Minitab

ACTIVITY SHEET: FIVE-NUMBER SUMMARY AND BOXPLOTS
This activity sheet includes exercises to assess students' understanding of important concepts presented in the *Boxplot* lesson.

Five-Number Summary and Boxplots

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

Exercise 1

The following table shows the running times (in minutes) of a sample of videotape versions of $n = 22$ movies directed by Alfred Hitchcock. Movies are listed in alphabetical order. The data are in the Minitab columns "Hitchcock Movies" and "Running Times."

Film	Time	Film	Time
The **B**irds	119	**P**sycho	108
Dial M for Murder	105	**R**ear Window	113
Family Plot	120	**R**ebecca	132
Foreign Correspondent	120	**R**ope	81
Frenzy	116	**S**hadow of a Doubt	108
I Confess	108	**S**pellbound	111
The **M**an Who Knew Too Much	120	**S**trangers on a Train	101
Marnie	130	**T**o Catch a Thief	103
North by Northwest	136	**T**opaz	126
Notorious	103	**U**nder Capricorn	117
The **P**aradise Cane	116	**V**ertigo	128

(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. The data below are ordered from minimum to maximum.

81 101 103 103 105 108 108 108 111 113 116 116 117 119
120 120 120 126 128 130 132 136

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

(c) By hand, calculate the lower and upper fences for a boxplot of this data using the IQR from part **(b)**. Recall that the lower fence is at position Q_1 – 1.5 * IQR with the upper fence at Q_3 + 1.5 * IQR.

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

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

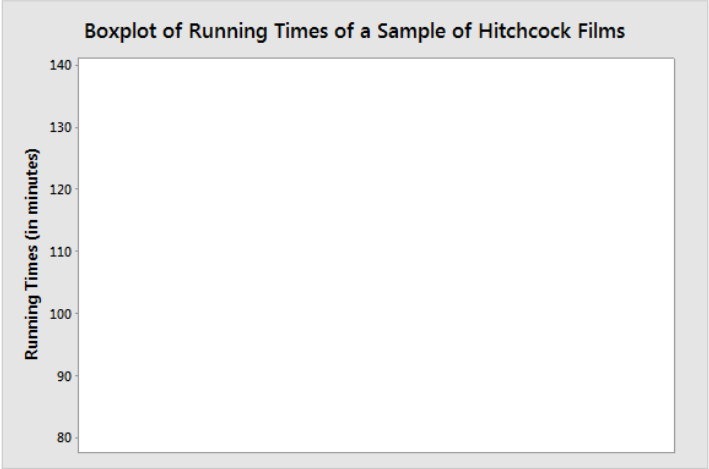
Minitab 17

1 Choose **Stat > Basic Statistics > Display Descriptive Statistics**.
2 In **Variables**, enter *'Running Times.'*
3 Click **Statistics** and check **Minimum**, **Maximum**, **First quartile**, **Median**, **Third quartile**, and **Interquartile range**.
4 Click **OK** in each dialog box.

Minitab Express

1 Open the descriptive statistics dialog box.
   - Mac: **Statistics** > **Summary Statistics** > **Descriptive Statistics**
   - PC: **STATISTICS** > **Descriptive Statistics**
2 In **Variable**, enter *'Running Times.'*
3 Click the **Statistics** tab, and then select **Minimum**, **Maximum**, **First quartile**, **Median**, **Third quartile**, and **Interquartile range**.
4 Click **OK**.

(f) By hand, construct a boxplot using the statistics from parts **(a)** – **(d)**. In the plot, make sure to include and label the following: Q_1, Q_2, Q_3, IQR, lower whisker endpoint, upper whisker endpoint, and outliers (if any).



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

Minitab 17

1. Choose **Graph > Boxplot**.
2. Under **One Y**, choose **Simple**, then click **OK**.
3. Under **Graph variables**, enter '*Running Times*.'
4. Click **OK**.

