Descriptive Statistics
Measures of Central Tendency and Variability

MARCELA FRAZIER, OD, MPH, FAAO

Statistics

- Set of mathematical tools used to...
  - Summarize
  - Describe
  - Compare
  - Interpret
  ...Data sets

Descriptive Epidemiology

- Characterizes the distribution in a population
  - Person
  - Place
  - Time

Person

- Age (most important determinant)
- Sex
- Race/Ethnicity
- Marital Status
- Religion
- Occupation

Place

- World-wide
- Regional
- Local
- Urban vs Rural

Time

- Short Term
- Recurrent/periodical/seasonal
- Long Term

Variables

- Nominal
- Ordinal
- Continuous

Nominal Variables

- Values are categorical and there is NO order associated with them. ie:
  - Male, Female
  - White, African American, Hispanic

Ordinal Variables

- Values are categorical, but there IS an order associated with them. ie:
  - A, B, C, F
  - 1,2,3,4 (for grading angles)
  - Mild, moderate, severe

Continuous Variables

- Values are continuous (decimals, fractions) and there IS an order associated with them. ie:
  - Hemoglobin A1C: 6.3, 7.8

Some can be ordinal or continuous

When data is being analyzed it may be more informative to classify the variables one way versus the other. ie:

Refractive errors for some studies have been analyzed as ordinal (-2.50,-3.00) and as continuous (-2.55,-2.80) for other studies.

Statistical Testing

- Statistical Testing is used to make inferences about a population based on information obtained from samples

Population vs Sample

- A sample is a subset of the population
- A sample is selected because in most cases it is impossible to test the entire population
- A sample must be chosen carefully to truly reflect the population it is intended to

Measures of Central Tendency

Describe the average of the distribution of values:
- Mean
- Median
- Mode

The Mean

- The sum of all values divided by the number of values
- $\mu = \dfrac{\sum x}{n}$

i.e.: Mean CCT and IOP was found to be 530.87±30.79 μm and 15.65±3.05 mm Hg respectively

Outliers can skew the mean in one direction

The Median

- The middle value when the values are organized in order

i.e.:Twenty-seven patients with lacrimal gland lymphoma were identified. Eight of the patients were men and 19 were women; the median (range) age was 69 (43-87) years

Outliers do not affect the median much

The Mode

- The most common value
- A set of data may or may not have a mode
(or have several modes)
Find the mode for this sample:
16,17,13,14,19,21,13,14,11,18,19,20,12,15,13

Find the mode for this sample :
20/15, 20/30, 20/15, 20/20, 20/40, 20/15, 20/50

Which one is used?

- The mean is preferred for normally-distributed samples (parametric data)
- The median is preferred for data sets with **high variability** (large Standard Deviation)
- The mode is not used much but may be mentioned in some studies

When comparing two groups, it's important to compare '*means to means*' or '*medians to medians*'

Normally Distributed Data (Parametric data)

- Bell-Shaped Curve: That in which the mean, the median, and the mode are equal, and 68% of the population lie within one SD

Mean,
Median,
Mode
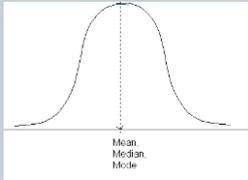
Measures of Variability

Summarize how much the values differ from the average
- Range
- Standard Deviation
- Variance
- Confidence intervals

Range

The difference between the lowest value and the largest value in a sample

Find the range for this sample:

16,17,13,14,19,21,13,14,11,18,19,20,12,15,13

The range is useful for finding out where values may lie

Standard Deviation (SD)

Average of the deviation of all values from the mean. In a bell-shaped curve:
- The mean+/- one SD includes ~68% of the values
- The mean+/- two SD includes ~95% of the values
- The mean+/- three SD includes~98% of the values

Standard Deviation (SD)

- If a study shows that the mean near CT is 3^ XP and the SD is 5, then...
- ~68% of the values are between 2EP & 8XP
- ~95% of the values are between 7EP & 13XP
- ~98% of the values are between 12EP & 18XP

12 11 10 9 8 7 6 5 4 3 2 1 0 1 2 **3** 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Variance

- SD squared
- Not often used in descriptive statistics

Confidence Intervals (CI)

- Most powerful measures of variability for normally distributed (parametric) data
- Contain the mean and the variability in one measure
- Most often used are 95% CI
- Expressed as P (low range, high range) ie:

Study results of IOP with Xalatan at 1month are P(12,19)

We can say with 95% confidence that the mean IOP is between 12 and 19

Example

- Calculate the mean IOP for this sample:
16,17,13,14,19,21,13,14,11,18,19,20,12,15,13,12,14,18, 19, 13

- What is the median?

- What is the mode?

- What is the range?

Hypothetical

- If mean IOP is 15mmHg and SD=2mmHg

What percentage of the ppn will have an IOP>17mmHg?

What percentage of the ppn will have an IOP>19mmHg?

What percentage of the ppn will have an IOP<13mmHg?

- http://www.youtube.com/watch?v=NRtltx_u8G0
- http://www.youtube.com/watch?v=PIJ8UYDAF3M