CSUSM
STEM Success Center

Math 101 Chapter 5/Section 5.1&5.2: Topic: Data Summary, Presentation, & Normal Distribution

Write the definition of the following terms:

1. Mean (and formula):

2. Median:

3. Mode:

4. Distribution:

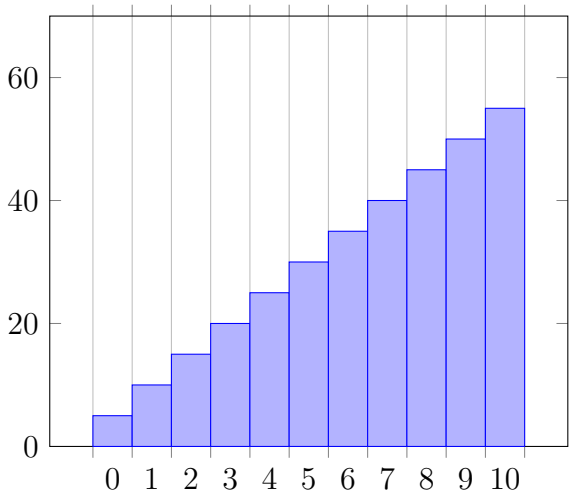
5. The 68-95-99.7% Rule:

State the four properties of a normally distributed data set:

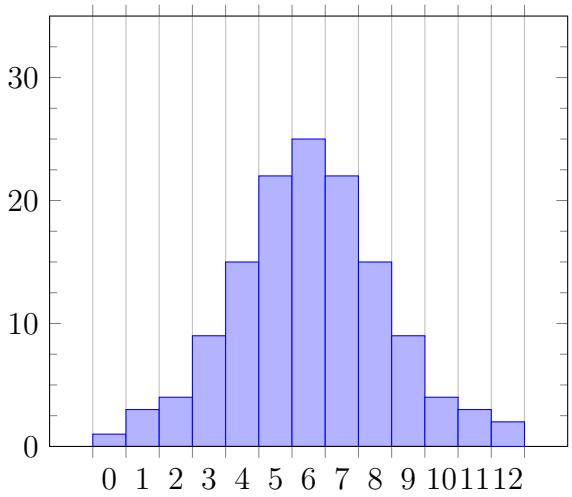
-
-
-
-

Given the following distributions $A-D$, indicate whether or not each of these are normally distributed.

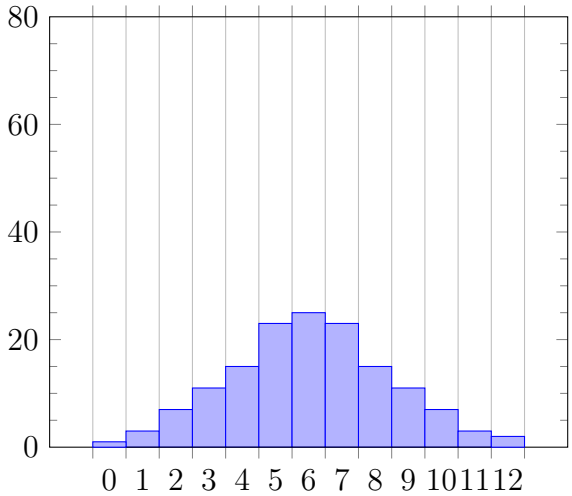
A



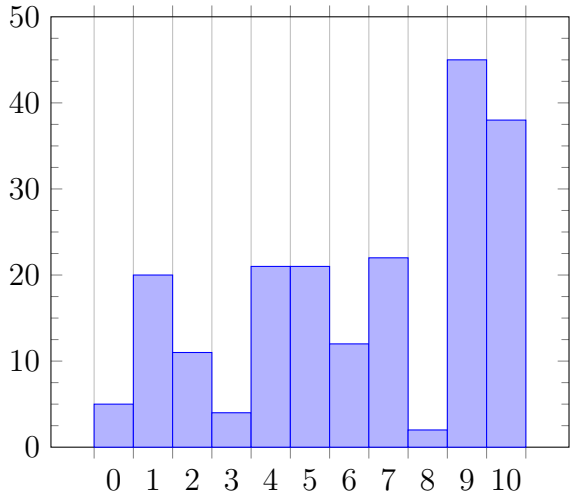
B



C



D



Given the data set: 6, 2, 3, 5, 4, 9, 12, 27.
Find the following:

1. Mean:

2. Median:

3. Mode:

Given the data set:

Number	1	2	3	7	8
Frequency	1	2	5	4	2

Find the following:

1. Mean:

2. Median:

3. Mode:

Answer the following questions:

1. Suppose that a study on the number of hours college students spend watching Netflix per month is normally distributed with a mean of 48 hours and standard deviation of 3.24 hours. What does the 68-95-99.7% rule tell us about the number of hours spent on Netflix in a given month?

