Question 1. (20 points)

1. (6 points) Draw a histogram for the exam scores (out of 100) given in the stem-and-leaf display. Label the axes.

2. (14 points) Consider the following table for the number of aftershocks for earthquakes that occurred last week around the world.

<table>
<thead>
<tr>
<th>Number of shocks</th>
<th>5</th>
<th>12</th>
<th>16</th>
<th>20</th>
<th>25</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>6</td>
<td>11</td>
<td>18</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

a) (1 point) What is the variable in the above data set?

b) (2 points) Find the 75th percentile.

\[ n = 6 + 11 + 18 + \ldots + 1 = 43 \]

\[ (0.75) \times 43 = 32.25 \rightarrow 33^{rd} \text{ observation: } 16 \]

c) (5 points) Find the mean and the standard deviation.

\[ \bar{x} = \frac{5 \times 6 + 12 \times 11 + 16 \times 18 + \ldots + 33 \times 1}{43} = 14.8 \]

\[ s^2 = \frac{6 \times (5-14.8)^2 + 11 \times (12-14.8)^2 + \ldots + (33-14.8)^2}{42} = 34.3 \]

\[ s = \sqrt{34.3} \approx 5.8 \]

d) (1 point) Are the quantities you found in part c), \( \bar{x} \) and \( s \) or \( \mu \) and \( \sigma \)?

\( \bar{x} \) and \( s \) because they are from a sample.

e) (2 points) Determine the exact percentage of the measurements within two standard deviations of the mean in the data set.

\[ 14.8 - 2(5.8), 14.8 + 2(5.8) = (3, 26.6) \]

\[ \frac{6 + 11 + 18 + 14 + 3}{45} = 26.6 \]

\[ 87.7\% \]

g) (3 points) Use either the empirical rule or Chebyshev's rule (which one is appropriate, why?) to approximate the percentage of the measurements within two standard deviations of the mean.

Chebyshev's rule because the distribution is skewed (to the right) \Rightarrow larger than 75%
**Question 2: (25 points)**

Zeynep is participating in a wine tasting contest. Contestants are given three glasses numbered 1 through 3, each with a different type of red wine. The contestants are also given labels with the names of the three types and are asked to attach one to each glass. Based on past experience, Zeynep correctly identifies the type in the first glass (glass 1) she tastes 40% of the time, and that in the second glass (glass 2) 30% of the time. She correctly identifies both the first and the second 25% of the time. She is completely wrong 20% of the time.

Let $B_i$ denote the event that she correctly identifies the type in the $i^{th}$ glass.

1. (6 points) Find $P(B_2 | B_1)$. Are $B_1$ and $B_2$ independent? Show.

   \[ P(B_2 | B_1) = \frac{P(B_2 \cap B_1)}{P(B_1)} = \frac{25\%}{40\%} = 0.625. \text{ Since } P(B_2 | B_1) \neq P(B_2) \text{ they are not independent.} \]

2. (3 points) Find $P(B_1 \cup B_2)$.

   \[ P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2) = 0.40 + 0.30 - 0.25 = 0.45 \]

3. (3 points) It is claimed that $B_2 \cap B_3$ and $B_1^c$ (the complement of $B_1$) are mutually exclusive. Do you agree? Justify your answer in at most two sentences.

   If two glasses are identified correctly then the 3rd one must also be correctly identified $\Rightarrow B_1$ must occur $\Rightarrow B_1 \cap B_2 \cap B_3 = \emptyset$

4. (7 points) What is the probability that

   i. Zeynep identifies all glasses correctly?

   \[ P(B_1 \cap B_2 \cap B_3) = P(B_1 | B_1 \cap B_2) P(B_2 \cap B_3) = (1)(0.25) = 0.25 \]

   ii. someone who randomly labels the glasses identifies all three glasses correctly?

   \[ P(\text{labels}) = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27} \text{ ways, but only one is correct} \Rightarrow \frac{1}{6} \approx 0.17 \]

Does Zeynep have expertise in wine tasting or not, do you think?

She has because her probability (0.25) is larger than 0.17.

5. (6 points) Let $X$ denote the number of glasses she correctly identifies. Find the probability distribution of $X$.

   \[ P(2) = P(X = 2) = P(B_1 \cap B_2 \cap B_3) = 0.25 \quad \text{ (from part 4)} \]

   \[ P(0) = P(X = 0) = P(B_1^c \cap B_2^c \cap B_3^c) = 0.20 \quad \text{ (from question)} \]

   \[ P(1) = P(X = 1) = 1 - 0.25 - 0.20 = 0.55 \]

   \[ P(1) = 1 - 0.25 - 0.20 = 0.55 \]
**Question 3: (20 points)**
An instructor gives a 20 question multiple choice exam to the students. Each question has four choices: a, b, c, d.

1. (6 points) If a student who is unprepared for the exam simply guesses the answer to each question, independently from the others, what is the probability that s/he will get
   
i) exactly 5 questions correct?
   \[
   \binom{20}{5} (0.25)^5 (0.75)^{15} = 0.202
   \]

   ii) at least 2 questions correct?
   \[
   1 - p(0) - p(1) = 1 - \binom{20}{0} (0.25)^0 (0.75)^{20} - \binom{20}{1} (0.25)^1 (0.75)^{19}
   = 1 - 0.0032 - 0.021 = 0.976
   \]

2. (5 points) Consider another student who is prepared for the exam and can answer questions correctly 80% of the time. (You must use the table in this part)
   
i. What is the probability that s/he will answer more than 16 questions correctly in this exam?
   \[
   1 - F(16) = 1 - 0.589 = 0.411
   \]

   ii. What is the probability that s/he will answer between 13 and 17 questions (inclusive) correctly in this exam?
   \[
   F(17) - F(12) = 0.794 - 0.032 = 0.762
   \]

