Sequences

A sequence is a list of numbers in some order with a rule to determines the next number. Each number in the sequence is called an "element." An ellipsis, indicated with three dots (. . .) indicates that the sequence continues infinitely in that direction.

For example, the counting numbers can be written as a very basic sequence: 1, 2, 3, 4 … When a number is left out of the sequence, a blank space is inserted:

1, 2, 3, ___, 5, 6, 7…

The pattern before and after the blank reveals that the missing number is 4. This sequence follows the pattern, or rule, that each element is one more than the number before it. We write that as (n + 1).

That kind of sequence, where the numbers increase by the same amount every time, is called an <u>Arithmetic Sequence</u>. The numbers in an arithmetic sequence can increase by any amount, as long as it is always the same amount. So the following is an arithmetic sequence:

7, 12, 17, 22, 27, 32, 37…
Each number is 5 greater than the previous number (n + 5)

We may need to use trial and error to identify the correct pattern in a sequence. Generally, we can start to solve a sequence by determining if the difference is always the same. If it is, fill in the blank to continue the pattern.

If the difference between elements isn't constant, multiplying (or dividing) is another likely pattern. In a <u>Geometric Sequence</u>, each number is multiplied by a constant number to determine the next number in the sequence. That constant number is called the "common ratio." For example, consider the sequence:

2, 4, 8, _____, 32, 64, 128 …
Each number is twice the previous number (2n)

Take the first number and multiply by 2: 2 x 2 = 4. That result is then multiplied by 2: 4 x 2 = 8. When 8 is multiplied by 2, the result is <u>16</u>, so that is the missing number. The common ratio is 2. Continue the pattern (multiply by 2) through the whole sequence, and confirm that our answer is correct.

If a sequence isn't a simple arithmetic or geometric sequence, it may be a combination of operations. Consider this sequence:

4, 6, 10, _____, 34, 66, 130 …

It is important to consider the entire sequence, because often the pattern is most obvious in the upper numbers. Notice that 130 is almost twice as much as 66. And notice that 66 is almost twice as much as 34. Once you notice that there may be some doubling involved, experiment with the options. You will notice that each following number in the sequence can be found by subtracting one and then doubling the previous number. The missing number in the pattern is 18.

4, 6, 10, 18, 34, 66, 130 …

Subtract 1 from previous number, then double.

$2(n - 1)$

Central Tendencies

Students' grades on a test, the scores of a sports team, movie box office sales, or the cost of groceries – these are all groups of data. To describe trends or changes, we need ways to summarize data.

Measurements of central tendency, such as an average, describe data. Instead of listing every student's grade in a class of 30, we might just say that the "average" grade was 82. But did everyone in the class get an 82? How many students failed? There are several ways to summarize data. Knowing how each method is calculated helps us understand the data they describe.

Mean

The "mean" of data is also known as the average. The formula for mean is:

$$\text{Mean} = \frac{total\ of\ all\ values}{quantity\ of\ values}$$

For example, assume that nine students took a math test, and their scores were: 100, 89, 52, 73, 86, 94, 62, 70, and 94. To find the mean, add the scores to get the total. Next divide by the quantity of scores, which is 9 because nine students took the test. The mean, or average score, is an 80.

Mean = $\frac{100+89+52+73+86+94+62+70+94}{9} = \frac{720}{9}$ = 80.

Median

"Median" is another way to describe the center of the data. To find the median, arrange the data from low to high and find the middle number. In our data sample above that would be 86. Half the students scored higher than 86; half the students scored lower than 86.

Data = 52, 62, 70, 73, 86, 89, 94, 94, 100. The median, or middle number, is 86.

If there is an even number of scores in your set of data, the median is the average of the two middle numbers. So, if a tenth student took the test and got a 92, we would take the average of 86 and 89, which would be 87.5.

Data = 52, 62, 70, 73, 86, 89, 92, 94, 94, 100. Median = average of middle numbers: $\frac{86+89}{2}$ = 87.5.

Mode

The "mode" is the value that appears most often in the data set. The mode, therefore, might not actually describe the center of the data. If every value of the data set occurs only once, there is no mode. In our data set the mode is 94, because that number occurs more than the others.

Range

The "range" of the data is the difference between the highest and the lowest value in the set of data. The formula is:

Range = Highest Value – Lowest Value

Once again using our set of math grades, the highest score is 100 and the lowest score is 52, so the range = 100 -52 = 48.

