Quiz 1

Do you know your 5 Number Summary?

Notice the dataset for the next 4 problems all use the same dataset.

What is the range associated with the following dataset?

1, 5, 10, 3, 8, 12, 4, 1, 2, 8

Enter your response here

max = 12
min = 1
range = 12 - 1 = 11

SUBMIT

Provided the values below, what is the value of the first quartile?

Remember the first quartile is the median number (middle number) of the first half of the numbers when put in ranked order. Ranked order for this dataset is 1,1,2,3,4...

1, 5, 10, 3, 8, 12, 4, 1, 2, 8 **1 , 1 , 2 , 3 , 4 , 5 , 8 , 8 , 10 , 12**

Enter your response here

Q1 = 2

SUBMIT

Provided the values below, what is the value of the third quartile?

1, 5, 10, 3, 8, 12, 4, 1, 2, 8

Enter your response here

Q3 = 8

SUBMIT

Provided the values below, what is the value of the median? (this is the same as the second quartile)

1, 5, 10, 3, 8, 12, 4, 1, 2, 8

Enter your response here

Q2 = 4.5

SUBMIT

Do you know your 5 Number Summary?

Notice the dataset for the next 4 problems all use the same dataset.

Provided the values below, what is the value of the median? (this is the same as the second quartile)

5, 10, 3, 8, 12, 4, 1, 2, 8

Enter your response here

Q2 = 5

SUBMIT

Provided the values below, what is the range? (The range should just be a single number)

5, 10, 3, 8, 12, 4, 1, 2, 8

Enter your response here

range = 12 - 1 = 11

SUBMIT

Provided the values below, what is the the value for the first quartile?

5, 10, 3, 8, 12, 4, 1, 2, 8

Enter your response here

Q1 = 2.5

SUBMIT

Provided the values below, what is the the value for the third quartile?

5, 10, 3, 8, 12, 4, 1, 2, 8

Enter your response here

Q3 = 9

SUBMIT

Quiz 2

QUESTION 1 OF 3

If we measure the variance associated with our sales in dollars for each month for 3 years, what are the units associated with the variance?

- Dollars
- Years
- Dollars per Year
- Dollars Squared
- Dollars per Month

For the following set of data provide the value of the **variance**.

Remember to find the variance we first find the mean average of the values, then subtract the mean from each value, then square each of these values, then add them up, then divide by the number of values. (Round your answer to two decimal places at the end of your calculation - don't round along the way.)

1, 5, 10, 3, 8, 12, 4

Enter your response here

V = 13.55

SUBMIT

For the following set of data provide the value of the **standard deviation**.

Remember the standard deviation is the square root of the variance (Round your answer to two decimal places at the end of your calculation.)

1, 5, 10, 3, 8, 12, 4

Enter your response here

S = 3.681

Quiz 3

QUESTION 1 OF 3

Assume d1 and d2 are datasets both measured in the same units. We know that the standard deviation of d1 is 5 and the variance of d2 is 36, which of the following are certainly true. Mark all that apply.

Remember the Standard Deviation is the square root of the variance. So if the Variance is 4 the Standard Deviation would be 2

- [ ] The mean is larger for d1 than for d2.
- [x] The variance for d2 is larger than for d1.
- [x] The standard deviation for d2 is larger than for d1.
- [ ] The median for d2 is larger than for d1.
- [ ] The range for d2 is larger than for d1.

QUESTION 2 OF 3

If a dataset has a standard deviation of zero, which of the following MUST be true?

- All the data points must be zero.
- All the data points must be the same.
- We made a calculation error because it is not possible for the standard deviation to be zero.

QUESTION 3 OF 3

For each of the below: If the statement is **true**, mark the box next to the statement.

- [x] If two datasets have the same variance, they will also have the same standard deviation.
- [ ] If I have two investment options with the same mean return, it really doesn't matter which I invest in.
- [ ] If I have two investment options with the same standard deviation associated with the return, they will also have the same max possible return.

Quiz 4

Investment Data

Consider we have two investment opportunities:

	Returns					
	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**Year 6**
Investment 1	5%	5%	5%	5%	5%	5%
Investment 2	12%	-2%	10%	0%	7%	3%

The returns for 6 consecutive years for each investment are shown above. Use this information to answer the questions below.

What is the mean return for Investment 1? (Write your answer as a percentage without the percentage sign - Ex: 9% should be reported as 9 or 93% is reported as 93.)

Enter your response here

mean = 5

What is the mean return for Investment 2? (Write your answer as a percentage without the percentage sign - Ex: 9% should be reported as 9 or 93% is reported as 93.)

