

2.2: Quantifying the Center of a Distribution

A stimulant is a type of drug that is often found in weight loss medications. We will examine the effect of a stimulant on the weight gains of a treatment group of rats. These are compared to a control group of rats who receive no stimulant treatment.

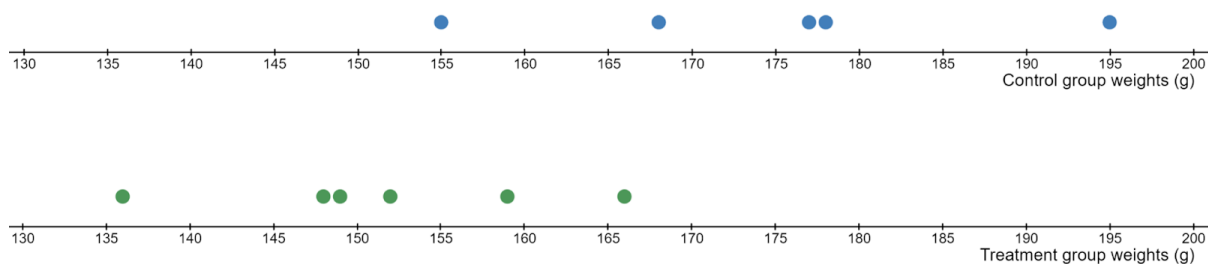
What is a typical value in the data set?

Suppose we observe the following weight gains (in grams) for twelve adolescent lab rats over a one-month period. The weight gain for the rats in the treatment and control groups are given below:

Control group weights in grams (no stimulant)	168	155	178	203	195	177
Treatment group weights in grams (stimulant)	136	159	152	149	166	148

To determine whether there might be an effect on weight gain due to the stimulant, we will determine representative (or central) values of the two groups, namely, the **sample mean** and the **sample median**.

Below are dotplots for the control and treatment groups of rats:



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

Sample Means

1. Imagine the dotplot as a scale that can tip left or right or stay balanced. Where do you think the control group's dotplot balances? That is, on the number line, where would you set a balance point so that the distribution does not tip to the left or right?

This value is an estimation of the **mean** or average. A mean is one way we could describe a typical data value in a set. To calculate the exact mean, we add all data values to find a sum and divide the sum by the number of data values in the set.

2. Compute the average weight gain for the rats in the control group.

This value is the exact **sample mean** since it is the mean of a sample of six rats in the control group. Luckily, this set is very small and therefore, the computation is not too difficult to do by hand. For most data sets, we will use technology to compute the sample mean for a set. The mathematical symbol we use to denote a sample mean is \bar{x} (pronounced “x-bar”). Formulaically, we say

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\text{sum of all values in the set}}{\text{number of values in the set}}$$

For the control group,

$$\bar{x} = \frac{168 + 155 + 178 + 203 + 195 + 177}{6} = 179.\bar{3} \approx 179.3 \text{ grams}$$

3. Compute the mean weight gain for the rats in the treatment group and call this \bar{y} (“y-bar”). Round to one decimal place.

$$\bar{y} = \frac{\quad}{\quad} = \quad \approx$$

4. Compare \bar{x} and \bar{y} . Which sample mean is larger? Is the difference between the sample means large enough to make you believe that the stimulant has an effect on weight gain in adolescent rats? Why or why not?

Sample Medians

A **sample median** is the middle number of a sorted list of data values. Here is the process for computing a median applied to the control group values.

- First we sort the data values from smallest to largest:

unsorted	168	155	178	203	195	177
sorted	155	168	177	178	195	203

- Notice that the middle number in this ordered set is between the 3rd and 4th values. This will always be the case when we have an even number of values in our set. To find the location of the median for a set that has an even number of elements, we can divide the sample size by 2 and the median will be between this quotient and the next number on the list. (For example, let’s say we have a set of 8 data values. The median will be exactly between the 4th and 5th values in the sorted list). For our control group, the median falls between 177 and 178. This means that the sample median is halfway between these two values or in other words, the median is the average of these two middle numbers:

$$\text{median} = \frac{177 + 178}{2} = 177.5 \text{ grams}$$

- If there are an odd number of values in the set, the median is the data value exactly in the middle of the sorted list, and there will be an equal number of data values on each side of the median. For example, consider the set $A = [1, 1, 12, 15, 17, 21, 22, 25, 40]$ which has 9 values in it (which is odd). The middle number or median of the sorted set is 17. There are 4 data values to the left of 17 and 4 data values to the right of 17. When the sample size is odd, we can divide by 2 as we did before, however, this will result in a number that is not whole. 9 divided by 2 is 4.5. To find the location of the median for an odd sized set, we can divide by 2 and round up to the nearest whole number. 4.5 rounds to 5, so the 5th value in this set is the median.
5. Compute the sample median for the treatment group.

