Example 10

Data is available for sales (number) over 70 weeks of two products manufactured by a company. The data are analysed with a view to identifying differences in sales. To this end histograms and summary statistics are produced in Excel. Use these to comment on sales of the two products. Is there a better graphical representation of data that makes comparison easier?

(a) Histograms.

(b) Summary statistics.

<table>
<thead>
<tr>
<th></th>
<th>Product 1</th>
<th>Product 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>446</td>
<td>Mean</td>
</tr>
<tr>
<td>Standard Error</td>
<td>7</td>
<td>Standard Error</td>
</tr>
<tr>
<td>Median</td>
<td>438</td>
<td>Median</td>
</tr>
<tr>
<td>Mode</td>
<td>#N/A</td>
<td>Mode</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>57</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>3214</td>
<td>Sample Variance</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0</td>
<td>Kurtosis</td>
</tr>
<tr>
<td>Skewness</td>
<td>0</td>
<td>Skewness</td>
</tr>
<tr>
<td>Range</td>
<td>248</td>
<td>Range</td>
</tr>
<tr>
<td>Minimum</td>
<td>321</td>
<td>Minimum</td>
</tr>
<tr>
<td>Maximum</td>
<td>569</td>
<td>Maximum</td>
</tr>
<tr>
<td>Sum</td>
<td>31196</td>
<td>Sum</td>
</tr>
<tr>
<td>Count</td>
<td>70</td>
<td>Count</td>
</tr>
</tbody>
</table>
Analysis

General

Product 1

Histogram shows a fairly symmetric distribution. Summary statistics show mean sales of 446 per week, standard deviation of 57 and a range of 248. Minimum sales are 321 per week and maximum 569 per week.

Product 1

Histogram shows a slightly negatively skewed distribution. The mean sales are 551 per week, standard deviation 27 and a range of 120. The minimum sales are 481 per week and the maximum 601 per week.

Conclusion

Product 2 has higher mean sales per week and less variability.

Improvements to Histogram

It would be better to plot the histograms on a common scale so that comparison is easier.

Exercise

How do you Excel to plot histograms on a common scale?
Example 11

A stem-and-leaf plot for data on the number of hours spent on studies per fortnight be a random sample of students is given below.

\[ N = 36 \]

Stem uni is Tens, leaf unit is Units. (Thus the minimum data value is 2 and the maximum is 150.)

0 : 2223457
1 : 0124558
2 : 0135779
3 : 00188
4 : 01
5 :
6 : 05
7 : 0
8 : 0
9 : 00
10 : 0
11 :
12 :
13 :
14 :
15 : 0

(i) Find the median, the lower quartile and the upper quartile.

**Solution**

\[ \text{Median} = (25+27)/2 = 26 \]
\[ \text{LQ} = (11+12)/2 = 11.5, \quad \text{UQ} = (40+41)/2 = 41.5 \]

(ii) Comment on the distribution of the data.

**Solution**

Most of the data lies between 2 and 41, but there is a very long right tail. The data indicates the presence of two groups of students – the majority spend less than 40 hours a fortnight, but about a fifth spend between 60 and 100 hours. There is an outlier at 150 hours.
Example 12

The mean and standard deviation of the annual cost of exports to the US from Australia from 1999-2003 (inclusive) is $US 6.22 million and $US 0.52 million respectively. Find these quantities in $AUD. The exchange rate (at 10:38 am on Wednesday 25 July 2007) is $US 1 = $AUD 1.13595.

Solution

Let x and y denote the cost of exports in $US and $AUD respectively. Then

\[ y = 1.27x \]

so

\[ \bar{y} = 1.27\bar{x} = 1.13595(6.22) = $AUD7.0656 million \]

\[ s_y = 1.27(s_x) = 1.13595(0.52) = $AUD0.5907 million \]