

Chapter Three Part Two: Numeric Summaries and Measures of Center and Spread

Continuing on our discussion about quantitative variables, I mentioned last class that _____ . We call such mathematical calculations, _____ .

We will divide our discussion of summary statistics into two categories:

- 1.

- 2.

Measures of Center

Measures of center are said to locate the _____ .

When looking for measures of center we ask ourselves: Where is the typical value located in this distribution? There are two measures of center that we will consider.

- 1.

- 2.

Median

The _____ is the value with exactly half of the data values above it, and half the data values below it. It is the _____. To find the median we follow the following procedure:

- 1.

- 2.

- 3.

•

•

In other words, when the sample size (n) is an odd number, the median is the $\frac{n+1}{2}$ number in the ordered list of observations.

Example If $n = 11$, what is the location of median in this dataset?

Example Let's consider the homerun records of professional base-ball player Hank Aaron. The data is shown below, please find the median number of homeruns for Hank Aaron. In this case, Hank Aaron played for 23 years, so the sample size is 23.

Year	HR	Year	HR	Year	HR
54	13	62	45	70	38
55	27	63	44	71	47
56	26	64	24	72	34
57	44	65	32	73	40
58	30	66	44	74	20
59	39	67	39	75	12
60	40	68	29	76	10
61	34	69	44		

When the sample size (n) is even, the $\frac{(n+1)}{2}$ place in the list is inbetween two numbers.

Example If $n = 10$, what is the location of the median?

Example Now let's consider the baseball career of Barry Bonds, he played a total of 22 years. Find the median number of homeruns in Barry Bonds' career.

Year	HR	Year	HR	Year	HR
86	16	94	37	02	46
87	25	95	33	03	45
88	24	96	42	04	45
89	19	97	40	05	5
90	33	98	37	06	26
91	25	99	34	07	28
92	34	00	49		
93	46	01	73		

Note: It is important to note that the median is a measure of center that is _____. We will discuss more on this topic later.

Mean

The mean is the _____ of all the data values in a distribution. We often refer to the mean using the mathematical notation _____. To calculate the mean we follow a familiar process:

- 1.
- 2.
- 3.

Example *Using the previous data for Hank Aaron and Barry Bonds, calculate the mean number of homeruns for each player.*

Note: Unlike the median, the mean is _____.
The reason for this is if we have an extremely high value in our dataset, it will artificially pull the mean higher than we would normally expect.

Measures of Spread

Other than learning where the most common numerical response is, it is often useful to know how spread out our data is. For example, data that is very tightly clustered around one point may be indicative of some relationship whereas data that is extremely spread out may give us less information. There are three major measures of spread that we will discuss today:

- 1.
- 2.
- 3.

Range

The range is the difference between the _____ and _____ values in the dataset. To find the range we do the following:

Example Using the same data from the Barry Bonds example, find the range of the dataset.

5	16	19	24	25	25	26	28	33	33	34
34	37	37	40	42	45	45	46	46	49	73

One thing to note is that when we add observations that change the maximum or the minimum of the dataset, the range _____. This means that the range is also _____.

Inter-Quartile Range

Before we can talk about the Inter-Quartile Range we need to establish a few other concepts. The first of which is _____. The k^{th} percentile is the value that $k\%$ of the data is below.

Example What does it mean if a student obtains a score in the 60th percentile on an exam?

Example How would you find the 50th percentile?

To calculate the IQR we need to know how to calculate _____. Quartiles divide the data into four equally sized groups, where each group contains 25% of the data.

1. **First Quartile(Q1):**

2. **Second Quartile(Q2):**

3. Third Quartile(Q3):

Example *Once again, let's return to the Barry Bonds example. Calculate the Q1 and Q3 quartile of this dataset.*

5	16	19	24	25	25	26	28	33	33	34
34	37	37	40	42	45	45	46	46	49	73

Finally, we can define the IQR. The IQR or the inter-quartile range refers to the difference between the first and third quartiles. It measures the _____.
To calculate the IQR:

Example *Using your answer to the previous example, calculate the IQR for Barry Bonds' home run records.*

Note: One thing that is beneficial about using the IQR, is that it largely ignores extreme values.

In summary:

- **Range:**
- **IQR:**

Note: When the median is reported as a measure of center, we often also report the _____ and the _____. This often takes the form of the Five Number Summary.

Five Number Summary

The five number summary summarizes the distribution of a quantitative variable. The five number summary includes:

-
-
-
-
-

Standard Deviation

The last measure of spread we will discuss today is the _____. This measures the _____ between each observation and the sample mean. We refer to the standard deviation from a sample using the notation s . (Sometime you may see it referred to as σ when it comes from a population). To calculate the standard deviation we use the following formula.

$$s = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}$$

More simply put, we follow the following steps:

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.

To calculate the variance, we simply _____ the standard deviation.

Example Calculate the standard deviation and the variance of the following dataset: 11, 5, 6, 6, 7.

We interpret the standard deviation as the _____.
Note that the standard deviation can _____ be less than zero. If the standard deviation is equal to zero then, _____. If we have two samples of the same size and one has a higher standard deviation than the other one. We can say that that distribution has a larger _____ from the mean.

Since the standard deviation relies on the sample mean, then standard deviation is _____.