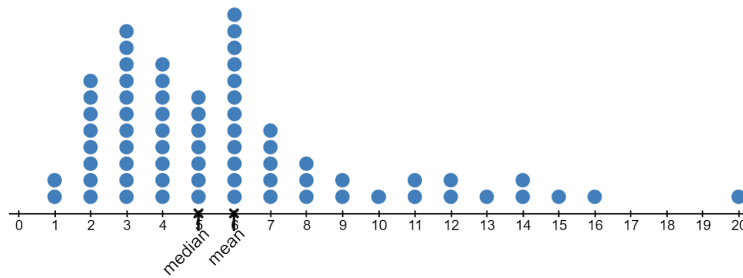
The sample median is another way to describe a central/representative/typical value in a set of data.

6. Compare the median weight gains of the two groups. Which sample median is larger? Is the difference between the sample medians large enough to make you believe that the stimulant has an effect on weight gain in adolescent rats? Why or why not?
7. Suppose we made an error when we recorded the largest weight gain in the treatment group. Instead of writing 166, we wrote 616. Recalculate the sample mean and sample median for the treatment group with this new value. In what way(s) did this error impact the mean? In what way(s) did this error impact the median?

Resistant Measures of Center

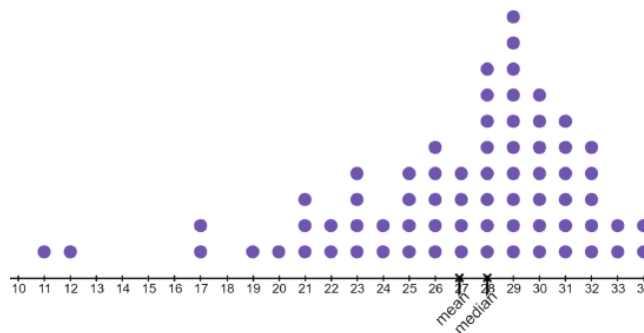
When should we avoid choosing a mean as a representative value for a data set? We call the mean and median measures of center. Measures of center are single values that represent a typical value for a given set of data. You've just seen that the mean is strongly affected by extreme values (values that are far away from most of the other data). We say that the mean is *not* a **resistant** measure of center. You also saw in the last example that the median was unaffected by the existence of an extreme value. We say that the median is a **resistant** measure of center. Extreme values are often present when a distribution is **skewed**. A distribution is skewed when it is not symmetric and one side has a long tail of values. When a distribution is skewed, the mean is pulled in the direction of the tail.

The following dotplot is an example of a distribution that is **skewed to the right**. When the distribution is right-skewed, the mean tends to be greater than the median.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

The following dotplot is an example of a distribution that is **skewed to the left**. When the distribution is left-skewed, the mean tends to be less than the median.



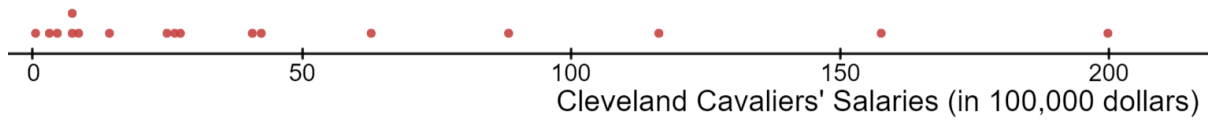
Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

8. Below are the salaries of the seventeen players on the Cleveland Cavaliers basketball team during [the 2009-2010 season](#) (access this website using QR code below). The 2009-2010 season was LeBron James' last season playing for the Cavaliers (until he later returned to the team). LeBron James was not the highest paid player on the team that season, Shaquille O'Neal was.



1	Shaquille O'Neal	20000000	10	Sebastian Telfair	2500000
2	LeBron James	15779912	11	J.J. Hickson	1429200
3	Antawn Jamison	11641095	12	Leon Powe	855189
4	Mo Williams	8860000	13	Darnell Jackson	736420
5	Anderson Varejao	6300000	14	Jawad Williams	736420
6	Delonte West	4254250	15	Danny Green	457588
7	Daniel Gibson	4088500	16	Coby Karl	311896
8	Jamario Moon	2750000	17	Cedric Jackson	53834
9	Anthony Parker	2644230			

A dotplot of the salaries is given below:



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

- Calculate the mean salary for the Cavaliers during the 2009-2010 season.
- Calculate the median salary for the Cavaliers during the 2009-2010 season.
- Would the mean or the median be most representative of the Cleveland Cavalier players' salaries in the 2009-2010 season? Justify your answer.
- How does Shaquille O'Neal's salary impact the mean? You can examine this question by computing the mean without Shaq's salary included and compare.

This page titled [2.2: Quantifying the Center of a Distribution](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Hannah Seidler-Wright](#).

- [Current page](#) by Hannah Seidler-Wright is licensed [CC BY-NC-SA 4.0](#).
- [1.2: The Statistical Analysis Process](#) by Hannah Seidler-Wright is licensed [CC BY-NC-SA 4.0](#).