

DATA DISTRIBUTION SHAPES

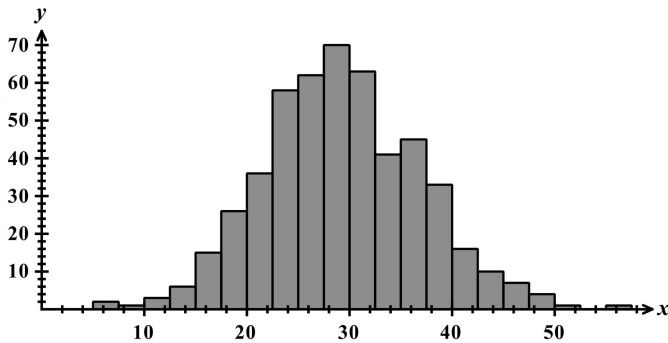
N-GEN MATH[®] ALGEBRA I



When we look at the values of a data set on a dot plot, histogram, or box plot, we are looking at how the **data distribution** is shaped across its range.

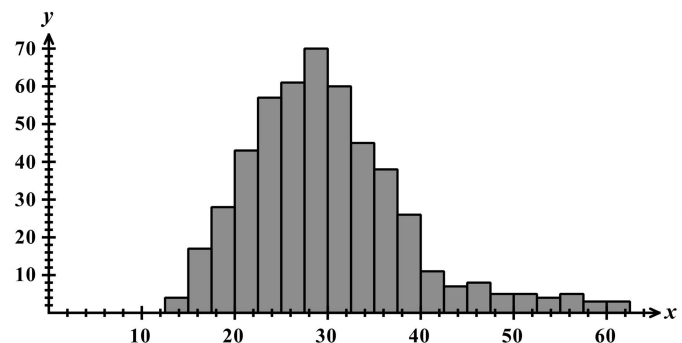
Exercise #1: Two data distributions are shown below that each contain 500 data points.

mean = 29.1, median = 28.9, $s_x = 7.5$, IQR = 10.7



Data Distribution A

mean = 30.1, median = 28.9, $s_x = 9.2$, IQR = 10.6



Data Distribution B

- (a) Although neither are perfectly **symmetric**, which of the two data distributions is more symmetric? Explain your choice.
- (b) What is true about the mean and median of the more symmetric distribution?
- (c) What effect does the **right skew** of the **asymmetric distribution** have on its mean?
- (d) How can you describe the variation within the two data sets?

Very often when we discuss statistics for a data set, we need to pay attention to whether the data set is symmetric or whether it is **skewed** to the right or left.

Exercise #2: Fill in the following.

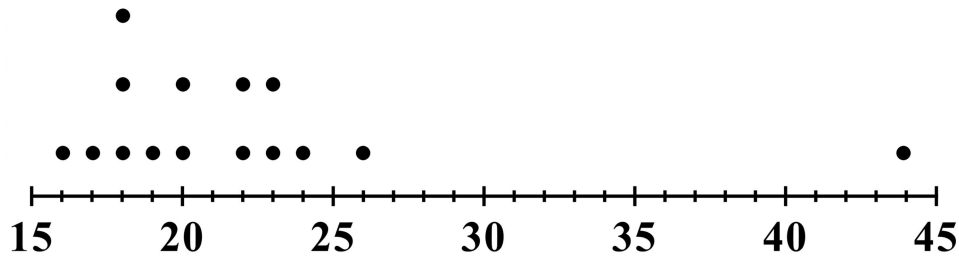
- (a) Use the **mean** and **standard deviation** to represent a data set when the data distribution is _____.
- (b) When the data distribution is **skewed** then the _____ and _____ should be used.



When data sets are smaller, sometimes there can be values that lie far away from the rest of the data set, known as **outliers**. Outliers can cause issues similar to what we saw with the **skewed distribution**.

Exercise #3: Lucas harvests 15 tomatoes from his garden and finds their weights, in ounces, shown below.

Distribution of Tomato Weights (ounces)



(a) Use your calculator to find all of the following measures. Round any non-whole numbers to the nearest tenth.

mean = _____ median = _____ sample standard deviation = _____

first quartile = _____ third quartile = _____ IQR = _____

(b) Which value is an outlier in this distribution? (c) What effect does it have on both the mean and standard deviation?

(d) Draw a box plot below the number line for this data set. (e) Construct an interval for values that are within one standard deviation of the mean.

(f) How many of the values in the data set fall within one standard deviation of the mean? (g) Why is the answer to (f) unusual?

(h) When specifying typical values within this data set, would it be better to use the interquartile range or the one standard deviation interval? Explain.

