Instructions: 1. WORK ALL PROBLEMS. Please, give details and explanations and SHOW ALL YOUR WORK so that partial credits can be given.
2. You may use four pages of notes, tables and a calculator but no other reference materials.

1. The following data were obtained from a random sample of eight oak trees on their age in years (x), and the diameter of their trunk in inches (y).

   | x: 22 | 14 | 31 | 36 | 9 | 41 | 19 | 26 |
   | y: 20 | 16 | 30 | 39 | 17 | 42 | 23 | 21 |

(4) a. Write down the regression model and explain its components.

   \[ y = \beta_0 + \beta_1 x + \epsilon \]

   Response \[ y \- \text{Intercept} \] \[ x \- \text{Explanatory} \] \[ \epsilon \- \text{Error} \]

(10) b. Obtain the estimated least-squares regression equation, and explain the meaning of the estimated slope in the context of this problem.

<table>
<thead>
<tr>
<th>x</th>
<th>y</th>
<th>(x-\bar{x})</th>
<th>(y-\bar{y})</th>
<th>(x-\bar{x})^2</th>
<th>(y-\bar{y})^2</th>
<th>(x-\bar{x})(y-\bar{y})</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>20</td>
<td>-2</td>
<td>-6</td>
<td>4</td>
<td>36</td>
<td>12</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>-10</td>
<td>-10</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>31</td>
<td>30</td>
<td>-10</td>
<td>-10</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>36</td>
<td>30</td>
<td>10</td>
<td>10</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>7</td>
<td>7</td>
<td>49</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>41</td>
<td>42</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>19</td>
<td>23</td>
<td>-4</td>
<td>-8</td>
<td>16</td>
<td>64</td>
<td>80</td>
</tr>
<tr>
<td>\sum</td>
<td>192</td>
<td>208</td>
<td>0</td>
<td>0</td>
<td>852</td>
<td>738</td>
</tr>
</tbody>
</table>

\[ \bar{x} = \frac{\sum x}{8} = 24 \]
\[ \bar{y} = \frac{\sum y}{8} = 26 \]

\[ b_1 = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sum (x-\bar{x})^2} = \frac{738}{852} = 0.8662 \]

\[ b_0 = \bar{y} - b \bar{x} = 26 - 0.8662(24) = 5.2112 \]

The estimated reg. eq. is

\[ \hat{y} = 5.2112 + 0.8662x \]

Slope: For every additional one year in age, on average the trunk diameter will increase 0.8662 inches.
(8) c. Test the null hypothesis that the correlation coefficient is equal to zero versus the alternative hypothesis that the correlation coefficient is positive at 0.01 level of significance and find the p-value.

\[ H_0: \rho = 0, \quad H_a: \rho > 0, \quad \text{where} \quad \rho = \text{Prf. Corr. Coeff.} \]

\[ r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{7.38}{\sqrt{65.2} \sqrt{70}} = -0.9611 \]

\[ t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} = -0.9611 \sqrt{6/(1-0.9611)} = 8.5235 \]

\[ df = 8-2 = 6 \quad t_{0.01} = 3.143 \]

We can reject \( H_0 \) at 0.01 level.

\[ p-value < 0.0005 \]

(8) d. Find the 95% prediction interval for the diameter of the trunk when the age of the tree is 30 years. Explain the meaning of this prediction interval.

The 95% prediction interval is

\[ \hat{y} \pm t_{\alpha/2} s_e \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \]

\[ d = 0.05 \quad t_{0.05} = 2.47 \quad df = 6 \quad t_{0.05} = 2.47 \]

\[ SS(\text{resid}) = \sum (y_i - \hat{y})^2 = 692 - \frac{733}{85} = 52.7465 \]

\[ S_e = \sqrt{SS(\text{resid})/(n-2)} = 2.9650 \]

\[ \hat{y} = 31.112 + 0.8642(30) = 31.1972 \]

The 95% prediction interval is:

\[ 31.1972 \pm 2.47(2.9650) \sqrt{1 + \frac{1}{8} + \frac{(30-26)^2}{85}} \]

\[ 31.1972 \pm 7.8386 \Rightarrow (23.3526, 39.0358) \]

We are 95% confident that if we have a 30 years old tree, the diameter of the trunk will be between 23.3526 and 39.0358 inches.

