Statistics 2 for AQA contents

5 Hypothesis testing

In this chapter you will learn

- what is meant by null hypothesis, alternative hypothesis, significance level, test statistic, critical value, critical region, Type I and Type II errors, one-tailed and two-tailed tests
- how to test for the mean of a normal distribution with either known or unknown variance
- how to test for the mean of a distribution using a normal approximation

A Basic ideas of hypothesis testing (answers p 128)

A person claims to have the ability to predict the outcome of a throw of a coin.

In order to test this claim, the person is asked to predict the outcomes of four throws. All four of his predictions turn out to be correct.

A1 If the person has no special ability but is just guessing, then the probability that each individual prediction is correct is $\frac{1}{2}$.

(a) What is the probability that all four predictions will be correct?

(b) Do you think that there is sufficient evidence that the person has some ability to predict the outcome of a throw?

Four is a small number of trials. A new test is to be carried out in which the person will be asked to predict the outcomes of 20 throws.

If the person is just guessing, then as before the probability of predicting an outcome correctly is $\frac{1}{2}$. So we would expect about 10 predictions to be correct. If he gets a lot more than 10 correct, then that would be evidence of some ability to predict.

The number of correct predictions out of 20 follows the binomial distribution with $n = 20$ and $p = \frac{1}{2}$.

Part of the table of cumulative probabilities for this distribution is shown on the right.

x	Probability that number of correct predictions $\leq x$
10	0.5881
11	0.7483
12	0.8684
13	0.9432
14	0.9793
15	0.9941
16	0.9987
17	0.9998
18	1.0000
19	1.0000
20	1.0000

From the table we can find the probability of getting, for example, 12 or more predictions correct, like this:

$$\text{P(12 or more correct)} = 1 - \text{P(up to 11 correct)}$$
$$= 1 - 0.7483 = 0.2517$$

Before the test is carried out, we have to decide how many predictions will need to be correct for us to be convinced that the result is not due to just guessing.

A2 Would you be convinced if the number of correct predictions were

(a) 12 or more **(b)** 13 or more **(c)** 14 or more **(d)** 15 or more

A **hypothesis** is a statement which may be true or may be false – we do not yet know.

In the example discussed so far, there are two hypotheses we have to decide between:

(1) The person has no special ability but is just guessing.

(2) The person has some ability to predict.

In terms of probability, these two hypotheses can be restated as:

(1) The probability of correctly predicting the outcome of a throw is $\frac{1}{2}$.

(2) The probability of correctly predicting the outcome of a throw is greater than $\frac{1}{2}$.

Unless the evidence is very convincing, then it is the first hypothesis that we will accept. This hypothesis is the **null hypothesis**. It says that this person is no different from anyone else.

The other hypothesis is the **alternative hypothesis**.

If the null hypothesis is true, then from the table on the opposite page the probability of correctly predicting 15 or more outcomes out of 20 is $1 - 0.9793 = 0.0207$, or approximately 2%.

We may decide that getting a result with a probability of only 2% is enough to make us reject the null hypothesis. If so, then 2% is called the **level of significance** of the result of the test.

If we decide to set the level of significance at 2%, then we will accept the null hypothesis if the person correctly predicts 14 outcomes or fewer, but reject it if the person correctly predicts 15 or more.

A3 The person is asked to predict the outcomes of 30 throws. Cumulative probabilities based on the null hypothesis are shown in the table on the right.

(a) The tester decides to reject the null hypothesis if the number of correct predictions is 19 or more. What level of significance (to the nearest 1%) is the tester using?

(b) If the level of significance is set at 5%, what is the smallest number of correct predictions that would lead to rejection of the null hypothesis?

x	Probability that number of correct predictions $\leq x$
15	0.5722
16	0.7077
17	0.8192
18	0.8998
19	0.9506
20	0.9786
21	0.9919
22	0.9974
23	0.9993
24	0.9998
25	1.0000
26	1.0000
27	1.0000
28	1.0000
29	1.0000
30	1.0000

The level of significance of a test shows how convincing the evidence is: the smaller the level of significance, the more convincing is the evidence.

B Mean of a normal distribution with known variance (answers p 128)

The next example of hypothesis testing introduces some other commonly used terms.

Imagine that a car engine is modified to try to increase the maximum speed of the car.

The maximum speed in m.p.h. of cars with unmodified engines is known to be distributed normally with mean 125 and standard deviation 3.5.

It will be assumed that the modification may affect the mean but will not affect the variance. So the standard deviation is unchanged.

The null hypothesis, denoted by $\mathbf{H_0}$, is that the mean maximum speed μ of the modified cars has not increased, so μ is still 125.

The alternative hypothesis $\mathbf{H_1}$ is that the mean maximum speed has increased, so that $\mu > 125$.

H_0: $\mu = 125$

H_1: $\mu > 125$

Test statistic

In order to test the effect of the modification, a random sample of 10 modified cars is to be selected and the cars' maximum speeds measured.

