1. (35 p) A random sample of 1556 people in country A were asked to respond to this statement: \textit{Increased world trade can increase our per capita prosperity}. Of these sample members, 38.4\% agreed with the statement. When the same statement was presented to a random sample of 1108 people in country B, 52.0\% agreed. Test the null hypothesis that the population proportions agreeing with this statement were the same in the two countries against the alternative that a higher proportion agreed in country B at the 5\% significance level.

2. (35 p) The following data give $X$, the price charged per piece of plywood, and $Y$, the quantity sold (in thousands) as follows:

<table>
<thead>
<tr>
<th>Price per piece ($X$, $$$)</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantity sold ($Y$, $10^4$)</td>
<td>80</td>
<td>60</td>
<td>70</td>
<td>40</td>
<td>0</td>
</tr>
</tbody>
</table>

a) Prepare a scatter plot of the data b) Compute covariance and correlation coefficient c) Compute and interpret $b_1$ and $b_0$ d) What quantity of plywood would you expect to sell if the price were $7$ per piece?

3. (35 p) University administrators have collected the following information concerning student grade point average (GPA) and the school of the student’s major as follows:

<table>
<thead>
<tr>
<th>School</th>
<th>GPA &lt; 3.0</th>
<th>GPA $\geq$ 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts and Sciences</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>Business</td>
<td>45</td>
<td>30</td>
</tr>
<tr>
<td>Music</td>
<td>15</td>
<td>25</td>
</tr>
</tbody>
</table>

Determine if there is any association between GPA and major at the 5\% significance level.

Recall:
\[
S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}}; \quad S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}; \quad S_{XY} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n-1};
\]
\[
r_{XY} = \frac{S_{XY}}{S_X S_Y}; \quad b_1 = \frac{\text{Cov}(X,Y)}{S_X^2}; \quad b_0 = \bar{Y} - b_1 \bar{X}; \quad \bar{X} = \frac{\sum_{i=1}^{n}X_i}{n}; \quad \bar{Y} = \frac{\sum_{i=1}^{n}Y_i}{n};
\]
\[
\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}; \quad E_{ij} = \frac{R_i C_j}{n}
\]
1. Let $x =$ agreed with the statement in country A; $y =$ agreed with the statement in country B
First identify the null and alternative hypothesis

$H_0 : P_x - P_y = 0$
$H_1 : P_x - P_y < 0$

To test $H_0 : P_x - P_y = 0$ against the alternative $H_1 : P_x - P_y < 0$ the decision rule is to reject $H_0$ and accept $H_1$ if

$$z < -z_{\alpha} \text{ (lower – tail test, From Appendix Guidelines Diagram E – 4)}$$

$$z = \frac{(\hat{p}_x - \hat{p}_y) - (P_x - P_y)}{\hat{p}_o (1 - \hat{p}_o) \left( \frac{1}{n_x} + \frac{1}{n_y} \right)} < -z_{\alpha} \text{ (Guidelines E – 4)}$$

$$\hat{p}_o = \frac{n_x \hat{p}_x + n_y \hat{p}_y}{n_x + n_y}$$

$\hat{p}_x = 0.384 \text{ (38.4%)}, \hat{p}_y = 0.52 \text{ (52.0%)}, n_x = 1556, n_y = 1108$

$$\hat{p}_o = \frac{n_x \hat{p}_x + n_y \hat{p}_y}{n_x + n_y} = \frac{1556(0.384) + 1108(0.52)}{1556 + 1108} = 0.44 \text{ (common proportion)}$$

Now substitute these values into the formula to find the test statistic, rounding to three decimal places.

$$z = \frac{(0.384 - 0.52) - 0}{\sqrt{0.44(1 - 0.44) \left( \frac{1}{1556} + \frac{1}{1108} \right)}} = -6.97$$

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

Compare the test statistic to $z_{0.05}$

If $z < -z_{0.05}, \text{Reject } H_0$

$-6.97 < -1.645$

Reject $H_0$ at the 5% significance level: $P_x - P_y < 0$

Conclusion: There is sufficient evidence that the two population proportions are not equal, i.e. a higher proportion agreed in country B compared to in country A.

