# Basic Statistics

Chapter 3
Central Tendency and Variability

1

## Chapter 3 Outline

- Measures of Central Tendency
  - mean
  - median
  - mode
- Measures of Variability
  - range (including interquartile)
  - standard deviation(s)
  - variance
- Graphical presentation

2

## Measures of central tendency

- mean
- median
- mode

3

## Mean

- a statistic calculated from a sample
- corresponding population parameter is $\mu$
- population parameter, we know the exact value with certainty
- statistic uncertainty is involved
- $\overline{X}$ is the best estimator of $\mu$

4

## Summation notation

- $\Sigma$ uppercase Greek sigma
- tells us to add up an entire group of numbers
- $\Sigma$X means add up all the Xs
- Formula for the mean
  - where:
  - $\overline{X}$ = the mean
  - $\Sigma$X = add up all the X values
  - N = number of scores

$$\overline{X} = \frac{\Sigma X}{N}$$

5

## Characteristics of the mean

- sum of differences between mean and each score in a distribution will always equal 0
- difference scores are called deviations from the mean
- stated mathematically: $\Sigma(X - \overline{X}) = 0$
- sum of squared deviations of the mean from each score represents a minimum
- no value can be used to make the sum of squares any smaller than using the mean
- So: $\Sigma(X - \overline{X})^2$ represents a minimum

6

## Median

- point at exact middle of the set of scores.
- list all scores in order, then locate the point in the center of the sample.
- if 499 scores in list, score #250 would be median if 500, the avg. of 250 and 251 would be the median
- odd number of scores, Median will be score in location (N+1)/2
- even number of scores, Median will be average of scores N/2, ((N/2)+1)
- even number of scores, the median may not be a score that actually exists in the distribution.

7

## Median (cont.)

- a point in the distribution, not necessarily a value observed
- simplest interpretation it is the score or value where half are higher and half lower
- also the $50^{th}$ percentile of a set of scores

8

## Mode

- simply the most frequently occurring value in a set of scores
- when giving a modal value should also give an idea of how often it occurred
- can be more than one mode

9

## Mean from frequency distribution

- Multiply each score by its frequency
- Add all of these scores up
- Divide by the total number of scores
- $\mu$ or $\overline{X} = \dfrac{\Sigma fX}{N}$
- where: $f$ = frequency of X
  X = value of X
  N = number of scores

10

## Median/mode from frequency distribution

- Median
  - example
- Mode
  - Just pick out the score with the highest *f*.

11

## Which should I be using?

- scale of measurement dictates measure of central tendency
- mean with interval/ratio level data, not ordinal/nominal.
- median with ordinal or higher, not nominal.
- mode with any level

12

## Skewness

- If the distribution is normal (i.e., bell-shaped), the mean, median and mode are all about equal
- positive skew: Mean>Median>Mode
- negative skew: Mean<Median<Mode

13

## Means of sets of means

- if means are based on the same number of cases can just average them
- if from different numbers of observations have to weight each according to number of cases it is based on

14

## Variability/Dispersion

- Another characteristic of a distribution
- Measures of central tendency: location
- Measures of dispersion: variation or spread
- Three estimates:
  - Range
  - Standard Deviation
  - Variance

15

## Variability/Dispersion (cont.)

- Standard Deviation
  - more accurate/detailed
  - outlier(s) can greatly extend the range
  - Standard deviation uses info from all of the data
  - not dictated solely by the extreme values

16

## Range (incl. Interquartile)

- Range = distance between high & low scores
- Administer the Beck Depression Inventory to a group of students
- Scores: 15,20,21,20,36,16, 25,15
- The range would be 36-15
- 21 points separate the highest and lowest scores
- range = $X_H$ - $X_L$
- Interquartile range
  - distance between the 25th and 75th percentiles
  - benefit relative to range?

