1. (35 p) A screening procedure was designed to measure attitudes towards minorities as managers. High scores indicate negative attitudes and low scores indicate positive attitudes. Independent random samples were taken of 151 male financial analysts and 108 female financial analysts. For the former group the sample mean and standard deviation scores were 85.8 and 19.13, whereas the corresponding statistics for the latter group were 71.5 and 12.2.

Test the null hypothesis that the two population means are equal against the alternative that the true mean score is higher for male than for female financial analyst at the 5% significance level.

2. (35 p) The president of a company has asked you to study the relationship between the market price and the tons of product supplied by his competitor company. He supplies you with the following observations of price per ton and number of tons as follows:

<table>
<thead>
<tr>
<th>Price per ton (X, $)</th>
<th>2</th>
<th>4</th>
<th>3</th>
<th>6</th>
<th>3</th>
<th>5</th>
<th>6</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Tons (Y, ton)</td>
<td>5</td>
<td>10</td>
<td>8</td>
<td>18</td>
<td>6</td>
<td>15</td>
<td>20</td>
<td>4</td>
</tr>
</tbody>
</table>

a) Prepare a scatter plot of the data
b) Compute covariance and correlation coefficient
c) Compute and interpret

d) How many tons would you estimate if the price per ton is 2.5 $.

3. (35 p) Following a presidential debate, people were asked how they might vote in the forthcoming election:

<table>
<thead>
<tr>
<th>Candidate Preference</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate A</td>
<td>Male: 150</td>
</tr>
<tr>
<td>Candidate B</td>
<td>Male: 100</td>
</tr>
</tbody>
</table>

Is there any association between one’s gender and choice of presidential candidate at the 5% significance level?

Recall:

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

\[ r_{XY} = \frac{S_{XY}}{S_X S_Y} \ ; \ b_1 = \frac{\text{Cov}(X,Y)}{S_X^2} \ ; \ b_0 = \bar{Y} - b_1 \bar{X} \ ; \ \hat{X} = \frac{\sum_{i=1}^{n}X_i}{n} \ ; \ \hat{Y} = \frac{\sum_{i=1}^{n}Y_i}{n} \ ; \]

\[ \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \ ; \ E_{ij} = \frac{R_i C_j}{n} \]
1. Let $x = \text{male financial analysts}; y = \text{female financial analysts}$
First identify the null and alternative hypothesis

\[
H_0: \mu_x - \mu_y = 0 \\
H_1: \mu_x - \mu_y > 0
\]

To test $H_0: \mu_x - \mu_y = 0$ against the alternative $H_1: \mu_x - \mu_y > 0$ the decision rule is to reject $H_0$ and accept $H_1$ if

\[
z > z_\alpha \text{ (upper - tail test, From Appendix Guidelines Diagram E - 1)}
\]

\[
z = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} > z_\alpha \text{ (Guidelines E - 4)}
\]

$\bar{x} = 85.8, \bar{y} = 71.5, n_x = 151, n_y = 108, \sigma_x = 19.13, \sigma_y = 12.2$
Now substitute these values into the formula to find the test statistic, rounding to three decimal places.

\[
z = \frac{(85.8 - 71.5) - 0}{\sqrt{\frac{(19.13)^2}{151} + \frac{(12.2)^2}{108}}} = 7.334
\]

\[
\alpha = 5\%, z_{0.05} = 1.645 \text{ (Read from Appendix Table 1, p.815)}
\]

Compare the test statistic to $z_{0.05}$

\[
\text{If } z > z_{0.05}, \text{ Reject } H_0
\]

\[
7.334 > 1.645
\]

Reject $H_0$ at the 5% significance level: $\mu_x - \mu_y > 0$

Conclusion: There is sufficient evidence that the true mean score is higher for male than for female financial analysts.

2.a) Prepare a scatter plot as follows:
b) Use tabulated method to compute covariance and correlation coefficient as follows:

