Measures of Central Tendency

James H. Steiger

Overview

- Optimality properties for measures of central tendency – Two notions of "closeness"
- Properties of the Median
- Properties of the Mean

Optimality Properties

- Imagine you were a statistician, confronted with a set of numbers like 1,2,7,9,11
- Consider a notion of "location" or "central tendency – the "best measure" is a single number that, in some sense, is "as close as possible to all the numbers."
- What is the "best measure of central tendency"?

Optimality Properties

- One notion of "closeness" is "lack of distance"
- According to this notion, the "best measure" of central tendency is a number that has the lowest possible sum of distances from the numbers in the list
- This number is the middle value, or *median*.

The Median

- The median is the middle value in the distribution.
- Why is it that point on the number line that minimize the sum of distances?

The Median

- Consider the list 1,2,7,9,11
- Compute the sum of distances from 7 the median, to the 5 numbers in the list. The distances are, respectively, 6,5,0,2,4, for a total of 17.
- Now, consider a number slightly to the right of 7, or slightly to the left of 7. What happens? (C.P.)

Calculating the Median

- Order the numbers from highest to lowest
- If the number of numbers is odd, choose the middle value
- If the number of numbers is even, choose the average of the two middle values.

The Mean

- Consider another notion of "closeness."
- According to this notion, a value is "most close" to a list of numbers if it has the smallest sum of *squared* distances to the list of numbers
- Notice that this measure penalizes long distances, so it is particularly sensitive to them.
- The *mean,* or arithmetic average, minimizes the sum of squared distances. (Proof ? C.P.)

The Mean

- The sample mean is defined as

$$\bar{X}_{\bullet} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Sensitivity to Outliers

- We say that the mean is "sensitive to outliers," while the median is not.
- Example: The Nova Scotia Albino Alligator Handbag Factory

Sensitivity to Outliers

- Incomes in Weissberg, Nova Scotia (population =5)

Person	Income (CAD)
Sam	5,467,220
Harvey	24,780
Fred	24,100
Jill	19,500
Adrienne	19,400
Mean	1,111,000

Which Measure is Better?

- In the above example, the mean is $1,111,000, the median is 24,100.
- Which measure is better? (C.P.)

Computing the Mean from a Frequency Distribution

- Consider the following distribution:

X	f
30	2
29	3
28	5
27	3
26	2

Computing the Mean from a Frequency Distribution

- How would you compute the mean?

$$\bar{X}_{\bullet} = \frac{\sum_{i=1}^{K} X_i f_i}{\sum_{i=1}^{N} f_i}$$

Estimating the Mean from a Grouped Frequency Distribution

- The estimated mean is 74.3

Interval	*f*	*MdPt*	Sum
81-90	7	85.5	598.5
71-80	11	75.5	830.5
61-70	4	65.5	262.0
51-60	3	55.5	166.5
Total	25		***1857.5***

Computing the Mean for Combined Groups

- Suppose you combine the following two groups into a single group. What will the mean of the combined group be?

$N =$ [illegible] $\bar{X} = 5$

$N =$ [illegible] $\bar{X} = 10$

Computing the Mean for Combined Groups

- Simply compute the combined sum, and divide by the combined *N*. We will derive this formula (given in Glass and Hopkins) in class.