(5) e. Find the coefficient of determination and explain its meaning.

The coefficient of determination is \( r^2 = (0.9611)^2 = 0.9237 \).

92.37% of the variation in trunk diameter can be explained by its linear relationship with age.
1. Researchers in an agricultural college want to test a new variety of corn seed, which they have
developed. They got eight farmers to plant the new seed and another seven farmers to plant old seed. The
number of bushels per acre each farmer obtained are:

| New Seed: 130 129 131 132 128 130 13 | Old Seed: 111 128 124 123 124 125 126 |

Is the new seed better than the old seed in terms of the yield?

(3) a. State the null and alternative hypotheses.

\[ H_0 : \mu_1 = \mu_2 \quad \mu_1 = \text{mean yield for the new seed} \]

\[ H_A : \mu_1 > \mu_2 \quad \mu_2 = \text{mean yield for the old seed} \]

(7) b. Assume the yields for the new and old seeds are normal, test at 0.05 level of significance.

\[ \bar{y}_1 = 130, \quad \bar{y}_2 = 123 \]

\[ s_1 = 1.3093, \quad s_2 = 5.5377 \]

\[ \text{large } s^2 / \text{small } s^2 = 17.9 > 2 \quad \text{use unpoled procedure} \]

\[ SE_1^2 = \frac{s_1^2}{n_1} = \frac{1.3093^2}{8}, \quad SE_2^2 = \frac{s_2^2}{n_2} = \frac{5.5377^2}{7} \]

\[ d.f. = \frac{(SE_1^2 + SE_2^2)^2}{\frac{SE_1^4}{n_1 - 1} + \frac{SE_2^4}{n_2 - 1}} \]

\[ \bar{y}_1 - \bar{y}_2 = 7 = 1.943 \]

\[ t = \frac{7}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \]

\[ = \frac{130 - 123}{\sqrt{\frac{1.3093^2}{8} + \frac{5.5377^2}{7}}} \]

\[ = 3.2654 \]

We can reject \( H_0 \) at 0.05 level.

(3) c. Find the p-value.

\[ .005 < p-value < .01 \]
(7) d. Suppose the distribution of the yields for the new and old seeds are not normal. State the null and alternative hypotheses and test at 0.05 level.

\[ H_0 : \text{The number of bushels/acre are the same for both seeds} \]

\[ H_a : \text{The yield of new seed is more than old seed} \]

\[
\begin{array}{ccc}
\text{Old} & \text{New} & \text{Diff} \\
7 & 136 & 111 & 0 \\
7 & 129 & 123 & 6 \\
7 & 131 & 124 & 7 \\
7 & 132 & 123 & 0 \\
6 & 128 & 124 & 0 \\
7 & 136 & 125 & 11 \\
7 & 131 & 126 & 5 \\
\end{array}
\]

\[ k_1 = 55.5 \]

The critical value from Table 6 for \( \alpha = 0.05 \) one sided is 4.3.

4.3 Reject

We can reject \( H_0 \) at 0.05 level.

3. A digital instrument is used to test the acid level of a particular drug under three different temperatures. Six replicates of each experiment are run in the laboratory and the results are shown below:

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Replicate 1</th>
<th>Replicate 2</th>
<th>Replicate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>155</td>
<td>130</td>
<td>188</td>
</tr>
<tr>
<td>60</td>
<td>75</td>
<td>80</td>
<td>136</td>
</tr>
<tr>
<td>60</td>
<td>106</td>
<td>98</td>
<td>106</td>
</tr>
<tr>
<td>80</td>
<td>174</td>
<td>104</td>
<td>98</td>
</tr>
<tr>
<td>80</td>
<td>150</td>
<td>82</td>
<td>106</td>
</tr>
</tbody>
</table>

The between sum of square is 10,474 and the within sum of squares is 12,670.

