1G – The five-number summary and the boxplot

**THE FIVE-NUMBER SUMMARY**

The five-number summary is a list of the following values for a data set:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**THE BOX PLOT**

* The boxplot is a graphical display of a five number summary
* 25% of the data values are located in each of the four sections of the boxplot
* 50% of the data values are located in the interquartile range (Q1 to Q3)



Write down the five-number summary and create a boxplot for the following data:

*The distribution of life expectancies (in years) for 23 countries is shown below:*

54 55 56 64 66 66 67 69 71 72 72 73 73 74 74 74 74 75 75 76 76 77 77

Median: n = \_\_\_\_ = \_\_\_\_\_ Q1 and Q3: n = \_\_\_\_ = \_\_\_\_\_

Five-number summary: \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Boxplot:

*Now check your boxplot using both your calculator and Mathematica*

1G – The five-number summary and the boxplot

**USING A BOXPLOT TO DISPLAY OUTLIERS**

* An OUTLIER is a data point that is \_\_\_\_\_\_\_\_\_\_ than the upper fence or \_\_\_\_\_\_ than the lower fence:

Upper fence = \_\_\_\_\_\_ + 1.5 × \_\_\_\_\_\_

Lower fence = \_\_\_\_\_\_ – 1.5 × \_\_\_\_\_\_

* Outliers are represented on boxplots as \_\_\_\_\_\_ that are \_\_\_\_\_\_\_\_\_\_\_\_\_ from the box and whiskers

Write down the five-number summary, determine whether there are outliers, and create a boxplot for the following data:

*19 marks from an assessment task are recorded below:*

28 21 21 3 22 31 35 26 27 33 43 31 30 34 48 36 35 23 24

Median: n = \_\_\_\_ = \_\_\_\_\_ Q1 and Q3: n = \_\_\_\_ = \_\_\_\_\_

Five-number summary: \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |

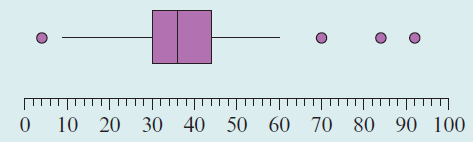
Boxplot:

*Now check your boxplot using both your calculator and Mathematica*

1G – The five-number summary and the boxplot

**INTERPRETING BOXPLOTS**

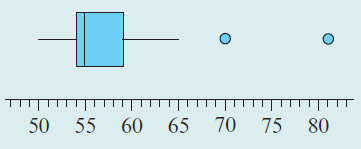
1. *Analyse the boxplot below and write down the specified values:*



Median: \_\_\_\_\_\_ Q1: \_\_\_\_\_\_ Q3: \_\_\_\_\_\_ IQR: \_\_\_\_\_\_ Fences: \_\_\_\_\_\_ , \_\_\_\_\_\_

Minimum: \_\_\_\_\_\_ Maximum: \_\_\_\_\_\_ Outliers: \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_

1. *Analyse the boxplot below to estimate the specified percentages:*



Less than 54: \_\_\_\_\_\_ Less than 55: \_\_\_\_\_\_ Less than 59: \_\_\_\_\_\_ Greater than 59: \_\_\_\_\_\_

Between 54 and 59: \_\_\_\_\_\_ Between 54 and 86: \_\_\_\_\_\_

1G – The five-number summary and the boxplot

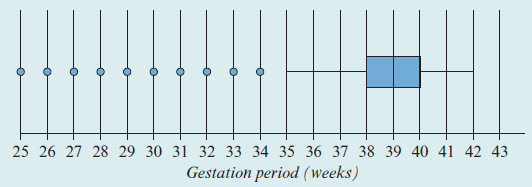
**DESCRIBING THE SHAPE OF DISTRIBUTIONS**

|  |  |
| --- | --- |
|  |  |
|  |  |

**WRITING A REPORT TO DESCRIBE TO DESCRIBE A DISTRIBUTION**

When writing a report, specify **shape, centre, spread and outliers**:

*Use the boxplot below to describe the distribution of gestation period:*



The distribution of gestation period is slightly \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_ with several

\_\_\_\_\_\_\_\_\_\_\_\_\_.

The distribution is centred at 39 weeks, the \_\_\_\_\_\_\_\_\_\_\_\_\_ value.

The range of the distribution is \_\_\_\_\_\_ weeks, but the interquartile range is only \_\_\_\_\_\_ weeks.

Any gestation period less than \_\_\_\_\_\_ weeks is considered unusual, with outliers at \_\_\_\_\_\_ ,

\_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , \_\_\_\_\_\_ , and \_\_\_\_\_\_ weeks.