Statistics Five Number Summary and More

California State University SAN MARCOS

Five Number Summary: Lowest Value, Q_1, Median, Q_3, Highest Value
And More: Range, Mean, Standard Deviation
A ***five-number summary*** comes from a given set of data values.
*The following set of data values are assumed to be a sample; therefore, sample notation is used. A sample is a **subset** of measurements from the population. Population data is a **complete** set of measurements or counts.*

Example 1: 1, 2, 4, 5, 5, 6, 8, 9, 10, 10, 10.

Example 2: 1, 2, 3, 4, 5, 6

Note: Example 1 has an odd set of data values(11) while Example 2 has an even set of data values(6). This affects how the set of data values are divided into *quartiles.*

Lowest Value(or Minimum): The smallest value of the data set

Using Example 1, 1 is the lowest value of the data set.

Median(or Quartile 2): The middle number in the set of given data values(after being ordered from least to greatest)

Note that the first example is an odd set of data values so the only middle number(circled) is the median:

1, 2, 4, 5, 5, 6, 8, 9, 10, 10, 10

Consider the second ordered example: 1, 2, 3, 4, 5, 6 . Note that the second example is an even set of data values. Then the median is found by finding the average of the 2 middle numbers(circled):

1, 2, 3, 4, 5, 6

$$Median = \frac{3+4}{2} = \frac{7}{2} = 3.5$$

Q_1 **(or Quartile 1):** The median of the lower half of the data set

Example 1: 1, 2, 4, 5, 5, $\underline{6}$, 8, 9, 10, 10, 10. (*The median has been underlined*)

The Lower Half of the data set is: 1, 2, 4, 5, 5
Note: *For an odd set of data values, the median is excluded when dividing the data set into a lower and upper half*
Q_1 is the circled number: 1, 2, 4, 5, 5

Example 2: 1, 2, 3, 4, 5, 6

The Lower Half of the data set is: 1, 2, 3
Note: *For an even set of data values, all data values are included when dividing the data set into a lower and upper half*
Q_1 is the circled number: 1, 2, 3

Q_3 **(or Quartile 3):** The median of the upper half of the data set

Example 1: The Upper Half of the data set is: 8, 9, 10, 10, 10

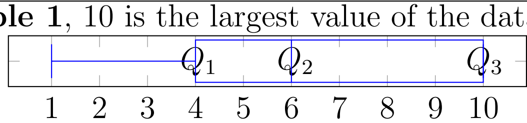
Q_3 is the circled number: 8, 9, 10, 10, 10

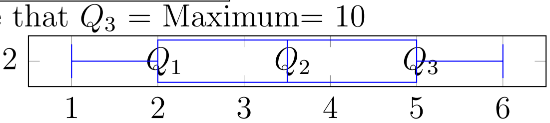
Example 2: The Upper Half of the data set is: 4, 5, 6

Q_3 is the circled number: 4, 5, 6

Highest Value(or Maximum): The largest value of the data set

For **Example 1**, 10 is the largest value of the data set. Note that Q_3 = Maximum= 10

Example 1


Example 2


The figures above are box and whisper plots for Examples 1 and 2 using the *five-number summary.* Example 1 has 1 whisker because Q_3 is the same as the maximum. Example 2 has two whiskers.

Range: The difference between the largest and smallest values of a data distribution

Example 1: 1, 2, 4, 5, 5, 6, 8, 9, 10, 10, 10.

Smallest Value: 1 Largest Value: 10 *Difference:* Largest-Smallest = 10-1 = 9 therefore **Range** = 9

Mean(sometimes called the arithmetic mean): An average that uses the exact value of each entry

$$\textbf{Mean} = \frac{\sum x}{n} = \frac{Sum\ of\ all\ data\ values}{Number\ of\ data\ values}$$

Notation: $\overline{x}$(x bar) for sample and μ(mu) for population

Note: It's important to understand that the mean does not represent an individual data value but *is* a summary of the entire data.

By using the first example given, we get:

$$Mean = \frac{1+2+4+5+5+6+8+9+10+10+10}{11} = \frac{70}{11}$$

You may keep the fraction as is or you may simplify it into decimal form.

Simplifying would then give us Mean$\approx 6.\overline{36}$ **OR** Mean $\approx$ 6.4 by rounding to the nearest tenth **OR** Mean $\approx$ 6.36 by rounding to the nearest hundredth.

Sample Standard Deviation(s): Measure of variability of distribution(or spread) of data values

Note: The more spread out a data distribution is, the greater its standard deviation.

Notation: s for sample and σ(sigma) for population

$$Formulas: \text{Computational: } s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} \text{ Defining: } s = \sqrt{\frac{\sum(x-\overline{x})^2}{n-1}}$$

Example 4: 3, 4, 5, 5, 6, 7, 8, 9, 12, 23

The following steps are for the defining formula. Both formulas stated above give the same results.

Step 1: Find the mean of the set of data values

$$Mean = \frac{3+4+5+5+6+7+8+9+12+23}{10} = \frac{81}{10} = 8.2 \text{ therefore } \overline{x} = 8.2$$

Step 2: Calculate $(x-\overline{x})^2$

Note: x is each data value and $\overline{x}$ is the mean

x	$x-\overline{x}$	$(x-\overline{x})^2$
3	3-8.2 = -5.2	$(-5.2)^2 = 27.04$
4	4-8.2 = -4.2	$(-4.2)^2 = 17.64$
5	5-8.2 = -3.2	$(-3.2)^2 = 10.24$
5	5-8.2 = -3.2	$(-3.2)^2 = 10.24$
6	6-8.2 = -2.2	$(-2.2)^2 = 4.84$
7	7-8.2 = -1.2	$(-1.2)^2 = 1.44$
8	8-8.2 = -.2	$(-.2)^2 = 0.04$
9	9-8.2 = .8	$(.8)^2 = 0.64$
12	12-8.2 = 3.8	$(3.8)^2 = 14.44$
23	23-8.2 = 14.8	$(14.8)^2 = 219.04$

Step 3: Calculate the sum of all $(x-\overline{x})^2$ terms in the last column of the table found in Step 2.

$$\sum(x-\overline{x})^2 = 305.6$$

Step 4: Divide by the number of data values minus 1

$$\frac{\sum(x-\overline{x})^2}{n-1} = \frac{305.6}{10-1} = \frac{305.6}{9} = 33.9555555555...$$

Step 5: Take the square root of the fraction from Step 4

$$\sqrt{\frac{\sum(x-\overline{x})^2}{n-1}} = \sqrt{\frac{305.6}{9}} = 5.8271395689...$$

Note: *For accuracy purposes, it's best not to round until Step 5.*

By rounding to the hundredths place, we get s= 5.83.