

V2 - Descriptive Statistics

Part I

Course: Statistical Testing & Regression
Dr. Renee Clark
Swanson School of Engineering
Industrial Engineering
University of Pittsburgh



Descriptive Statistics – Part I

- ❑ Population vs. Sample
- ❑ Descriptive statistics
 - ❑ Central Tendency
 - ❑ Variability



Population vs. Sample

A **population** consists of All observations we are concerned with or interested in

A **sample**: Subset from population



All observations we are concerned with ??

Examples of what we mean:

- All voters in Pennsylvania
- All females who played golf last year
- All plastic parts made in 2019 by manufacturer
- All students taking IE-1071 this semester



What is a Descriptive Statistic?

- Number that describes a sample of data
- Also known as Statistic
- Parameter characterizes the population



Descriptive Statistic

- ❖ Numerically summarizes sample of data
 1. Location/central tendency
 2. Variability

Location/central tendency

Average/mean: most familiar

$$\bar{X} = \frac{4+5+5+6+6}{5} = \underline{5.2}$$

Problem with average: affected by outliers

$$\bar{X} = \frac{4+5+5+6+6+145}{6} = \underline{28.5}$$



Location/Central Tendency

Median

- Less sensitive to outliers
- Middle observation (of ordered data)

- Data: 3, 5, 9
- Median = 5

- Data: 2, 3, 4, 6

- Median = $\frac{3+4}{2} = \frac{7}{2} = 3.5$



Weighted Average/Mean

- Weighted based on the number of times each value occurs

Value	Number of Occurrences
4	2
6	6
9	2
11	1

→ 4, 4, 6, 6, 6, 6, 6, 6, 9, 9, 11

$$\frac{73}{11}$$

$$= 6.636$$

- Multiply each value by the number of times it occurs
- Add these products together
- Divide this sum by total number of occurrences

→ Numerator = $(4 * 2) + (6 * 6) + (9 * 2) + (11 * 1) = 73$

→ Denominator = $2 + 6 + 2 + 1 = 11$

Weighted average = $73 / 11 = 6.636$



What is *Variability*?

Variability is extent data differs or is spread or stretched out

Chance of rain (n=15)	
→ 99.5%	✓
90.3%	✓
83.6%	✓
72.11%	✓
66.33%	
65.9%	
53.1%	✓
50.9%	
49.88%	
48.11%	
33.33%	✓
20.9%	
20.0%	
19.7%	
→ 2.01%	

High Variability:



Descriptive Statistics for Variability

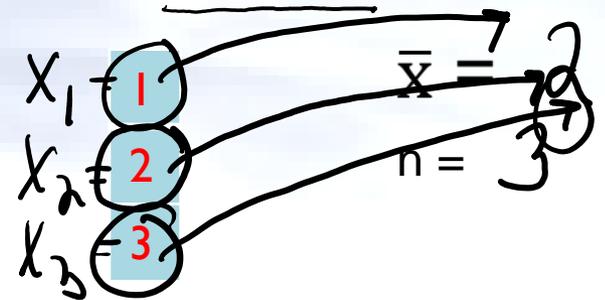
Variance & Standard Deviation

- Most common
- Captures variability of data around mean
 - Sensitive to outliers

* Variance = $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

Average squared deviation from mean

Large deviations \rightarrow large s^2



Standard deviation (s): square root of var

$$s = \sqrt{s^2}$$



Degrees of Freedom

sample
variance

Formula for _____

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

$n-1$ is degrees of freedom

$n-1$ independent pieces of information

$$\sum_{i=1}^n (x_i - \bar{x}) = 0$$

Deviations
around mean





Acknowledgement

This material is based upon work partially supported by the National Science Foundation under Grant# 2335802. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.





You are free

- to **Share** – to copy, distribute, display and perform the work
- to **Remix** – to make derivative works

Under the following conditions

- **Attribution** — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- **Noncommercial** — You may not use this work for commercial purposes.
- **Share Alike** — If you alter, transform, or build upon work, you may distribute the resulting work only under the same or similar license to this one.





THE END