Enter your response here

mean = 5

Investment Data

In the previous two questions, you should have found that these investments have the same mean! That is, regardless of which investment opportunity you choose, you are expected to earn the same amount. So how are they different? Let's look at some additional questions to see if we can find some differences.

The same data as above is provided again (to minimize scrolling).

	Returns					
	Year 1	**Year 2**	**Year 3**	**Year 4**	**Year 5**	**Year 6**
Investment 1	5%	5%	5%	5%	5%	5%
Investment 2	12%	-2%	10%	0%	7%	3%

The returns for 6 consecutive years for each investment are shown above. Use this information to answer the questions below.

QUESTION 3 OF 4

Using the information above, mark all of the below that are **true** statements.

- [x] The risk associated with investment 1 is lower than the risk associated with Investment 2.
- [x] The standard deviation associated with Investment 1 is smaller than the standard deviation associated with Investment 2.
- [ ] Knowing the mean return amount across all the years for each investment provides us with all of the information necessary to understand which investment we should choose.

QUESTION 4 OF 4

Based on the observed data, which of the above two investments has the best opportunity of earning more than 7%?

- Investment 1
- Investment 2
- Neither.
- We cannot tell.

Useful Insight

The above example is a simplified version of the real world, but does point out something useful that you may have heard before. Notice if you were not fully invested in either **Investment 1** or fully invested in **Investment 2**, but instead you were diversified across both investment options, you could earn more than either investment individually. This is the benefit of diversifying your portfolio for long term gains. For short term gains, you might not need or want to diversify. You could get lucky and hit short term gains associated with the upswings (12%, 10%, or 7%) of Investment 2. However, you might also get unlucky, and hit a down term and earn nothing or even lose money on your investment using this same strategy.

Quiz 5

QUESTION 1 OF 2

For the following dataset, match each value to the appropriate label:

15, 4, 3, 8, 15, 22, 7, 9, 2, 3, 3, 12, 6

2 , 3 , 3 , 3 , 4 , 6 , 7 , 8 , 9 , 12 , 15 , 15 , 22

13 | 7 | 3 | 13.5 | 8.4 | 3

TERM	VALUE
n	**13**
median	**7**
first quartile	**3**
third quartile	**13.5**
mean	**8.4**
mode	**3**

QUESTION 2 OF 2

For the following dataset, match each value to the appropriate label:

15, 4, 3, 8, 15, 22, 7, 9, 2, 3, 3, 12, 6

10.5 | 20 | 33.9 | 5.8 | 2 | 22

TERM	VALUE
interquartile range	10.5
range	20
variance	33.9
standard deviation	5.8
minimum	2
maximum	22

Quiz 6

QUESTION 1 OF 2

Match the distribution shape with the correct relationship in comparing the mean to the median.

Cannot be Determined. | Mean is greater than the Median.

Mean is less than the Median. | Mean is equal to the Median.

SHAPE	COMPARISON
Right-skewed	**Mean Is greater than the median**
Left-skewed	**mean is less than median**
Symmetric	**mean is equal to median**

QUESTION 2 OF 2

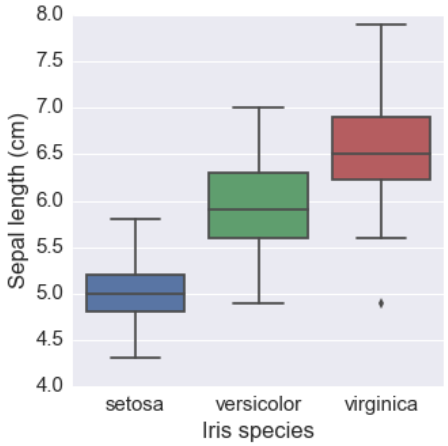
Check all of the below that must be true.

- [ ] For every dataset the mean equals the median, so every data set is normally distributed.
- [x] Normally distributed data must have a mean equal to the median.
- [ ] All data sets are normally distributed.
- [ ] Data must be distributed either right-skewed or left-skewed.
- [x] Histograms and box plots are both used to plot quantitative data. They cannot be used to plot categorical data.
- [x] A box plot relates directly to the 5 number summary.

Quiz 7

Image Summary

In the below image, we have three box-plots. Each box-plot is for a different Iris flower: `setosa`, `versicolor`, or `virginica`. On the y-axis, we are given the sepal length. Notice that `virginica` has an **outlier** towards the bottom of the plot. Therefore, the **minimum** is not given by the bottom line here; rather, it is provided by this point.