Minitab Express

1. Open the boxplot of a single y variable dialog box.
    - Mac: **Graphs > Boxplot > Single Y Variable: Simple**
    - PC: **GRAPHS > Boxplot > Single Y Variable: Simple**
2. In **Variable**, enter '*Running Times*.'
3. Click **OK**.

Include a more descriptive title with the boxplot, such as "Boxplot of a Sample of Hitchcock Film Lengths." You can do this by double-clicking on the current title and entering a new title in the **Edit Title** dialog box. Similarly, include the units on the vertical axis.

To check Minitab's values for Q_1, Q_2, Q_3, IQR, lower whisker endpoint, and upper whisker endpoint, hover over the boxplot. If there are any outliers, hover over them to see their values.

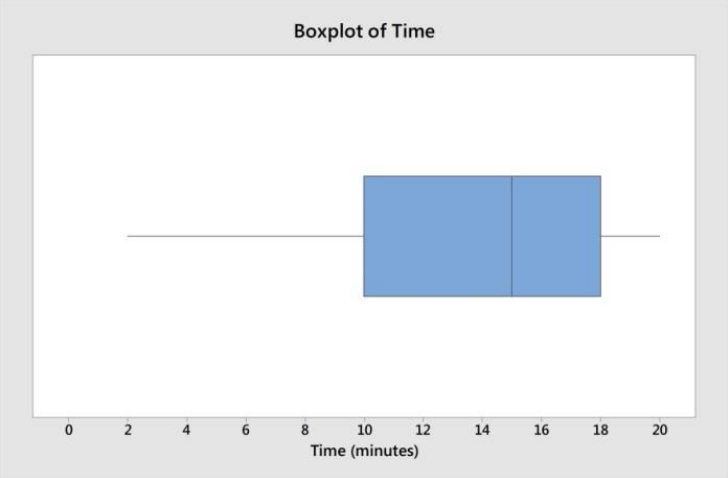
Exercise 2

Consider the boxplot to the right. 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.

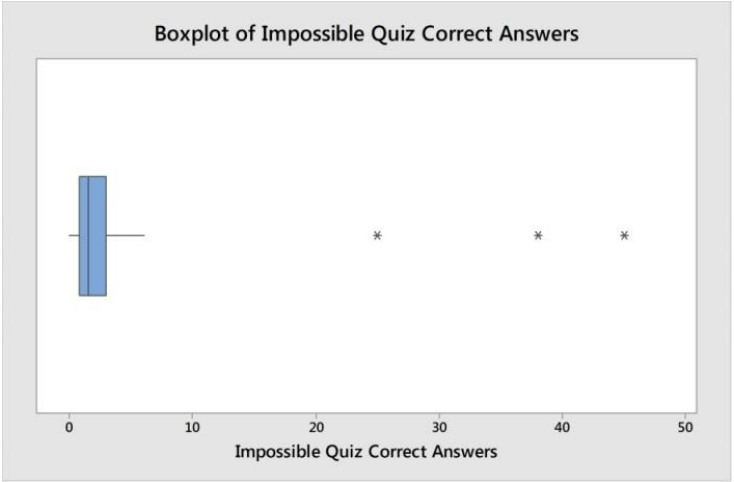


(A) I only (B) II only (C) III only (D) I and III (E) II and III

Exercise 3

Go to the website http://www.notdoppler.com/theimpossiblequiz.php. This is a fun quiz with answers that don't always make sense the first time around. You are given 3 lives at the start of the quiz, but lose a life every time you answer a question incorrectly. If this is the first time you are trying this quiz, don't expect to get many problems correct before losing your 3 lives.

In class, students take the quiz and are asked to record the number of questions they answer correctly before losing their 3 lives. A boxplot of the numbers is shown below.



Three students had already seen this quiz prior to class, which explains the highly unusual outliers.

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

Exercise 4

Below are the ages, ordered from youngest to oldest, at which U.S. presidents began their (non-consecutive) first terms, from George Washington to Barack Obama. The data are in the Minitab column "Presidents Ages."

(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. See **Exercise 1** for instructions on how to obtain these values in Minitab.

