Statistical Vocabulary #1

Statistics is a collection of methods for planning experiments, obtaining data, and then organizing, presenting, analyzing, interpreting, and drawing conclusions based on the data.

Definitions

- A population is the complete collection of all elements (scores, people, measurements, and so on) to be studied.
- A sample is a subcollection of elements drawn from the population.
- A parameter is a numerical measurement describing some characteristic of a population.
- A statistic is a numerical measurement describing some characteristic of a sample.
- In an observational study, we observe and measure specific characteristics, but we don't attempt to manipulate or modify the subjects being studied.
- In an experiment, we apply some treatment and then proceed to observe its effects on the subjects.

Determine whether each of the following is an observational study or an experiment.

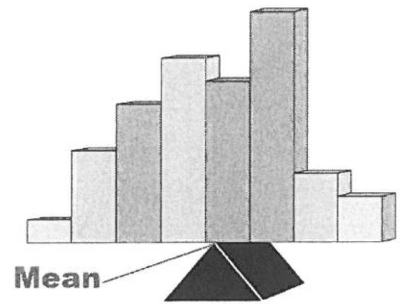
1. Different brands of cigarettes are measured for tar, nicotine, and carbon monoxide. observational study
2. People who smoke are asked to halve the number of cigarettes consumed each day so that any effect on pulse rate can be measured. experiment
3. In a PE class, the effect of exercise on blood pressure is studied by requiring that half of the students walk a mile each day while the other students run a mile each day. experiment
4. The relationship between weights of bears and their lengths is studied by measuring bears that have been anesthetized. observational study

Measures of Central Tendency

- A measure of central tendency is a value at the center or middle of a data set.

Mean

- The arithmetic mean of a set of scores is the value obtained by adding the scores and dividing the total by the number of scores.
- Referred to as "mean"
- $\bar{x} = \frac{\sum x}{n}$
  - Pronounced "x-bar"
  - the mean of all values in a <u>sample</u>
  - Σ: represents the total or summation of the scores
  - x: is the variable used to represent the individual data values
  - n: represents the number of values in a sample
- $\mu = \frac{\sum x}{N}$
  - the mean of all values in a <u>population</u>
  - N: represents the number of values in a population



Example:

A. Listed below are the times (in years) that the first ten presidents survived after inauguration. Find the mean for this sample.

$$\frac{10 + 29 + 26 + 28 + 15 + 23 + 17 + 25 + 0 + 20}{10} = 19.3 \text{ yrs}$$

Median

- The median of a set of scores is the middle value when the scores are arranged in order of increasing (or decreasing) magnitude.
- $\tilde{x}$
  - Pronounced "x-tilde"
- To find the median:
  - Arrange the scores in order (increasing or decreasing)
    - If the number of scores is odd, the median if the number that is located in the exact middle of the list.
    - If the number of scores is even, the median is found by computing the mean of the two middle numbers.

Example:

a. The following values are the incomes (in thousands of dollars) that performers received for one rock concert. Find the median.

500 60 80 50000 1000 400

60 80 400 500 1000 50000

$$\frac{400 + 500}{2} = 450$$

Mode
➢ The mode of a data set is the score that occurs most frequently. ➢ M ▪ Represents mode ➢ When two scores occur with the same greatest frequency, each one is a mode and the data set is bimodal. ➢ When more than two scores occur with the same greatest frequency, each is a mode and the data set is said to be multimodal. ➢ When no score is repeated, we say that there is no mode. ➢ To find the mode: ▪ Arrange the scores in order (increasing or decreasing) ▪ Determine which score(s) has the greatest frequency

Example:

b. Find the mode(s) of each data set.

1. 5 5 5 3 1 5 1 4 3 5 — 5
2. 1 2 2 2 3 4 5 6 6 7 7 7 9 — 2 and 7
3. 1 2 3 6 7 8 9 10 — none

The Best Measure of Central Tendency

Comparison of Mean, Median, and Mode						
Average	Definition	How Common?	Existence	Takes Every Score into Account?	Affected by Extreme Scores?	Advantages and Disadvantages
Mean	$\bar{x} = \frac{\sum x}{n}$	Most familiar "average"	Always exists	Yes	Yes	A: most common and takes all scores into account D: Can be affected by extreme scores
Median	$\tilde{x}$	Commonly used	Always exists	No	No	A: Unaffected by extreme scores D: Extreme scores unreported
Mode	M	Sometimes used	May not exist or may be more than one mode	No	No	A: Useful with tightly grouped data or nominal data D: may not exist or reflect data

Which is the best measure of central tendency for each situation?

- Determining the "average" cost of a house in a particular area median
- Determining the "average" eye color in the school mode
- Determining the "average" test score mean

Practice:
For questions 1-2 find the mean, median, and mode.

1. The ages (in years) of students taking a Calculus III class in college.

17	20	21	18	20	20	20	18	19	19
20	19	21	20	18	20	20	19	18	19

$\bar{x} = 19.3$

med = 19.5

mode = 20

2. Digits selected in the Iowa Pick Three lottery:

1 6 8 6 9 5 2 1 5 0 3 9 9 0 7

$\bar{x} = 4.7\bar{3} = 4.7$

median = 5

mode = 9

Practice:
For questions 3 find the mean, median, and mode of each sample, and then compare the two sets of results.

3. Samples of the ages (in years) of student cars and faculty/staff cars at a particular college.

Students	10	4	5	2	9	7	8	8	16	4	13	12

Faculty/Staff	7	10	4	13	23	2	7	6	6	3	9	4

a. Mr. Clark decides to trade in his 23 year old car and buy a brand new car. How does this effect the results?

Students

mean = 8.2

median = 8

mode = 4 and 8

Faculty/Staff

mean = 7.8

median = 6.5

mode = 4, 6, 7

Faculty/Staff after new car purchase

mean = 5.9

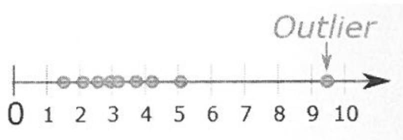
median = 6

mode = 4, 6, 7

Unit 4 1.1

Outlier

- A value that "lies outside" (is much smaller or larger than) most of the other values in the set of data.



Example #1: 25, 29, 3, 32, 85, 33, 27, 28 Outlier(s): 3 and 85

How do Outliers affect the measures of central tendency?

Example #2: A new coach has been working with the long jump team this month, and the athletes' performance has changed. Here are the results:

Athlete	Augustus	Tom	June	Carol	Bob	Sam
Result	$+0.15m$	$+0.11m$	$+0.06m$	$+0.05m$	$+0.12m$	$-0.56m$

Find the mean of the data set: $\frac{.15+.11+.06+.05+.12+(-5.6)}{6} = -.07m$

Is there an outlier? -.56

Recalculate the mean.

$$\frac{.15+.11+.06+.05+.12}{5} = .098\ m$$

Example #3:The following data represents the math scores of a group of friends:

Albert	Beth	Cindy	David	Emily	Frank	Gary	Helen	Ida	Jeremy
96%	92%	85%	81%	37%	88%	95%	84%	96%	78%

Calculate the mean: $\frac{96+92+85+81+37+88+95+84+96+78}{10}$

mean = 83.2%

Is there an outlier? 37%

Recalculate the mean.

$$\frac{96+92+85+81+88+95+84+96+78}{9}$$

mean = 88.3%