1. (30 p) It is hypothesized that the total sales of a corporation should vary more in an industry with active price competition than in one with duopoly and tacit collusion. In a study of the merchant ship production industry it was found that in 4 years of active price competition, the variance of company A’s total sales was 114.09. In the following 7 years, during which there was duopoly and tacit collusion, this variance was 16.08. Assume that the data can be regarded as an independent random sample from two normal distributions. Test at the 5% level, the null hypothesis that the two population variances are equal against the alternative that the variance of total sales is higher in years of active price competition.

2. (35 p) The sample data are reported as follows:

<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>7</th>
<th>3</th>
<th>8</th>
<th>11</th>
<th>5</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>16</td>
<td>50</td>
<td>22</td>
<td>59</td>
<td>63</td>
<td>46</td>
<td>43</td>
</tr>
</tbody>
</table>

a) Prepare a scatter plot of the data and show the regression line fitted
b) Compute covariance and correlation coefficient
c) Compute and interpret least square coefficients $b_1$ and $b_0$
d) Determine the predicted $Y$-value when $X = 10$.

3. (35 p) A marketing research firm is conducting a study to determine if there is a relationship between an individual’s age and the individual’s preferred source of news. The research firm asked 1000 individuals to list their preferred source for news: newspaper, radio and television, or the Internet. The following results were obtained:

<table>
<thead>
<tr>
<th>Preferred News Source</th>
<th>Age of Respondent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20-30</td>
</tr>
<tr>
<td>Newspaper</td>
<td>19</td>
</tr>
<tr>
<td>Radio/TV</td>
<td>27</td>
</tr>
<tr>
<td>Internet</td>
<td>104</td>
</tr>
</tbody>
</table>

At the 0.01 level of significance, can the marketing research firm conclude that there is a relationship between the age of the individual and the individual’s preferred source for news?

Recall:

\[
S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}};\quad S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}};\quad S_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1};
\]

\[
r_{XY} = \frac{S_{XY}}{S_X S_Y};\quad b_1 = \frac{\text{Cov}(X,Y)}{S_X}; b_0 = \bar{Y} - b_1 \bar{X}; \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}; \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n};
\]

\[
\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}; E_{ij} = \frac{R_i C_j}{n}
\]
1. Let 1-industry with active price competition; 2-industry with duopoly and tacit collusion. Use the Diagram E-5. First identify the null and alternative hypothesis:

\[ H_0: \sigma_1^2 \leq \sigma_2^2 \]
\[ H_1: \sigma_1^2 > \sigma_2^2 \text{ (one - sided)} \]

Use F test statistic with the definition given at the diagram E-5:

\[ F = \frac{s_1^2}{s_2^2} = \frac{114.09}{16.08} = 7.095 \]

Given:
\[ \alpha = 0.05, \ n_1 = 4 \text{ years (with active price competition)}, n_2 = 7 \text{ (duopoly and tacit collusion)} \]

Now determine \( F_{n_1-1,n_2-1,\alpha} \) such that \( P(F_{n_1-1,n_2-1,\alpha} > F_{n_1-1,n_2-1,\alpha}) = \alpha \) with the numerator degrees of freedom \( (n_1-1) \) and denominator degrees of freedom \( n_2 - 1 \). Use table of cutoff points for the F distribution (Appendix Table 9/p.845) as follows:

\[ F_{3,6,0.05} = 4.76 \text{ (Read from Table 9, p. 846)} \]

Compare with the test statistic F. Reject \( H_0 \), if the computed F test statistic \((7.095)\) is greater than \( F_{3,6,0.05} \).

Since 7.095 < 4.76, reject the null hypothesis and conclude as follows:

Conclusion: There is evidence of higher variance in years of active price competition.

