Statistics vs. statistics

Question: What is Statistics (with a capital S)?

Definition: Statistics is the science of collecting, organizing, summarizing and interpreting data.

Note: There are 2 main ways to summarize data…

1. With a picture or a graph
2. With a number or small collection of numbers that tell you something about the data

Statistics vs. statistics

Note: There are 2 main ways to summarize data…

1. With a picture or a graph
2. With a number or small collection of numbers that tell you something about the data

Question: What is a statistic (with a lowercase s)?

Definition: A statistic is a number that can be calculated from a set of data that summarizes the data in some way

Example of statistics

Ex: The following data gives the ages of the students in one of my previous statistics classes.

Student Ages									
36	21	19	26	29	29	22	30	22	21
24	20	26	21	20	26	20	18	24	22
19	40	19	24	16	20	23	21	46	21
19	20	21	46	24	20	19	21	22	34

Come up with some statistics for this data set.

Chapter 3: Numerically Summarizing Data

Sec. 3.1: Measures of Central Tendency (measures of center)

Sec. 3.2: Measures of Dispersion (measures of spread)

Sec. 3.1: Measures of Central Tendency (measures of center)

Sec. 3.1: Measures of Central Tendency (measures of center)

Why would you want to measure the center of data? Story…

Sec. 3.1: Measures of Central Tendency (measures of center)

There are 4 ways of measuring the center…

1) Name: MEAN (or AVERAGE)

   Symbol: $\bar{x}$

   How to calculate: Add all of the data points and divide by the total number of data points

Sec. 3.1: Measures of Central Tendency (measures of center)

There are 4 ways of measuring the center…

2) Name: MEDIAN

Symbol: $\tilde{x}$

How to calculate:

1. Put the data in increasing order
2. If n is odd, the median is the middle number
3. If n is even, the median is the average of the 2 middle numbers

Sec. 3.1: Measures of Central Tendency (measures of center)

There are 4 ways of measuring the center…

3) Name: MODE

How to calculate:

- The number that appears the most often
- There can be more than one mode
- If no number repeats, there is no mode

Sec. 3.1: Measures of Central Tendency (measures of center)

There are 4 ways of measuring the center…

4) Name: MIDRANGE

How to calculate:

- The average of the lowest and highest data points (i.e. add the lowest and highest numbers and divide by 2)

Sec. 3.1: Measures of Central Tendency (measures of center)

Ex 1: Find the mean, median, mode and midrange of the following data set…

Data: 12, 4, 15, 12, 10, 10, 12, 2, 19, 10, 4

Sec. 3.1: Measures of Central Tendency (measures of center)

Ex 2: Find the mean, median, mode and midrange of the following data set…

Data: 23, 17, 8, 19, 4, 11, 31, 2

Sec. 3.2: Measures of Dispersion (measures of spread)

Sec. 3.2: Measures of Dispersion (measures of spread)

Why would you want to measure the spread of data? Story…

Sec. 3.2: Measures of Dispersion (measures of spread)

There are 3 ways of measuring the spread…

1) Name: RANGE

   How to calculate: Highest data point – lowest data point

Sec. 3.2: Measures of Dispersion (measures of spread)

There are 3 ways of measuring the spread…

2) Name: STANDARD DEVIATION

Symbol: *s*

How to calculate: The square root of the "almost average" of the squares of the distance each data point is from the mean of the entire data set

Sec. 3.2: Measures of Dispersion (measures of spread)

There are 3 ways of measuring the spread…

3) Name: VARIANCE

Symbol: s^2

How to calculate: The square of the standard deviation

Sec. 3.2: Measures of Dispersion (measures of spread)

Ex 3: "Derive" the formula for the standard deviation s of a data set by finding the standard deviation of the data set below…

Data: 1, 2, 4, 11, 12

Formulas

If N stands for the population size and n stands for the sample size, then...

Population Mean

$$\mu = \frac{\sum x}{N}$$

Sample Mean

$$\bar{x} = \frac{\sum x}{n}$$

Formulas

If N stands for the population size and n stands for the sample size, then...

Population Standard Deviation

$$\sigma = \sqrt{\frac{\sum (x - \mu)^2}{N}}$$

Sample Standard Deviation

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

Formulas

If N stands for the population size and n stands for the sample size, then...

Population Variance

$$\sigma^2 = \frac{\sum(x-\mu)^2}{N}$$

Sample Variance

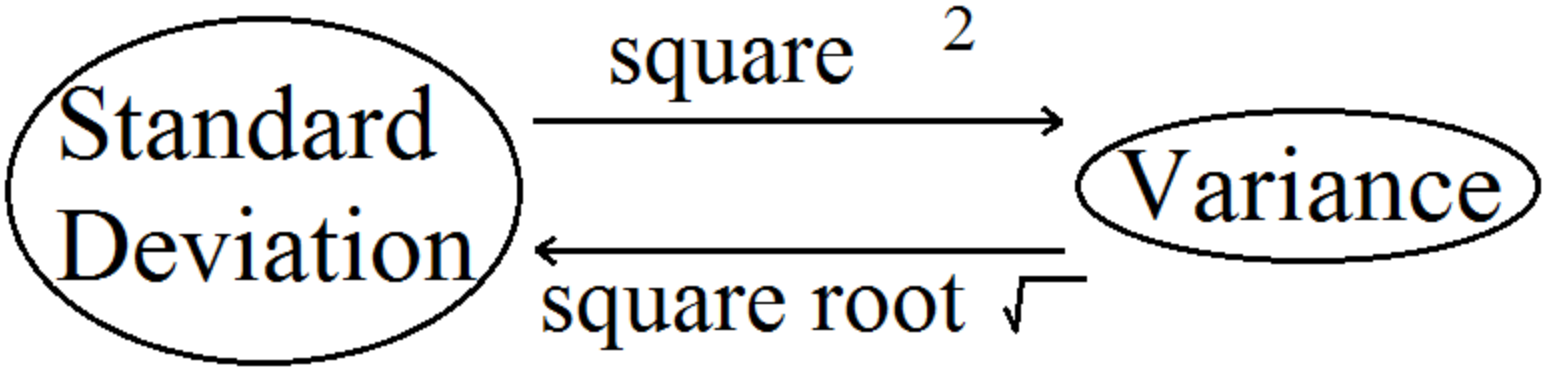
$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1} = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)}$$

Sec. 3.2: Measures of Dispersion (measures of spread)

Ex 4: Find the range, standard deviation, and variance of the following data set…

Data: 23, 17, 8, 19, 4, 11, 31, 2

Going Between Standard Deviation and Variance



Ex 5: If the standard deviation of a data set is 9, what is the variance of the data set?

Ex 6: If the variance of a data set is 9, what is the standard deviation of the data set?

Some Other Questions

Ex 7: Suppose you have 2 data sets. Let s_1 be the standard deviation of data set 1 and let s_2 be the standard deviation of data set 2. If $s_1 = 8.3$ and $s_2 = 4.1$, what can you say about the data sets?

Some Other Questions

Ex 8: Below are 2 data sets.

Data set 1: 755, 753, 756, 757, 751

Data set 2: 1, 3, 12, 76, 163

Let s_1 be the standard deviation of data set 1 and let s_2 be the standard deviation of data set 2. Without calculating them, which one is bigger: s_1 or s_2 ? Why?