1. (35 p) A random sample of the miles driven by 20 rental car customers is shown as follows:

<table>
<thead>
<tr>
<th>Miles Driven</th>
<th>90</th>
<th>85</th>
<th>100</th>
<th>125</th>
<th>75</th>
<th>50</th>
<th>100</th>
<th>75</th>
<th>60</th>
<th>35</th>
<th>90</th>
<th>100</th>
<th>125</th>
<th>75</th>
<th>85</th>
<th>50</th>
<th>100</th>
<th>50</th>
<th>80</th>
</tr>
</thead>
</table>

a) Compute the mean, median and mode for these data.
b) Indicate whether the data are skewed or symmetrical.
c) Develop a box and whisker plot for these data.

2. (35 p) Suppose a quality manager for Dell computers has collected the following data on the quality status of disk drives by supplier. Quality manager inspected a total of 700 disk drives as given in the table below:

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Driving Status</th>
<th>Working</th>
<th>Defective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company A</td>
<td></td>
<td>120</td>
<td>10</td>
</tr>
<tr>
<td>Company B</td>
<td></td>
<td>180</td>
<td>15</td>
</tr>
<tr>
<td>Company C</td>
<td></td>
<td>50</td>
<td>5</td>
</tr>
<tr>
<td>Company D</td>
<td></td>
<td>300</td>
<td>20</td>
</tr>
</tbody>
</table>

a) Based on these inspection data, what is the probability of randomly selecting a disk drive from Company B?
b) What is the probability of a defective disk drive being received by Dell computers?
c) What is the probability of a defect given that Company B supplied the disk drive?

3. (35 p) The owner of an electronics repair business has read an advertisement from a local competitor that guarantees all high-definition television (HDTV) repairs within four days. Based on his company’s past experience, he wants to know if he can offer a similar guarantee. His past service records are used to determine the following probability distribution

<table>
<thead>
<tr>
<th>Number of Days</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>0.30</td>
</tr>
<tr>
<td>4</td>
<td>0.18</td>
</tr>
<tr>
<td>5</td>
<td>0.12</td>
</tr>
</tbody>
</table>

a) Calculate the mean number of days his customers wait for an HDTV repair.
b) Calculate the variance and standard deviation.
c) Based on the calculations in parts a and b, what conclusion should the owner reach regarding his company’s repair times?

Formulas:

\[ E(x) = \sum xP(x) \]
\[ \sigma^2 = \sum [x - E(x)]^2 P(x) \]

Lower limit = \( Q_1 - 1.5(Q_3-Q_1) \); Upper Limit = \( Q_3 + 1.5(Q_3-Q_1) \)
STAT 203 STATISTICS I FALL 11-12 MIDTERM EXAM SOLUTIONS

1. a) The sample mean is computed using the following steps:
   Step 1: Collect the sample data.
   Step 2: Add the values in the sample:
   \[ \sum x = 1700 \]
   Step 3: Divide the sum by the sample size.
   \[ x = \frac{1700}{20} = 85 \]
   
   Median location: Step 1: Sort the data from low to high.
   The sorted data are shown below:
   
   | 35 | 50 | 50 | 50 |
   | 60 | 75 | 75 | 75 |
   | 80 | 85 | 85 | 90 |
   | 90 | 100| 100| 100|
   | 100| 125| 125| 150|
   
   The median location is \((50/100)*20 = (1/2)*20 = 10\). Because the location point is an integer the median is the average of the values in the 10th and 11th location.
   Median = \((85 + 85)/2 = 85\).  
   Mode = 100 (occurs 4 times which is the most of any value in the sample). 
   Mean = Median = 85 (The data are symmetrical)

c) Step 1: Sort data from low to high (sorted above for median location)

Step 2: Calculate the 25th percentile \((Q_1)\), the 50th percentile (median), and the 75th percentile \((Q_3)\).
   The 25th percentile location is \((25/100)*20 = 5\). So \(Q_1\) is the average of the values in the 5th and 6th position of the sorted array. \(Q_1 = (60+75)/2 = 67.5\).
   The 50th percentile (median) = 85 (found in part b).
   The 75th percentile location is \((75/100)*20 = 15\). So \(Q_3\) is the average of the values in the 15th and 16th position of the sorted array. \(Q_3 = (100 + 100)/2 = 100\).

Step 3: Draw the box so the ends correspond to \(Q_1\) and \(Q_3\).
   Step 4: Draw a vertical line through the box at the median.

Step 5: Compute the upper and lower limits:
   Lower limit = \(Q_1 - 1.5(Q_3-Q_1) = 67.5 - 1.5*32.5 = 18.75\)
   Upper Limit = \(Q_3 + 1.5 (Q_3-Q_1) = 100 + 1.5*32.5 = 148.75\)

Any value outside these limits will be labeled an outlier.