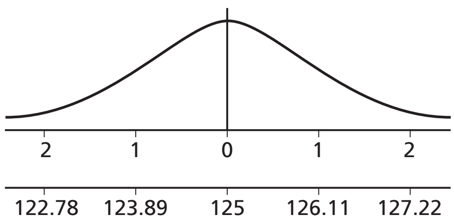
We know that the sample mean $\overline{X}$ is a random variable that is normally distributed with mean μ, variance $\frac{\sigma^2}{n}$ and standard deviation, or standard error, $\left(\frac{\sigma}{\sqrt{n}}\right)$.

We shall use the value of $\overline{X}$ to decide between the two hypotheses.
$\overline{X}$ is the **test statistic**.

Critical value, critical region, acceptance region

If hypothesis H_0 is true, then $\mu = 125$ and $\sigma = 3.5$.
So $\overline{X}$ will be normally distributed with mean 125 and standard error $\frac{3.5}{\sqrt{10}} = 1.11$ (to 2 d.p.).

The diagram shows the distribution of $\overline{X}$ in relation to the standard normal distribution.



We will reject H_0 if the value of $\overline{X}$ that we get is so large that it would be very unlikely to happen if H_0 is true.

'Very unlikely' is usually taken to mean 'with probability 5%'.
5% is the level of significance of the test that we shall use.

Let Z be the standardised variable corresponding to $\overline{X}$. So $Z = \dfrac{\overline{X} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)} = \dfrac{\overline{X} - 125}{\left(\dfrac{3.5}{\sqrt{10}}\right)}$.

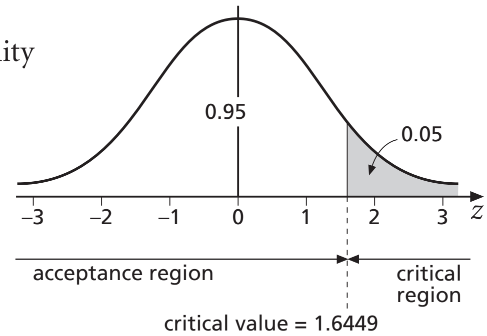
Z has the standard normal distribution.

The percentage points table of the standard normal distribution tells us that there is a 0.05, or 5%, probability that Z will be greater than 1.6449.

So if the value of Z turns out to be greater than 1.6449, we will reject H_0.

1.6449 is called the **critical value** of Z.

The region $Z > 1.6449$, which contains all the values of Z for which H_0 will be rejected, is called the **critical region** of the test.



The region $Z < 1.6449$, which contains all the values of Z for which H_0 will be accepted, is called the **acceptance region**. Both regions are shown in the diagram.

The critical value and critical region of a hypothesis test are worked out before the actual sample values are considered.

z-statistic

Suppose that a random sample of 10 modified cars gives the following set of maximum speeds:

123 128 132 125 123 128 131 130 123 126

We use lower-case letters for values obtained from a particular sample.

The value of $\overline{X}$ for this sample is denoted by $\overline{x}$.

$$\overline{x} = \frac{123 + 128 + \ldots + 126}{10} = 126.9$$

The value of Z for this sample is denoted by z, where $z = \dfrac{\overline{x} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$.

So $z = \dfrac{126.9 - 125}{\left(\dfrac{3.5}{\sqrt{10}}\right)} = 1.7167$.

We now compare the value of z with the critical value, which is 1.6449.

$z > 1.6449$, so the value of z is in the critical region. So we reject H_0 and accept H_1.

So there is evidence, at the 5% level, that the the mean maximum speed has increased.

Note

The critical region may also be expressed in terms of the non-standardised $\overline{X}$.

$Z > 1.6449$ becomes $\overline{X} > \mu + 1.6449\left(\dfrac{\sigma}{\sqrt{n}}\right)$,

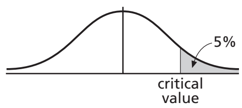
that is $\overline{X} > 125 + 1.6449\left(\dfrac{3.5}{\sqrt{10}}\right)$, or $\overline{X} > 126.82$.

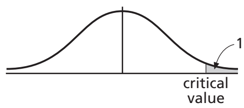
K The process of carrying out a hypothesis test is summarised below.
The car engines test is used as an example.

State the null hypothesis H_0 and the alternative hypothesis H_1.	$H_0: \mu = 125$ $H_1: \mu > 125$
Decide on the level of significance for the test.	Level of significance = 5%
Decide on the test statistic.	Test statistic is $\overline{X}$, the sample mean of a random sample of size 10. This is standardised to give Z, which has the standard normal distribution.
Calculate the critical value and critical region, using the fact that P(test statistic is in critical region) = significance level	Critical value of Z is 1.6449. Critical region is $Z > 1.6449$.
Find the value of test statistic for the random sample that has been taken.	$\bar{x} = 126.9$, leading to $z = 1.7167$
Is the value of the test statistic in the critical region? Yes → Reject H_0. Accept H_1. No → Accept H_0. Reject H_1.	1.7167 is greater than the critical value, so is in the critical region. Reject H_0. Accept H_1.
Restate the conclusion in words. Include the level of significance of the test.	There is sufficient evidence at the 5% level of significance that the modification increases the mean maximum speed.

The 5% level of significance is the one most often used in practice. Using it is equivalent to saying that if something happens whose probability, according to the null hypothesis, is only 5%, then that counts as convincing evidence against the hypothesis. Of course, the null hypothesis may still be true, in which case the probability of rejecting it wrongly is 5%.

