Bozeman Science: Statistics for Science Video Sheet

www.bozemanscience.com/statistcs-for-science

1. According to Paul Andersen, statistics is basically, ___collecting___, ___organizing___, ___analyzing___, ___interpreting___, and ___presenting___ data so people can use it.

2. Please define each of the statistical terms below:

(a) sample size (*n*) - ___number of observations___

(b) mean / (x) - ___average___

(c) median - ___the midpoint___

(d) range - ___difference between extremes (highest value and lowest value)___

3. Complete the basic statistic calculations Paul Andersen uses in the video using the data set below. *Show all of your work and round your answer to the nearest tenth.*

Data: 2, 6, 3, 2, 5, 9, 8

(a) sample size (*n*) - ___7___

(b) mean / (x) - ___5___

(c) median - ___5___

(d) mode = ___2___

(e) range - ___7___

4. The formula for Degrees of Freedom = ___n-1___

In your own words, what does Degrees of Freedom mean?

Degrees of freedom is the number of values in the final calculation of a statistic that are free to vary.

Basic Statistics

Formulas:

Mean $\bar{x}$ $\bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i$ = = sum of all data points divided by the number of data points

Median = middle value that separates the greater and lesser halves of a data set

Mode = value that occurs most frequently in a data set

Range = value obtained by subtracting the smallest observation (sample minimum) from the greatest (sample maximum)

Standard Deviation = $\sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$ where $\bar{x}$ = mean and n = size of the sample

- measures the average of the deviation between each measurement in the sample and the sample mean.

Standard Error = $SE_{\bar{x}} = \frac{s}{\sqrt{n}}$ where S = the standard deviation and n = size of sample

- measures the accuracy with which a sample represents a population.

Chi-Square = $\chi^2 = \Sigma \frac{(o - e)^2}{e}$ = where *o* is the observed results and *e* is the expected results

- used to determine whether there is a *significant difference* between the observed and expected results.

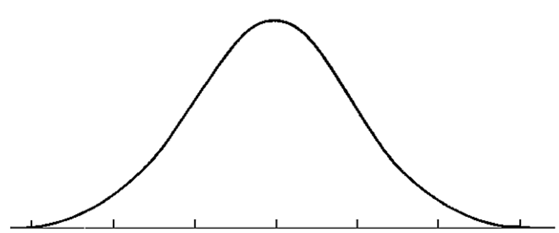
Background Information

Does the data tend to be more or less symmetrically distributed across a range with most measurements around the center of the distribution.?

YES or **NO**

If so, this is a characteristic of a **normal distribution**. Most statistical methods covered in this activity apply to data that are **normally distributed**. Normally distributed data usually take on the shape of a "bell curve" as indicated in the diagram below. Other types of distributions require either different kinds of statistics or transforming data to make them normally distributed.

Figure 1. *The Bell Curve*



The first part of this activity involves determining the **mean**, **median**, **mode**, **range**, **standard deviation** and **standard error** along with introducing you to the Grid-In style answer sheets the College Board uses for numerical answers. Some of this information you should remember from middle school = ***prior knowledge***.

Measures of Average: Mean, Median, and Mode

Mean $\bar{x}$

You calculate the sample mean (*also referred to as the average or arithmetic mean*) by summing all the data points in a data set (Σx) and then dividing this number by the total number of data points (*n*):

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i$$

Median

When the data are ordered from the largest to the smallest, the **median** is the *midpoint of the data*. It is not distorted by extreme values, or even when the distribution is not normal. For this reason, it may be more useful for you to use the median as the main descriptive statistic for a sample of data in which some of the measurements are extremely large or extremely small. To determine the median of a set of values, you first arrange them in numerical order from lowest to highest. The middle value in the list is the median. If there is an even number of values in the list, then the median is the mean of the middle two values.

Mode

The mode is the value that appears most often in a sample of data.

