Minitab

ANSWER KEY: FIVE-NUMBER SUMMARY AND BOXPLOTS
This answer key provides solutions to the corresponding student activity sheet.

Five-Number Summary and Boxplots

The data for these exercises are in the Minitab file ***Boxplot_Activity.mtw***.

Exercise 1

(a) By hand, calculate the five-number summary statistics (minimum, maximum, first quartile (Q_1), second quartile (Q_2), and third quartile (Q_3)) for this data.

Solution: The five-number summary for the data is:

- **Minimum** = **81 minutes**
- **Maximum** = **136 minutes**
- Since n = 22, then Q_2 is in position (22 + 1)/2 = 11.5 in the data sorted from minimum to maximum. Since the 11th and 12th data points are both 116, then $\boldsymbol{Q_2}$ **= 116 minutes**.
- Since n = 22, then Q_1 is in position (22 + 1)/4 = 5.75 in the data sorted from minimum to maximum. The position 5.75 means that Q_1 is the value that is 0.75 of the way between the 5th and 6th data points in sorted order from minimum to maximum. The value of $\boldsymbol{Q_1}$ is 105 + 0.75 * (108 – 105) = **107.25 minutes**.
- Since n = 22, then Q_3 is in position 3 * (22 + 1)/4 = 17.25 in the data sorted from minimum to maximum. The position 17.25 means that Q_3 is the value that is 0.25 of the way between the 17th and 18th data points in sorted order from minimum to maximum. The value of $\boldsymbol{Q_3}$ is 120 + 0.25 * (126 – 120) = **121.5 minutes**.

(b) By hand, calculate the interquartile range (IQR) for this data.

Solution: **IQR** = $Q_3 - Q_1$ = 121.5 – 107.25 = **14.25**.

(c) By hand, calculate the lower and upper fences for a boxplot of this data, using the IQR from part **(b)**.

Solution: The **lower fence** is 107.25 – 1.5 * 14.25 = **85.875 minutes**. The upper fence is 121.5 + 1.5 * 14.25 = **142.875 minutes**.

(d) Use the fences from part **(c)** to determine the data values for the lower whisker endpoint, upper whisker endpoint, and outliers (if any).

Solution: There is one **outlier**, **81 minutes**, since it's beyond the lower fence at 85.875 minutes. The **lower whisker** is the smallest data value within the lower fence, and it is **101 minutes**. The **upper whisker** is the largest data value within the upper fence, and it is **136 minutes.**

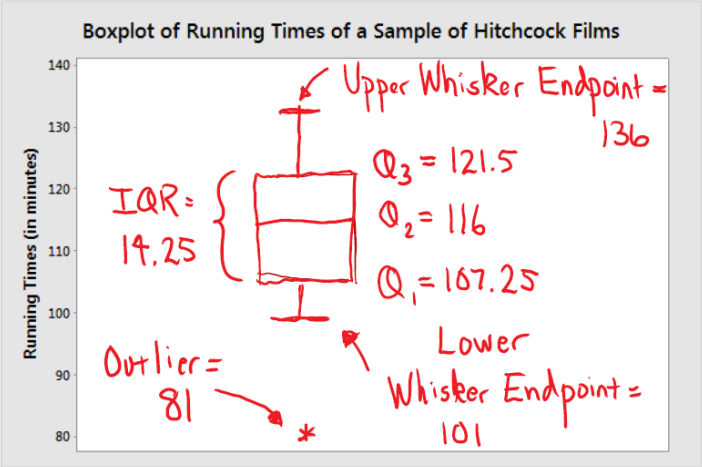
(e) Verify the statistics you calculated in parts **(a)** and **(b)** in Minitab.