If the consequences of making a wrong decision are very serious, then the level of significance may be made more stringent, say 1%. In this case the evidence needed to reject the null hypothesis has to be more convincing. A more extreme value of the test statistic would be required.





B1 The weights of tomatoes grown in a market garden are normally distributed with mean 55.2 g and standard deviation 3.8 g.

A new kind of plant treatment is claimed to increase the weights of the tomatoes. In order to test this claim, some of the plants are treated and then a random sample of 50 of the tomatoes from these plants is selected.

The mean weight of the sample is 56.3 g.

Assume that the standard deviation of the weights is unchanged.

(a) The null hypothesis for the test is $H_0: \mu = 55.2$. State the alternative hypothesis.

If H_0 is true, then the sample mean $\overline{X}$ is normally distributed with mean 55.2.

(b) Find the standard deviation (standard error) of $\overline{X}$.

The level of significance for the test is to be 5%.

(c) Find the appropriate percentage point of the standard normal distribution. This is the critical value of Z.

(d) Find the value of z for the sample.

(e) Compare the value of z with the critical value and state the test conclusion.

B2 A factory produces batteries whose lifetimes are normally distributed with mean 94 hours and standard deviation 3.3 hours.

A modification to the manufacturing process is carried out which is intended to increase the mean lifetime. It has no effect on the standard deviation.

In order to test whether the modification has been effective, a random sample of 20 batteries is selected. The mean lifetime of this sample is 95.5 hours.

The level of significance is to be **1%**.

(a) State the null and alternative hypotheses.

(b) Find the standard error of $\overline{X}$.

(c) Find the appropriate percentage point of the standard normal distribution for the 1% level of significance.

(d) Carry out the test using the z-statistic and state your conclusion.

B3 Redo question B2 but at the 5% level of significance. Do you come to a different conclusion? If so, explain why the conclusions are different.

B4 A firm makes elastic bands. The maximum length to which a band can be stretched before breaking is normally distributed with mean 235 mm and standard deviation 48 mm. A researcher claims that the mean maximum length is increased when a different material is used. It is planned to test this claim at the 5% level of significance by taking a random sample of 15 bands and using $\overline{X}$ as the test statistic.

(a) Write down the critical region of this test in the form $Z > \ldots$, where Z is the standardised variable corresponding to $\overline{X}$.

(b) Find the standard error of $\overline{X}$ and hence express the critical region in the form $\overline{X} > \ldots$

One-tailed and two-tailed tests

In the previous examples, the purpose of the modification or treatment was to increase the maximum speed, weight or lifetime. So the purpose of the test was to see if there was evidence that μ had increased.

Suppose instead that car engines are modified in a way that might lead to a higher or a lower maximum speed. So the purpose of the test is to see whether μ has changed – either increased or decreased.

If the null hypothesis is $H_0: \mu = 125$, then the alternative hypothesis is $H_1: \mu \neq 125$.

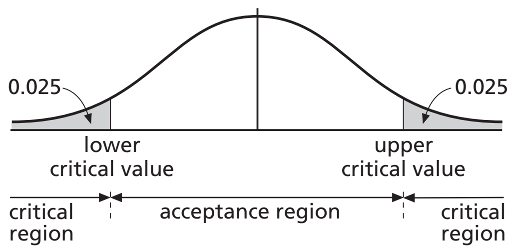
In this case either very low or very high values of the test statistic will indicate that μ has changed as a result of the modification.

There are two critical values, an upper value and a lower value.

The critical region consists of two parts, a high value part and a low value part.

The acceptance region lies between the two critical values.

If the level of significance is to be 5%, the probability of 0.05 is split equally between the two parts, as shown here.



The test in this case is called a **two-tailed test**. (The 'tails' are the parts of the distribution beyond each critical value.)

The test used in the previous examples is a **one-tailed test**.

K

Use a one-tailed test when H_1 is of the form $\mu > \ldots$ or of the form $\mu < \ldots$

Use a two-tailed test when H_1 is of the form $\mu \neq \ldots$

B5 Ten years ago it was found that the weights in grams of animals of a certain species were normally distributed with mean 84.0 and standard deviation 3.2.

Since that time climatic conditions have changed and this may have affected the mean weight. It is assumed that the standard deviation has not changed.

In order to test whether the mean weight has changed, a random sample of 20 animals will be selected and the mean weight of the sample will be used as test statistic. The level of significance will be 5%.

(a) The null hypothesis is $H_0: \mu = 84.0$. What is the alternative hypothesis?

(b) Let $\overline{X}$ be the mean weight of a random sample of 20 animals. If the null hypothesis is true, $\overline{X}$ is normally distributed with mean 84.0. What is its standard error?

(c) If Z is the standardised variable corresponding to $\overline{X}$, then the upper critical value of Z is such that 2.5% of the distribution is above it.
Use the percentage points table to find the upper critical value.

(d) Find the lower critical value and state the acceptance region.

(e) When the random sample is taken, the sample mean $\bar{x}$ is 82.4.
Find the value of z for this sample.

(f) What conclusion do you draw?

K

To test a hypothesis about the mean of a normal distribution with known variance, the standard normal distribution is used to find the critical region.

