Section 1: Using the Normal distribution

Notes and Examples

These notes contain subsections on:

The distribution of sample means

Suppose you use a random number generator to choose three numbers at random from the integers 1 – 100, and find the average of the three numbers you have chosen. There are a very large number of possible results you could obtain for the mean of your sample of three, ranging from 1 (if the numbers you obtain are all 1's) to 100 (if the numbers you obtain are all 100's). Clearly, it is quite unlikely that the mean would be 1 or 100 – it is much more likely to be fairly close to 50.

You could work out the probability distribution for the sample means, by calculating the probability of each possible value for the mean. What sort of shape would this probability distribution have, and what would be the mean and standard deviation of the distribution?

You can investigate the distribution of sample means using a simple example: throwing an ordinary, fair die. This means that you are dealing with the population {1, 2, 3, 4, 5, 6}. Throwing one die is equivalent to taking a sample of size 1 from the population; throwing two dice is equivalent to taking a sample of size 2 from the population, and so on.

Samples of size 1

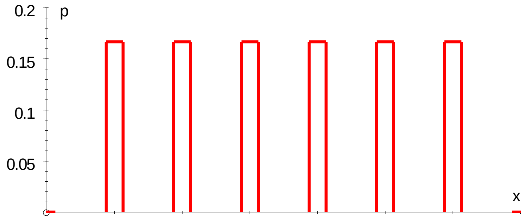
If you throw one die, then there are six possible samples you could obtain:

$$\{1\} \quad \{2\} \quad \{3\} \quad \{4\} \quad \{5\} \quad \{6\}$$

Each of these samples is equally likely to occur. The sample mean in each case is, of course, just the value of the score on the die.

So the probability distribution of the sample means for a sample of size 1 is:

$\bar{x}$	1	2	3	4	5	6
$P(\bar{X}=\bar{x})$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



It can be shown that $\mathrm{E}(X) = 3.5$ and $\mathrm{Var}(X) = \frac{35}{12}$

Samples of size 2

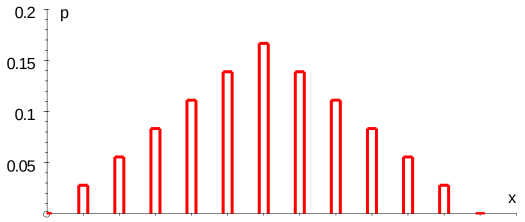
If you throw two dice, then there are 36 possible samples you could obtain (some of which are the same, e.g. {1, 2} and {2, 1}).

The table below shows the possible values of the sample mean.

	1	2	3	4	5	6
1	1	1.5	2	2.5	3	3.5
2	1.5	2	2.5	3	3.5	4
3	2	2.5	3	3.5	4	4.5
4	2.5	3	3.5	4	4.5	5
5	3	3.5	4	4.5	5	5.5
6	3.5	4	4.5	5	5.5	6

So the probability distribution of the sample means for a sample of size 2 is:

$\bar{y}$	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
$\mathrm{P}(\bar{Y} = \bar{y})$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



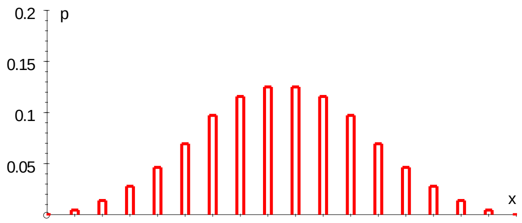
It can be shown that $\mathrm{E}(Y) = 3.5$ and $\mathrm{Var}(Y) = \frac{35}{24}$

Samples of size 3

If you throw three dice, then there are 216 possible samples you could obtain (again, some are the same, such as {1, 1, 2}, {1, 2, 1} and {2, 1, 1}).
If a complete list is made of all the possible samples, and the sample mean calculated for each, you can find the probability distribution of the sample mean in the same way as for samples of size 2.