Solution: The Minitab output is:

Descriptive Statistics: Running Times

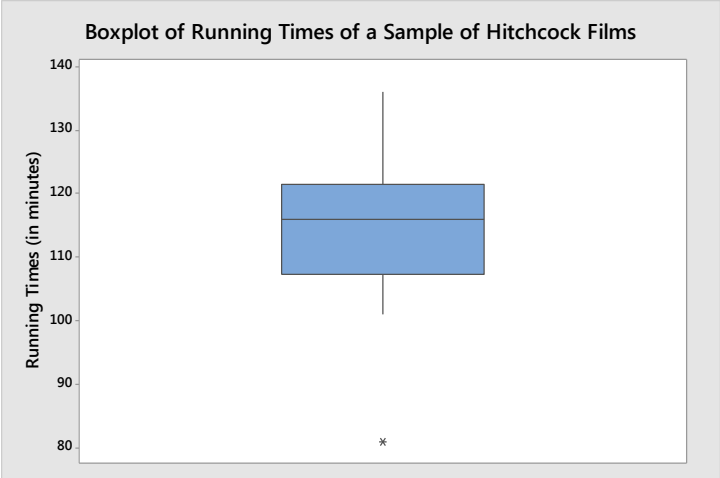
Variable	Minimum	Q1	Median	Q3	Maximum	IQR
Running Times	81.00	107.25	116.00	121.50	136.00	14.25

(f) By hand, construct a boxplot using the statistics from parts **(a) – (d)**.



(g) Verify the boxplot in part **(f)** with the boxplot that Minitab produces.

Solution: Minitab produces the following boxplot. Minitab does not put the fences in the plot.



Exercise 2

Which of the following statements are true?

I. The boxplot is positively skewed.
II. The interquartile range is about 8.
III. The median is about 10.

(A) I only **(B) II only** (C) III only (D) I and III (E) II and III

Solution: The median is closer to Q_3 than Q_1 and the lower tail is longer than the upper tail, so the boxplot is negatively skewed. The interquartile range is indicated by the length of the box, which is 8. The median is indicated by the vertical line running through the box, which is 15.

Exercise 3

Which of the following statements are true?

I. The boxplot is positively skewed.
II. The upper fence must be above 20.
III. The median is less than 5.
IV. There are 3 outliers at approximately 25, 38, and 45.
V. The horizontal axis indicates that approximately 50 students took the quiz in class.

(A) I, III, IV only (B) I, II, III, IV only (C) I, III, IV, V only (D) All of them (E) III, IV only

Solution: The median is closer to Q_1 than Q_3 and the upper tail is longer than the lower tail (whisker and outliers), so the boxplot is positively skewed. The upper fence could be any number between the upper whisker endpoint, at 6, and the first outlier, at 25. Using the approximate values for Q_3 (~3) and the IQR (~2 or 3), we can estimate the upper fence to be between about 6 and 7.5. The horizontal axis indicates the number of correct answers by students. We can't tell from the boxplot alone how many students took the quiz.

Exercise 4

(a) By hand or in Minitab, calculate the five-number summary statistics (minimum, maximum, first quartile (Q_1), second quartile (Q_2), and third quartile (Q_3)) for this data.

Solution: Here are the Minitab calculations:

Descriptive Statistics: Presidents Ages

Variable	Minimum	Q1	Median	Q3	Maximum
Presidents Ages	42.000	51.000	54.500	57.750	69.000

Here are the calculations by hand, which reveal the same values:

- Since $n = 44$, then Q_2 is in position $(44 + 1)/2 = 22.5$ in the data sorted from minimum to maximum. Since the 22nd and 23rd data points are 54 and 55, respectively, then **Q_2 = 54.5 years**.
- Since $n = 44$, then Q_1 is in position $(44 + 1)/4 = 11.25$ in the data sorted from minimum to maximum. Its value is **51 years** since both the 11th and 12th data points are 51.
- Since $n = 44$, then Q_3 is in position $3 * (44 + 1)/4 = 33.75$ in the data sorted from minimum to maximum. The position 33.75 means that Q_3 is the value that is 0.75 of the way between the 33rd and 34th data points in sorted order from minimum to maximum. The value of **Q_3** is $57 + 0.75 * (58 – 57) =$ **57.75 years**.

(b) By hand or in Minitab, calculate the interquartile range (IQR) for this data.

Solution: The **IQR** is $Q_3 – Q_1 = 57.75 – 51 =$ **6.75 years**.