2. Suppose a class of 45 students take an exam and the recorded scores are normally distributed with a mean score of 76 points and a standard deviation of 5.6 points. How many scores will fall between 59.2 and 92.8?

Solutions

Write the definition of the following terms:

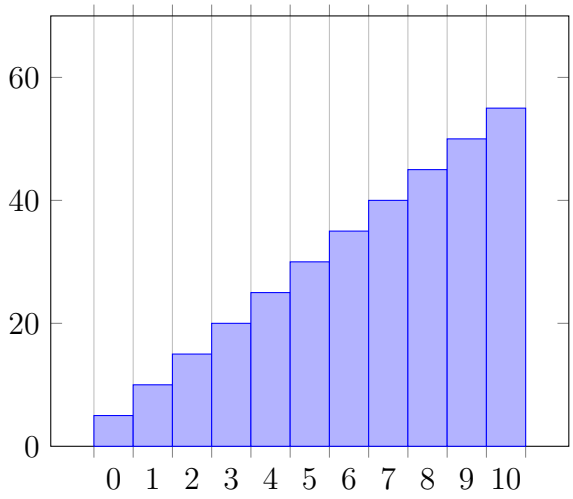
1. **Mean (and formula):** The average of a set of data points. To compute the mean, find the sum of all data points and divide by the number of data points in the set. In other words, add up all of the numbers and divide by the number of numbers.
2. **Median:** The middle data point. To find the middle number, list all data points from smallest to largest. The number in the middle is the median. If the median falls between two data points, take the average of the two middle points (*i.e.* add the two data points and divide by 2).
3. **Mode:** The data point that occurs more frequently than any other data point.
4. **Distribution:** A set of data that measures the frequency that a data point occurs
5. **The 68-95-99.7% Rule:** If a data set is normally distributed then
   - 68% of the data falls within one standard deviation from the mean. This means that 68%/2=34% of the data falls within one standard deviation above the mean and 34% of the data falls within one standard deviation below the mean.
   - 95% of the data falls withing two standard deviations from the mean. This means that 95%/2=47.5% of the data falls within two standard deviations above the mean and 47.5% of the data falls within two standard deviations below the mean.
   - 99.7% of the data falls withing three standard deviations of the mean. This means that 99.7%/2=49.85% of the data falls within three standard deviations above the mean and 49.85% of the data falls within three standard deviation below the mean.

State the four properties of a normally distributed data set:

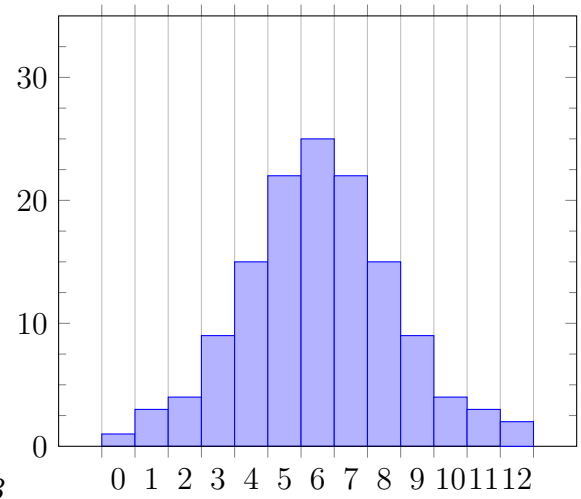
- The graph has a bell-shaped curve.
- The graph is symmetrically shaped below and above the mean.
- The mean is equal to the median.
- Most of the data tends to be clustered near the mean.

Given the following distributions $A - D$, indicate whether or not each of these are normally distributed.