If the value of z for the sample, where $z = \dfrac{\bar{x} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$, is in the critical region, H_0 is rejected.

Example 1

The amount of liquid that a filling machine pours into a bottle is normally distributed with mean 155 ml and standard deviation 2.5 ml.

After a breakdown and repair of the machine, it is suspected that the mean amount put into a bottle may have changed.

A random sample of 50 bottles yields a mean amount of 154.2 ml per bottle.

Assuming that the standard deviation of the amount is unchanged, investigate at the 1% level of significance the claim that the mean amount put into a bottle by the machine has not changed.

Solution

Let μ be the mean amount in ml after the breakdown and repair.

The two hypotheses are

$H_0 : \mu = 155$

$H_1 : \mu \neq 155$

Let $\bar{X}$ be the mean amount for a sample of size 50.

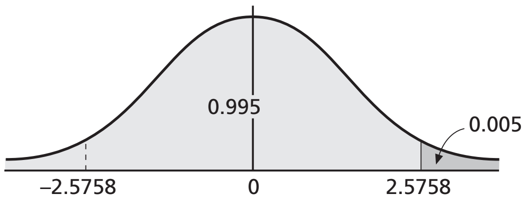
If H_0 is true, then $\bar{X}$ is normally distributed with mean 155 and standard deviation $\dfrac{2.5}{\sqrt{50}} = 0.3536$ (to 4 d.p.).

So Z, which is $\dfrac{\bar{X} - 155}{0.3536}$, has the standard normal distribution.

Because the mean amount may have increased or decreased, a two-tailed test is needed.

Each tail will contain 0.5% of the distribution, or probability 0.005.

The diagram on the right shows that we need the percentage point with 0.995 below it.



From the percentage points table, this is 2.5758.

So the upper critical value is 2.5758 and the lower critical value is –2.5758.

For the sample, $z = \dfrac{\bar{x} - 155}{0.3536} = \dfrac{154.2 - 155}{0.3536} = -2.262$ (to 3 d.p.)

This is not in the critical region.
So there is insufficient evidence, at the 1% level, to reject H_0.

Conclusion: Accept, at the 1% level of significance, that the mean amount poured by the machine has not changed.

Exercise B (answers p 128)

1 The heights in cm of five-year-old pine trees are normally distributed with mean 125.6 and standard deviation 12.2. A supplier claims to sell seedlings that grow to be taller than average. A gardener buys 12 seedlings from the supplier and grows them for five years. Their heights are then

131 145 128 118 136 152 163 124 171 145 138 137

Test the supplier's claim at the 5% significance level.

2 A machine is set up to make bolts whose diameter is normally distributed with mean 5 mm and standard deviation 0.4 mm. After repairs to the machine, a random sample of 10 bolts is taken and these are found to have a mean diameter of 5.28 mm.

Test at the 5% significance level whether the mean diameter of the bolts made by the machine has changed.

3 A rope manufacturer claims that its ropes have a breaking strain that is normally distributed with mean 50 newtons and standard deviation 1.2 newtons. A random sample of rope sections is taken and the breaking strain of each is measured, with the following results:

52.0 50.5 49.4 49.3 49.7 48.7 52.8 51.1 51.3 50.7 51.2

Test at the 2% level of significance the manufacturer's claim that the mean breaking strain of the rope is 50 newtons. Assume that the standard deviation is 1.2.

4 The wavelength in microns of a radioactive source A is known to be normally distributed with mean 1.622 and standard deviation 0.004. Nine random measurements of an unidentified radioactive sample are taken and give wavelengths of

1.612 1.622 1.613 1.621 1.609 1.618 1.624 1.625 1.622

Test at the 10% significance level whether this sample could be from source A.

5 The lightbulbs produced by a company have lifetimes that are normally distributed with mean 1250 hours and standard deviation 62 hours. A modification to the manufacturing process is claimed to increase the mean lifetime. To test this claim the lifetimes of a random sample of 15 of the modified bulbs are measured. The mean lifetime of the sample is 1274 hours. Investigate the claim.

6 A random variable X is normally distributed with mean μ and standard deviation 3.6. The null hypothesis $H_0: \mu = 30$ is to be tested against the alternative hypothesis $H_1: \mu \neq 30$ using the 5% level of significance.

The mean $\overline{X}$ of a random sample of 20 observations is to be used as the test statistic.

(a) Write down the distribution of $\overline{X}$ assuming H_0 is true.

(b) Find the acceptance region for $\overline{X}$, giving its limits to two decimal places.

(c) When the sample is taken, the mean value of the sample is found to be 36.2. What conclusion do you draw?

C Mean of a normal distribution with unknown variance

(answers p 129)

Suppose that the weight of potatoes grown in ordinary soil is normally distributed with mean 48 g but with unknown variance.

A horticulturalist develops a new kind of soil and claims that it increases the mean weight of potatoes grown in it.

Let μ be the mean weight in grams of potatoes grown in the new soil.

The null hypothesis is that $\mu = 48$ and the alternative hypothesis is that $\mu > 48$. So a one-tailed test is appropriate.

The test will be based on a random sample of 20 potatoes grown in the new soil.

However, we cannot use the test statistic $\overline{X}$ or Z in the same way as before, because we do not know the variance.