(4) a. Write down the analysis of variance model and explain its components.

\[ Y_{ij} = \mu + \tau_i + \epsilon_{ij} \]

\( Y_{ij} \) = \( i \)-th obs. mean 
\( \mu \) = Grand mean 
\( \tau_i \) = Effect of \( i \)-th treat. 
\( \epsilon_{ij} \) = Error
b. Construct the ANOVA table.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp.</td>
<td>2</td>
<td>16.974</td>
<td>52.37</td>
<td>6.1976</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>12.676</td>
<td>8.45</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>17</td>
<td>23.144</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(5) c. Perform a global F test at 0.05 level and explain the result.

\[ F = \frac{\text{MS (between)}}{\text{MS (within)}} = \frac{52.37}{8.45} = 6.1976 \]

\[ df = 2 \text{ and } 15 \]

\[ F_{15} = 3.68 \]

We can reject \( H_0: \mu_1 = \mu_2 = \mu_3 \) at .05 level.

(3) d. Find the p-value for the F-test.

\[ .01 < p \text{-value} < .02 \]

(8) e. Construct a 95% confidence interval for the difference between means in temperatures 50 and 80. Explain the meaning of this confidence interval.

The 100(1-\( \alpha \))% confidence interval is

\[ \bar{Y}_1 - \bar{Y}_2 \pm t_{1,025} \sqrt{\text{MS} \left( \frac{1}{15} + \frac{1}{15} \right)} \]

\[ \bar{Y}_1 = 144.833 \] \[ \bar{Y}_2 = 36 \] \[ \text{SE} = 8.45 \]

\[ \alpha = .05 \] \[ d.f. = 15 \] \[ t_{1,025} = 2.131 \]

The 95% confidence interval is

\[ (144.833 - 36) \pm 2.131 \sqrt{8.45 \left( \frac{1}{15} + \frac{1}{15} \right)} \]

\[ 58.833 \pm 35.7644 \Rightarrow (23.0689, 94.5974) \]

We are 95% confident that the difference in mean acidity levels for temperatures 50 and 80 is between 23.0689 and 94.5974.
4. A study was conducted to observe the relationship between the emotional stability and talent. One hundred individuals were selected randomly and the following cross-classification was obtained.

<table>
<thead>
<tr>
<th>Emotional Stability</th>
<th>Stable</th>
<th>Unstable</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Talent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Artist</td>
<td>10</td>
<td>16</td>
<td>26</td>
</tr>
<tr>
<td>Musical genius</td>
<td>20</td>
<td>14</td>
<td>34</td>
</tr>
<tr>
<td>Math</td>
<td>30</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>60</td>
<td>40</td>
<td>100</td>
</tr>
</tbody>
</table>

(2) a. State the null and alternative hypotheses in words.

\[ H_0 : \text{Emotional stability is independent of talent} \]

\[ H_A : \text{H}_0 \text{ is not true} \]

(10) b. Use the chi-squares procedure to test the hypothesis stated in (a) at 0.01 level.

\[
\chi^2 = \sum \frac{(O - E)^2}{E}
\]

\[
= \frac{(10 - 15.6)^2}{15.6} + \frac{(16 - 10.4)^2}{10.4} + \frac{(20 - 22.4)^2}{22.4} + \frac{(10 - 13.6)^2}{13.6} + \frac{(14 - 24)^2}{24} + \frac{(10 - 16)^2}{16}
\]

\[= 8.795 \]

\[df = (r-1)(c-1) = (3-1)(2-1) = 2\]

\[\alpha = .01\]

\[\chi^2_{o1} = 9.21\]

We cannot reject \( H_0 \) at .01 level.

(3) c. What is the p-value?

\[.01 < p-value < .02\]

A Brief solution of the regression example given in Lecture 24:

\[x-bar = 2, y-bar = 14, \Sigma (x - \bar{x})^2 = 8, \Sigma y = 58, \Sigma x = 32, b_1 = 4, b_0 = 6, \]

\[y^\prime = 6 + 4x, SS(resid) = 16, Se = 1.633, Sb_1 = .5774, t = 6.9276, r = .9428.\]