Introduction to One-Sample t-Tests

Consider this hypothetical situation.

The UK government is planning the construction of a new airport that is 2 miles from a small village. They produce a report which states that the average noise level of planes overhead for residence of this village should be 50dB which is considered to be an acceptable level of noise pollution.

Residents of this village are suspicious of this 50dB figure, and decide to travel to an already existing airport (of the same size as the proposed airport) and stand 2 miles away with a decibel meter. They take a decibel reading every 5 minutes for an hour, producing the following 13 readings.

54, 44, 59, 61, 53, 68, 55, 43, 52, 52, 51, 53, 57

The residents find that the average of these 13 readings is 54dB. Do they have enough evidence to challenge the UK government on the 50dB figure?

It’s an interesting question; whether 54 is significant different to 50. It might be, it might not be. There is where statistics can help us.

One could argue that since 54dB > 50dB, the residents can definitely challenge the government’s figure. However, would this still be true if the average of the 13 readings was 51dB? Perhaps not, as 51dB is very close to 50dB, and these results might well be consistent with the government’s estimate. If, however, these 13 readings gave an average of 75dB, then the evidence would be very strong that the government’s figure is wrong. Clearly **the difference between the government’s figure of 50dB and the residents’ figure** influences the strength of the evidence.

Suppose now that the average of the resident’s observations is 54dB. Might you still feel you have enough evidence if they only took one or two observations? Perhaps not, as this is not very much data. If, however, the average of 54dB resulted from 200 observations, then the evidence would be much stronger. Clearly **the number of observations the residents take** will also influence the strength of the evidence.

Suppose now that the residents do have 13 observations with an average of 54dB. If you were to learn that all 13 of observations were between 53dB and 55dB, might this alter the strength of evidence? Surely yes, as it seems now almost impossible that the true average could be as low as 50dB. Suppose, however, you were to learn that the 13 observation were well spread between 34dB and 64dB. How might this affect the strength of evidence? Surely there is more of a possibility now that 50dB could be the true mean. Clearly **the spread of the residents’ data** will influence the strength of the evidence.

We have argued that there are three factors that will influence the strength of the evidence against the government’s claim; the average of the residents’ decibel readings, the number of readings and the spread of these values. These are called, respectively; the sample mean, the sample size and the sample standard deviation; and we use the respective generic symbols; , and .

As for the government’s claim that the true mean is 50dB, we call this the *null hypothesis* and label this as . We use the symbol to represent the true mean, and so the null hypothesis is

In addition to the *null hypothesis* will be the *alternative hypothesis*, . This describes the hypothesis that the residents are actually trying to prove. In this case, they wouldn’t mind if the true mean level is less than 50dB, as this would imply that the village is not affected by the noise from the airport; it’s only if the mean is *larger* than 50dB that the residents would be able to complain. Hence, in this case,

The question remains. How much evidence do we have to reject and and hence accept ? To answer this, we use what is called the one-sample -test. This is a simple procedure with three steps.

## Step 1

Calculate the *test statistic*, . This value incorporates all three factors discussed earlier; , and .

The test statistic will be larger if

* is far away from (data is different)
* is small (consistent data)
* is large (many data points)

In all three of these cases, this should equate to *more* evidence against .

The test statistic will be closer to if

* is close to (data is similar)
* is large (inconsistent data)
* is small (few data points)

In all three of these cases, this should equate to *less* evidence against .

## Step 2

Calculate the -*value*. The -value is the over-all probability of obtaining a test statistic of , or more extreme, if is true. In other words, it is the chance of getting our data as extreme as this *even if* .

To calculate the -value, we use the *Student -distribution* with degrees of freedom, sometimes shortened to . If then

-value

Calculating the -value is similar to finding the area under the Normal distribution (the bell curve). You can use computer software, phone apps, the internet, or probability tables. Regardless of which method you use, you will need to provide the *degrees of freedom* which will always be one fewer than the sample size, .

## Step 3

Make your judgment on . The -value will be your guide. In essence, if the -value is small than the *test statistic* was too large (hence, very unlikely to happen by chance) and so we conclude that there is sufficient evidence against the null hypothesis. If, however, the -value is not small then the *test statistic* was not too large (hence, quite a reasonable value to expect) and so we conclude that there is insufficient evidence against the null hypothesis.

In summary, if the -value is small, reject . If the -value is not small, don’t reject .

What counts as small? This will depend on the field of study, as different fields will have different standards. An often-quoted cross-over value is -value .

Let’s look at our example. With and (we can also calculate ) we have

We now calculate the -value using one of the methods described earlier (below, *GeoGebra* has been used). We find that -value .



If we choose to use a cut-off -value of , we would declare that there is sufficient evidence to suggest that the government’s figure of 50dB is incorrect; the true mean is larger.

### Activity

In each of the following situations, use a *one-sample* -*test* to determine if there is enough evidence to reject the claim. Use -value as a cut-off.

1. A local dairy farmer manufactures and produces one-litre milk bottles but recently some customers have been suspicious that they are purchasing less milk than what is advertised. They confront the farmer who replies with “while you sometimes may have slightly less or slightly more than one litre of milk, the average value is definitely one litre”.

The customers are not convinced and so purchase 11 bottles and accurately measure the exact amount of milk in each bottle. The measurements (in mL) are 997, 995, 1001, 998, 1000, 1001, 994, 998, 995, 1000, 999, giving an average of and a standard deviation of .

Do the customers have enough evidence to claim that they are being under-sold on these milk bottles?

1. Twelve plants of a certain variety, grown under uniform conditions and treated with a *new brand of fertiliser*, grow to the following heights (in cm); 25, 28, 24, 23, 27, 30, 24, 21, 28, 30, 26, 27; giving a mean of and a standard deviation of .

It is well-known that plants of this variety growing with the *standard fertiliser* grow to an average height of .

Is there evidence that the new brand of fertiliser increases the average height of these plants?

1. A management of a coffee shop wants to guarantee that all of their lattes are consistent. They believe that the perfect latte should have of espresso, but accept that there will always be a slight error when staff are making lattes. As long as the average is close to then the management will be happy, otherwise they will need to take action to either increase or decrease the spoon sizes that the staff are using.

They order their staff to create a random sample of 25 lattes but by precisely measuring the amount of espresso used. The sample produces a mean of and a standard deviation of .

Is there any evidence that the staff are not using the correct amount of espresso on average?