Quick Refresher: The measures of center and spread we can determine from a Box Plot are as follows. Let's use Setosa for these examples.

Median is the center line inside the box and is 5

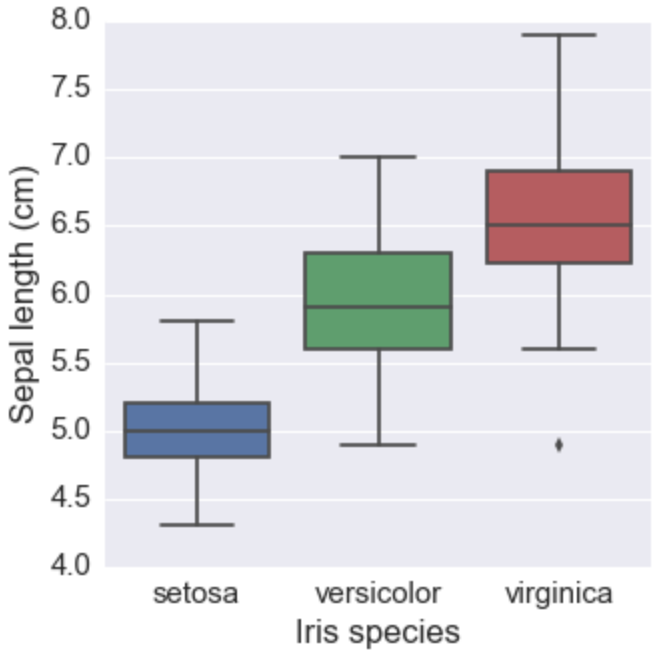
IQR is space between the first and third quartile which are the edges of the box. They are about 4.8 for the first quartile and 5.2 for the third

QUESTION 1 OF 2

Match the appropriate Iris type to the statement(s) that are true for its Sepal Length.

None | Virginica | Setosa | Setosa | Versicolor | All | Virginica

SEPAL LENGTH	IRIS TYPE
The largest Range	Vigrinica
The smallest Interquartile Range	setosa
Median is approximately 5	setosa
Third quartile is approximately 6.3	versicolor
Approximately Symmetric	all
The largest sepals on average.	virginica



QUESTION 2 OF 2

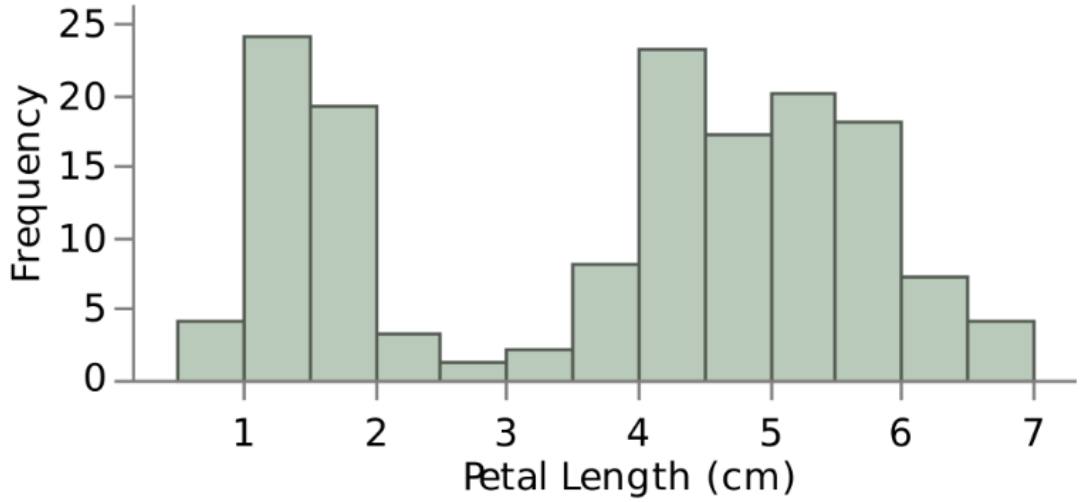
Using the same flower data, select all of the below statements that **MUST** be true.

- [ ] **All** setosa flowers have a shorter sepal length than versicolor flowers.
- [ ] **All** virginica flowers have larger sepal length than setosa flowers.
- [ ] More than 75% of the virginica flowers have a larger sepal length than the largest setosa flower.
- [ ] More than 50% of setosa flowers have larger sepal length than the shortest versicolor flower.
- [ ] More data was collected on versicolor flowers than on virginica flowers.
- [ ] More data was collected on setosa flowers than on virginica flowers.

