Significance testing

Significance testing aims to make **statements** about a **population parameter**, or parameters, on the basis of **sample evidence**. It is sometimes called **point estimation**. The emphasis here is on testing whether a set of sample results support, or are consistent with, some fact or supposition about the population. We are thus concerned to get a "Yes" or "No" answer from a **significance test**; either the sample results do support the supposition, or they do not.

Since we are dealing with samples, and not a census, we can never be 100% sure of our results. However, by testing an idea (or hypothesis) and showing that it is very, very unlikely to be true we can make useful statements about the population which can form the basis for decision making.

Significance testing is about putting forward an idea or **testable proposition that captures the properties of the problem** that we are interested in. The idea we are testing may have been obtained from a past census or survey, or it may be an assertion that we put forward, or has been put forward by someone else, for example, an assertion that a majority of people support legislation to enhance animal welfare. Similarly, it could be a critical value in terms of product planning; for example, if a new machine is to be purchased but the process is only viable with a certain level of throughput, we would want to be sure that the company can sell, or use, that quantity before commitment to the purchase of that particular type of machine. Advertising claims could also be tested using this method by taking a sample and testing if the advertised characteristics, for example strength, length of life, or percentage of people preferring this brand, are supported by the evidence of the sample.

OBJECTIVES:
After working through this chapter, you will be able to:

> ➢ understand and apply the concept of a significance test;

> ➢ use a hypothesis test on a population mean or percentage;

> ➢ construct one-sided tests of hypotheses

1. Hypothesis testing for single samples

Hypothesis testing is merely an alternative name for **significance tests**, the two being used interchangeably. This name does stress that we are testing some supposition about the population, which we can write down as a hypothesis. In fact, we will have **two hypotheses** whenever we conduct a significance test: one relating to the supposition that we are testing, and one which describes *the alternative situation*.

The first hypothesis relates to the claim, supposition or previous situation and is usually called the **null hypothesis** and labeled as H_0. It implies that there has been no change in the value of the parameter that we are testing from that which previously existed; for example, if the average spending per week on beer last year amongst 18 to

25 year olds was £15.53, then the null hypothesis would be that it is still £15.53. If it has been claimed that 75% of consumers prefer a certain flavour of ice-cream, then the null hypothesis would be that the population percentage preferring that flavour is equal to 75%.

The null hypothesis could be written out as a sentence, but it is more usual to abbreviate it to:

$$H_0: \mu = \mu_0$$

for a mean where μ_0 is the claimed or previous population mean.

For a percentage:

$$H_0: \pi = \pi_0$$

where π_0 is the claimed or previous population percentage.

The second hypothesis summarizes what will be the case if the null hypothesis is **not true**. It is usually called the **alternative hypothesis (*fairly obviously!*)**, and is labeled as H_A or as H_1 depending on which text you follow: we will use the H_1 notation here. This alternative hypothesis is **usually not specific**, in that it does not usually specify the exact alternative value for the population parameter, but rather, it just says that some other value is appropriate on the basis of the sample evidence; for example, the mean amount spent is not equal to £15.53, or the percentage preferring this flavour of ice-cream is not 75%.

As before, this hypothesis could be written out as a sentence, but a shorter notation is usually preferred:

$$H_1: \mu \neq \mu_0$$

for a mean where μ_0 is the claimed or previous population mean.

For a percentage:

$$H_1: \pi \neq \pi_0$$

where π_0 is the claimed or previous population percentage.

Whenever we conduct a hypothesis test, **we assume that the null hypothesis is true whilst we are doing the test**, and then come to a conclusion on the basis of the figures that we calculate during the test.

Since the sampling distribution of a mean or a percentage (for large samples) is given by the Normal Distribution, to conduct a test, we compare the sample evidence (sample statistic) with the null hypothesis (what is assumed to be true). This difference is measured by a **test statistic**. A test statistic of zero or reasonably near to zero would suggest the null hypothesis is correct. The larger the value of the test statistic the larger the difference between the sample evidence and the null hypothesis. If the difference is sufficiently large we conclude that the null hypothesis is incorrect and accept the alternative hypothesis. To make this decision we need to decide on a **critical value** or **critical values**. The critical value or values define the point at which the chance of the null hypothesis being true is at a small, predetermined level, usually 5% or 1% (called the significance level). In this chapter you will see the z-value being used as a test statistic. It is more usual to divide the Normal distribution diagram into sections or areas, and to see whether the z-value falls into a particular section; then we may either accept or reject the null hypothesis.