42	43	46	46	47	47	48	49	49	50	51	51	51
51	51	52	52	54	54	54	54	54	55	55	55	55
56	56	56	57	57	57	57	58	60	61	61	61	62
64	64	65	68	69								

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

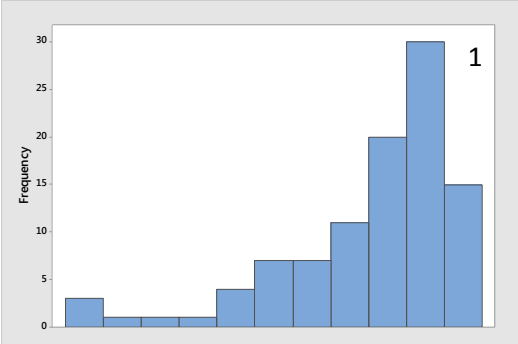
(c) By hand, calculate the lower and upper fences for a boxplot of this data, using the IQR from part **(b)**. Recall that the lower fence is at $Q_1 - 1.5 * IQR$ with the upper fence at $Q_3 + 1.5 * IQR$.

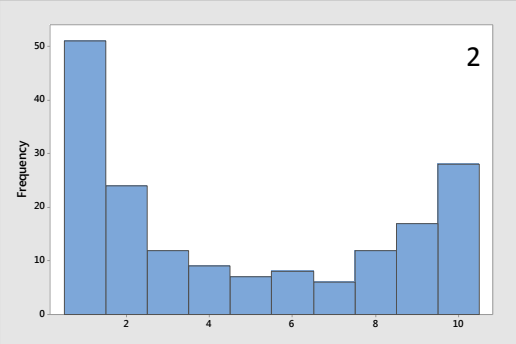
(d) Use the fences from part **(c)** to determine the data values for the lower whisker endpoint, upper whisker endpoint, and outliers (if any). Outliers are points beyond the fences.

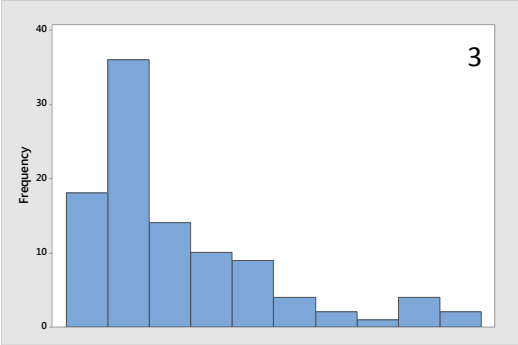
(e) In Minitab, construct a boxplot using the statistics from parts **(a) – (d)**. See **Exercise 1** for instructions on how to obtain this plot in Minitab.

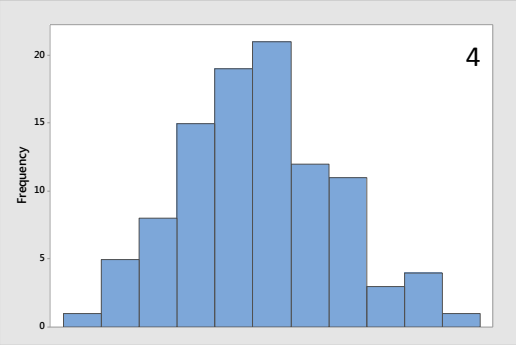
Exercise 5

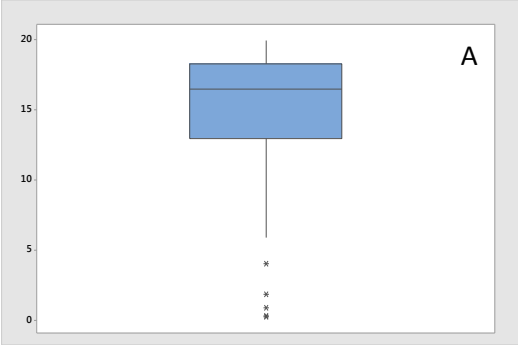
Match each histogram to the boxplot that represents the same data set.