Quiz 8

Questions 1 - 2: Petal Length

The below plot will be used to answer the first two questions in this section.



QUESTION 1 OF 5

What is the name of the above plot?

- Bar Chart
- Box Plot
- Histogram
- Pie Chart

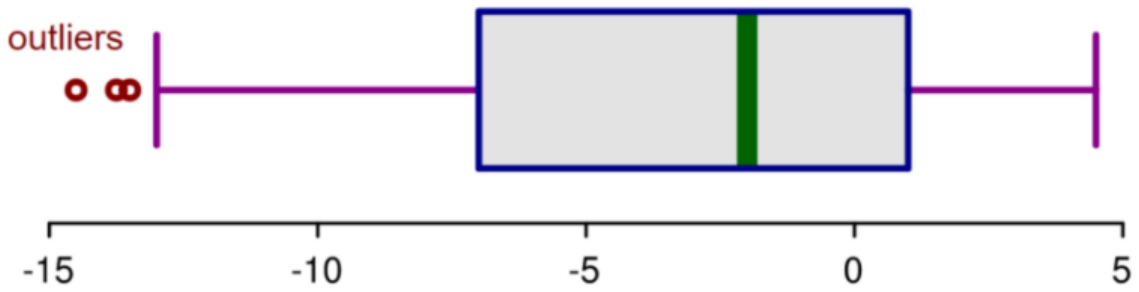
QUESTION 2 OF 5

What is the shape of the above distribution?

- Right skewed
- Left skewed
- Symmetric
- Bi-modal

Questions 3 - 6: Shape and Outliers

Use the below image to assist with answering the next three questions.



QUESTION 3 OF 5

What is the name of the above plot?

- Bar Chart
- Box Plot
- Histogram
- Pie Chart

QUESTION 4 OF 5

What is the shape of the distribution?

- Right skewed
- Left skewed
- Symmetric
- Bi-modal

QUESTION 5 OF 5

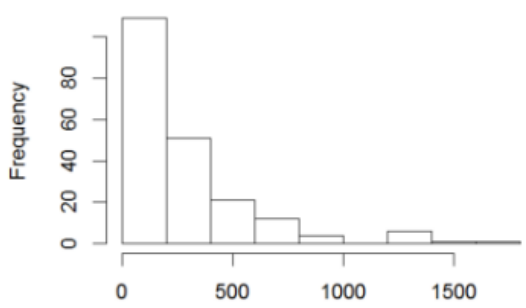
Select the true statement for the box-plot above.

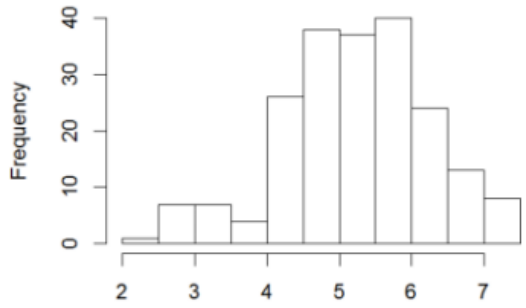
- The mean is less than the median.
- The mean is greater than the median.
- The mean is approximately equal to the median.
- It is impossible to tell the relationship between the mean and median.

Quiz 9

Histograms

Let the histogram on the left be **Histogram 1** and the histogram on the right be **Histogram 2**.





Quick Notes

Pay attention to the scale of these two graphs. The first is dealing with much higher numbers.

The median is the middle number and is not affected by outliers.

The average factors in all the numbers so outliers will bring the average towards them.

Left Skewed is when the graphs start with a low frequency and then slopes up. Right Skewed is when the graph starts with a high frequency and slopes down.

QUIZ QUESTION

Correctly match the histograms to the statements that are true about each.

Both | Histogram 2 | Neither | Histogram 1 | Impossible to tell.