However, it is worth noting that some **statistical packages** just calculate the probability of the null hypothesis being true when you use them to conduct tests. The reason for doing this is to see how likely or unlikely the result is. If the result is particularly unlikely when the null hypothesis is true, then we might begin to question whether this is, in fact, the case. Obviously we will need to define more exactly the phrase "**particularly unlikely**" in this context before we can proceed with hypothesis tests.

Most tests are conducted at the 5% level of significance, and you should recall that in the normal distribution, the z-values of +1.96 and -1.96 cut off a total of 5% of the distribution, 2.5% in each tail. If a calculated z-value is between -1.96 and +1.96, then we accept the null hypothesis; if the calculated z-value is below -1.96 or above +1.96, we reject the null hypothesis in favour of the alternative hypothesis. This situation is illustrated in Figure 1.
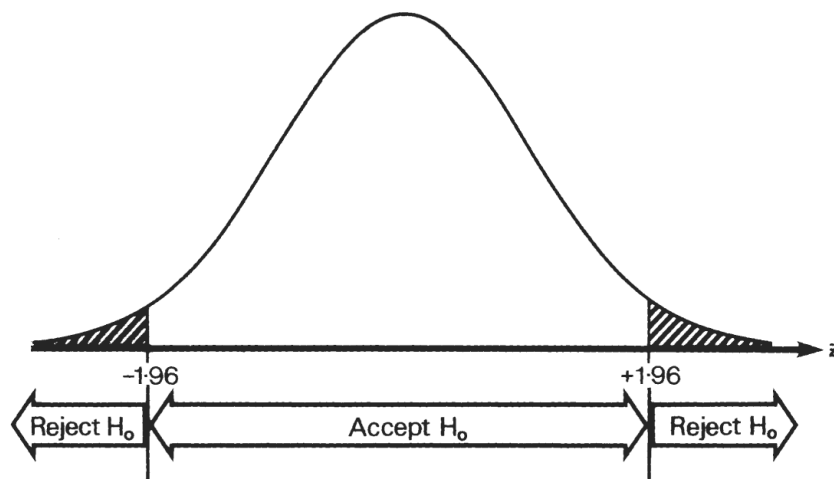
Figure 1   Acceptance and rejection regions.

If we were to conduct the test at the 1% level of significance, then the two values used to cut off the tail areas of the distribution would be +2.576 and -2.576.

For each test that we wish to conduct, the basic layout and procedure will remain the same, although some of the details will change, depending upon what exactly we are testing. A proposed layout is given below, and we suggest that by following this you will present clear and understandable significance tests (and not leave out any steps).

	Step	**Example**
1	State hypotheses.	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
2	State significance level.	5%
3	State critical (cut-off) values.	-1.96 +1. 96
4	Calculate the test statistic (z).	Answer varies for each test, but say

			2.5 for example.
5	Compare the z value to the critical values		In this case it is above +1.96.
6	Come to a conclusion.		Here we would reject H_0.
7	Put your conclusion into English.		The sample evidence does not support the original claim that the mean was the specified value.

Whilst the significance level may change, which will lead to a change in the critical values, the major difference as we conduct different types of hypothesis test is in the way in which we calculate the z value at Step 4

1.1 A test statistic for a population mean

In the case of testing for a particular value for a population mean, this formula used to calculate z in Step 4 and we use the sample standard deviation (s) in place of the population value (since it is not known). We are assuming that the null hypothesis is true, and so $\mu = \mu_0$ The formula will become:

$$z = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

We are now in a position to carry out a test.

EXAMPLE

A production manager claims that an average of 50 boxes per hour are filled with finished goods at the final stage of a production line. A random sampling of 48 different workers, at different times, working at the end of identical production lines shows an average number of boxes filled as 47.5 with a standard deviation of 0.7 boxes. Does this evidence support the assertion by the production manager at the 5% level of significance?

Step 1   The null hypothesis is based on the production manager's assertion:
     H_0: $\mu = 50$ boxes per hour
     The alternative hypothesis is any other answer:
     H_1: $\mu \neq 50$ boxes per hour

Step 2   As stated in the question, this level is 5%

Step 3   The critical values are -1.96 and $+1.96$ (from Appendix C)

Step 4   Using the formula given above, the z value can be calculated as:

$$z = \frac{47.5 - 50}{0.7 / \sqrt{48}} = \frac{-2.5}{0.101036} = -24.744$$

Step 5   The calculate value (-24.744) is **below** the lower critical value (-1.96)

Step 6   We may therefore **reject the null hypothesis** at the 5% level of significance

Step 7   The sample evidence does not support the production manager's assertion that the average number of boxes filled per hour is 50. The number is **significantly different** from 50.