<table>
<thead>
<tr>
<th>$i$</th>
<th>$X_i$</th>
<th>$Y_i$</th>
<th>$(X_i - \bar{X})$</th>
<th>$(X_i - \bar{X})^2$</th>
<th>$(Y_i - \bar{Y})$</th>
<th>$(Y_i - \bar{Y})^2$</th>
<th>$(X_i - \bar{X})(Y_i - \bar{Y})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>5</td>
<td>-1.875</td>
<td>3.515625</td>
<td>-5.75</td>
<td>33.0625</td>
<td>10.78125</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>10</td>
<td>0.125</td>
<td>0.015625</td>
<td>-0.75</td>
<td>0.5625</td>
<td>-0.09375</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>8</td>
<td>-0.875</td>
<td>0.765625</td>
<td>-2.75</td>
<td>7.5625</td>
<td>2.40625</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>18</td>
<td>2.125</td>
<td>4.515625</td>
<td>7.25</td>
<td>52.5625</td>
<td>15.40625</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>6</td>
<td>-0.875</td>
<td>0.765625</td>
<td>-4.75</td>
<td>22.5625</td>
<td>4.15625</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>15</td>
<td>1.125</td>
<td>1.265625</td>
<td>9.25</td>
<td>85.5625</td>
<td>19.65625</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>20</td>
<td>2.125</td>
<td>4.515625</td>
<td>9.25</td>
<td>85.5625</td>
<td>19.65625</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>4</td>
<td>-1.875</td>
<td>3.515625</td>
<td>-6.75</td>
<td>45.5625</td>
<td>12.65625</td>
</tr>
<tr>
<td>$\Sigma$</td>
<td>31</td>
<td>86</td>
<td>18.875</td>
<td>265.5</td>
<td>69.75</td>
<td>69.75</td>
<td>69.75</td>
</tr>
</tbody>
</table>

\[
\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{31}{8} = 3.875; \quad \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{86}{8} = 10.75
\]

\[
S_{XY} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} = \frac{18.875}{8-1} = 1.64208; \quad S_X^2 = 2.6964
\]

\[
S_Y = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1}} = \sqrt{\frac{265.5}{8-1}} = 6.15862; \quad S_Y^2 = 37.9286
\]

\[
r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{9.964286}{(1.64208)(6.15862)} = 0.9853
\]

c) \[
b_1 = \frac{\text{Cov}(X,Y)}{S_X^2} = \frac{9.964286}{2.6964} = 3.6954
\]

Interpretation of $b_1$: For a one dollar per ton increase in the price, we estimate that the number of tons would increase by 3.70 tons (number of tons has a positive effect on price per ton, i.e. as price per ton increases, the tons supplied also increase).

\[
b_0 = \bar{Y} - b_1 \bar{X} = 10.75 - (3.6954)(3.875) = -3.5697
\]

Interpretation of $b_0$: If the price per ton were $0, we would expect to have number of tons (-3.5697). 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 $r_{XY} = 0.9853$ shows good correlation of price per ton with the number of tons positively).

\[
\hat{Y}_{n+1} = b_0 + b_1 X_{n+1} = -3.5697 + 3.6954(2.5) = 5.6688 \text{ tons}
\]

3. Complete the contingency table (2x2 matrix) given to calculate expected counts ($E_{ij}$) from the corresponding formula written below observed counts ($O_{ij}$) as boldface and italic with the row and column total as follows:
<table>
<thead>
<tr>
<th>Candidate Preference</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate A</td>
<td>150</td>
<td>130</td>
<td>280</td>
</tr>
<tr>
<td></td>
<td>140.00</td>
<td>140.00</td>
<td></td>
</tr>
<tr>
<td>Candidate B</td>
<td>100</td>
<td>120</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td>110.00</td>
<td>110.00</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>250</td>
<td>500</td>
</tr>
</tbody>
</table>

\[
H_0: \text{No association exists between one's gender and choice of presidential candidate.}
\]

\[
H_1: \text{Association exists between one's gender and choice of presidential candidate.}
\]

\[
E_{ij} = \frac{R_i C_j}{n} \rightarrow E_{11} = \frac{R_1 C_1}{n} = \frac{(280)(250)}{500} = 140
\]

\[
E_{12} = \frac{R_1 C_2}{n} = \frac{(280)(250)}{500} = 140
\]

\[
E_{21} = \frac{R_2 C_1}{n} = \frac{(220)(250)}{500} = 110
\]

\[
E_{22} = \frac{R_2 C_2}{n} = \frac{(220)(250)}{500} = 110
\]

\[
\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}
\]

\[
= \frac{(150 - 140)^2}{140} + \frac{(130 - 140)^2}{140} + \frac{(100 - 110)^2}{110} + \frac{(120 - 110)^2}{110}
\]

\[
= 0.714 + 0.714 + 0.909 + 0.909 = 3.247
\]

Decision rule:

\[
\text{reject } H_0 \text{ if } \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} > \chi^2_{(r-1)(c-1), \alpha}
\]

\[
\chi^2 = 3.247 < \chi^2_{(2-1)(2-1), 0.05}
\]

\[
\chi^2_{(1,0.05)} = 3.84 \text{ (Read from Appendix Table 7)}
\]

Conclusion: Do not reject \(H_0\) of no association at the 5% significance level (\(p\text{-value} = P(\chi^2_{v=1} > 3.247) > 0.05\) from Table 7)