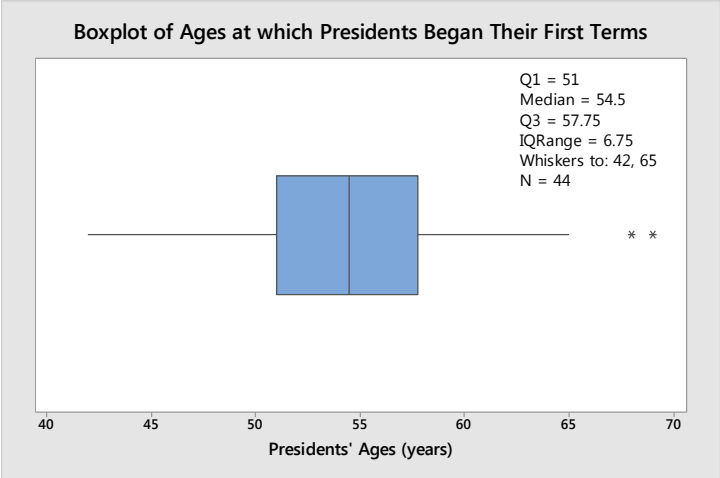
(c) By hand, calculate the lower and upper fences for a boxplot of this data , using the IQR from part **(b)**.

Solution: The **lower fence** is $51 – 1.5 * 6.75 =$ **40.875 years**. The upper fence is $57.75 + 1.5 * 6.75 =$ **67.875 years**.

(d) Use the fences from part **(c)** to determine the data values for the lower whisker endpoint, upper whisker endpoint, and outliers (if any).

Solution: There are two **outliers**, **68** and **69**, since both are beyond the upper fence of 67.875. The **lower whisker** is the smallest data value within the lower fence, and it is **42 years**. The **upper whisker** is the largest data value within the upper fence, and it is **65 years.**

(e) In Minitab, construct a boxplot using the statistics from parts **(a) – (d)**.



Exercise 5

Match each histogram to the boxplot that represents the same data set.

Boxplot A matches histogram **1**.

Boxplot B matches histogram **4**.

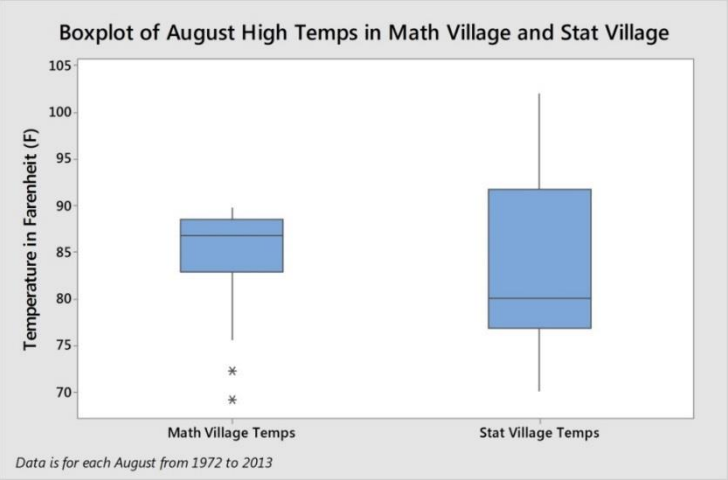
Boxplot C matches histogram **3**.

Boxplot D matches histogram **2**.

Exercise 6

(a) Create comparison boxplots for the highest temperature in Math Village versus Stat Village in August from 1972 to 2013.

Solution: Minitab produces the comparison boxplots below. We have retitled the plot and the axes, and we added a footnote. Editing text in a graph is discussed in the ***Describing Data Graphically*** lesson.



(b) Given the comparison boxplots in part **(a)**, answer the following true/false questions about the data from both villages.

A. The temperatures are more variable for Stat Village than Math Village. **T**

B. The temperatures in Stat Village are positively skewed. **T**

C. Stat Village has a greater median temperature for those 42 years than Math Village. **F**

D. Stat Village has a smaller IQR than Math Village. **F**

E. It is obvious from the boxplots that Stat Village's mean temperature for those 42 years is less than Math Village's temperatures. **F**

F. The lower whisker endpoint for Stat Village is less than the lower whisker endpoint for Math Village. **T**

G. Stat Village's second quartile is less than Math Village's first quartile. **T**

H. If you prefer August high temperatures consistently around 85 degrees Fahrenheit, then you should move to Stat Village. **F**