Section 3.2 — Measures of Dispersion

Chris Godbout

Outline

Populations vs Samples

Parameters and Statistics

Definition (Parameter)

A parameter is a numerical measurement describing some characteristic of a *population*.

Parameters and Statistics

Definition (Parameter)

A parameter is a numerical measurement describing some characteristic of a *population*.

Definition (Statistic)

A statistic is a numerical measurement describing some characteristic of a *sample*.

Notation

Table 1: Sample vs Population Notation

	Sample	Population
Count	n	N
Mean	$\bar{x}$	μ
Standard Deviation	s	σ

Measures of Dispersion

Rounding

All measures of dispersion are rounded to one more decimal place than the data.

Range

Definition (Range)

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

Standard Deviation of a Sample

Definition (Standard Deviation)

The standard deviation of a set of sample values, denoted by s, is a measure of how much the data values deviate from the mean.

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

Standard Deviation of a Sample

Definition (Standard Deviation)

The standard deviation of a set of sample values, denoted by s, is a measure of how much the data values deviate from the mean.

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

Definition (Variance)

The variance of a set of data is the square of the standard deviation.

Standard Deviation and Variance of a Population

Definition (Population Standard Deviation)

The standard deviation of a *population* is

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

Definition (Population Variance)

The variance of a population is σ^2.

Cookies!

In a sample of 4 Chips Ahoy cookies, the number of chocolate chips in each cookie was:

22	22	26	24

Properties of the Standard Deviation

Properties of the Standard Deviation

- $s \geq 0$.

Properties of the Standard Deviation

- $s \geq 0$.
- It is heavily influenced by outliers.

Properties of the Standard Deviation

- $s \geq 0$.
- It is heavily influenced by outliers.
- The units of s are same as the units of the data values.

Properties of the Standard Deviation

- $s \geq 0$.
- It is heavily influenced by outliers.
- The units of s are same as the units of the data values.
- s is biased

Properties of the Standard Deviation

- $s \geq 0$.
- It is heavily influenced by outliers.
- The units of s are same as the units of the data values.
- s is biased
- s^2 is unbiased

Biased and unbiased estimators

Definition (Biased and Unbiased)

An estimator (or statistic) is biased if the values of the sample do not target the value of the population. It is unbiased if they do.

Rules of thumb

Empirical Rule

The empirical rule says that for data that is roughly bell-shaped,

- About 68% of all values fall within 1 standard deviation of the mean.
- About 95% of all values fall within 2 standard deviations of the mean.
- About 99.7% of all values fall within 3 standard deviations of the mean.

Chebyshev's Theorem

The proportion of data that lie within K standard deviations of the mean is at least $1 - \frac{1}{K^2}$ for $K > 1$. So

$K = 2$ At least $1 - \frac{1}{2}^2 = \frac{3}{4} = 75\%$ of the data lie within 2 standard deviations.

$K = 3$ At least $1 - \frac{1}{3}^2 = \frac{8}{9} = 88.9\%$ of the data lie within 3 standard deviations.

Different Populations

Coefficient of Variationd

Definition (Coefficient of Variation)

The coefficient of variation for a set of nonnegative sample or population data, expressed as a percent, describes the standard deviation relative to the mean

Sample	Popluation
$CV = \frac{s}{\bar{\bar{x}}} \cdot 100\%$	$CV = \frac{\sigma}{\mu} \cdot 100\%$

Shoe Size and Age

Treating this class as a *sample*, we have the following

	Age	Shoe Size
$\bar{X}$		
s		