Practice Problems:

1. Scientists and researchers typically collect data on a sample of a population and use these data to draw *conclusions* or make *inferences* (*connections*). An example of such a data set is shown below. The data below represents student test scores on the first AP Biology exam given at North Salem University.

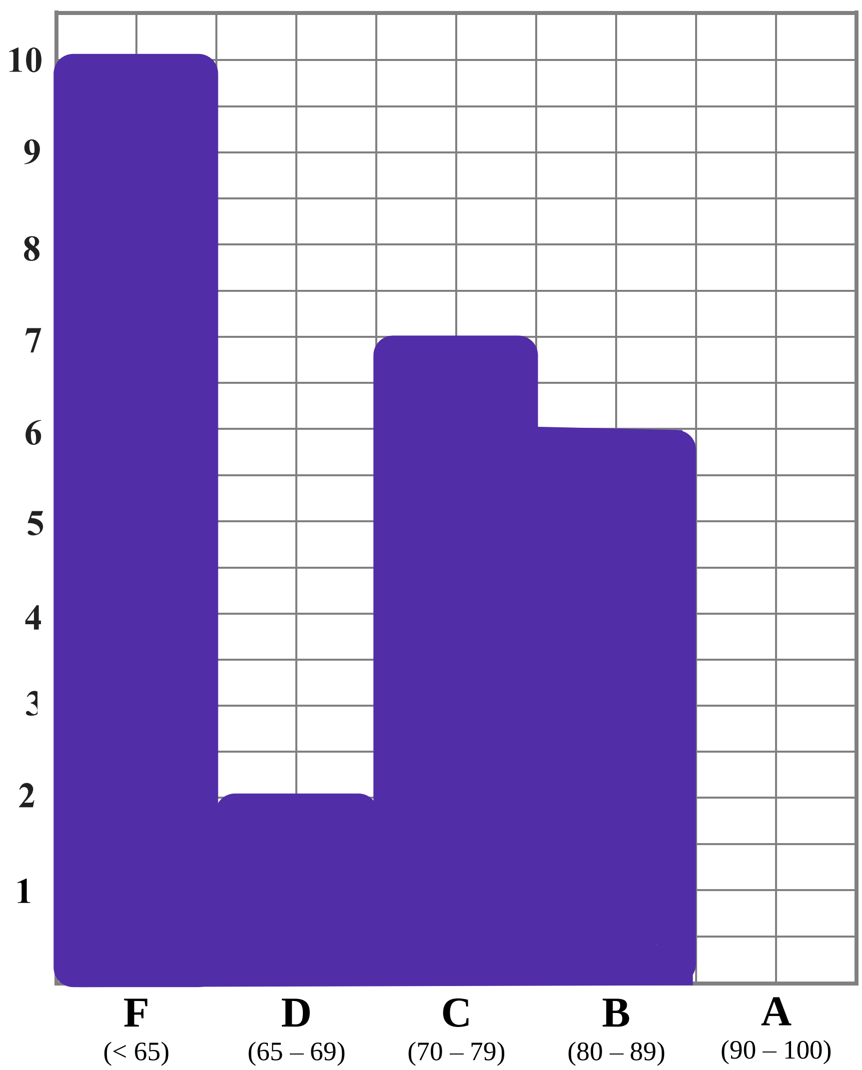
80	**61**	**60**	**53**	**59**	**62**	**71**	**62**	**86**
71	**76**	**80**	**63**	**82**	**62**	**75**	**87**	**75**
69	**52**	**86**	**74**	**54**	**69**	**71**		

How would you describe the data above, and what does it tell you about the population of AP Biology students at North Salem University? These are difficult questions to answer by simply looking at a set of numbers. One of the first steps in analyzing a data set like the one shown above is to *organize* it into a Data Table located on the next page and make it more appealing to the eye by graphing it.