We have met this situation when finding confidence intervals. We use the sample itself to give an estimate of the population variance.

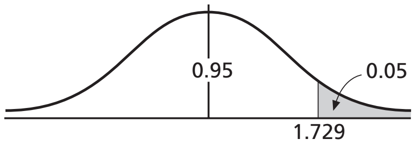
The random variable $S^2 = \dfrac{\sum X_i^2 - n\overline{X}^2}{n-1}$ is an unbiased estimator of σ^2.

The formula for Z, which we use when the variance is known, is $Z = \dfrac{\overline{X} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$.

If we replace σ by S, we get $\dfrac{\overline{X} - \mu}{\left(\dfrac{S}{\sqrt{n}}\right)}$, which has the t-distribution with $(n - 1)$ degrees of freedom.

We use the t-distribution instead of the normal distribution to find the critical value and the critical region.

If the level of significance is to be 5%, we need the corresponding percentage point of the t-distribution with 19 degrees of freedom.



From the table of percentage points, this is 1.729.

So the critical value is 1.729.

From the random sample we need to find both $\bar{x}$ and s^2.

s^2 is the particular value of S^2 for the sample, and is given by $s^2 = \dfrac{\sum x_i^2 - n\bar{x}^2}{n-1}$.

Instead of the z-statistic $\dfrac{\bar{x} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$ we use the ***t*-statistic** $t = \dfrac{\bar{x} - \mu}{\left(\dfrac{s}{\sqrt{n}}\right)}$.

Suppose a random sample of size 20 gives $\bar{x} = 48.92$ and $s^2 = 5.12$.

For this sample, $t = \dfrac{\bar{x} - \mu}{\left(\dfrac{s}{\sqrt{n}}\right)} = \dfrac{48.92 - 48}{\left(\dfrac{\sqrt{5.12}}{\sqrt{20}}\right)} = 1.8183$. This is greater than 1.729, so reject H_0.

C1 A machine puts vegetables into sacks that are meant to have a mean weight of 10 kg.

The supervisor suspects that the machine is not working properly and that the mean weight μ kg of the sacks has increased.

It may be assumed that the weights are normally distributed.

(a) State the null hypothesis and the alternative hypothesis.

(b) Is the appropriate test a one-tailed test or a two-tailed test?

The supervisor weighs a random sample of 20 sacks. Their weights are

10.5	10.8	9.8	9.7	10.5	9.9	9.8	10.1	9.9	10.1
9.8	9.8	10.8	10.0	10.3	9.8	10.2	10.1	10.2	9.7

(c) The supervisor wishes to test at the 5% level. Use the percentage points table for the t-distribution to find the critical value(s).

(d) Find the value of $\bar{x}$ for this sample.

(e) Find the value of s^2 for this sample.

(f) Find the value of t for this sample.

(g) State the conclusion of the test.

C2 A student reads that the mean weight of babies born in her local hospital ten years ago was 3.45 kg. She decides to investigate whether there has been any change in the mean weight since then.

She collects records of births and draws from them a random sample of 25 birth weights. For this sample, $\sum x_i = 88$ and $\sum x_i^2 = 310.4$.

(a) State the null and alternative hypotheses.

(b) Is the appropriate test a one-tailed test or a two-tailed test?

Assume that babies' birth weights are normally distributed and that the level of significance is 5%.

(c) Find the critical value(s) using the t-distribution.

(d) For the student's sample find

(i) the value of $\bar{x}$

(ii) the value of s^2

(iii) the value of t

(e) State the conclusion of the test.

K

To test a hypothesis about the mean of a normal distribution with unknown variance, the t-distribution is used to find the critical region.

The variance is estimated from the sample using the formula $s^2 = \dfrac{\sum x_i^2 - n\bar{x}^2}{n-1}$.

If the value of t for the sample, where $t = \dfrac{\bar{x} - \mu}{\left(\dfrac{s}{\sqrt{n}}\right)}$, is in the critical region, H_0 is rejected.

Example 2

15 snakes of a certain species were selected at random and their lengths measured. The lengths in cm were as follows.

135 141 132 138 136 129 142 130 131 133 136 138 135 134 135

Investigate, at the 5% level of significance, the claim that these snakes are drawn from a normally distributed population with mean 133 cm.

You are given that $\sum x_i = 2025$, $\sum x_i^2 = 273\,571$.

Solution

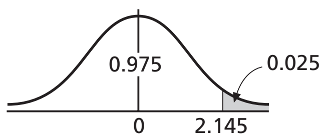
The null hypothesis is $H_0: \mu = 133$

The alternative hypothesis is $H_1: \mu \neq 133$

Because the null hypothesis would be rejected if the sample mean were either very high or very low, a two-tailed test is needed.

The variance of the population is unknown, so the *t*-distribution will be used to find the critical values and critical region.

To find the critical values we need the 97.5 percentage point of the *t*-distribution with 15 – 1 = 14 degrees of freedom.



From the tables, this is 2.145.

So the upper critical value is 2.145 and the lower –2.145.