2.a) The scatter plot and the regression line fitted are shown below:

![Scatterplot of x vs. y](image)

b) Use tabulated method to compute covariance and correlation coefficient as follows:

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
 i & X_i & Y_i & (X_i - \bar{X}) & (X_i - \bar{X})^2 & (Y_i - \bar{Y}) & (Y_i - \bar{Y})^2 & (X_i - \bar{X})(Y_i - \bar{Y}) \\
\hline
 1 & 1 & 16 & 1-5.57= -4.57 & 20.8849 & 16-42.71=-26.71 & 713.4241 & 122.0647 \\
 2 & 7 & 50 & 7-5.57=1.43 & 2.0449 & 50-42.71=7.29 & 53.1441 & 10.4247 \\
 3 & 3 & 22 & 3-5.57=-2.57 & 6.6049 & 22-42.71=-20.71 & 428.9041 & 53.2247 \\
 4 & 8 & 59 & 8-5.57=2.43 & 5.9049 & 59-42.71=16.29 & 265.3641 & 39.8547 \\
 5 & 11 & 63 & 11-5.57=5.43 & 29.4849 & 63-42.71=20.29 & 411.6841 & 110.1747 \\
 6 & 5 & 46 & 5-5.57=-0.57 & 0.3249 & 46-42.71=3.29 & 10.8241 & -1.8753 \\
 7 & 4 & 43 & 4-5.57=-1.57 & 2.4649 & 43-42.71=0.29 & 0.0841 & -0.4553 \\
\hline
\Sigma & 39 & 299 & 67.7143 & & 1883.4287 & & 333.1429 \\
\end{array}
\]

\[ \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{39}{7} = 5.57; \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{299}{7} = 42.71 \]
\[ S_{XY} = \text{Cov}(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1} = \frac{333.1429}{7-1} = 55.5238 \]

\[ S_x = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} = \sqrt{\frac{67.7143}{7-1}} = 3.3594; S_x^2 = 11.2857 \]

\[ S_y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{1883.4287}{7-1}} = \sqrt{313.9048} = 17.7174 \]

\[ r_{XY} = \frac{S_{XY}}{S_x S_y} = \frac{55.5238}{(3.3594)(17.7174)} = 0.9329 \]

c)

\[ b_1 = \frac{\text{Cov}(X, Y)}{S_x^2} = \frac{55.5238}{11.2857} = 4.92 \]

Interpretation of \( b_1 \): The slope of the equation indicates the change in the average value of \( Y \) when \( X \) variable increases by 1 unit. For a 1 unit increase in the value of \( X \) variable, we estimate that the value of \( Y \) would increase by 4.92 units.

\[ b_0 = \bar{Y} - b_1 \bar{X} = 42.71 - (4.92)(5.57) = 15.31 \]

Interpretation of \( b_0 \): If the value of \( X \) were 0 unit, we would expect to have the value of \( Y \) 15.31 units. Interpret with caution—note that we are extrapolating the results beyond the observed data.

d) Prediction (Estimation) from the least-square linear correlation (Value of \( r_{XY} = 0.9329 \) shows good correlation of \( Y \) with \( X \) positively.

\[ \hat{Y}_{n+1} = b_0 + b_1 X_{n+1} = 15.31 + (4.92)(10) = 64.51 \text{ units} \]

3. Null and alternative hypothesis are as follows:

\( H_0: \) Age of the individual is independent of the individual’s preferred source for news.

\( H_1: \) Age of the individual and the individual’s preferred source for news are dependent.

Complete the contingency table (3x4 matrix) given to calculate expected counts \( (E_{ij}) \) from the corresponding formula below. These values were written under the observed counts \( (O_{ij}) \) as **boldface** and *italic* with the row and column total as follows:

