Minitab

ANSWER KEY: DESCRIBING DATA *NUMERICALLY*
This answer key provides solutions to the corresponding student activity sheet.

Describing Data *Numerically*

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

Exercise 1

(a) By hand, calculate the **sample mean** $\bar{x}$ of these scores. Write down the formula that you use.

Solution:

$$\bar{x} = \frac{212 + 232 + 250 + 204 + 261}{5} = \frac{1159}{5} = \mathbf{231.8}$$

(b) By hand, calculate the **sample median** *m* of these scores.

Solution: Ordered data: 204, 212, 232, 250, 261; the middle value is **232.**

(c) By hand, calculate the **sample range** *R* of these scores.

Solution: **Range** = maximum – minimum = 261 – 204 = **57**

(d) By hand, calculate the **sum of the *squared* deviations** from the mean of these scores. Write down the formula you are using. Note: If you didn't square the deviations, what would the sum of just the deviations be?

Solution: $(204 - 231.8)^2 + (212 - 231.8)^2 + \cdots + (261 - 231.8)^2 = \mathbf{2348.8\ pins^2}$

If you didn't square each term, the sum would be 0.

(e) Calculate the **sample variance** s^2 of these scores by using the sum of the squared deviations from part **(d)**.

Solution: $s^2 = \frac{2348.8}{4} = \mathbf{587.2\ pins^2}$

(f) Calculate the **sample standard deviation**, *s*, for the scores.

Solution: $s = \sqrt{587.2} \cong 24.2$ **pins**

(g) Verify the statistics you computed in parts **(a)** – **(c)**, **(e)** and **(f)** using Minitab. Put the five bowling scores in a column in Minitab, such as C1, and name it "Bowling Scores."

Solution:

Statistics

Variable	Total Count	Mean	StDev	Variance	Median	Range
Bowling Scores	5	231.8	24.2	587.2	232.0	57.0

(h) Suppose we randomly select another one of his scores from the past month; it is 101. In Minitab, calculate sample statistics for the mean, median, range, variance, and standard deviation using this additional data value. Provide their values below.

Solution: $\bar{x} = 210.0$, $m = 222.0$, R = 160, $s^2 \cong 3321.2$, $s \cong 57.6$

Statistics

Variable	Total Count	Mean	StDev	Variance	Median	Range
Bowling Scores	6	210.0	57.6	3321.2	222.0	160.0

(i) In general, do outliers have a larger impact on the **mean** or **median**?

Solution: On the mean.

(j) Given these 6 data points, what additional 7th data point could you add that would keep:

- The median at the same value?

Solution: The median is the ordered data value, and the ordered data that we currently have is:

101, 204, 212, 232, 250, 261

The current median is the average of the data values 212 and 232. By adding a 7th data point, the median will be the middle ordered data value. Thus, we can only add the data point 222 to keep the median as 222:

101, 204, 212, **222**, 232, 250, 261

- The mean at the same value?

Solution: The mean is the average of the data values, and the current mean is 210. The only way to keep the mean exactly the same is to add the data point **210**.

$$\bar{x} = \frac{101 + 212 + 232 + 250 + 204 + 261}{6} = \frac{1260}{6} = 210$$

$$\bar{x} = \frac{101 + 212 + 232 + 250 + 204 + 261 + \boldsymbol{x_7}}{7} = \frac{1260 + \mathbf{210}}{7} = \frac{1470}{7} = 210$$

Exercise 2

(a) By hand or with Minitab, calculate sample statistics for the mean, median, range, variance, and standard deviation using the reaction time data.

Solution: The sample statistics will depend on the student's data. The mean and median can be no larger than 3, since this is the maximum time that can be recorded for a reaction time.

(b) Look at the mean and the median for your data. What is the relationship between the mean and the median (e.g. is the mean greater than the median)? Provide a brief explanation as to why you think this has happened.

Solution: Again, this answer depends on the student's data. But, if the student had several "penalty" shots that were recorded as 3 second reaction times, then the mean will likely be greater than the median.

Exercise 3

(a) By hand or with Minitab, calculate sample statistics for the mean, median, range, variance, and standard deviation using the running time data.

Solution:

Statistics

Variable	Total Count	Mean	StDev	Variance	Median	Range
Running Times	22	114.59	12.43	154.54	116.00	55.00

(b) The sample range is almost 1 hour. What data point is the main influence behind this large range?

Solution: There is one "unusually" short running time of 81 minutes – the movie "Rope." Since the range is the maximum minus the minimum, then the short running time is contributing to the large range.

(c) When plotted, the movie running times are fairly symmetric. How can you tell this without graphing the running times?

Solution: Since the mean and median are close in value, then a graph (e.g. histogram, boxplot) of the data will produce a fairly symmetric shape. Again, the value 81 minutes is an unusually low running time for the sample of Hitchcock movies.

Exercise 4

(a) Using only the data (without Minitab), what is the median salary of the top 20 player salaries?

Solution: Since there is an even number of data points, then the median is the average of the 10^{th} and 11^{th} points. The average of 19.0 and 19.3 is 19.15. The median salary is **19.15 million dollars**.

(b) What is the mode salary of the top 20 player salaries?

Solution: The most frequent occurring data value in the top 20 player salaries is **15.7 million dollars**.

Exercise 5

(a) By hand or in Minitab, find the mean, median, and mode ages of Presidents when they began their first term in office.

Solution:

Statistics

Variable	Total Count	Mean	Median	Mode	N for Mode
Presidents Ages	44	54.659	54.500	51, 54	5

The mean is approximately 54.7 years and the median is 54.5 years. The data is bimodal since it has two modes: 51 years and 54 years. Five presidents started their presidency at age 51 and five started their presidency at age 54.

(b) Considering the results that you obtained in part **(a)**, is the shape and center of the histogram surprising to you?

Solution: Not at all. The mean and median are almost identical at 54 years, and one of the modes is 54 years. These data support the symmetric shape of the histogram.

(c) Using only the histogram from part **(b)**, can you determine the age of President John F. Kennedy when he first took office? Why or why not?

Solution: Unfortunately, we cannot determine the age of any one of the presidents using the histogram alone. Histograms do not show data in time order. Histograms take the element of order out of the data display. Also, histograms display the data using bins, so we only know the range for a given president's age.