Table 1

Score Range	Number of Students
A (90 – 100)	0
B (80 – 89)	6
C (70 – 79)	7
D (65 – 69)	2
F (< 65)	10

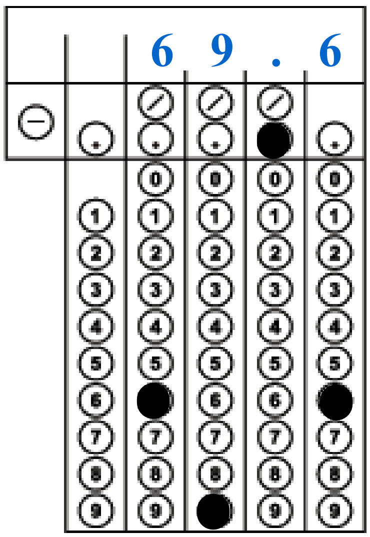
(b) In the grid below, construct a **histogram** (*a bar graph that groups data into ranges*) based upon the table above.



(c) Determine the **mean**, **median**, **mode** and **range** for the data representing the test scores on the first AP Biology exam given at North Salem University and fill you answers in on the sample Grid-In answer sheet provided.

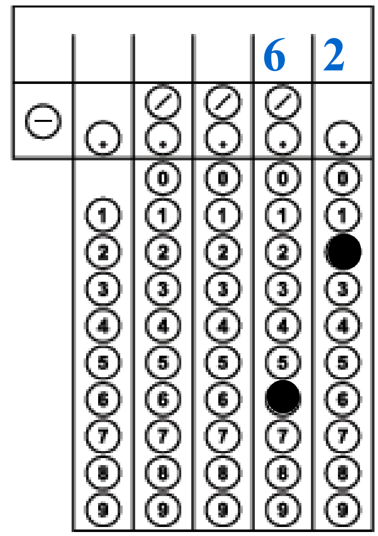
Mean: 69.6

Round answer to the nearest tenth.



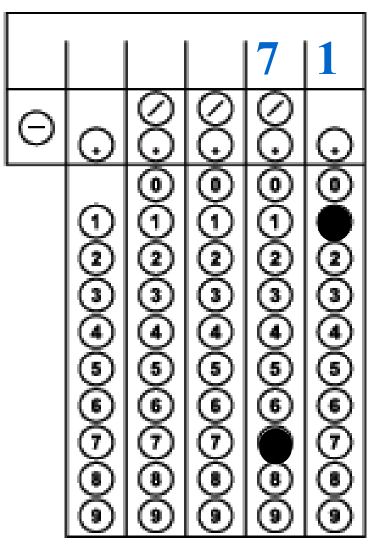
Mode: 62, 71

Round answer to the nearest whole number



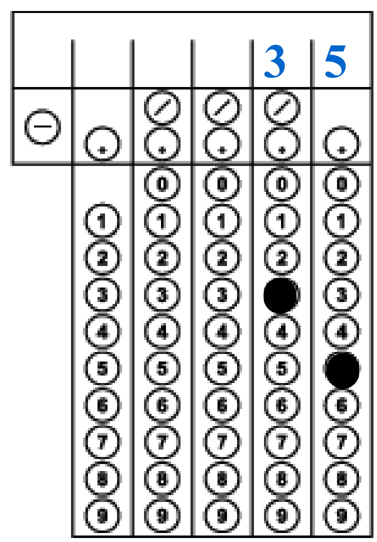
Median: 71

Round answer to the nearest whole number.



Range: 35

Round answer to the nearest whole number.



