Population Standard Deviation

\[ \sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N}} \]

Sample Standard Deviation

\[ S = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N - 1}} \]

A normal distribution is a set of data that follows a symmetrical, bell-shaped curve. Most of the data is relatively close to the mean. As the distance from the mean increases on both sides of the mean, the number of data points decreases. The empirical rule states that, for data that is distributed normally,

- approximately 68% of the data will be located within one standard deviation on either side of the mean;
- approximately 95% of the data will be located within two standard deviations on either side of the mean; and
- approximately 99.7% of the data will be located within three standard deviations on either side of the mean.
Jesse is the manager of a guitar shop. He recorded the number of guitars sold each week for a period of 10 weeks. His data is shown in this table.

<table>
<thead>
<tr>
<th>Week</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td># Guitars Sold</td>
<td>12</td>
<td>15</td>
<td>20</td>
<td>8</td>
<td>15</td>
<td>18</td>
<td>17</td>
<td>21</td>
<td>10</td>
<td>24</td>
</tr>
</tbody>
</table>

a. What is the mean of Jesse’s data?

\[
\bar{x} = \frac{12 + 15 + \ldots + 24}{10} = 16
\]

b. What are the variance and the standard deviation of Jesse’s data?

\[
\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}
\]

\[
\sum = 228 \quad \text{variance} = \frac{228}{10} = 22.8
\]

\[
\sigma = \sqrt{22.8} = 4.78
\]

(c) \( \frac{\bar{x} + 2\sigma}{\bar{x} - 2\sigma} \) \( \text{outlier is beyond these values} \)

\[
16 + 2(4.78) = 25.56 \quad \text{No outlier}
\]

\[
16 - 2(4.78) = 6.4 \quad \text{No outlier}
\]
12) Anne is the regional sales manager for a chain of guitar shops. She recorded the number of
guitars sold at two stores in her region each week for one year (52 weeks). These histograms
show the data Anne collected.

\[ \bar{x} = 17.5 \]

\[ \sigma \approx 6.5 \]

\[ \text{Range} = 8.75 \]

\[ \text{Range} = \frac{35 - 0}{4} = 8.75 \]

a. Estimate the mean and standard deviation for each store.

b. What is the range of possible means for each store?

   \[ 15 - 20 \]

c. Explain why the empirical rule is or is not a good fit for each store.

   Store 1 is a good fit because it's bell shaped.
This frequency table shows the heights for Mrs. Quinn's students.

<table>
<thead>
<tr>
<th>Height (in inches)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>1</td>
</tr>
<tr>
<td>43</td>
<td>2</td>
</tr>
<tr>
<td>44</td>
<td>4</td>
</tr>
<tr>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>46</td>
<td>4</td>
</tr>
<tr>
<td>47</td>
<td>2</td>
</tr>
<tr>
<td>48</td>
<td>1</td>
</tr>
</tbody>
</table>

What is the approximate standard deviation of these data?

A. 1.0 inches
B. 1.5 inches
C. 2.5 inches
D. 3.5 inches
Katrina took 10 random samples of the winning margins for each of two professional basketball teams. The sample size was 4. The distributions of the sample means are shown in these histograms.

Which is the best estimate of the standard deviation for both samples?

A. Team 1: 3.75 points; Team 2: 2.2 points
B. Team 1: 7.4 points; Team 2: 4.4 points
C. Team 1: 15 points; Team 2: 8.8 points
D. Team 1: 10 points; Team 2: 10 points
John took 10 random samples of the winning margins for each of two professional basketball teams. The distributions of the sample means are shown in these histograms.

Based on John’s data, which statement is MOST likely true?