Although there is only a 2.5 difference between the claim and the sample mean in this example, the test shows that the sample result is significantly different from the claimed value. In all hypothesis testing, it is **not the absolute difference** between the values which is important, but the number of standard errors that the sample value is away from the claimed value. Statistical significance should not be confused with the notion of **business significance** or **importance**.

1.2 A test for a population percentage

Here the process will be identical to the one followed above, except that the formula used to calculate the z-value will need to be changed. (You should recall from Chapter 11 that the standard error for a percentage is different from the standard error for a mean.) The formula will be:

$$z = \frac{p - \Pi_0}{\sqrt{\dfrac{\Pi_0 (100 - \Pi_0)}{n}}}$$

where p is the sample percentage and Π_0 is the claimed population percentage (remember that we are assuming that the null hypothesis is true).

EXAMPLE

An auditor claims that 10% of invoices for a company are incorrect. To test this claim a random sample of 100 invoices are checked, and 12 are found to be incorrect. Test, at the 5% significance level, if the auditor's claim is supported by the sample evidence.

Step 1: The hypotheses can be stated as:

$H_0 : \pi_0 = 10\%$

$H_1 : \pi_0 \neq 10\%$

Step 2: The significance level is 5%.

Step 3: The critical values are -1.96 and +1.96.

Step 4: The sample percentage is $\dfrac{12}{100} \times 100 = 12\%$.

$$z = \frac{12 - 10}{\sqrt{\dfrac{10(100 - 10)}{100}}} = 0.67$$

Step 5: The calculated value falls between the two critical values.

Step 6: We therefore cannot reject the null hypothesis.

Step 7: The evidence from the sample is consistent with the auditor's claim that 10% of the invoices are incorrect.

In this example, when the calculated value falls between the two critical values. The answer is to **not reject** the claim of 10% rather than to accept the claim. This is because we only have the sample evidence to work from, and we are aware that it is subject to sampling error. The **only way** to firmly accept the claim would be to check every invoice (i.e. carry out a **census**) and work out the actual population percentage which are incorrect. We have already made the **assumption** that the null hypothesis is true (whilst we are conducting the test), and so we can use the hypothesized value of the population percentage in the calculations. The formula for the sampling error will therefore be

$$\sqrt{\left[\frac{\Pi_0\left(100-\Pi_0\right)}{n}\right]}$$

CASESTUDY

It has been claimed on the basis of census results that 87% of households in Tonnelle now have exclusive use of a fixed bath or shower with a hot water supply. In the Arbour Housing Survey of this area, 246 respondents out of the 300 interviewed reported this exclusive usage. Test at the 5% significance level whether this claim is supported by the sample data.

Step 1: The hypothesis can be stated as:
      H0: $\pi_0 = 87\%$ H1: $\pi_0 \neq 87\%$
      We are assuming that the claim being made is correct.
Step 2: The significance level is 5% (but could have been set at a different level if required).
Step 3: The critical values are -1.96 and +1.96.
Step 4: The sample percentage is (246/300) x 100 = 82%

$$z = \frac{82-87}{\sqrt{\left[\frac{87\left(100-87\right)}{300}\right]}} = -2.58$$

Step 5: The calculated value falls below the critical value of -1.96.
Step 6: We therefore **reject** the null hypothesis at the 5% significance level.
Step 7: The sample evidence does not **support** the view that 87% of households in the Tonnelle area have exclusive use of a fixed bath or shower with a hot water supply. Given a sample percentage of 82% we could be tempted to conclude that the percentage was lower - the sample does suggest this, but the test was not structured in this way and we must accept the alternative hypothesis that the population percentage is not likely to be 87%. We will consider this issue again when we look at one-sided tests.

2. One-sided significance tests

In general, we want to specify whether the real value is **above or below** the claimed value in those cases where we are able to reject the null hypothesis. One-sided tests will allow us to do exactly this. The method employed, and the appropriate test statistic which we calculate, **will remain exactly the same**; it is the **hypotheses** and the **interpretation** of the answer which will change. Suppose that we are investigating the purchase of cigarettes, and know that the percentage of the adult population who regularly purchased last year was 34%. If a sample is selected, we do not want to know only whether the percentage purchasing has changed, but rather whether it has decreased (or increased). **Before** carrying out the test it is necessary to decide which of these two propositions you wish to test.