Bozeman Science: Standard Deviation Video Sheet

www.bozemanscience.com/standard-deviation

1. Normal Distribution is also referred to as a ___bell___ curve.

2. In statistics, how is the mean represented? $\bar{x}$

3. Standard deviation measures the ___spread___ or ___variation___ in a bell-shaped curve.

4. About ___68___% percentage of data should be **ONE** standard deviation above and below the mean.

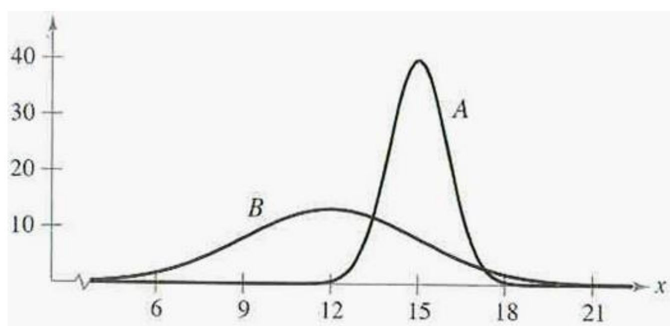
5. About ___95___% percentage of data should be **TWO** standard deviations above and below the mean.

6. About ___99___% percentage of data should be **THREE** standard deviations above and below the mean.

7. Standard Deviation will vary depending on ___the data that you collect___.

8. According to the graph to the right, which data set has a smaller standard deviation? ___A___

9. According to the graph to the right, which data set has a higher standard deviation? ___B___



10. State what each element represents in the standard deviation formula.

(a) n = ___sample size / total number of data points___

(b) $\bar{x}$ = ___mean (average)___

(c) å = ___summation symbol___

(d) $n - 1$ = ___degrees of freedom___

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$$

11. Complete the standard deviation question Paul Andersen uses in the video using the data set below.
Show all of your work and round your answer to the nearest tenth.

Data: 0 2 4 5 7

S = 2.7

Sometimes calculating the standard deviation of a set of numbers can be made easier by using a table.

12. Grades from recent AP Biology exam are located to the right: 96, 96, 93, 90, 88, 86, 86, 84, 80, 70

(a) Calculate the mean, median, mode and range for the data above. *Give your answer to the nearest whole number.*

Mean ($\bar{x}$): 87 **Median:** 87 **Mode:** 86, 96 **Range:** 26

(b) Calculate the standard deviation by filling in the table below. Give your answer to the nearest tenth.

Exam Score (x)	Mean ($\bar{x}$)	$x-\bar{x}$	$(x-\bar{x})^2$
96	87	9	81
96	87	9	81
93	87	6	36
90	87	3	9
88	87	1	1
86	87	-1	1
86	87	-1	1
84	87	-3	9
80	87	-7	49
70	87	-17	289
		$\Sigma(x-\bar{x})^2 \rightarrow$	$\boxed{557}$

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

$$\sqrt{\frac{589}{9}}$$

$$\sqrt{61.88}$$

S = 7.86 = 7.9

Measures of Variance: Standard Deviation

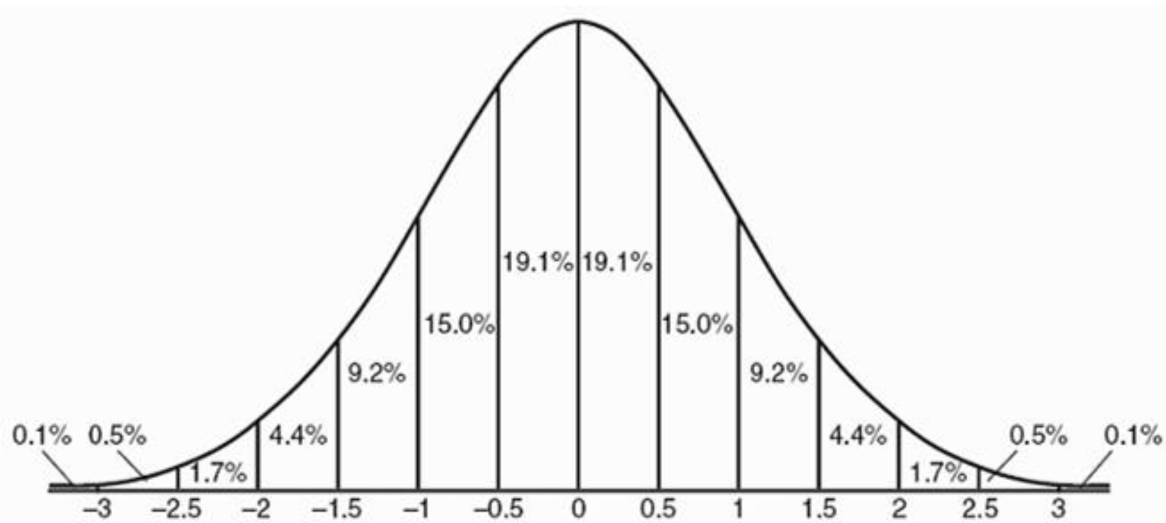
Standard deviation is essentially the average of the deviation between each measurement in the sample and the sample mean (x). The formula for calculating the sample standard deviation follows:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