Step 6: Draw the whiskers.

Step 7: Plot the outliers. Outliers are typically indicated by an asterisk, *.
   The box and whisker plot is shown below.

---

Prof. Dr. Hüseyin Oğuz

Page 2 of 4

STAT203_MTEXam_18-11-11
2. a. The probability of a disk drive coming from company B can be found using the relative frequency assessment method as:

\[ P(B) = \frac{\text{number of drives from B}}{\text{total drives}} = \frac{195}{700} = 0.28 \]

b. The probability of a defective disk drive is:

\[ P(\text{Defect}) = \frac{\text{number of defective drives}}{\text{total drives}} = \frac{50}{700} = 0.07 \]

c. The probability of a defect given that company B supplied the disk drive is found using the following steps. The key word here is given which means that we are dealing with conditional probability.

Step 1: Define the experiment.
A quality manager for a Dell Computers has collected the following data on the quality status of disk drives by supplier.

Step 2: Define the events of interest
The two events of interest are:

\[ E_1 = \text{Company B} \]
\[ E_2 = \text{Defective Drive} \]

Step 3: Define the probability statement of interest
We are interested in the following:

\[ P(\text{Defect} \mid B) = \text{probability of a defective drive given company B} \]

Step 4: Convert the data to probabilities using the relative frequency assessment method

\[ P(B) = \frac{\text{number of drives from B}}{\text{total drives}} = \frac{195}{700} = 0.28 \]
\[ P(\text{Defect}) = \frac{\text{number of defective drives}}{\text{total drives}} = \frac{50}{700} = 0.07 \]
\[ P(\text{Defect and B}) = \frac{\text{number of defective drives from B}}{\text{total drives}} = \frac{15}{700} = 0.021 \]

Step 5: Use the rule for conditional probability

\[ P(\text{Defect} \mid B) = \frac{P(\text{Defect and B})}{P(B)} = \frac{0.02}{0.28} = 0.076 \]

Note, you can also find the conditional probability from the data table by

\[ P(\text{Defect} \mid B) = \frac{\text{number of defective drives from B}}{\text{number of drives from B}} = \frac{15}{195} = 0.076 \]

3. a. The expected value of a discrete probability distribution is determined using:

\[ E(x) = \sum xP(x) \]

The calculations are shown as follows:

<table>
<thead>
<tr>
<th>x</th>
<th>P(x)</th>
<th>xP(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td>4</td>
<td>0.18</td>
<td>0.72</td>
</tr>
<tr>
<td>5</td>
<td>0.12</td>
<td>0.6</td>
</tr>
</tbody>
</table>

\[ \text{Sum} = 2.87 \]

Thus, the expected number of days that will be required to repair HDTVs is 2.87 days.

b. The variance for a discrete random variable is computed using the following equation:

\[ \sigma^2 = \sum [x - E(x)]^2 P(x) \]
The calculations are shown as follows:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(x)$</th>
<th>$xP(x)$</th>
<th>$[x - E(x)]$</th>
<th>$[x-E(x)]^2$</th>
<th>$[x-E(x)]^2P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.15</td>
<td>0.15</td>
<td>-1.87</td>
<td>3.4969</td>
<td>0.5245</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.5</td>
<td>-0.87</td>
<td>0.7569</td>
<td>0.1892</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td>0.9</td>
<td>0.13</td>
<td>0.0169</td>
<td>0.0051</td>
</tr>
<tr>
<td>4</td>
<td>0.18</td>
<td>0.72</td>
<td>1.13</td>
<td>1.2769</td>
<td>0.2298</td>
</tr>
<tr>
<td>5</td>
<td>0.12</td>
<td>0.6</td>
<td>2.13</td>
<td>4.5369</td>
<td>0.5444</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Sum = 2.87</strong></td>
</tr>
</tbody>
</table>

The variance is 1.4931 days squared. The standard deviation is the square root of the variance as follows:

$$\sigma = \sqrt{\sum [x - E(x)]^2 P(x)}$$

$$\sigma = \sqrt{1.4931} = 1.22$$

The standard deviation is 1.22 days.

c. Based on the probability distribution, the chances of exceeding 4 days for repair time is 0.12. Thus, not all repair times can be expected to be done in 4 days or less. However, using the Empirical rule from Chapter 3, if the distribution is approximately bell shaped, about 68% of the occurrences will be within one standard deviation of the mean so the manager can expect about two-thirds of the time to repairs within the range:

$$2.87 \pm 1.22$$

1.65 days to 4.09 days.