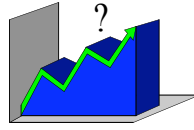


Ps2 Research Methods and Data Analysis



DESCRIPTIVE STATISTICS PART 1: Measures of central tendency

General Introduction

Descriptive statistics are usually the first step in the analysis of the data you have gathered in your research, particularly when your data are quantitative (numbers). They are very straightforward to calculate and provide an easy-to-manage summary of your data. They will often form the basis for the calculations you might perform later on, when you use inferential statistics. Despite their simplicity, descriptive statistics are very useful because they can extend your understanding of the data you have gathered in a number of ways (e.g. they can tell you how spread out your data are).

There are two different types of descriptive statistics:

- Measures of central tendency
- Measures of dispersion (on another handout)

In the box below, describe what is meant by Measures of central tendency (use Coolican to research your answer).

Measures of central tendency are

Measures of central tendency

There are three main measures of central tendency:

Mean

This is the *arithmetic average* of a set of data. It is represented in statistics by the symbol \bar{x} (pronounced *ex-bar*). It is calculated by adding up all the scores in the set of data and then dividing by the *number* of scores in the set of data. The formula for the mean is $\Sigma X/n$, which means "add up (Σ) the set of scores (X) and divide by ($/$) the number of scores (n). To see how this works, after looking at the example in the box below, calculate the mean for Data set 2, then check your answer with other students in your group.

Data set 1: 27 39 65 47 95 32 76 85 94 40 ($n=10$)

$$\Sigma X = (27 + 39 + 65 + 47 + 95 + 32 + 76 + 85 + 94 + 40) = 600$$

$$\bar{x} = 600/10 = 60$$

Data set 2: 120 130 110 154 173 150 176 125 164 100 137 105 ($n=$)

$$\Sigma X =$$

$$\bar{x} =$$

□ **Median**

This is the *positional average* of a set of data. It represents the central tendency of a set of data in a literal way. If a set of data has been placed into ascending or descending order then the median is that value which occupies the precise mid-point of the order. Note that the actual size of individual values does not matter. Exactly half the scores in any set of data will lie above the median and exactly half will lie below it. To see how this works, after looking at the example in the box below, calculate the median for Data set 2, then check your answer with other students in your group.

Data set 1: 27 39 65 47 95 32 76 85 94 40 (n=10)

In ascending order: 27 32 39 40 47 65 76 85 94 95

The central value(s): 27 32 39 40 47 65 76 85 94 95

The median is half way between these two values $((47+65)/2)$, at **56**

Data set 2: 120 130 110 154 173 150 176 125 164 100 137

In ascending order:

The central value is:

□ **Mode**

This is used less often than the other two measures of central tendency because it is less informative. It is the easiest of the three measures to find. It is simply the value that occurs most frequently in your set of data - the most popular value. Some sets of data will not have a mode because no one value in the data occurs more frequently than any other. Data sets 1 & 2 above are examples of this and neither would have a mode. Some sets of data have more than one mode - these are called **bimodal** (if there are two modes) or **multi-modal** (if there are three or more modes). An example of each of these appears in the box below. For Data set 4, work out the mode(s) and say whether the data set has a single mode, is bimodal or is multi-modal.

Data set 1: 10 11 12 13 14 14 14 17 19 21

Mode = 14

Data set 2: 10 11 12 13 14 14 17 17 19 21

Bimodal, modes = 14 and 17

Data set 3: 10 10 10 11 12 12 12 13 14 14 14 17 17 17 19 21

Multi-modal, modes = 10 and 12 and 14 and 17

Data set 4: 25 57 86 33 46 25 42 57 25 49 55 57

Mode, bimodal or multi-modal?

Value(s) of mode(s)?

How do you choose the measure of central tendency that you should use? First of all, you need to consider the advantages and disadvantages of each. Use Coolican to find this out and complete the box below

Mean	
Advantages:	
Disadvantages:	
Median	
Advantages:	
Disadvantages:	
Mode	
Advantages:	
Disadvantages:	

After considering the pros and cons of each measure of central tendency you would look at your set of data and decide which would be the best to use. As a rough guide: if you had a set of data with no extreme values (values a lot bigger or smaller than the other values in the data set) then you could use the mean; if you had a set of data with extreme values or that you collected from a questionnaire you should probably use the median; if you had a set of data where it makes no sense to use the mean or median (for example, if you had counted numbers of smokers and non-smokers who had a cough) then you should use the mode. There is a little more to it than this, but you will find out what in due course!