ANSWERS TO QUIZ #2

1. If the data comprise a sample, the sample mean is given by

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

\[
= \frac{2 + 6 + 23 + 29 + 33 + 37 + 40 + 40 + 45 + 10,001}{10}
\]

\[
= \frac{10,256}{10}
\]

\[
= 1,025.6
\]

<table>
<thead>
<tr>
<th>(x_i)</th>
<th>(x_i - \bar{x})</th>
<th>((x_i - \bar{x})^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-1,023.6</td>
<td>1,047,757</td>
</tr>
<tr>
<td>6</td>
<td>-1,019.6</td>
<td>1,039,584</td>
</tr>
<tr>
<td>23</td>
<td>-1,002.6</td>
<td>1,005,207</td>
</tr>
<tr>
<td>29</td>
<td>-996.6</td>
<td>993,211.6</td>
</tr>
<tr>
<td>33</td>
<td>-992.6</td>
<td>985,254.8</td>
</tr>
<tr>
<td>37</td>
<td>-988.6</td>
<td>977,330</td>
</tr>
<tr>
<td>40</td>
<td>-985.6</td>
<td>971,407.4</td>
</tr>
<tr>
<td>40</td>
<td>-985.6</td>
<td>971,407.4</td>
</tr>
<tr>
<td>45</td>
<td>-980.6</td>
<td>961,576.4</td>
</tr>
<tr>
<td>10,001</td>
<td>8,975.4</td>
<td>80,557,805</td>
</tr>
<tr>
<td>Total</td>
<td>10,256</td>
<td>0</td>
</tr>
</tbody>
</table>

The sample variance is given by

\[
s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2
\]

\[
= \frac{89,510,540}{9}
\]

\[
= 9,945,615.556
\]

Thus, the sample standard deviation is given by

\[
s = \sqrt{9,945,615.556}
\]

\[
= 3,153.667001
\]

Since the number of data points is even \((n = 10)\), the median is the average of the two middle values when the data points are ordered from lowest to highest. Thus, the median is given by

\[
\text{median} = \frac{\text{fifth value} + \text{sixth value}}{2}
\]
\[ \frac{33 + 37}{2} = 35 \]

The coefficient of variation (c.v.) is calculated as

\[
\text{c.v.} = \frac{s \times 100}{x} = \frac{3,153.667001}{1,025.6} \times 100 = 307.4948
\]

Since mean > median, the distribution is most likely skewed to the right (positively skewed).

To determine whether we have outliers, we need first to determine the first and third quartiles (Q1 and Q3, respectively) to get the interquartile range (IQR).

Q1 (or the 25th percentile)

\[
L = 10 \times \frac{25}{100} = 2.5 \Rightarrow 3\text{rd value} = 23
\]

Q3 (or the 75th percentile)

\[
L = 10 \times \frac{75}{100} = 7.5 \Rightarrow 8\text{th value} = 40
\]

Thus,

\[
\text{IQR} = Q3 - Q1 = 40 - 23 = 17
\]

The outlier detection limits are given by

(upper limit): \( Q3 + 1.5 \times \text{IQR} \)

\[
\Rightarrow 40 + 1.5 \times 17 = 65.5
\]

(lower limit): \( Q1 - 1.5 \times \text{IQR} \)

\[
\Rightarrow 23 - 1.5 \times 17 = -2.5
\]

We have an outlier if there is a value that is either greater than the upper outlier detection limit or less than the lower outlier detection limit. Since 10,001 > 65.5 (the upper outlier detection limit), then 10,001 is an outlier.
2.

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>Class Mark</th>
</tr>
</thead>
<tbody>
<tr>
<td>−10 − 0 (dummy)</td>
<td>0</td>
<td>−5</td>
</tr>
<tr>
<td>0 − 10</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>10 − 20</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>20 − 30</td>
<td>21</td>
<td>25</td>
</tr>
<tr>
<td>30 − 40</td>
<td>9</td>
<td>35</td>
</tr>
<tr>
<td>40 − 50</td>
<td>4</td>
<td>45</td>
</tr>
<tr>
<td>50 – 60 (dummy)</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>50</strong></td>
<td></td>
</tr>
</tbody>
</table>

The mode is the class mark or midpoint of the class with the highest frequency; thus, the mode of the above grouped data is 25.