

Social Studies 201
September 22-26, 2003

Measures of Central Tendency

1. Introduction

For much of the rest of this course, frequency or percentage distributions describing a sample or population are not listed in their entirety. Rather than presenting the whole frequency or percentage distribution in a table, bar chart, or histogram, statistical analysts summarize the distribution using measures of central tendency and variation. These are discussed in Chapter 5 of the text.

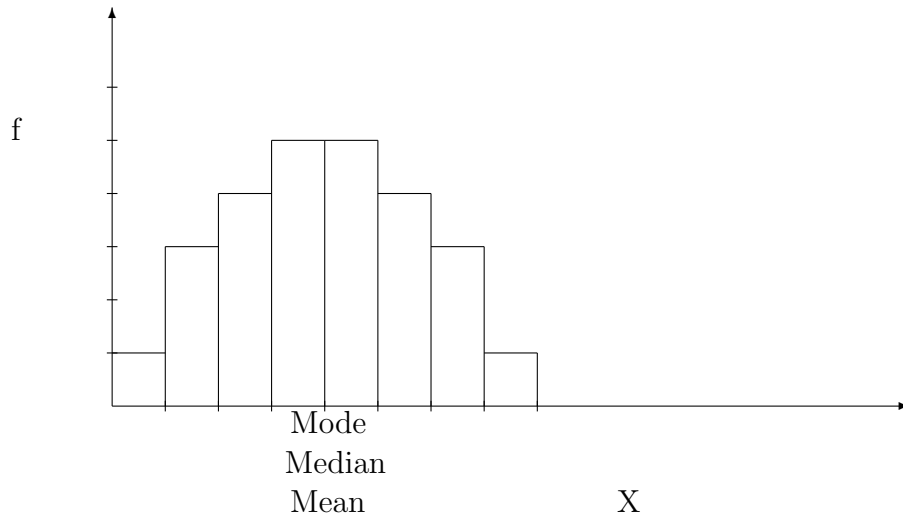
Measures of central tendency, or measures of centrality, are more commonly referred to as averages. These measures of central tendency are single values of the variable that are located at the centre of a distribution. Measures of variation are discussed later in Chapter 5 and are single values that describe how varied or concentrated a particular distribution is.

There are three common measures of central tendency, or centrality, of a distribution. These are the mode, median, and mean. An informal definition of each is as follows:

- **Mode.** The mode is the value of the variable that occurs most frequently.
- **Median.** The median is the middle value, with one half of the cases less than the median and the other half greater than the median.
- **Mean.** The mean is what is commonly termed the average – it is the sum of all the values of the variables divided by the number of cases.

All three coincide when the distribution is symmetrical around the central value of the variable. This is the case in Figure 1, where one half of the distribution is the mirror image of the other. In this case, the mode is at the value of X at the centre. At this same centre, one-half of the cases are less than this and the other half are greater than this. The mean, or common average, is also at this same value of X .

Figure 1: Symmetrical frequency distribution



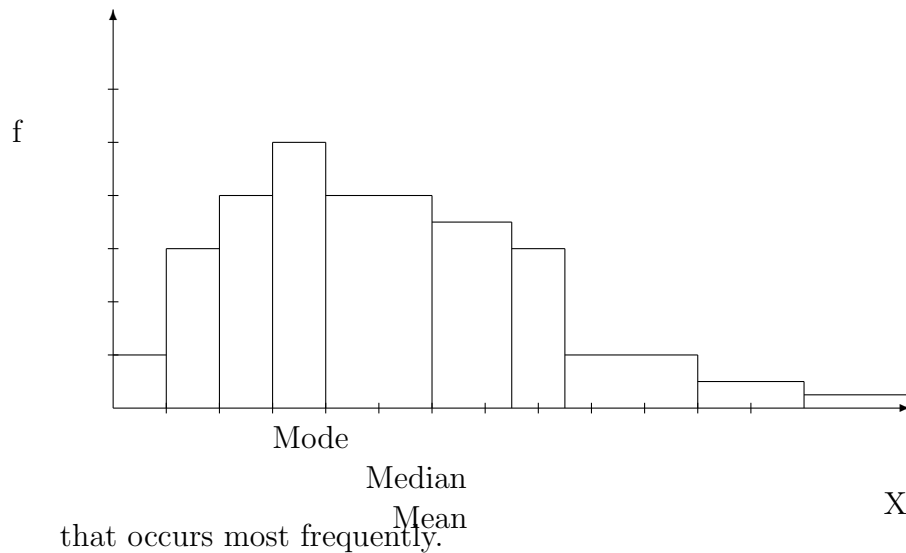
When a frequency of percentage distribution is not symmetrical, as in the case of Figure 2, the three measures generally diverge. The mode is again the value of X where the distribution reaches its peak, since this is the most common value. In a distribution like Figure 2, the median generally lies to the right of the mode. Finally, the large values of X , even if there are relatively few of them, can weight heavily in the sum of all the values, so when this sum is divided by the number of cases, the mean will lie even further to the right of centre in a distribution such as Figure 2.

Procedures for calculating each of these measures differ depending on whether the researcher has only a list of all possible values (ungrouped data) or whether the data have been grouped into a frequency or percentage distribution table (grouped data). In the following notes, procedures for calculating each of the mode, median, and mean are discussed for the case of both ungrouped and grouped data. These notes also discuss ways of interpreting the meaning of each of these three measures and which measure is most appropriate for different types of data sets.

Mode. Section 5.2, pp. 163-171.

Definition. The mode is the most common value of the variable or the value

Figure 2: Asymmetrical frequency distribution – skewed to the right



In the case of a list of numbers, or what we term ungrouped data, the mode is the value (or values) of the variable that occur most frequently. In order to determine this, all that is needed is to count the number of times each value occurs and the mode is the one with the greatest count or tally.

Example. The ages for the first eleven respondents from the SSAE98 data set are 27, 19, 18, 18, 18, 29, 20, 39, 24, 18, and 19. In order to make it easier to see which value occurs most frequently, it is best to place them in order and the number of cases can easily be counted. In increasing order by age, these eleven values are: 18, 18, 18, 18, 19, 19, 20, 24, 27, 29, and 38. From this it can easily be seen that the mode is 18. That is, there are four 18 year old respondents, two 19 year olds, and one of each of the other ages. Since 18 occurs more frequently than any of the other values, the mode is 18 years of age.

Some characteristics of the mode.

- Mode can mean fashion or what is most common or popular. That is, it refers only to the value or characteristic adopted by more people than any other value or characteristic.

- The mode may tell us little about the average, as the average is commonly understood. We often consider the average to be the centre of a distribution, but in the above example the mode is the smallest value. In that example, the mode is 18 since there are likely more students at the University aged 18 than any other single value of age. While the mode may not indicate the centre in examples like that above, it is useful as a measure when there is some distinctive value.
- One use for the mode is in terms of peak use or capacity utilization. In terms of having sufficient electric power capacity to ensure that there are no power blackouts, it is important for a power generating company to have sufficient power to meet peak needs. The company should thus plan its resources so it can meet this mode or peak use.
- There may be more than one mode. It is possible for there to be two or more values that occur more frequently than other values. The following example illustrates this:

Example. In the tally or count for the variable “Priority,” what should be the priority for using the federal government budget surplus (September 19), there were 4 respondents who favoured reducing the debt, 2 respondents who favoured reducing taxes, and 4 respondents who favoured social spending. In this case, there are two modes: reducing debt (category 1) and social spending (category 4)

Mode for grouped data.

Where the data have been grouped into categories or intervals, the mode is the value of the variable that occurs most frequently, or the value of the variable where the histogram reaches its peak.