The next step is to calculate the *t*-statistic $\dfrac{\bar{x}-\mu}{\left(\dfrac{s}{\sqrt{n}}\right)}$ for the given sample.

$$\bar{x} = \frac{2025}{15} = 135 \qquad s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = \frac{273571 - 15\times 135^2}{15-1} = \frac{196}{14} = 14$$

So $t = \dfrac{135-133}{\left(\dfrac{\sqrt{14}}{\sqrt{15}}\right)} = 2.0702$

s may also be found by entering the data 135, 141, … *into a calculator and using the* s_{n-1} *key.*

This value is not in the critical region, so H_0 is accepted.

Conclusion: At the 5% level of significance, the evidence shows that the sample could have been drawn from a population whose mean is 133 cm.

Exercise C (answers p 129)

1 Suppose the test in example 2 above is to be at the 10% level of significance.

(a) Find the new upper and lower critical values and state the acceptance region of the test.

(b) State the conclusion of the test.

2 The weight in kilograms of a certain breed of dog is normally distributed with mean 16.2. A breeder claims that he has developed this breed to produce dogs that are on average heavier.

To test this claim the editor of a dog fanciers' magazine arranges for a random sample of ten of the breeder's dogs to be weighed. The level of significance is to be 5%.

(a) State the null and alternative hypotheses for the mean weight μ.

(b) Is a one-tailed test or a two-tailed test needed in this case?

(c) Use the t-distribution with the appropriate number of degrees of freedom to find the critical value(s) and state the critical region.

The weights of the dogs in the sample are

15.9 16.8 17.5 18.3 19.2 17.5 15.8 18.9 20.6 21.5

(d) Find the values of $\bar{x}$ and s^2 for this sample and hence calculate the value of the t-statistic.

(e) State the conclusion of the test.

3 Measurements of a water supply over a long period have shown that the mean concentration of a chemical, measured in parts per million, is 16.8. The concentration can be assumed to be normally distributed.

After a storm it is decided to test the water to see whether the mean concentration μ of the chemical has changed. Twenty samples of the water are taken at random and the concentration x measured. For this sample, $\sum x = 307.0$ and $\sum x^2 = 4811.25$.

(a) State appropriate null and alternative hypotheses for μ.

(b) Determine, at the 1% level of significance, whether there is evidence that the mean concentration of the chemical changed after the storm.

4 The 'lifetime' of a car tyre is measured by the number of miles it travels before needing replacement. The mean lifetime of tyres made by company A is 24 150 miles. Company B claims that its tyres are better and publishes the lifetimes of what it describes as 'ten typical tyres' made by the company:

25 100 24 250 23 700 24 350 24 350 24 110 24 320 24 190 24 300 24 180

(a) State all the assumptions you need to make in order to carry out a hypothesis test using the data provided by company B.

(b) Making these assumptions, state the null and alternative hypotheses.

(c) Carry out the test at the 5% level of significance.

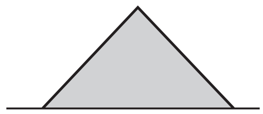
5 The weight of a certain species of rodent is known to be normally distributed with mean 0.675 kg. It is suspected that the rodents living on an island are heavier than average. The weights, w kg, of a random sample of 30 of the island rodents were measured. For this sample, $\sum w = 20.7$ and $\sum w^2 = 14.335$.

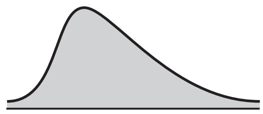
Test at the 5% level of significance whether the island rodents are heavier than average.

D Using a normal approximation

Up to now we have assumed that the random sample used in a hypothesis test has been drawn from a normally distributed population. This has justified using the z-statistic when the variance is known, and the t-statistic when the variance is unknown.

The central limit theorem states that if the sample size n is large enough, then the sample mean $\overline{X}$ is approximately normally distributed even if the sample is drawn from a non-normal population.





The size of sample needed to get a good approximation depends on the shape of the population distribution. If it is roughly similar to a normal distribution, with a peak at the centre, then even with fairly small samples the approximation will be good. If it is very asymmetric, larger samples are needed for a good approximation.

If nothing is known about the shape of the distribution, then the normal approximation can be used when $n \geq 30$.

As before, the z-statistic is used when the variance is known and the t-statistic when the variance is unknown.

K If the distribution from which a random sample is taken is unknown, then provided the sample is sufficiently large ($n \geq 30$) the z-statistic (variance known) or the t-statistic (variance unknown) can be used to test a hypothesis about the population mean.

Example 3

An airport claims that the mean time for its security checks is 20 minutes. A researcher thinks it is longer than this and records the queuing times of 35 randomly selected passengers.

The mean $\bar{x}$ of the sample is 22.4 and the estimate s^2 of the population variance is 41.4.

Test at the 5% significance level whether the researcher's claim is justified.

Solution

The two hypotheses are $H_0: \mu = 20$
$H_1: \mu > 20$

A one-sided test is appropriate.
The sample is large enough ($n = 35$) to assume that $\overline{X}$ is normally distributed.
The variance is unknown, so the t-distribution must be used, with $35 - 1 = 34$ d.f.

From tables, the percentage point for 95% below/5% above is 1.691. This is the critical value.

For the given sample, $t = \dfrac{22.4 - 20}{\left(\dfrac{\sqrt{41.4}}{\sqrt{35}}\right)} = 2.207$. This is greater than the critical value, so reject H_0.

