Statistical Reasoning
Standard Deviation

Name: m Date: Class:

The Meaning of Standard Deviation

Standard deviation is a measure of spread (like Range, Variance, or IQR). It tells you how wide or concentrated your data is. You can interpret it as how far away the typical/average observation is from the mean.

Consider the following

You have ten people and you are given that their mean age is 30 → $\bar{x} = 30$

Why are you given the average age and not the age of each person separately? Because, the mean is convenient for its ease of analysis. When we try to represent the information contained in ten values (the ages) by a single value (the mean), there is a tradeoff - you lose the accuracy of the information Research analysis is a representative summarization of a data set and it prevents a researcher from staring blankly at all 10 of the people's ages.

If the average age of ten people is 30, there can be many different combinations of their ages. For example...

1. - They could be 1, 1, 1, 1, 1, 59, 59, 59, 59, 59 $\bar{x} = 30$
- They could be 30, 30, 30, 30, 30, 30, 30, 30, 30, 30 $\bar{x} = 30$
- They could be 15, 15, 15, 15, 15, 45, 45, 45, 45, 45 $\bar{x} = 30$
- They could be ...

In the first case, the actual values are spread far away from the mean.
In the second case, all of them are exactly on the mean.
In the third case, the values are moderately separated away. from the mean.
The mean, alone, is not a good representation, or summarization, of this data set.

Here comes the need for **standard deviation.** A high standard deviation signifies high deviation of data points from the mean. A moderate standard deviation signifies a moderate deviation of data points from the mean and a low (can b even zero) signifies that the data points are close to the mean.
Adding more information yields a more accurate standard deviation and better analyzes the spre of information.

S_x = sample st. dev. σ_x = population st. dev.

Standard deviation on its own is of no utility. Given the mean and standard deviation it can clearly show the average age and the spread of the ages.

- With the data set $\{1, 1, 1, 1, 1, 59, 59, 59, 59, 59\}$
  - There is a mean of 30 → $\bar{x} = 30$
  - Here is a standard deviation of 29 → $\sigma = 29$
  - This means the average age is 30, but the data is very spread out and the average distance away from the mean of 30 is 29 years.

- With the data set $\{30, 30, 30, 30, 30, 30, 30, 30, 30, 30\}$
  - There is a mean of 30 → $\bar{x} = 30$
  - Here is a standard deviation of 0 → $\sigma = 0$
  - This means the average age is 30, but the data is very close and the average distance away from the mean of 30 is 0 years.

- With the data set $\{15, 15, 15, 15, 15, 45, 45, 45, 45, 45\}$
  - There is a mean of 30 → $\bar{x} = 30$
  - Here is a standard deviation of 15 → $\sigma = 15$
  - This means the average age is 30, but the data is somewhat spread out and the average distance away from the mean of 30 is 15 years.