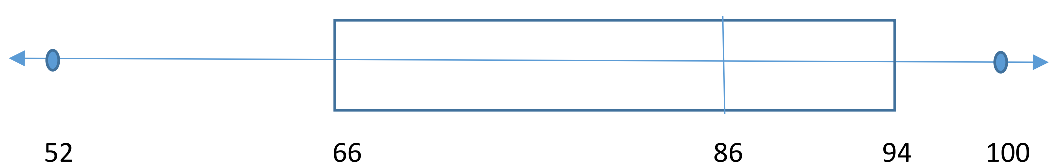
Box and Whisker Plots

The data set in our examples (100, 89, 52, 73, and 86, 94, 62, 70, and 94) has a mean of 80, a median of 86, a mode of 94, and a range of 48. Even though the mean is 80, no one taking the test actually scored an 80. And a large range like 48 means there's a lot of variation in the test scores. A box and whisker plot is a visual aid to help us understand the results. Although these plots are not used as often as the other measures in this section, they are another tool for understanding data sets.

Box and whisker plots use the concept of "quartiles," which are the values that can separate our data values into four equal groups. We already know where the halfway point is, because by definition that is the median. To divide the lower half into quarters, find the median of the lower numbers; to divide the upper half into quarters, find the median of the higher numbers.

So, in our example with (52, 62, 70, 78, 86, 89, 94, 94, 100), 86 is the median and therefore the score that determines the second quartile. There are four numbers lower than 86, and their median is 66 (the mean of 62 and 70). There are four numbers higher than 86, and their median is 94 (the mean of 94 and 94). So, the first quartile is from 52 to 66, second quartile is from 66 to 86, third quartile is from 86 to 94, and the fourth quartile is from 94 to 100.

To draw this, start by marking the lowest value, highest value, and the values marking the quartiles along a number line, so that the distance between these points is to scale. Draw a box around the quartile scores, and mark the high and low scores with an endpoint.



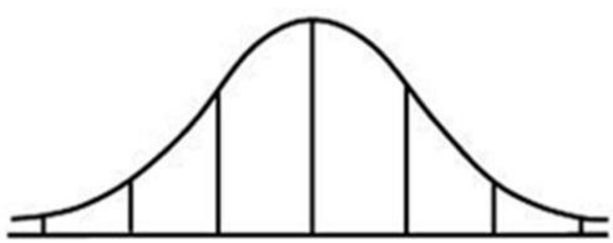
So, now, how to interpret the plot? We know that half the students scored above 86 and half scored below 86. We also know that one-fourth of the students scored between 94 and 100, and one-fourth scored between 52 and 66.

Visually, the median is a lot closer to the highest point than it is to the lowest point, so the scores were clustered more tightly together in that range. There's a large range from highest to lowest, but the larger distances between quartile values show that the scores were not clustered so closely together in the low ranges.

Normal Distribution

A set of data that fits the "normal distribution" is centered around one point in a completely symmetrical way forming a "bell curve." The normal distribution is divided into sections based on the concept of standard deviation, which is more advanced statistics than we'll cover here. There are a few definitions to know about normal distribution, however, that may help you interpret a set of data.

In a normal distribution, the mean, median and mode are all the same value. The data can be clustered very close together under a spikey curve, or more spread out under a flatter curve. However tightly the data is clustered, though, it is distributed symmetrically around the mean. A normal distribution curve can look like this:



Collecting Data

Data can be collected in several ways, and the method used can affect the data's quality and reliability. The sampling method is often the first decision for research. Data should be collected from subjects who are randomly chosen from a large population of representative individuals. As an example of how this can go wrong, suppose a political party wants to predict a presidential election outcome, but they choose to only gather data from Republicans who all live on one street. The results are much different than data gathered randomly from all likely voters in the election.

Experiments, observational studies and surveys all collect data. An experiment includes a treatment group and a control group. A lab researcher can experiment on people with colds by giving one "treatment" group a new cold medicine while the "control" group receives a placebo. In an observational study, a group of subjects is randomly selected and then information is gathered about them. Once the information is gathered, researchers look for correlations, such as whether tall people are more likely to attend basketball games.

Data is often collected by surveys, which can be a written questionnaire or a personal, phone, or social media interview. Surveys may be quicker and cheaper than the other methods. Surveys have to be carefully designed and implemented however. Respondents may overstate their income levels, for example. The field of market research studies in great depth the best ways to collect and analyze information about behavior.

Practice Problems

1. Find the number missing from this sequence: 3, 7, 11, _____, 19, 23, 27.

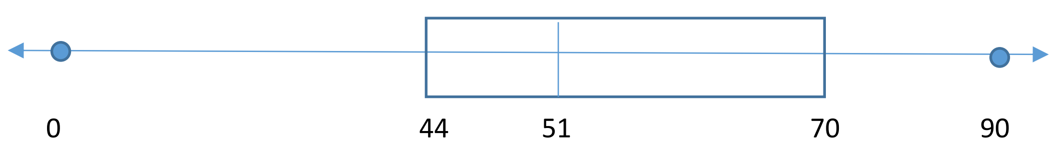
2. Find the number missing from this sequence: $1\frac{9}{16}$, $3\frac{1}{8}$, $6\frac{1}{4}$, _____, 25, 50, 100.