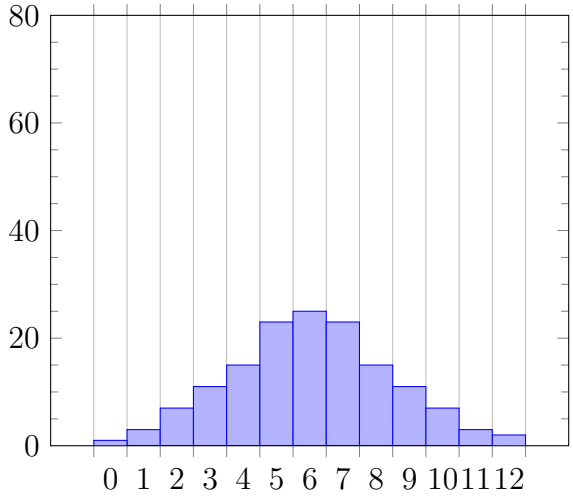
A



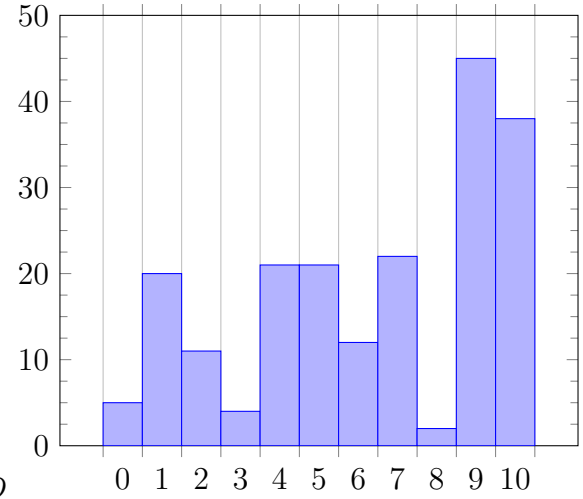
B



C



D



- A is not normally distributed since it does not have a bell shaped curve
- B is normally distributed since it has a bell shaped curve
- C is normally distributed since it has a bell shaped curve
- D is not normally distributed since it does not have a bell shaped curve

Given the data set:

6, 2, 3, 5, 4, 9, 12, 27

Find the following:

1. **Mean:**

$$\text{Mean} = \frac{6+2+3+5+4+9+12+27}{8} = \frac{17}{2} = 8.5$$

2. **Median:** We will start by ordering the numbers from smallest to largest as follows:

2, 3, 4, 5, 6, 9, 12, 27

We have 8 numbers on this list, that means our median will fall between two numbers. In this case, our median is between 5 and 6. The median will be the average of the two. The average of the two is $5 + 6/2 = 11/2 = 5.5$. Thus, 5.5 is the median.

3. **Mode:** Since no number occurs more frequently than the other, we have no mode here.

Given the data set:

Number	1	2	3	7	8
Frequency	1	2	5	4	2

Find the following:

Remark. For the following calculations, it might help to write out the numbers as a list (with each number listed as many times its frequency) as follows:

1, 2, 2, 3, 3, 3, 3, 3, 7, 7, 7, 7, 8, 8

The frequency for each number tells us how many times that number appears in the data set. This tells us that if we add up the frequencies, we will get the number of numbers. So there are 1+2+5+4+2=14 numbers. You can equivalently find this out by listing out the numbers (as seen above) and counting all of the numbers. However, with more numbers and larger frequencies, this could be tedious.

1. **Mean:** Since 3 has a frequency of 5, it contributes to our list 5 times, so when computing the average, instead of adding each of the numbers, we take each number, multiply it by its corresponding frequency, add them all up, and then divide by the number of numbers which we found to be 14. Observe

$$\frac{1\times 1+2\times 2+3\times 5+7\times 4+8\times 2}{1+2+5+4+2} = \frac{64}{14} \approx 4.57$$

2. **Median:** Recall that we have 14 numbers, so since that's even, the median will fall in between two numbers, the $7th$ and $8th$ numbers which are both 3, so then the median is 3.

3. **Mode:** The number that appears the most is the number with the highest frequency. Thus, the mode is 3.

Answer the following questions:

1. **Suppose that a study on the number of hours college students spend watching Netflix per month is normally distributed with a mean of 48 hours and standard deviation of 3.24 hours. What does the 68-95-99.7% rule tell us about the number of hours spent on Netflix in a given month?**

   - Since our data is normally distributed, 68% of the data lies withing one standard deviation of the mean. So 68% of our data lies within

   $$48 - 3.24 = 44.76 \qquad \text{and} \qquad 48 + 3.24 = 51.24$$

   - Since our data is normally distributed, 95% of the data lies withing two standard deviations of the mean. So 95% of our data lies within

   $$48 - 2 \times 3.24 = 41.52 \qquad \text{and} \qquad 48 + 2 \times 3.24 = 54.48$$

   - Since our data is normally distributed, 99.7% of the data lies withing three standard deviations of the mean. So 99.7% of our data lies within

   $$48 - 3 \times 3.24 = 38.28 \qquad \text{and} \qquad 48 + 3 \times 3.24 = 57.72$$

2. **Suppose a class of 45 students take an exam and the recorded scores are normally distributed with a mean score of 76 points and a standard deviation of 5.6 points. How many scores will fall between 59.2 and 92.8?**

   First observe that scores of 59.2 points are 76-59.2=16.8 points below the mean and scores of 92.8 are 92.8-76=16.8 points above the mean. Note that 16.8 points represents $16.8/5.6 = 3$ standard deviations. This tells us that the range of scores between 59.2 points to 92.8 points lie within 3 standard deviations of the mean. Given that the data is normally distributed, using the 68-95-99.7% rule, 99.7% of the scores will fall between 59.2 points to 92.8 points.