A. Both the sample mean and the sample standard deviation are greater for Team 1 than for Team 2.
B. The sample means for both teams are equal, but the sample standard deviation for Team 1 is greater.
C. The sample means for both teams are equal, but the sample standard deviation for Team 2 is greater.
D. Both the sample mean and the sample standard deviation are greater for Team 2 than for Team 1.
The mean and standard deviation of the sample means can be used to estimate the mean and standard deviation of the population. If the mean of the sample means is \( \bar{X} \) and the standard deviation of the sample means is \( S_{\bar{X}} \), the population mean can be estimated by using \( \bar{X} \) and the standard deviation can be estimated by using \( S_{\bar{X}} \sqrt{n} \).

In a set of 10 random samples of winning scores for games played in a professional basketball league, the sample size is 6, the sample mean is 97.5 points, and the sample standard deviation is 5.2 points. Which expression represents the estimated standard deviation of all the winning scores?

A. \( \frac{5.2}{\sqrt{10}} \)
B. \( 5.2\sqrt{10} \)
C. \( \frac{5.2}{\sqrt{6}} \)
D. \( 5.2\sqrt{6} \)
The median-median line is a way to estimate a line of best fit that involves relatively simple calculations. Since it involves using medians, it is also somewhat resistant to the effect of outliers in the data. To calculate a median-median line, order the data from the lowest to the greatest value of the \( x \)-coordinate. Order data points that have the same value of \( x \) from the least to the greatest value of the \( y \)-coordinate. Next, use this ordering to divide the data into three equal groups. Find the median \( x \)-coordinate value for each of the three groups (low, middle, high values). Then find the \( y \)-value associated with each of these data points. Find the equation of the line containing the two outside points (the points from the low- and high-value sets). Then adjust the position of the line by moving it \( \frac{1}{3} \) of the way toward the middle point.

Tina collected the height in inches and the shoe size of a random sample of nine boys in her class. Her data is shown in this list.

\[
\begin{array}{c|c}
(64, 7) & (69, 11) \\
(61, 7) & (67, 8.5) \\
(65, 8) & (68, 10) \\
(69, 9.5) & (69, 11.5) \\
(73, 11) & \\
\end{array}
\]

In each ordered pair, the \( x \)-coordinate is the height in inches and the \( y \)-coordinate is the shoe size.

Find the equation of the median-median line for Tina’s data.

\[
m = \frac{11 - 7}{69 - 64} = \frac{4}{5} \quad y = \frac{4}{5} x + b
\]

\[
11 = \frac{4}{5} (69) + b
\]

\[
y = \frac{4}{5} x - 44.2
\]

\[
\frac{11}{5} = \frac{55}{5} + b
\]

\[
-44.2 = -55 \frac{2}{5}
\]

\[
b = 2
\]

\[
\text{Step 5: Adjust the line:}
\]

\[
\begin{array}{c|c}
(67, 8.5) & y = \frac{4}{5} (67) - 44.2 \\
(67, 9.4) & \end{array}
\]

\[
8.5 - 9.4 = \frac{-1}{3}
\]

\[
y = \frac{4}{5} x - 44.2 - 1.3
\]

\[
y = \frac{4}{5} x - 44.5
\]
12) This graph plots the number of wins in the 2006 and 2007 seasons for a sample of professional football teams. What is the equation of the median-median line?

\[(x_{md}, y_{md}) = (9, 10) \quad (10, 7) \quad (11, 13)\]

\[m = \frac{13 - 10}{11 - 9} = \frac{3}{2} = 1.5\]

\[b = 13 - 1(10) = 3\]

\[y = 1.5x + 3\]

13) For the same scatterplot in #12, the equation of the linear regression model is \[y = 1.10x - 2.29\]. Based on this model, what is the predicted number of 2007 wins for a team that won 5 games in 2006?

\[y = 1.10(5) - 2.29 = 2.29\]

\[y = 3.29\]
14) This graph shows the expected income from sales vs. price per issue for a new magazine.

Which equation models these data?