The probability distribution of the sample means for a sample of size 3 is:

$\bar{z}$	1	$1\frac{1}{3}$	$1\frac{2}{3}$	2	$2\frac{1}{3}$	$2\frac{2}{3}$	3	$3\frac{1}{3}$	$3\frac{2}{3}$	4	$4\frac{1}{3}$	$4\frac{2}{3}$	5	$5\frac{1}{3}$	$5\frac{2}{3}$	6
$\mathrm{P}(\bar{Z} = \bar{z})$	$\frac{1}{216}$	$\frac{3}{216}$	$\frac{6}{216}$	$\frac{10}{216}$	$\frac{15}{216}$	$\frac{21}{216}$	$\frac{25}{216}$	$\frac{27}{216}$	$\frac{27}{216}$	$\frac{25}{216}$	$\frac{21}{216}$	$\frac{15}{216}$	$\frac{10}{216}$	$\frac{6}{216}$	$\frac{3}{216}$	$\frac{1}{216}$



It can be shown that $\mathrm{E}(Z) = 3.5$ and $\mathrm{Var}(Z) = \frac{35}{36}$

Comparing the distributions for samples of size 1, 2 and 3, you can see that whereas a sample of size 1 has a uniform distribution, for samples of size 2 and 3 the distribution has a peak in the centre corresponding to the mean value of 3.5.

In addition, the distribution for sample size 2 is triangular, whereas the one for sample size 3 is more “bell-shaped”, suggesting that the standard deviation is smaller. In fact, this trend continues with larger sample sizes.

We have used the theoretical distribution of throwing a die to model the outcomes of sampling from a very simple population (the numbers 1, 2, 3, 4, 5 and 6). The mean (3.5) and standard deviation ($\sqrt{\frac{35}{12}}$) are the same as the population mean, μ, and standard deviation, σ (the population standard deviation is calculated using divisor n, since we are dealing with a complete population). All three probability distributions have mean 3.5, which is the same as the population mean μ.

The standard deviation of the distribution for sample size 2 is $\sqrt{\frac{35}{24}}$, which can be written as $\frac{\sigma}{\sqrt{2}}$. The standard deviation of the distribution for sample size 3 is $\sqrt{\frac{35}{36}}$, which can be written as $\frac{\sigma}{\sqrt{3}}$.

Generalising: given a population with a mean of μ and a standard deviation of σ, the sampling distribution of the mean has a mean of μ and a standard deviation of $\frac{\sigma}{\sqrt{n}}$, where n is the sample size.

Notice that the standard deviation of the distribution of sample means (sometimes called the standard error of the mean) is smaller than the population standard deviation and decreases as the sample size increases.

As the distribution of the sample means is so important, it is often abbreviated to just the sampling distribution. However, this does not mean other sampling distributions are not possible: the sampling distribution of the median is possible of course.

In this topic we are assuming that the underlying distribution has a Normal distribution.
Given a population X with a mean of μ and a standard deviation of σ

i.e. $X \sim \mathrm{N}(\mu, \sigma^2)$, and a sample of size n is taken, the distribution of the sample means is given by $\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

You can therefore use the skills learnt when working with the Normal distribution to calculate probabilities with a sample mean.

Note you can become confused between the theoretical distribution and a practical experiment. If you are conducting a biology experiment you will normally be collecting one sample of data. When analysing the results you are using the theory from the theoretical distribution.

Standardising the distribution of the sample means

As you saw in the work on the normal distribution, any normal distribution X ~ N(μ, σ) can be transformed to the standard normal distribution Z ~ N(0, 1). The variable X has mean μ and standard deviation σ so x, a particular value of X, is transformed into z by the formula:

$$z = \frac{x - \mu}{\sigma}$$

So for the distribution of the sample means, $\bar{X}$, you can standardise by using

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}.$$

Hypothesis tests

You have already met hypothesis tests involving the binomial distribution $\mathrm{B}(n, p)$, in which you investigate whether a hypothesised value for the population parameter p takes a particular value.

You will now look at hypothesis tests using the Normal distribution N(μ, σ), in which you test whether the population mean takes a particular value.

In the test, you are assuming that the value of the population mean is the one given in the null hypothesis, and then considering the value of the sample mean. If your sample mean is too far away from the assumed population mean, then you conclude that as it is very unlikely that a randomly chosen sample would have such a high (or low) sample mean, the population mean does not in fact have the value that you assumed it to have. This means that you are rejecting the null hypothesis.

There are two main approaches that can be used in the hypothesis test. They are equivalent but you should know both.

Suppose that you are using the hypotheses

H_0: $\mu = m$
H_1: $\mu > m$

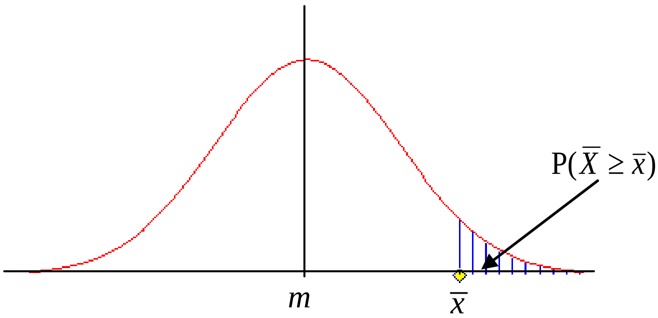
where μ is the true population mean
and you are testing at the 5% level.