If we wish to test whether or not the percentage has decreased, then our hypotheses would be:

|	
null hypothesis	$H_0: \pi = \pi_0$
alternative hypothesis	$H_1: \pi < \pi_0$

where π_0 is the actual percentage in the population last year. This may be an appropriate hypothesis test if you were working for a health lobby.

If we wanted to test if the percentage had increased, then our hypotheses would be:

null hypothesis	$H_0: \pi = \pi_0$
alternative hypothesis	$H_1: \pi > \pi_0$

This could be an appropriate test if you were working for a manufacturer in the tobacco industry and were concerned with the effects of improved packaging and presentation.

To carry out the test, we will want to concentrate the chance of rejecting the null hypothesis **at one end** of the Normal distribution. Where the significance level is 5%, then the critical value will be **-1.645** (i.e. the cut off value taken from the Normal distribution tables) for the hypotheses $H_0: \pi = \pi_0$, $H_1: \pi < \pi_0$; and **+1.645** for the hypotheses $H_0: \pi = \pi_0$, $H_1: \pi > \pi_0$. Where the significance level is set at 1%, then the critical value becomes either **-2.33 or +2.33**. In terms of answering examination questions, it is important to *read the wording very carefully* to determine which type of test you are required to perform.
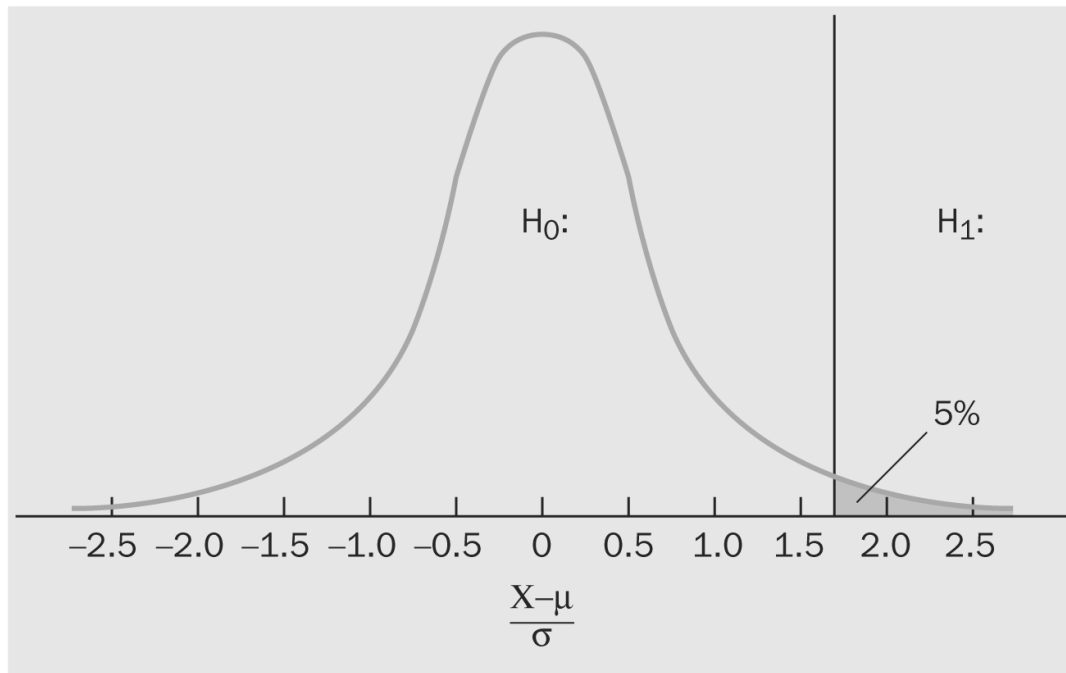
Figure 2   One-tailed test.

The interpretation of the calculated z-value is now merely a question of deciding into which of two sections of the Normal distribution it falls.

3.1 A one-sided test for a population mean

Here the hypotheses will be in terms of the population mean, or the claimed value. Consider the following example.

EXAMPLE

A manufacturer of batteries has assumed that the average expected life is 299 hours. As a result of recent changes to the filling of the batteries, the manufacturer now wishes to test if the average life has increased.

A sample of 200 batteries was taken at random from the production line and tested. Their average life was found to be 300 hours with a standard deviation of 8 hours. You have been asked to carry out the appropriate hypothesis test at the 5% significance level.

Step 1: $H_0 : \mu = 299$

   $H_1 : \mu > 299$

Step 2: The significance level is 5%.

Step 3: The critical value will be + 1.645.

Step 4:

$$z = \frac{300 - 299}{8/\sqrt{200}} = 1.77$$

(Note that we are still assuming the null hypothesis to be true while the test is conducted.)