Calculation Steps:

1. Calculate the mean (x) of the sample.
2. Find the difference between each measurement (x) in the data set and the mean (x) of the entire set: $(xi - x)$
3. Square each difference to remove any negative values: $(xi - x)^2$
4. Add up (sum, Σ) all the squared differences: $\Sigma\ (xi - x)2$
5. Divide by the **degrees of freedom** (df), which is 1 less than the sample size (n – 1): $\frac{\Sigma\ (xi - x)^2}{(n - 1)}$
6. Take the square root to calculate the standard deviation (s) for the sample.

Standard Deviations



So what does this mean?

For normally distributed data, 68% of data points lie between **±1** standard deviation of the mean, 95% of data points lie between **±2** standard deviations of the mean and 99% of data points lie between **±3** standard deviation of the mean.

Bozeman Science: Standard Error Video Sheet

www.bozemanscience.com/standard-error

1. Why is standard error one of Paul Andersen's favorite statistics?

Standard error shows how good your data is.

2. State what each element represents in the standard error formula.

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

(a) s = **standard deviation**

(b) $\sqrt{n}$ = **square root of the sample size**

3. *Review*: What does standard deviation measure?

Standard deviation measures the spread or how far the data is spread out.

4. Complete the standard error question Paul Andersen uses in the video using the data sets and table below to help you calculate the standard deviation first for Data Set 2 and then Data Set 1.
Show all of your work and round your answer to the nearest thousandth.

(a) Data Set 2: 10.9, 11.9, 12.2, 12.2, 12.9, 12.6, 12.3, 12.3, 12.5, 10.2

Data Set 2 (x)	Mean ($\bar{x}$)	$x-\bar{x}$	$(x-\bar{x})^2$
10.9	12	-1.1	1.21
11.9	12	-0.1	0.01
12.2	12	0.2	0.04
12.2	12	0.2	0.04
12.9	12	0.9	0.81
12.6	12	0.6	0.36
12.3	12	0.3	0.09
12.3	12	0.3	0.09
12.5	12	0.5	0.25
10.2	12	-1.8	3.24
120		$\Sigma(x-\bar{x})^2$	$\boxed{6.14}$

Standard Deviation:

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}$$

$$\sqrt{\frac{6.14}{9}}$$

$$\sqrt{0.68}$$

S = 0.8259 = 0.826

Standard Error:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$\frac{0.826}{\sqrt{10}}$$

$$\frac{0.826}{3.16}$$

SE_X = 0.2614 = 0.261

Standard Deviation:

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$$

$$\sqrt{\frac{18}{1}}$$

(b) Data Set 1: 9, 15

Data Set 1 (x)	Mean ($\bar{x}$)	$x-\bar{x}$	$(x-\bar{x})^2$
9	12	-3	9
15	12	3	9
		$\Sigma(x-\bar{x})^2$	18

S = 4.2426 = 4.243

Standard Error:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

$$SE_x = 3.000 = 3.00$$

5. Common practice is to add *standard error bars* to graphs, marking one standard error above & below the sample mean (see graph to the right). These give an impression of the *precision* of estimation of the mean, in each sample. Construct a bar graph with standard error bars of the mean for data sets 1 and 2 on the graph below.