Method 1: Using a p-value
You need to look at the probability a sample of size n taken from a distribution with mean m and standard deviation σ, has a value at least as extreme (in this case, at least as large) as $\bar{x}$, the mean of the given sample. If this probability is less than the significance level you will reject H_0. In such a case you are saying that it is so unlikely that a sample from a distribution with mean m would give this value for $\bar{x}$, that you conclude that in fact the distribution does not have mean m, but a larger mean.

The distribution of the sample means is $\mathrm{N}\left(m, \dfrac{\sigma^2}{n}\right)$.

You use your calculator to find $\mathrm{P}(\bar{X} \geq \bar{x})$ (the p-value)

You reject H_0 if $\mathrm{P}(\bar{X} \geq \bar{x}) < 0.05$.



The diagram shows a Normal distribution with mean m and standard deviation σ. If the area shown is less than the significance level, we reject H_0.

Example 1
Test results are normally distributed with a mean of 65 and a standard deviation of 10. After the introduction of a dynamic new teacher the results for a group of 8 students had a mean of 72. Is there evidence that the results have significantly improved at a 5% level of significance?

Solution
$H_0 : \mu = 65$
$H_1 : \mu > 65$
where μ is the population mean test score.

You want to see if the results could have come from a distribution where the population mean has remained unchanged.

Remember to define μ as the population mean – there is often a mark awarded for this.

Let X be the distribution of test scores.
$X \sim \mathrm{N}(65, 10^2)$
$\bar{X} \sim \mathrm{N}\left(65, \dfrac{10^2}{8}\right)$
$\mathrm{P}(\bar{X} > 72) = 0.02$

This is the p-value – the probability that (if the mean is 65) the sample mean is more than 72

Since 0.0239 < 0.05 (the required significance level of 5%) the null hypothesis is rejected. There is evidence to suggest that the mean score has increased, i.e. the teacher has had some effect.

Edexcel Maths Hypothesis testing 1 Notes & examples

Method 2: Finding a critical region

In this method, you find the range of values of the sample mean for which the null hypothesis would be rejected. This is the critical region. You can then simply look to see if the sample mean lies in the critical region.

You can use your calculator to find the critical value (the boundary of the critical region). For a null hypothesis of the form H_1: $\mu > m$, you are looking at the right-hand tail, so for a 5% significance level you need the inverse normal value for 0.95 for $N\left(m, \frac{\sigma^2}{n}\right)$.

Example 2

Test results are normally distributed with a mean of 65 and a standard deviation of 10. After the introduction of a dynamic new teacher the results for a group of 8 students had a mean of 72. Is there evidence that the results have significantly improved at a 5% level of significance?

Solution

$H_0 : \mu = 65$

$H_1 : \mu > 65$

where μ is the population mean test score.

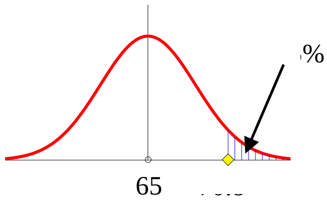
Let X be the distribution of test scores.

$X \sim N(65, 10^2)$

$\bar{X} \sim N\left(65, \frac{10^2}{8}\right)$

Using a calculator, inverse normal of 0.95 is 70.8

The critical region is $\bar{X} > 70.8$.



The sample mean $\bar{x} = 72$ lies in the critical region, so reject H_0.

There is evidence to suggest that the mean score has increased, i.e. the teacher has had some effect.

Notice from the example above that the conclusion should always be given in terms of the problem. First state whether H_0 is to be accepted or rejected, then make a statement beginning “there is evidence to suggest that …” or “there is not sufficient evidence to suggest that …”. You should **NOT** write “this proves that ….” or “so the claim is right”. You are not proving anything, only considering evidence.

The left-hand tail

In the examples above, you were looking at the right-hand tail of the distribution, since the alternative hypothesis suggested that the mean might have increased. If the alternative hypothesis suggests a possible decrease in the mean, then you will be looking at the left-hand tail of the distribution. This means that the critical region will be on the left-hand side of the distribution and so at a significance level of 5% you need to use the inverse normal of 0.05. If using the p-value, you will need to find P($\bar{X} < \bar{x}$).

