Measures of Central Tendency

Lesson: Central Tendency

Sample A: $Y = 4, 8, 2, 1, 8$

1. $N = 5$

2. $\Sigma Y = 4 + 8 + 2 + 1 + 8 = 23$

3. Mean = ?

4. Median = ?

5. Mode = ?

Sample B: $Y = 1, 5, 2, 7, 1, 4$

6. $M = ?$

7. Median = ?

8. Mode = ?

9. The mean:
   a. is the high point of the histogram
   b. divides the ordered set of scores in two
   **c. is the balance point of the histogram**

10. An extreme score (compared to the rest of the scores) is called an **outlier**.

11. Data that have extreme scores off in one direction only are said to be **skewed**.

12. Of the mean and the median, which is used the most in statistics?
   a. Mean
   b. Median

13. Of the mean and the median, which is more representative of the data when the data are skewed?
   a. Mean
   b. Median
14. Which of the following would be appropriate to compute for nominal data: (select all that apply)
   a: mean
   b: median
   c: mode

15. Which of the following would be appropriate to compute for ordinal data: (select all that apply)
   a: mean
   b: median
   c: mode

16. Which of the following would be appropriate to compute for cardinal data: (select all that apply)
   a: mean
   b: median
   c: mode

Lecture: Measures of Central Tendency

*Measures of Central Tendency* are various ways of measuring the average or most typical score in our data. In essence, we are making our sample simpler by making a generalization, taking the complexity of the data and saying ‘in general, the average score was…’

**The Mean**

Actually, all three measures of central tendency could be called averages. To find the *mean* you sum together all of the scores and divide that sum by the number of scores.

**Important Property #1 of the Mean:** The mean is the balance point of a histogram of the frequency distribution.

**Effect of Outliers**

Because the mean is the balance point, an outlier can have a dramatic effect on the mean. An outlier is a score that is very different (much higher or lower) than the others.

The effect of an outlier is less if it is just one score out of many.

If you have extreme scores off in both directions then the mean would return to the middle area of the scores (as the extreme scores on both ends would balance each other
out), the problem we are talking about arises when you have extreme scores out in one direction only.

Data that have extreme scores off in one direction only are said to be skewed. The mean is not a very good measure of central tendency when the data are skewed as it often does not accurately portray what most of the scores were like.

**The Median**
The second measure of central tendency we will look at is the median. The median is the central point in the list of scores once the scores have been put in numeric order (this will be easier to understand once you see how we compute the median).

*The median is a good measure of central tendency when you have skewed data.*

**The Mode**
The mode is simply the value which occurred most frequently in the sample.

Sometimes there is no mode (no score occurs more frequently than the others), sometimes there are two modes (other called a bimodal distribution), and sometimes there are more than two modes (a graph with two or more modes is called a multimodal distribution).

The term bimodal is used when there are two peaks in the distribution even if both peaks aren’t exactly the same height.

**Selecting a Measure of Central Tendency**
To select the one that makes the most sense given what you want to know about the sample. Are you interested in the ‘average’ (mean) score, the middle score that half the people scored less than and half scored more than (median), or the score that occurred the most (mode)?

By far, the mean is the measure of central tendency that is used the most in statistics. The mean has certain mathematical properties that make it easier to develop statistical tools for, and consequently the vast majority of statistical procedures involved the mean.

The median does a better job that the mean of describing the average when the data are skewed. If you want to perform a more sophisticated analysis of the data, however, you will find that there aren’t many statistical procedures that involved the median.
The main advantage of the mode is that it is the only measure that is appropriate for certain types of data.

**Measurement Scales and Measures of Central Tendency**

The measurement scale you use determines which measures of central tendency would be appropriate.

**Nominal data:** With nominal data the numbers serve simply as labels, they represent qualitative or categorical differences, rather than representing quantitative differences.

Thus, with nominal data the only measure of central tendency that we can use is the mode. Therefore, it would not make sense to compute the variance of nominal data.

**Ordinal Data:** with ordinal data numbers represent an increase or decrease in quantity but the sizes of the steps are not the same.

With ordinal data the median and mode make sense, but the mean does not.

When rank scores are used it is usually within somewhat more complicated experimental design.

There is no mode so you can’t use that.

Therefore, rank data it is going to be the median.

**Cardinal data (ratio and interval scales):** With cardinal data numbers reflect quantities and the size of the steps between numbers is always equal (think of these scales as direct measurements of some quantity). Examples of cardinal scales are: reaction times in milliseconds, number of books read in a month, number of children in the family, height in inches, etc.

All three measures of central tendency can be applied to cardinal data. Most of the variables we will use this semester will be cardinal so that we can bring in a large number of statistical tools to analyze the data. Therefore, it make sense to compute the variance of cardinal data.