Math 122
Introduction to Statistics
Variation

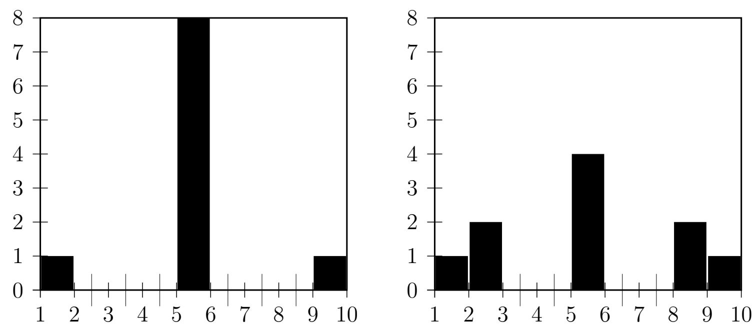
DATA SETS WITH DIFFERENT DISTRIBUTIONS

Consider these two data sets

1555555559

1225555889

Both data sets have mean=mode=median=midrange=5, but we can see in these histograms that the distributions of the two sets of data are different.



The second data set has more "variety" than the first.

RANGE

Range: The range of a set of quantitative data is the difference between the maximum and minimum values.

Example: The range of both data sets above is $9 - 1 = 8$.

STANDARD DEVIATION AND VARIANCE

Interpretation: When we speak of standard deviation, think "average distance from the mean." (This is not technically correct, but it is close enough.)

Notation:

- Σ (Greek Sigma) denotes a SUM
- x - variable to denote individual data values
- Σx - sum of all values of x
- n - number of data values in a sample
- N - number of data values in a population

- $\bar{x}$ - sample mean
- μ - population mean
- s or Sx - standard deviation of a sample
- s^2 - variance of a sample
- σ - standard deviation of a population
- σ^2 - variance of a population

Sample Standard Deviation: The standard deviation of a set of n sample values is

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

Population Standard Deviation: The standard deviation of a population of N values is

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

Variance: The variance of a set of data is the square of the standard deviation. For samples

$$\text{variance} = s^2.$$

For populations

$$\text{variance} = \sigma^2.$$

Calculations: We will NEVER use these formulas to calculate standard deviation or variance. We will always use a calculator.

Example: A calculator shows that the standard deviation of the first sample above is about 1.886. The variance of this sample would be $1.886 \times 1.886 \approx 3.557$. The standard deviation of the second sample is about 2.749. The variance is about $2.749 \times 2.749 \approx 7.557$. The second sample exhibits more variation.

RANGE RULE OF THUMB

Range Rule of Thumb: For many types of data (whose distribution is bell-shaped) about 95% of data values are within two standard deviations of the mean.

RRT Usual and Unusual Values: The maximum and minimum "usual" values of a data set are

$$\text{maximum usual value} = \text{mean} + 2 \times (\text{standard deviation})$$

$$\text{minimum usual value} = \text{mean} - 2 \times (\text{standard deviation})$$

According to the Range Rule of Thumb, about 95% of data in a bell-shaped distribution is between these values.

Example: IQ scores are normally distributed with a mean of $\mu = 100$ and a standard deviation of $\sigma = 15$. The maximum usual IQ score is

$$\mu + 2\sigma = 100 + 2 \times 15 = 130.$$

Any score higher than this is considered unusually high. The minimum usual score is

$$\mu - 2\sigma = 100 - 2 \times 15 = 70.$$

Any score below this is considered unusually low. Scores from 70 to 130 are not unusual.

RRT Estimate of s: The difference between the maximum and minimum usual values according to the Range Rule of Thumb is about four standard deviations. This means that the RRT gives this estimate of the standard deviation:

$$s \approx \frac{range}{4}.$$

Example: If the maximum value in a data set is 179 and the minimum value is 23, then the standard deviation could be approximated as $s \approx (179 - 23)/4 = 39$.

Empirical Rule: For many types of data (whose distribution is bell-shaped):

- About 68% of the data values are within one standard deviation of the mean.
- About 95% of the data values are within two standard deviations of the mean.
- About 99.7% of the data values are within three standard deviations of the mean.

Example: IQ scores are normally distributed with a mean of $\mu = 100$ and a standard deviation of $\sigma = 15$.

- About 68% of IQ scores should be between $100 - 15 = 85$ and $100 + 15 = 115$.
- About 95% of IQ scores should be between $100 - 2 \times 15 = 70$ and $100 + 2 \times 15 = 130$.
- About 99.7% of IQ scores should be between $100 - 3 \times 15 = 55$ and $100 + 3 \times 15 = 145$.