Example 3
The supplier of LITE light bulbs claims that the mean life of a LITE light bulb is 130 hours. Responding to customer complaints that the light bulbs did not last as long as expected, a training standards organisation tested 400 bulbs and found the mean to be 128.5 hours. The standard deviation may be assumed to be 13 hours.
Is there evidence at a 2% level that the mean is lower than 130 hours?

Solution
$H_0 : \mu = 130$
$H_1 : \mu < 130$.
where μ is the population mean lifetime.

Let X be the distribution of times of LITE light bulbs.
$X \sim N(130, 13^2)$

$$\bar{X} \sim N\left(130, \frac{13^2}{400}\right)$$

Method 1: Using the p-value
P($\bar{X} < 128.5$) = 0.0105
Since 0.0105 < 0.02 (the required significance level of 2%) the null hypothesis is rejected. There is evidence to suggest that that the lifespan of a LITE light bulb is less than 130 hours.

Method 2: Using critical regions
The inverse normal of 0.02 for $N\left(130, \frac{13^2}{400}\right)$ is 128.67

The critical region is $\bar{X} < 128.67$ ’
Since the sample mean of 128.5 is in the critical region, the null hypothesis is rejected. There is evidence to suggest that the lifespan of a LITE light bulb is less than 130 hours.

Two tailed tests

In the examples so far, the alternative hypothesis has been of the form $\mu > k$ (in which case you are looking at the right-hand tail) or $\mu < k$ (in which case you are looking at the left-hand tail). These are all one-tailed tests.

However, sometimes you will need to look at situations where the alternative hypothesis is of the form $\mu \neq k$ (in which you are testing whether the mean is as stated or not, without specifying in which direction it is likely to be wrong. A test like this is a two-tailed test, as you are looking at both tails of the distribution.

In a two-tailed test, there are two parts to the critical region. If you are asked to give the critical region for a test, you must give both parts. However, if you are just asked to carry out the hypothesis test, you need only look at the relevant tail, depending on whether the sample mean is higher or lower that the value given in the null hypothesis. At the 5% significance level, you find the lower tail critical region using the inverse normal of 0.025, and the upper tail critical region using the inverse normal of 0.0975, so that the two tails correspond to a total probability of 5%.

Similarly, if you are using p-values, you compare with half the significance level, since you are looking at just the relevant tail.

Example 4

The lengths of the leaves of a certain species of rare plant are Normally distributed with mean 8.6 cm and standard deviation 1.2 cm. A botanist finds a clump of plants and wants to find out whether they are of the rare species. She collects and measures 50 leaves and finds that the total of their lengths is 442 cm. Carry out a test at the 5% level. What should the biologist conclude?

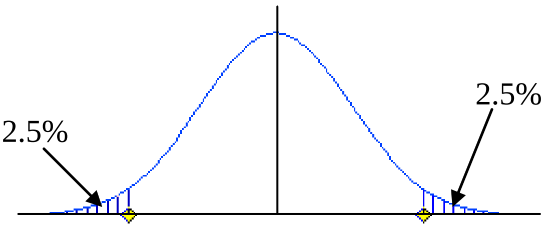
Solution

This is a two-tailed test, as the alternative hypothesis is that the mean is not 8.6, rather than being specifically more or less than 8.6.

In this test, we are looking for evidence that the plants are **not** of the rare species

H_0: $\mu = 8.6$
H_1: $\mu \neq 8.6$
where μ is the population mean leaf length.

$$\bar{x} = \frac{442}{50} = 8.84$$



Let X be the distribution of the lengths of the leaves.

$$\bar{X} \sim \mathrm{N}\left(8.6, \frac{1.2^2}{50}\right)$$

As the sample mean is greater than 8.6, we are looking at the right-hand tail.

Method 1: Using a p-value

P($\bar{X}$ > 8.84) = 0.0787

Since 0.0787 > 0.025 (the required significance level of 2.5% in each tail) the null hypothesis is accepted. There is not sufficient evidence to suggest that the plants are **not** of the rare species.

This is not the same as evidence that they are of the rare species

Method 2: Using critical regions

The critical value for the upper tail is found using the inverse normal of 0.0975.

For $\mathrm{N}\left(8.6, \frac{1.2^2}{50}\right)$ this is 8.93

The critical region is $\bar{X} > 8.93$

Since the sample mean of 8.84 is not in the critical region, the null hypothesis is accepted. There is not sufficient evidence to suggest that the plants are **not** of the rare species.