1. Overview

Pictures are nice, but people find numerical summaries very useful. The usual two numbers summarizing a distribution are the ***"center"*** *[the "typical" value] and the* ***"spread"*** *[how close or far the data are to each other].*

2. "Center"

Much of the time we use what is known as the "arithmetic average." Calculate it by summing all of the values in a list of data and dividing by the number of values in that list. The statistical term for this commonly used and widely understood statistic is the mean.

A. The mean or average is denoted: $\bar{x}$ ***(pronounced "x-bar")***

The mean is computed as follows: given a list of n *numbers:* $x_1, x_2, \ldots, x_n,$

their mean is calculated using the formula:

$$\bar{x} = \frac{x_1 + x_2 + x_3 \ldots + x_n}{n} = \frac{\sum_1^n x_n}{n}$$

B. The Median is often denoted M

This is an alternative measure of "center." It is the midpoint. Why an alternative? Each measure has its advantages and disadvantages when used to summarize data (see below). The median is the "middle value" of a list that has been sorted in order from lowest to highest. Half of the data are larger than (or equal to) the median, and half of the data are smaller than (or equal to) the median.

The median is computed as follows:

(1) Given a list of n *numbers* $x_1, x_2, \ldots, x_n,$

(2) Sort all the numbers

(3) Select the middle number from the list.

(4) If the list has an even number of elements, take the average of the two middle numbers.

3. Remarks

a. If you could place a histogram on a weightless bar and the bar on a fulcrum, the histogram would balance perfectly when the fulcrum is directly under the mean. By contrast, the median is the value where 50% of the area of the histogram lies above and 50% lies below this value.

b. If the histogram is symmetric, the mean and the median are the same. If the histogram is not symmetric, the mean and median can be quite different. Take a data set whose histogram is symmetric. Balance it on the fulcrum. Now take the largest observation and start moving it to the right. The fulcrum must move to the right with the mean, too, if the histogram is to stay balanced.

c. You can distort the mean with an outlier, but all this time the median stays the same! (see p.63)

d. It is not necessary to know HOW MANY numbers are in a list, only the RELATIVE FREQUENCY of the values; e.g., if we had 5 students with times 3, 7, 7, 14, 17, the mean of the list is 9.6. If we had 10 students with times: "3, 3, 7, 7, 7, 7, 14, 14, 17, 17" the mean is still 9.6. As long as the values in the list maintain the relative frequencies (in this example: 20% x_1*'s, 20%* x_2*'s, 20%* x_3*'s and so forth) the mean will be unchanged.*