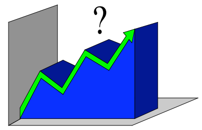
Research Methods and Data Analysis



DESCRIPTIVE STATISTICS PART 2: Measures of dispersion

Introduction

As outlined on the previous handout (Descriptive statistics Part 1: Measures of Central Tendency), there are two different types of descriptive statistics:

- Measures of central tendency (on the other handout)
- Measures of dispersion

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

Measures of dispersion are

Measures of dispersion

There are three main measures of dispersion:

- **Range**

  This is the simplest but least informative of all the measures of dispersion. It is calculated by finding the difference between the lowest and the highest values in a set of data. It tells us the gap between the top and bottom limits of the data, but it tells you nothing about how the data values are arranged within those top and bottom limits. Of course, the range is greatly affected by extreme values in the data. To see how the range works, after looking at the example in the box below, calculate the range 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

Range = 95 (highest value) - 27 (lowest value) = 68

Data set 2: 120 130 110 154 173 150 176 125 164 100 137 105

Range =

- **Variance**

  This is based on the mean and indicates the spread of scores around the mean. It takes the mean of your data set and compares each score making up your data set with the mean. It is (now bare with me!) *the mean of the squared differences between each of the values and their mean.* So, to unpack that sentence:

1. You take a score
2. Subtract (minus) the mean from that score.
3. Square the result of this (i.e. multiply it by itself)
4. Do this for each score in your data set
5. Add up all these squares
6. Divide by the number of scores in your data set
7. The result is the variance.

The mathematical formula for this is: $\sum d^2/n$, where d= the difference between a score from your data set and the mean. To see how this works, after looking at the example in the box below, calculate the variance 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) 🖱 = 60

The next bit can be calculated using the statistical function on your calculator, or you can do it as follows:

60-27=33;	33 x 33=	1089
60-39=21;	21 x 21=	441
60-65=-5;	-5 x –5=	25
60-47=13;	13 x 13=	169
60-95=-35;	-35 x –35 =	1225
60-32=28;	28 x 28 =	784
60-76=-16;	16 x 16 =	256
60-85=-25;	25 x 25 =	625
60-94=-34;	34 x 34 =	1156
60-40 =20;	20 x 20 =	400

1089 + 441 + 25 + 169 + 1225 + 784 + 256 + 625 + 1156 + 400 = 6170

$\sum d^2$ = 6170

$\sum d^2/n$= 6170/10 = 617

The variance = 617

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

n=

🖱 =

$\sum d^2$ =

$\sum d^2/n$=

The variance =

- **Standard Deviation**

  This is used more often than the other two measures of dispersion. It is related to the variance directly – it is simply the square root of the variance. So, the only difference between the variance and the standard deviation is that in the calculation of the variance the differences between the mean and score remain squared whereas in the standard deviation the effect of squaring is removed by taking the square root of the result as the last step in the calculation. The symbol for the standard deviation is ⇖, and standard deviation can be abbreviated to **sd** . The basic formula for the sd of a set of scores is ⇖ **=** $\sqrt{\sum d^2/n}$. So, the formula is exactly the same as the formula for the variance, but you square root the result of $\sum d^2/n$. To see how this works, after looking at the example in the box below, calculate the standard deviation 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) 🖱 = 60

The next bit can be calculated using the statistical function on your calculator, or:

60-27=33;	33 x 33=	1089
60-39=21;	21 x 21=	441
60-65=-5;	-5 x –5=	25

60-47=13;	13 x 13=	169
60-95=-35;	-35 x –35 =	1225
60-32=28;	28 x 28 =	784
60-76=-16;	16 x 16 =	256
60-85=-25;	25 x 25 =	625
60-94=-34;	34 x 34 =	1156
60-40 =20;	20 x 20 =	400

1089 + 441 + 25 + 169 + 1225 + 784 + 256 + 625 + 1156 + 400 = 6170

$\sum d^2 = 6170$ $\sum d^2/n = 6170/10 = 617$ $\sqrt{\sum d^2/n} = 24.84$ ⇖ = 24.84

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

n= 🖰 = $\sum d^2 =$ $\sum d^2/n =$ $\sqrt{\sum d^2/n} =$ ⇖ =

The sd gets a little more complicated than this. There are actually two forms of the sd:

- The one described above (denoted by the symbol ⇖) is, in fact, the standard deviation when the whole population has been sampled. When we say "the whole population" we do not mean that you are expected to sample every person in the world! The population simply means the particular type of people who are participating in your investigation. For example, if you are investigating the students in a class as your entire population of interest rather than a sample from some larger population then you can use the ⇖. If you are actually attempting to measure something about the whole population and generalise from your results then you should not use the ⇖.
- The alternative to ⇖, which you should use when the sample you have used are only part of some population and you want to say something about the whole population then you should use the following formula instead: $\sqrt{\sum d^2/(n-1)}$. So, it's almost the same as ⇖, but you take one away from the number of scores in your data set. This gives you the standard deviation of your data set and is denoted by the symbol ***s***. In fact, because you are almost always going to be trying to get behind the data set and generalise to the whole population then you will almost certainly use ***s*** rather than ⇖.

Just to check that you understand this, use Coolican to fill in the box below.

You would use ⇖ when

You would use ***s*** when

If any of this is unclear do not struggle on alone. Ask the other people in your class and/or ask your tutor. It is not worth worrying about this – help is at hand!