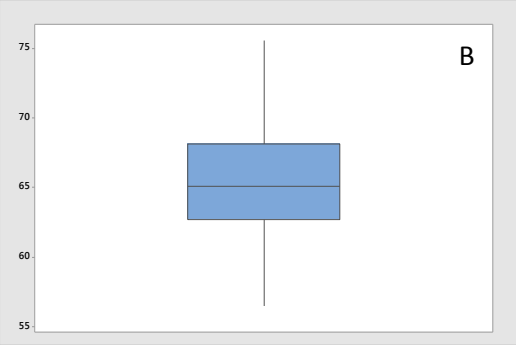


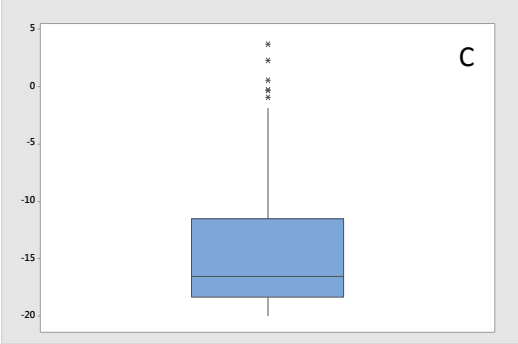


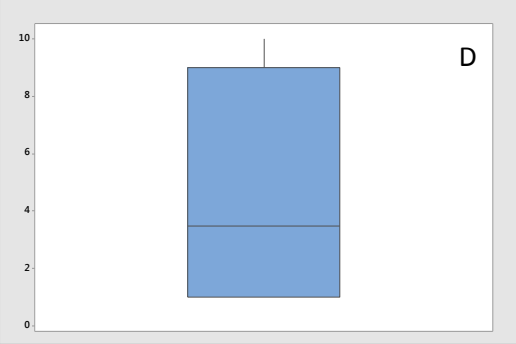












Boxplot A matches histogram _____.

Boxplot B matches histogram _____.

Boxplot C matches histogram _____.

Boxplot D matches histogram _____.

Exercise 6

Temperatures in the cities of Math Village and Stat Village are greatest in the month of August. The highest temperature, in degrees Fahrenheit, in Math Village for each August from 1972 to 2013 is given below. The temperatures are sorted from minimum to maximum over this 42-year period.

69.1	72.3	75.6	77.6	78.1	79.0	79.1	79.7	81.8	82.5	83.1	83.1	83.5	83.8
84.1	84.4	84.6	84.8	85.7	86.5	86.7	86.8	87.3	87.3	87.5	87.7	87.8	88.0
88.2	88.3	88.4	88.5	88.7	89.1	89.2	89.3	89.5	89.5	89.5	89.7	89.8	89.8

The highest temperature, in degrees Fahrenheit, in Stat Village for each August from 1972 to 2013 is given below. The temperatures are sorted from minimum to maximum over this 42-year period.

70.1	70.1	70.1	70.4	71.7	72.4	72.7	72.9	73.7	75.0	77.4	78.2	78.7	78.9
79.2	79.5	79.6	79.8	79.8	80.0	80.1	80.1	80.2	84.8	85.0	86.1	86.4	88.3
89.1	90.4	90.4	91.6	92.2	93.2	94.5	97.7	98.6	98.7	98.7	99.5	100.6	102.0

(a) Create comparison boxplots for the highest temperature in Math Village versus Stat Village in August from 1972 to 2013. Use a meaningful title and correctly label the axes with units.

Minitab 17

1. Choose **Graph > Boxplot.**
2. Under **Multiple Y's**, choose **Simple**, then click **OK**.
3. Under **Graph variables**, enter '*Math Village Temps*' and '*Stat Village Temps*.'
4. Click **OK**.

Minitab Express

1. Open the boxplot of multiple y variables dialog box.
    - Mac: **Graphs > Boxplot > Multiple Y Variables: Simple**
    - PC: **GRAPHS > Boxplot > Multiple Y Variables: Simple**
2. In **Variables**, enter '*Math Village Temps*' and '*Stat Village Temps*.'
3. Click **OK**.

(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 F

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

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

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

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

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

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