Histogram 1 | Histogram 1 | Histogram 2 | Histogram 2 | Neither

Histogram 2

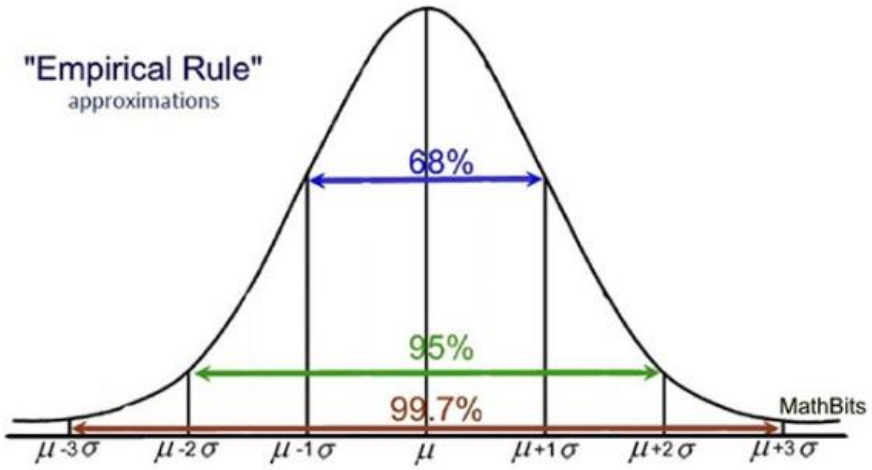
STATEMENT	HISTOGRAM
Mean is greater than the median.	Histogram1
Data has higher variance.	Histogram1
Binwidth is equal to 0.5.	Histogram2
The range is approximately 5.5.	Histogram2
Distribution is left-skewed.	Neither
The mean is approximately equal to the median.	Histogram2

Quiz 10

More On Center And Spread

When analyzing skewed data, it is common to report numeric summaries like the median and 5 number summary, as the mean and standard deviation may be misleading.

However, with symmetric data, the mean and standard deviation are commonly used, as we can understand what proportion of points might fall 1, 2, or 3 standard deviations away based on the empirical rule associated with normal distributions.



- 68% of the distribution lies within **one** standard deviation of the mean.
- 95% of the distribution lies within **two** standard deviations of the mean.
- 99.7% of the distribution lies within **three** standard deviations of the mean.

You can read more about this here.

Standard Deviation and Skewed Distributions

Standard Deviations can be calculated for any data set, whether it is normally distributed or skewed. So with the data below be careful what assumptions you are making about the underlying data.

Also the standard deviation basically provides which of two sets of data are more spread out.

QUIZ QUESTION

If we know the following about heights for the accountants in our company and the IT people in our company:

- The mean height of accountants is 67.84 in.
- The mean height of IT staff is 69.76 in.
- The standard deviation of heights for accountants is 4.5 in.
- The standard deviation of heights for IT staff is 3.2 in.

Which of the below statements are definitely true? **Select all that apply.**

- [ ] All accountants are taller than all IT staff.
- [ ] On average, accountants are shorter than IT staff.
- [ ] There is more variability in the heights of the accountants than the heights of the IT staff.
- [ ] There is less variability in the heights of the accountants than the heights of the IT staff.
- [ ] The distribution of heights for accountants is normally distributed.

Quiz 11

QUESTION 1 OF 2

Identify the population(s), parameter(s), sample(s), and statistic(s) for the below scenario:

Consider we own a bagel shop. We know that the average diameter of all of our bagels is 5.5 inches. A competitor moves right next door to us! We are interested in if they make larger bagels than us. We obtain 100 of their bagels, and we find they have an average diameter of 6 inches.

None | Statistic | Parameter | Sample | Parameter | Statistic

Population | Population | Sample

DESCRIPTION	TERM
5.5 inches	**Parameter**
6 inches	**Statistic**
All the bagels at our bagel shop.	**Population**
All the bagels at our competitor's bagel shop.	**Population**
The 100 bagels from the competitor's bagel shop.	**Sample**

QUESTION 2 OF 2

For the below, match the term to the correct description.

Statistic | Statistic | None | Population | Statistic | Parameter

Inference | None | Sample | Population | Parameter

DESCRIPTION	TERM
A numeric summary of a sample.	Statistic
A numeric summary of a population.	Parameter
Drawing conclusions regarding a population using information from a sample.	Inference
Drawing conclusions regarding a sample using information from a population.	None
A subset of a population.	Sample
Our entire group of interest.	Population
Frequently we do not know this value, so we must try and estimate.	parameter

Quiz 12

QUIZ QUESTION

Identify the population, parameter, sample, and statistic for the below scenario:

Consider we are interested in the average number of hours slept by all Udacity students (100,000 students). I send an email to all Udacity students, but I only receive 5,000 response emails. The average amount of sleep of those that responded was 6.8 hours of sleep.

6.8 hours of sleep | We cannot know for sure. | All Udacity students

We cannot know for sure. | 5,000 Udacity students | 6.8 hours of sleep

TERM	DESCRIPTION
Population	**All Udacity Students**
Parameter	**We Can Not Know Of Sure**
Sample	**5000 Udacity Students**
Statistic	**6.8 hours of sleep**