xx} = 14435.7 - (.82697)(17324.4) = 357.07$

$$\hat{\sigma} = s = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{357.07}{13}} = 5.24$$

e. $r^2 = 1 - \frac{SSE}{SST} = 1 - \frac{357.07}{14435.7} = .9753$. So 97.53\% of the observed variation in runoff volume can be attributed to the simple linear regression relationship between runoff and rainfall.

35.

a. $\hat{\beta}_1 \pm t_{\alpha/2,n-2} \cdot s_{\hat{\beta}_1} : 0.5549 \pm 2.306(.3101) = (-0.16,1.27)$. The interval contains the value 0, indicating 0 is a possible value for $\beta_1$, and showing little usefulness of the model.

b. The p-value associated with the model is 0.11 (based on t-test of $\beta_1$ or ANOVA), which exceeds standard levels $\alpha$. This indicates least square is not a good way to predict age from transparent dentive content.

36.

a. We reject $H_0$ if $t > t_{0.01,13} = 2.650$. With $S_{xx} = \sum x_i^2 - \frac{(\sum x_i)^2}{n} = 324.40$.

$$t = \frac{1.7035 - 1}{3.725/\sqrt{324.40}} = \frac{.7035}{.2068} = 3.40$$

Since 3.40 > 2.650, $H_0$ is rejected in favor of $H_a$.

b. $t_{0.005,13}$, so the CI is

$$1.7035 \pm \frac{(3.012)(3.725)}{\sqrt{324.40}} = 1.7035 \pm .6229 = (1.08,2.32)$$

45.

a. $\hat{\beta}_0 + 25\hat{\eta}_1 = 50.80$, $\bar{x} = 26.80$, $\sqrt{\frac{1}{15} + \frac{18(26.80-25)^2}{4866}} = .2769$, $s = 3.725$ and $t_{0.05,13} = 1.771$, so the 90\% CI is

$$50.80 \pm (1.771)(3.725)(.2769) = 50.80 \pm 1.83 = (48.97,52.63)$$

b. Utilizing the computation of a, $s_{\hat{\beta}_0+\hat{\beta}_1} = (3.725)(.2769)=1.031$, so

$$t = \frac{50.80 - 50}{1.031} = .78.$$
$H_0$ is rejected if $t \geq t_{0.13} = 1.350$. Since, $.78 < 1.350$, so $H_0$ cannot be rejected.

52.

a. We wish to test $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$. The test statistic $t = \frac{10.6026}{0.9985}$ leads to a p-value of < .006 ($2P(t > 4.0)$ from the 7th row of table A.8), and $H_0$ is rejected since p-value is smaller than any reasonable $\alpha$. The data suggests this model does specify a useful relationship between chlorine flow and etch rate.

b. A 95% confidence interval for $\beta_1$: $10.6026 \pm (2.365)(.9985) = (8.24, 12.96)$. We can be highly confident that when the flow rate is increased by 1 SCCM, the associated expected change in etch rate will be between 824 and 1296 A/min.

c. A 95% CI for $\mu_{Y:30}$:

$$38.256 \pm 2.365 (2.546 \sqrt{\frac{1}{9} + \frac{9(3.0 - 2.667)^2}{58.50}}}$$

$$= 38.256 \pm 2.365 (2.546)(1.06) = 38.256 \pm 6.398 = (36.100, 40.412)$$

d. The 95% PI is

$$38.256 \pm 2.365 (2.546 \sqrt{\frac{1}{9} + \frac{9(3.0 - 2.667)^2}{58.50}}}$$

$$38.256 \pm 2.365 (2.546)(1.06) = 38.256 \pm 6.398 = (31.859, 44.655)$$

e. The intervals for $x^*$ will be narrower than those above because 2.5 is closer to the mean than 3.0.

f. No. a value of 6.0 is not in the range of observed $x$ values, therefore predicting at that point is meaningless.