So the researcher's claim is justified at the 5% level of significance.

Exercise D (answers p 130)

1 A brand of matches has on its boxes 'Average contents 48 matches'. A student suspects that the mean number in a box is actually less than 48. He buys 30 boxes from different sources and counts the number of matches in each box. The results are

47	51	45	46	48	47	43	50	49	45	47	48	47	46	51
46	44	47	46	49	44	48	48	47	49	50	49	48	47	46

(a) Test at the 5% level of significance the student's suspicion.

(b) Explain why your method is appropriate, even though the distribution of the number of matches in a box is unknown.

2 A local historian reads that the mean age at death of inhabitants of his area ten years ago was 64.8 years. In order to find out whether this has changed, he selects a random sample of 40 recent deaths and notes the age at death, x years, for each one. For the sample, $\sum x = 2608$ and $\sum x^2 = 170\,120$.

(a) Test at the 5% level of significance whether the mean age of death has changed.

(b) Repeat the test, but at the 10% level of significance.

3 A student has a theory that the mean number of words in a sentence can be used to help identify the writer of a story. She analyses the sentences in a book by Alf Abita and finds that the mean sentence length is 14.8 words.

She has another book, written under a pen name, which she suspects is by the same writer. She selects 35 sentences at random and counts the number, x, of words in each sentence. She finds that $\sum x = 421$ and $\sum x^2 = 7206$.

Investigate at the 5% level of significance whether the mean number of words per sentence in the second book could be 14.8.

4 The loaves made by a bread factory are labelled as weighing 0.8 kg. However, to prevent accusations that loaves are underweight, the mean weight of loaves made by the factory is actually 0.82 kg.

A consignment of these loaves is sent to a supermarket. As a quality control exercise, 40 loaves are selected at random from the consignment and the weight, x kg, of each loaf is recorded. It is found that $\sum x = 32.32$ and $\sum x^2 = 26.1787$.

Investigate at the 5% level of significance whether the mean weight of the loaves in the consignment is different from 0.82 kg.

5 A food company claims that the mean meat content of its pies is 270 g and the standard deviation 15 g. A laboratory analyses the meat content of a random sample of 50 pies and finds that the mean meat content is 265.5 g.

(a) Assuming that the standard deviation is 15 g, test at the 2% level of significance whether there is evidence that the mean meat content is different from what the company states.

(b) Repeat the test, but at the 5% level of significance.

(c) How would you report the outcomes of the two tests in a way that would make sense to a person who was not trained in statistics?

E Type I and Type II errors

Imagine that two species of animal, A and B, are very similar in appearance. The weights in grams of species A are normally distributed about a mean of 30 with a variance of 40. The weights of species B are normally distributed about a mean of 36 with a variance of 90.

A zoologist has found a colony of animals but does not know whether they are A or B. In her previous experience in the area, animals have all been of species A. So her null hypothesis is that the new colony is also A. Her two hypotheses are

$H_0: \mu = 30, \sigma^2 = 40 \qquad H_1: \mu = 36, \sigma^2 = 90$

The zoologist plans to take a sample of size 10; the significance level will be 5%.

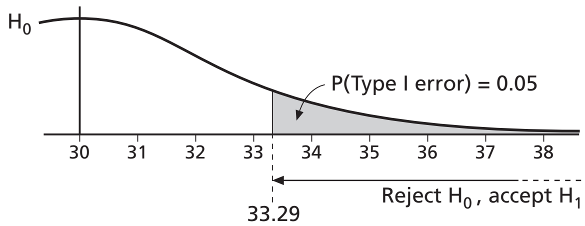
If H_0 is true, then the sample mean, $\overline{X}$, of a sample of size 10 will be normally distributed with mean 30, variance $\frac{40}{10} = 4$ and standard error $\sqrt{4} = 2$.

The standardised variable corresponding to $\overline{X}$ is $Z = \dfrac{\overline{X} - 30}{2}$.

The critical value of Z for the 5% level of significance is 1.6449.

The corresponding value of $\overline{X}$ is $30 + 1.6449 \times 2 = 33.29$ (to 2 d.p.).

The diagram shows the distribution of $\overline{X}$ if H_0 is true.



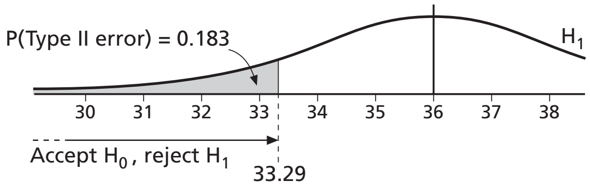
If her sample mean $\bar{x}$ is greater than 33.29, the zoologist will reject H_0, because a value as large as this is unlikely to happen if H_0 is true.

But 'unlikely' does not mean impossible, and the zoologist might reject H_0 when it is true. This error – rejecting H_0 when it is true – is called a **Type I error**.

The probability of making this error if H_0 is true is 0.05. This is called the **size** of the Type I error. It is equal to the significance level.

If, however, H_1 is true, then $\overline{X}$ will be normally distributed with mean 36, variance $\frac{90}{10} = 9$ and standard error 3.