\[
\begin{align*}
E_{ij} &= \frac{R_iC_j}{n} \\
E_{11} &= \frac{R_1C_1}{n} = \frac{(323)(150)}{1000} = 48.45 \\
E_{12} &= \frac{R_1C_2}{n} = \frac{(323)(300)}{1000} = 96.9 \\
E_{13} &= \frac{R_1C_3}{n} = \frac{(323)(300)}{1000} = 96.9 \\
E_{14} &= \frac{R_1C_4}{n} = \frac{(323)(250)}{1000} = 80.75 = 323 - (48.45 + 96.9 + 96.9) \\
E_{21} &= \frac{R_2C_1}{n} = \frac{(408)(150)}{1000} = 61.2 \\
E_{22} &= \frac{R_2C_2}{n} = \frac{(408)(300)}{1000} = 122.4 \\
E_{23} &= \frac{R_2C_3}{n} = \frac{(408)(300)}{1000} = 122.4
\end{align*}
\]
\[ E_{24} = \frac{R_2C_4}{n} = \frac{(408(250)}{1000} = 102 \]
\[ E_{31} = \frac{R_3C_1}{n} = \frac{(269(150)}{1000} = 40.35 \]
\[ E_{32} = \frac{R_3C_2}{n} = \frac{(269(300)}{1000} = 80.7 \]
\[ E_{33} = \frac{R_3C_3}{n} = \frac{(269)(300)}{1000} = 80.7 \]
\[ E_{34} = \frac{R_3C_4}{n} = \frac{(269)(250)}{1000} = 67.25 \]

<table>
<thead>
<tr>
<th>Preferred News Source</th>
<th>Age of Respondent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20-30</td>
<td>31-40</td>
</tr>
<tr>
<td>Newspaper</td>
<td>19</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td><strong>48.45</strong></td>
<td><strong>96.9</strong></td>
</tr>
<tr>
<td>Radio/TV</td>
<td>27</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td><strong>61.2</strong></td>
<td><strong>122.4</strong></td>
</tr>
<tr>
<td>Internet</td>
<td>104</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td><strong>40.35</strong></td>
<td><strong>80.7</strong></td>
</tr>
<tr>
<td>Total</td>
<td>150</td>
<td>300</td>
</tr>
</tbody>
</table>

\[ \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \]
\[ = \frac{(O_{11} - E_{11})^2}{E_{11}} + \frac{(O_{12} - E_{12})^2}{E_{12}} + \frac{(O_{13} - E_{13})^2}{E_{13}} + \frac{(O_{14} - E_{14})^2}{E_{14}} + \frac{(O_{21} - E_{21})^2}{E_{21}} \]
\[ + \frac{(O_{22} - E_{22})^2}{E_{22}} + \frac{(O_{23} - E_{23})^2}{E_{23}} + \frac{(O_{24} - E_{24})^2}{E_{24}} + \frac{(O_{31} - E_{31})^2}{E_{31}} + \frac{(O_{32} - E_{32})^2}{E_{32}} \]
\[ + \frac{(O_{33} - E_{33})^2}{E_{33}} + \frac{(O_{34} - E_{34})^2}{E_{34}} \]
\[ \chi^2 = \frac{(19 - 48.45)^2}{48.45} + \frac{(62 - 96.9)^2}{96.9} + \frac{(95 - 96.9)^2}{96.9} + \frac{(147 - 80.75)^2}{80.75} + \frac{(27 - 61.2)^2}{61.2} \]
\[ + \frac{(125 - 122.4)^2}{122.4} + \frac{(168 - 122.4)^2}{122.4} + \frac{(88 - 102)^2}{88 - 102} + \frac{(104 - 40.35)^2}{40.35} \]
\[ + \frac{(113 - 80.7)^2}{80.7} + \frac{(37 - 80.7)^2}{80.7} + \frac{(15 - 67.25)^2}{67.25} \]
\[ = 17.901 + 12.570 + 0.0373 + 54.354 + 19.112 + 0.0552 + 16.988 + 1.922 + 100.405 + 12.928 + 23.664 + 40.596 = 300.531 \]

Decision rule:

\[ reject \ H_0 \ if \ \chi^2 > \chi^2_{(r-1)(c-1),\alpha} \]
\[ \chi^2 = 300.531 > \chi^2_{(3-1)(4-1),0.01} = 16.81 (Read \ from \ Appendix \ Table \ 7) \]

Since the calculated chi-square value is higher than the critical chi-square value (300.531 > 16.81), we reject the null hypothesis and conclude that age of the individual and the individual’s preferred source for news are not independent.