Step 5: The calculated value is larger than the critical value.

Step 6: We may therefore reject the null hypothesis.

(Note that had we been conducting a two-sided hypothesis test, then we would have been unable to reject the null hypothesis, and so the conclusion would have been that the average life of the batteries had not changed.)

Step 7: The sample evidence supports the supposition that the average life of the batteries has increased by a significant amount.

Although the significance test has shown that there has been a significant increase in the average length of life of the batteries, this **may not be an important conclusion** for the manufacturer. For instance, it would be quite misleading to use it to back up an advertising campaign which claimed that "our batteries now last even longer!" Whenever hypothesis tests are used it is important to **distinguish between statistical significance and importance**. This distinction is often ignored.

CASESTUDY

It has been argued that, because of the types of property and the age of the population in Tonnelle, average mortgages are likely to be around £200 a month. However, the Arbour Housing Trust believe the figure is higher because many of the mortgages are relatively recent and many were taken out during the house price boom. Test this proposition, given the result from the Arbour Housing Survey that 100 respondents were paying an average monthly mortgage of £254. The standard deviation calculated from the sample is £72.05 (it was assumed to be £70 in an earlier example). Use a 5% significance level.

1       $H_0: \mu_0 = £200$
        $H_1: \mu_0 > £200$

2   The significance level is 5%.

3   The critical value will be +1.645.

4       $z = \dfrac{£254 - £200}{£72.05/\sqrt{100}} = 7.49$

5   The test statistic value of 7.49 is greater than the critical value of +1.645.

6   On this basis, we can **reject the null hypothesis** that monthly mortgages are around £200 in favour of the alternative that they are likely to be more.

7   A test provides a means to clarify issues and resolve different arguments. In this case the average monthly mortgage payment is higher than had been argued but this **does not necessarily** mean that the reasons put forward by the Arbour Housing Trust for the higher levels are correct. In most cases, further

investigation is required. It is always worth checking that the sample structure is the same as the population structure. In an area like Tonnelle we would expect a range of housing and it can be difficult for a small sample to reflect this. The conclusions also refer to the average. The pattern of mortgage payments could vary considerably between different groups of residents. It is known that the Victorian housing is again attracting a more affluent group of residents, and they could have the effect of pulling the mean upwards. Finally, the argument for the null hypothesis could be based on dated information, e.g. the last census.

3.2 A one-sided test for a population percentage

Here the methodology is exactly the same as that employed above, and so we just provide two examples of its use.

EXAMPLE

A small political party expects to gain 20% of the vote in by-elections. A particular candidate from the party is about to stand in a by-election in Derbyshire South East, and has commissioned a survey of 200 randomly selected voters in the constituency. If 44 of those interviewed said that they would vote for this candidate in the forthcoming by-election, test whether this would be significantly above the national party's claim. Use a test at the 5% significance level.

1. $H_0 : \pi = 20\%$

   $H_1 : \pi > 20\%$

2. The significance level is 5%.

3. The critical value will be =+1.645.

4. Sample percentage = $\frac{44}{200}$ X100 =22%

$$z = \frac{22 - 20}{\sqrt{\left[\frac{20(100 - 20)}{200}\right]}} = 0.7071$$

5. 0.7071 <1.645.

6. Cannot reject H_0 .

7. There is no evidence that the candidate will do better than the national party's claim.

Here there may not be evidence of a statistically significant vote above the national party's claim (a sample size of 1083 would have made this 2% difference statistically significant), but if the candidate does manage to achieve 22% of the vote, and it is a close contest over who wins the seat between two other candidates, then the extra 2% could be very important.

EXAMPLE

A fizzy drink manufacturer claims that over 40% of teenagers prefer its drink to any other. In order to test this claim, a rival manufacturer commissions a market research company to conduct a survey of 200 teenagers and ascertain their preferences in fizzy drinks. Of the randomly selected sample, 75 preferred the first manufacturers drink. Test at the 5% level if this survey supports the claim.

Sample percentage is: $\dfrac{75}{200} \times 100 = 37.5\,\%$

1. $H_0: \pi = 40\%$; $H_1: \pi < 40\%$
2. The significance level is 5%
3. The critical value will be -1.645
4. $z = \dfrac{37.5 - 40}{\sqrt{\dfrac{40(100 - 40)}{250}}} = 0.807$
5. $-1.645 < -0.807$
6. Cannot reject H_0
7. There is insufficient evidence to show that less than 40% of teenagers prefer that particular fizzy drink.