3. Find the number missing from this sequence: 1, 2, 5, _____, 41, 122, 365.

4. The high temperatures for the first ten days in January were: 20, 24, 12, 18, 24, 32, 24, 19, 20 and 17. Find the mean, median, mode and range.

5. Joe's son played in five basketball games and scored the following points per game: 19, 10, 42, 15, and 19. Find his mean, median, mode and range.

6. The following box and whisker plot shows the scores for Mr. Smith's students on a science test.
a) What percentage of students scored above 70? b) What percentage of students scored below 44?
c) What is the median score on the science test? d) What is the highest score on the test?



7. Which of the following statements are true about data in a normal distribution?
a) The mean is the same as the median.
b) The data does not have a mode.
c) The data is distributed symmetrically around the mean.
d) The range is twice the mean.

Answer Key for Practice Problems

1. 15 2. $12\frac{1}{2}$ 3. 14

4. Mean = 21; Median = 20; Mode = 24; Range = 20

5. Mean = 21; Median = 19; Mode = 19; Range = 32

6. a) 25% b) 25% c) 51 d) 90 7. a) and c)

Practice Problems Solved with Explanation

1. Start by finding the difference between the first and second numbers: 7 – 3 = 4. This is also the difference between the next numbers: 11 – 7 = 4. Add 4 to the third number: 11 + 4 = 15, so that may be the missing number. Confirm that the pattern continues: 15 + 4 = 19; 19 + 4 = 23; and 23 + 4 = 27. The missing number is 15.

2. This sequence starts with fractions and ends with whole numbers. That means there isn't a constant amount being added to each number. However, it could mean each number is multiplied by a constant amount. The pattern is most obvious in the higher numbers. 100 = 50 x 2. And 50 = 25 x 2. Divide 25 by 2 and get $\frac{25}{2} = 12\frac{1}{2}$. Notice that as we move from the higher number to the lower number we *divide* by 2, and as we move from lower to higher numbers we *multiply* by 2. We can check each sequence and confirm that 2 is the common ratio, and the missing number is $12\frac{1}{2}$.

3. This sequence starts to increase slowly, then increases more quickly, so there is probably some multiplication. The numbers aren't multiplied by the same number each time, though, so there must be some combination of operations. Again, the pattern is more obvious in the larger numbers, where the intervals seem about triple the previous number. If we start at the beginning and multiply each number by 3, we get 1 x 3 = 3, but the next number is 2, so perhaps we should try (3n -1).

> Try that on the next number: (3n -1) = (3 x 2) - 1 = 5.
> And (3 x 5) – 1 = <u>14</u>, so that may be the missing number.
> Then start with 14: (3 x 14) – 1 = 41.
> (3 x 41) – 1 = 122
> (3 x 122) – 1 = 365

The pattern works throughout the whole sequence, so the missing number is 14.

4. Mean = $\frac{total\ of\ all\ values}{quantity\ of\ values}$. For this set of data, that is $\frac{20+24+12+18+24+32+24+19+20+17}{10} = \frac{210}{10}$ = 21.

To find the median, arrange the data from low to high: 12, 17, 18, 19, <u>20, 20</u>, 24, 24, 24, 32. The median is the middle value. Since there is an even number of values, the median is the average of the two middle values, 20 and 20. Therefore the median is 20.

The mode is the value that occurs most often. Since three days had temperature values of 24, that is the most frequent value. The mode is 24.

The range is the difference between the highest and lowest value. The highest value is 32; the lowest value is 12. The range is therefore 32 – 12 = 20.

5. The mean for this data set is $\frac{19+10+42+15+19}{5} = \frac{105}{5}$ = 21.

Arranging the values from low to high gives us: 10, 15, <u>19</u>, 19, 42. The middle value, the median, is 19.

The value 19 occurs twice, more than any other value, and so it is the mode.

The range is the highest value minus the lowest value: 42 – 10 = 32.

6. a) The box and whisker plot charts the data around the data points that are the lowest, highest, middle and the two quartiles. The data point at 70 is marks the upper quartile, which means that 25% of the scores are above 70.

6. b) Likewise, the data point 44 marks the lower quartile, so 25% of the scores are below 44.

6. c) The median score is marked by the point in the middle, which is 51.

6. d) The highest score on the test is 90.

7. For data in a normal distribution, the mean is the same as the median, and the data is distributed symmetrically around that value. That makes statements a) and c) true. Although some data does not have a mode, in a normal distribution the mode will also be the same as the mean. There is no relationship between the range and the mean, so only a) and c) are true.