17

## Standard Deviation

- The Standard Deviation is a value that shows the relation that individual scores have to the mean of the sample.
- Can be thought of as roughly the average deviation from the mean

18

## Standard Dev. (con't)

- Three different SDs - different purposes
  - σ measure population variability
  - ŝ estimate population variability
  - S measure sample variability
- σ & S are considered descriptive indices
  - calculated the same way
  - use different terms in equations

19

## Standard Deviation (con't)

$$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}} \qquad S = \sqrt{\frac{\sum (X-\overline{X})^2}{N}}$$

- **σ = population standard deviation**
- **S = sample standard deviation**
- **X = a data point (observation)**
- **N = number of deviations (observations)**

20

## Standard Dev. (con't)

- Generally interested in standard deviation as estimator
- Dividing the sum of squared deviations by N has been shown to result in a significantly biased estimate of σ
- This estimate is generally too small, underestimating the population parameter
- Using N-1 results in a better estimate

21

## Standard Dev. (con't)

- The standard deviation is the square root of: the sum of the squared deviations from the mean of all the scores divided by the number of scores - 1

$$\hat{s} = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n-1}}$$

*Definitional Formula*

22

## Example Standard Deviation

- Lets take the set of scores: 15,20,21,20,36,16, 25,15
- The mean of this sample is 21, first line up the scores
- 15,15,16,20,20,21,25,36.
- 15-21=-6, 15-21=-6, 16-21=-5, 20-21=-1, 20-21=-1, 21-21=0, 25-21=4, 36-21=15
- Square these values: -6, -6, -5, -1, -1, 0, 4, 15
- Add them all up: 36+36+25+1+1+0+16+225 = 340

23

## Working Ex. (Stan. dev. con't)

- Divide 340 by 7 = 48.57
- square root of 48.57 = 6.97
- 6.97 = Standard Deviation

*Computational Formula*

$$\sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}}$$

24

## SD Computational Formula

- Take values, square each
- Sum both columns
- Square the summed X's
- Plug into formula

$$\hat{s} = \sqrt{\frac{3868 - \frac{28224}{8}}{8-1}}$$

| | X | $X^2$ |
|---|---|---|
| | 15 | 225 |
| | 15 | 225 |
| | 16 | 256 |
| | 20 | 400 |
| | 20 | 400 |
| | 21 | 441 |
| | 25 | 625 |
| | 36 | 1296 |
| $\Sigma =$ | 168 | 3868 |
| | 28224 | |

25

## SD Computational Formula

$$= \sqrt{\frac{3868 - 3528}{7}}$$

$$= \sqrt{\frac{340}{7}}$$

$$= \sqrt{48.57}$$

$$\mathbf{= 6.9693}$$

26

## Graphical presentation of central tendency & variability

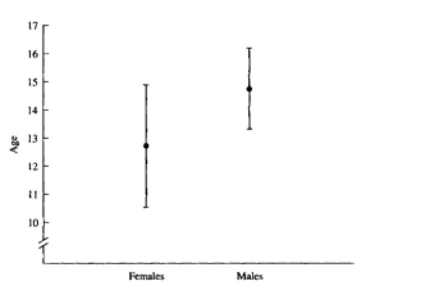


**FIGURE 3.1** Puberty in females and males—means and standard deviations

27

## Variance

- The square of the standard deviation
- If we subtract each value from the mean and take their average, always zero, just as many values are above as below the mean
- Squaring differences eliminates the problem of summing to zero
- Average squared deviation is the variance

$$\frac{\Sigma(X-\overline{X})^2}{n-1}$$

28

## Variance Computational Formula

| X | X² |
|---|---|
| 0 | 0 |
| 1 | 1 |
| 1 | 1 |
| 1 | 1 |
| 2 | 4 |
| 2 | 4 |
| 2 | 4 |
| 2 | 4 |
| 2 | 4 |
| 4 | 16 |
| ΣX=17,(ΣX)²=289 | ΣX² = 39 |

$$\hat{s}^2 = \frac{\sum X^2 - \frac{(\sum X)^2}{N}}{N-1}$$

29

## Variance Computational Formula

| X | X² |
|---|---|
| 0 | 0 |
| 1 | 1 |
| 1 | 1 |
| 1 | 1 |
| 2 | 4 |
| 2 | 4 |
| 2 | 4 |
| 2 | 4 |
| 2 | 4 |
| 4 | 16 |
| ΣX=17,(ΣX)²=289 | ΣX² = 39 |

$$\hat{s}^2 = \frac{39 - \frac{289}{10}}{n-1}$$

$$\hat{s}^2 = \frac{39-28.9}{9} = \frac{10.1}{9} = 1.12$$

30