2.a) Scatter plot—Quantity sold ($10^3$) vs. Price per piece ($\$)
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>6</td>
<td>80</td>
<td>-2</td>
<td>4</td>
<td>30</td>
<td>900</td>
<td>-60</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>60</td>
<td>-1</td>
<td>1</td>
<td>10</td>
<td>100</td>
<td>-10</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>400</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>40</td>
<td>1</td>
<td>1</td>
<td>-10</td>
<td>100</td>
<td>-10</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>-50</td>
<td>2500</td>
<td>-100</td>
</tr>
<tr>
<td>$\Sigma$</td>
<td>40</td>
<td>250</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>4000</td>
<td>-180</td>
</tr>
</tbody>
</table>

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{40}{5} = 8; \quad \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{250}{5} = 50$$

$$S_{XY} = \text{Cov}(X, Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{n - 1} = -180 \quad \frac{5 - 1}{1} = -45$$

$$S_X = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n - 1}} = \sqrt{\frac{10}{5 - 1}} = 1.5811; \quad S_X^2 = 2.5$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n - 1}} = \sqrt{\frac{4000}{5 - 1}} = 31.623; \quad S_Y^2 = 1000$$

$$r_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\text{Cov}(X, Y)}{S_X S_Y} = \frac{-45}{(1.5811)(31.623)} = -0.900$$

c)

$$b_1 = \frac{\text{Cov}(X, Y)}{S_X^2} = \frac{-45}{2.5} = -18$$

Interpretation of $b_1$: For a one dollar increase in the price per piece of plywood, the quantity sold of plywood is estimated to decrease by 18 thousand pieces.

$$b_0 = \bar{Y} - b_1 \bar{X} = 50 - (-18)(8) = 194.00$$

Interpretation of $b_0$: If the price per piece of plywood were $0, we would expect to sell 194 thousand pieces of plywood. Interpret with caution—note that we are extrapolating the results beyond the observed data.

d) Prediction (Estimation) from the least-square linear correlation ($Value\ of\ r_{XY} = 0.900$ shows good correlation of quantity sold of plywood with the price per piece of plywood negatively.

$$\hat{Y}_{n+1} = b_0 + b_1 X_{n+1} = 194.00 + (-18)(7) = 68\ \text{thousand pieces}$$

3. Complete the contingency table (3x2 matrix) given to calculate expected counts ($E_{ij}$) from the corresponding formula. These expected counts ($E_{ij}$) were written below observed counts ($O_{ij}$) as boldface and italic with the row and column total as follows:

<table>
<thead>
<tr>
<th>Age</th>
<th>GPA $&lt; 3.0$</th>
<th>GPA $\geq 3.0$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arts and Sciences</td>
<td>50</td>
<td>35</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>$\textbf{46.75}$</td>
<td>$\textbf{38.25}$</td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>45</td>
<td>30</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>$\textbf{41.25}$</td>
<td>$\textbf{33.75}$</td>
<td></td>
</tr>
<tr>
<td>Music</td>
<td>15</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>$\textbf{22.00}$</td>
<td>$\textbf{18.00}$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>110</td>
<td>90</td>
<td>200</td>
</tr>
</tbody>
</table>
\[ H_0: \text{No association exists between GPA and major.} \]
\[ H_1: \text{Association exists between GPA and major.} \]

\[ E_{ij} = \frac{R_i C_j}{n} \Rightarrow E_{11} = \frac{R_1 C_1}{n} = \frac{(85)(110)}{200} = 46.75 \]
\[ E_{12} = \frac{R_1 C_2}{n} = \frac{(85)(90)}{200} = 38.25 \]
\[ E_{21} = \frac{R_2 C_1}{n} = \frac{(75)(110)}{200} = 41.25 \]
\[ E_{22} = \frac{R_2 C_2}{n} = \frac{(75)(90)}{200} = 33.75 \]
\[ E_{31} = \frac{R_3 C_1}{n} = \frac{(40)(110)}{200} = 22.00 \]
\[ E_{32} = \frac{R_3 C_2}{n} = \frac{(40)(90)}{200} = 18.00 \]

\[ \chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \]
\[ \chi^2 = \frac{(50 - 46.75)^2}{46.75} + \frac{(35 - 38.25)^2}{38.25} + \frac{(45 - 41.25)^2}{41.25} + \frac{(30 - 33.75)^2}{33.75} + \frac{(15 - 22.00)^2}{22.00} + \frac{(25 - 18.00)^2}{18.00} \]
\[ \chi^2 = 0.226 + 0.276 + 0.341 + 0.417 + 2.227 + 2.722 = 6.209 \]

Decision rule:

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

\[ \chi^2 = 6.209 > \chi^2_{(3-1)(2-1),0.05} \]

Conclusion: Reject \( H_0 \) of no association at the 5% significance level (\( p-value = P(\chi^2_{v=2} > 6.209) < 0.05 \))