3. (5 points) Find the mean and variance of the questions answered correctly by the student in part 2 above.
   \[
   \mu(X) = np = 20 \times (0.8) = 16
   \]
   \[
   \sigma^2(X) = npq = 20 \times (0.8)(0.2) = 3.2
   \]

4. (4 points) Find the expected value of the total number of questions answered by 6 independent students, if 2 of them are unprepared and guess the answers, 3 of them are 80% prepared and and 1 of them is 90% prepared.
   \[
   E(X_1 + X_2 + X_3 + X_4 + X_5 + X_6) = \frac{2 \times (20)(0.25)}{p=0.25} + \frac{3 \times (20)(0.80)}{p=0.80} + (20)(0.90)
   = 7.6
   \]
Question 4: (15 Points)

6 data points are given as follows:

<table>
<thead>
<tr>
<th>$x$</th>
<th>12</th>
<th>8</th>
<th>14</th>
<th>11</th>
<th>16</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>45</td>
<td>56</td>
<td>41</td>
<td>47</td>
<td>39</td>
<td>?</td>
</tr>
</tbody>
</table>

$\sum x_i^2 = 1105$ and $\bar{x} = 13.1667$.

The fitted least squares line (regression line) of the data set is $\hat{y} = 71.015 - 2.077x$.

(a) (3 Points) Find $S_{xx}$.

$$S_{xx} = \sum x_i^2 - \left(\frac{\sum x_i}{n}\right)^2$$

$$= 1105 - \left(\frac{79}{6}\right)^2 = 64.83$$

(b) (4 Points) Find $S_{xy}$.

$$\hat{b}_1 = \frac{S_{xy}}{S_{xx}} \Rightarrow -2.077 = \frac{S_{xy}}{64.83}$$

$$\Rightarrow S_{xy} = -134.66$$

(c) (3 Points) Find $\bar{y}$.

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} \Rightarrow \bar{y} = 71.015 + (-2.077) \left(\frac{79}{6}\right)$$

$$= 43.67$$

(d) (2 Points) Find the predicted $y$-value for $x = 10$

$$\hat{y} = 71.015 - 2.077(10) = 50.245$$

(e) (3 Points) Why is the regression line also called the least squares fitted line?

Because the sum of squares of the (vertical) errors is minimized by this line.
Question 5 (30 Points):

A bag has one white, two blue, and three red balls. Two balls are drawn in succession, at random, without replacement, from the bag. The random variable $X$ represents the number of blue balls among the two drawn balls, and the random variable $Y$ represents the number of red balls in the second draw. The joint probability distribution of $X$ and $Y$ is given below:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$\frac{1}{15}$</td>
<td>$\frac{1}{15}$</td>
<td>$\frac{1}{15}$</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{15}$</td>
<td>$\frac{3}{10}$</td>
<td>$\frac{3}{5}$</td>
</tr>
</tbody>
</table>

$p(X=1, Y=0) = p(\text{one B, 2nd not R})$

$A = p(WB) + p(BW) + p(RB)$

$A = \frac{1}{6} \cdot \frac{2}{5} + \frac{1}{3} \cdot \frac{1}{5} + \frac{1}{2} \cdot \frac{2}{5}$

$A = \frac{1}{15} + \frac{1}{15} + \frac{1}{5} = \frac{1}{3}$

a) (5 Points) Show that the number $A$ is $\frac{1}{3}$ by calculating this probability using the probability tree.

b) (2 Points) Show that the number $A$ is $\frac{1}{3}$ by using a property of the joint probability distribution.

c) (4 Points) Find the marginal probabilities.

d) (5 Points) Find $E(X)$, and the variance of $Y$, $\text{Var}(Y)$.

e) (2 Points) Are $X$ and $Y$ independent? Explain!

f) (4 Points) Find the variance of $2X + 3Y$, $\text{Var}(2X+3Y)$.

g) (5 Points) Find $\text{Cov}(X,Y)$.

h) (3 Points) At least how many times must the experiment of drawing two balls be done so that two blue balls will appear at least once with probability equal or greater than 0.9?
e) \(X\) and \(Y\) are not independent because \((4/10)(1/2) = P(X=0)P(Y=0) \neq P(X=0,Y=0) = 1/10\) for example.

f) \(\text{Var}(2X + 3Y) = 4 \text{Var}(X) + 9 \text{Var}(Y) + 2 (2)(3) \text{Cov}(X,Y)\)

\[
= 4 (1^2(8/15) + 2^2(1/15) - (2/3)^2) + 9 (1/4) + 12 (-2/15) \quad \text{(See g) for Cov)}
\]

\[
= 2.07
\]

\(g) \quad \text{Cov}(X,Y) = (1.1.(1/5) + 1.2.(0)) - (2/3)(1/2) = 1/5 - 1/3 = -2/15\)

h) \(P(\text{two Blue balls will appear in the first trial of this experiment}) = (1/3)(1/5) = 1/15\)

\(P(\text{they will not appear in the first trial – but later}) = 14/15\)

\(P(\text{they will not appear in the first 2 trials – but later}) = (14/15)^2 = 0.87\)

\(P(\text{they will appear in the first or second trial}) = 1 - (14/15)^2 = 0.13\)

Similarly,

\(P(\text{they will not appear in the first } k \text{ trials – but later}) = (14/15)^k\)

\(P(\text{they will appear at least once before the } k^{th} \text{ trial}) = 1 - (14/15)^k > 0.9\)

That is, \((14/15)^k < 0.1\). We find \(k > 33.4\). We can select at least \(k = 34\).