A. \( y = -5.1x^2 + 34.4x - 3.0 \)
B. \( y = 5.1x^2 - 34.4x + 3.0 \)
C. \( y = -34.4x^2 + 5.1x - 3.0 \)
D. \( y = 34.4x^2 - 5.1x + 3.0 \)

\[ y = a(x-h)^2 + k \quad (h, k) \approx (3, 55) \]

\[ y = -5(x-3)^2 + 55 \quad a \approx -5 \]
All students in a gym class were required to complete an obstacle course. The times to complete the course were normally distributed with a mean of 46 seconds and a standard deviation of 3.5 seconds. Fill in the normal distribution chart.

Find the probability that a student completed the course between 42.5 and 49.5 seconds.

\[ \text{prob.} = 0.68 \]
Which set has the largest standard deviation?

<table>
<thead>
<tr>
<th>Set 1</th>
<th>Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 6 6 6 11</td>
<td>1 5 6 7 11</td>
</tr>
</tbody>
</table>

A) Set 1, because 5 and 7 in set 1 are farther from 6 than 6 and 6 in set 2.
B) Set 2, because 6 and 6 in set 2 are farther from 6 than 5 and 7 in set 1.
C) Set 1, because 6 and 6 in set 1 are farther from 6 than 5 and 7 in set 2.
D) Neither, because set 1 and set 2 have the same standard deviation.
E) Set 2, because 5 and 7 in set 2 are farther from 6 than 6 and 6 in set 1.

Title: May 5-10:13 AM (12 of 19)
Which is true of the data shown in the histogram?

I. The distribution is approximately symmetric.  
II. The mean and median are approximately equal.  
III. The median and IQR summarize the data better than the mean and standard deviation.

A) I only  
B) III only  
C) I and II  
D) I and III  
E) I, II, and III
Two sections of a class took the same quiz. Section A had 15 students who had a mean score of 80, and Section B had 20 students who had a mean score of 90. Overall, what was the approximate mean score for all of the students on the quiz?

A) It cannot be determined.
B) 85.0
C) 85.7
D) 84.3
E) none of these

\[
\frac{15(80) + 20(90)}{35} = 85.7
\]
The best estimate of the standard deviation of the men's weights displayed in this dotplot is

\[ \frac{930 - 100}{4} = 32.5 \]

\[ \frac{230 - 100}{6} = 21.6 \]

C) 25
A town's average snowfall is 33 inches per year with a standard deviation of 11 inches. How many standard deviations from the mean is a snowfall of 66 inches?

A) About 0.33 standard deviations below the mean
B) About 3.00 standard deviations above the mean
C) About 2.00 standard deviations above the mean
D) About 0.33 standard deviations above the mean
E) About 3.00 standard deviations below the mean

\[ 33 + 11 + 11 + 11 = 66 \]

3 standard above

\[ \frac{x - \mu}{\sigma} = \frac{66 - 33}{11} = 3 \]
A town's average snowfall is 48 inches per year with a standard deviation of 6 inches. Using a Normal model, what values should border the middle 68% of the model?

A) 60 inches and 36 inches
B) 51 inches and 45 inches
C) 48 inches and 45.96 inches
D) 34 inches and 42 inches
E) 49.2 inches and 46.8 inches

Empirical rule: 68% within 1 std. dev. of mean
The weights of people in a certain population are normally distributed with a mean of 158 lb and a standard deviation of 25 lb. What is standard deviation of the sampling distribution of the mean for samples of size 9?

\[ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{25}{\sqrt{9}} = 8.33 \text{ lb} \]

a) 2.78 lb  b) 8.33 lb  c) 25 lb

25) The mean annual income for adult women in one city is $28,520 and the standard deviation of the incomes is $5700. The distribution of incomes is skewed right. Determine the standard deviation of the sampling distribution of the mean for samples of size 110.

\[ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5700}{\sqrt{110}} = \$543.5 \]

a) $543  b) $5700  c) $52
26) Circle the graph with the larger standard deviation