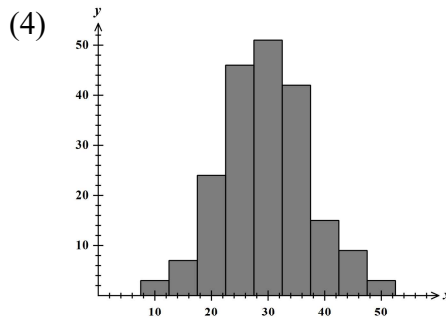
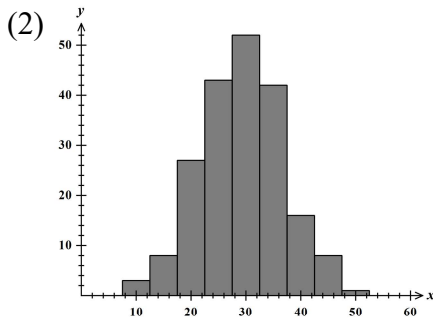
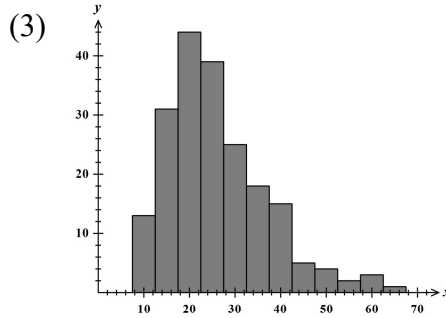
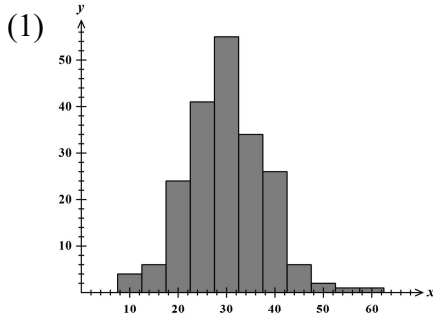


DATA DISTRIBUTION SHAPES

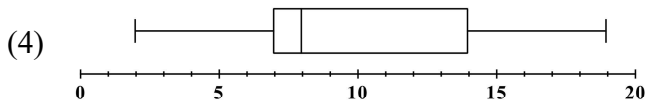
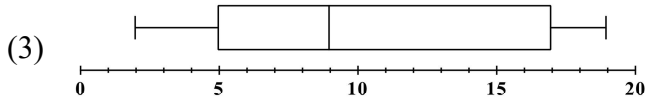
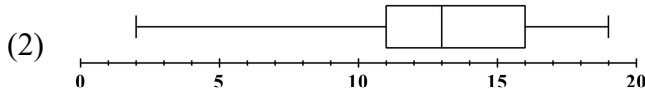
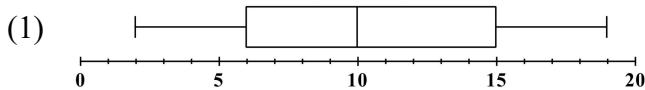
N-GEN MATH[®] ALGEBRA I HOMEWORK

FLUENCY

1. Which of the following distribution is least symmetric?



2. Which box plot below indicates a distribution that most likely has an outlier?



3. When a distribution is not symmetric it is best to use which of the following measures to represent the range of typical values within the distribution?

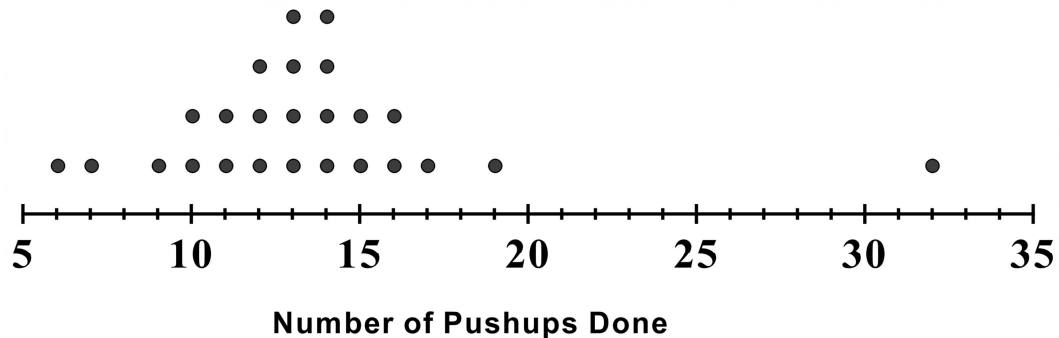
(1) the mean (3) the standard deviation

(2) the median (4) the interquartile range



APPLICATIONS

4. Carla is trying to determine how many pushups a typical 9th grade student can do in a row. She takes a random sample of 25 9th grade students and records the number of pushups they can do. The distribution is shown on the dot plot below.



- (a) Carefully enter the data into your calculator and find each of the following measures. Round answers to the nearest tenth when needed.

mean = _____ median = _____ standard deviation = _____

first quartile = _____ third quartile = _____ interquartile range = _____

- (b) How many of the distribution's values lie within one standard deviation of the mean? Show how you found your answer.

- (c) When discussing the typical number of pushups done by students, would it be better to use the interquartile range or the one standard deviation from the mean interval? Explain.

5. Remove the outlier from the data set above and do the following:

- (a) Recalculate the following statistics. Round again to the nearest tenth.

mean = _____ median = _____ standard deviation = _____

first quartile = _____ third quartile = _____ interquartile range = _____

- (b) Which measures changed the most when the outlier was removed?