This diagram shows the distribution of $\overline{X}$ if H_1 is true.



If the value of $\bar{x}$ from her sample is less than 33.29, the zoologist will accept H_0 and reject H_1. But H_1 might still be true. Rejecting H_1 when it is true is a **Type II error**.

The size of this error is shown by the shading. It can be found from tables to be 0.183. (The size can be found only when H_1 specifies a value for μ.)

K

A Type I error is to reject H_0 when it is true. Its size is equal to the significance level.

A Type II error is to reject H_1 when it is true.

Key points

- In a hypothesis test, the null hypothesis is denoted by H_0 and the alternative hypothesis by H_1. (p 78)
- The critical region is the set of values of the test statistic for which H_0 will be rejected. The acceptance region is the set for which H_0 will be accepted.
  If the level of significance is, say, 5% then the critical region is chosen so that the probability that the test statistic falls within it if H_0 is true is 0.05. (pp 78–80)
- Use a one-tailed test when H_1 is of the form $\mu > \ldots$ or of the form $\mu < \ldots$
  Use a two-tailed test when H_1 is of the form $\mu \neq \ldots$ (p 82)
- To test a hypothesis about the mean of a normal distribution with known variance, the standard normal distribution is used to find the critical region.
  If the value of z for the sample, where $z = \dfrac{\bar{x} - \mu}{\left(\dfrac{\sigma}{\sqrt{n}}\right)}$, is in the critical region, H_0 is rejected. (p 83)
- To test a hypothesis about the mean of a normal distribution with unknown variance, the t-distribution is used to find the critical region.
  If the value of t for the sample, where $t = \dfrac{\bar{x} - \mu}{\left(\dfrac{s}{\sqrt{n}}\right)}$, is in the critical region, H_0 is rejected. (p 86)
- If the distribution from which a random sample is taken is unknown, then if $n \geq 30$ the z-statistic (population variance known) or the t-statistic (population variance unknown) can be used to test a hypothesis about the population mean. (p 89)
- A Type I error is to reject H_0 when it is true. Its size is equal to the significance level.
  A Type II error is to reject H_1 when it is true. (p 91)

Mixed questions (answers p 130)

1 The fuel economy of a car is measured in litres per 100 km, so a lower figure indicates better fuel economy.

It is known that the fuel economy of cars of a certain model is normally distributed with mean 10.7 and **variance** 3.24. A garage owner modifies the engine and claims that modified cars have better fuel economy.
A consumers' organisation measures the fuel economy of a random sample of 15 modified cars and records the mean economy of the sample as 9.9.

Investigate, at the 5% significance level, the claim that the modification improves the mean fuel economy.

2 A casualty department claims that waiting times have fallen. The mean waiting time used to be 54.5 minutes. For a random sample of 35 waiting times, x minutes, it is found that $\sum x = 1820$ and $\sum x^2 = 96\,130$. Test the department's claim at the 1% level of significance.

3 Scores on IQ tests are designed to be normally distributed with mean 100 and standard deviation 15. A polling organisation uses a method of selecting people that is claimed to be random. However, a researcher suspects that the method is biased towards people with higher IQs.

She plans to use the method to choose a sample of 20. Her null hypothesis will be that the sample is a random sample from a population with mean 100 and standard deviation 15. She will use the sample mean $\overline{X}$ as test statistic.

(a) State the distribution of $\overline{X}$.

(b) Will a one-tailed test or a two-tailed test be appropriate?

The researcher will test at the 2% level of significance.

(c) Let Z be the standardised variable corresponding to $\overline{X}$. Find the critical value of Z.

(d) Hence find the critical value of $\overline{X}$ and state the acceptance region in terms of $\overline{X}$.

(e) When the researcher takes the sample she finds that the sample mean $\bar{x}$ is 107.9. What conclusion should she draw?

Test yourself (answers p 130)

1 The lengths of fish of a particular species are normally distributed with a mean of 56 cm and a standard deviation of 4.2 cm. There is a suspicion that, due to overfishing, the mean length of these fish has changed.

A random sample of 50 fish was measured and was found to have a mean length of 54.8 cm.

Investigate, at the 5% level of significance, whether this indicates a change from 56 cm in the mean length of this species of fish. AQA 2001

2 A random variable X is normally distributed with mean μ and standard deviation 0.8. The null hypothesis $H_0: \mu = 40$ is to be tested against the alternative hypothesis $H_1: \mu \neq 40$ using the 5% level of significance.

The mean, $\overline{X}$, of a random sample of 50 observations is to be used as the test statistic.

(a) Write down

(i) the distribution of $\overline{X}$ assuming H_0 is true

(ii) the probability of a Type I error

(b) Calculate the acceptance region for $\overline{X}$, giving its limits to two decimal places.

(c) Explain what is meant by a Type II error. AQA 2002

3 The lengths of snakes of a particular species are normally distributed with mean 74.6 cm. A zoologist suspects that the snakes of this species that live in a marshland are longer than average. He measures the lengths in cm of a random sample of 10 snakes from the marshland, with these results.

75.1 77.2 75.2 74.6 73.9 76.1 77.0 75.0 72.8 76.1

Test, at the 5% level of significance, the zoologist's suspicion.