Section 2: More about hypothesis testing

Notes and Examples

These notes contain subsections on
- Using critical regions
- Examples using larger samples
- Examination style question
- Two-tailed tests

Using critical regions

Using a critical region is an alternative approach to carrying out a hypothesis test.

A hypothesis test using critical regions is carried out in much the same way as the hypothesis tests in section 1.

- Establish the null and alternative hypotheses
- Decide on the significance level
- Find the critical region
- Collect suitable data using a random sampling procedure that ensures the items are independent, and look at whether the result lies in the critical region.
- Interpret the results in terms of the original claim.

In the previous section, you calculated the p-value for the collected data. What you were doing is asking
- What is the probability of getting this result or a more extreme one?
- Is this probability greater or less than the significance level?

When you use a critical region, what you are doing is asking
- What set of values would give a probability less than the significance level (i.e. would cause the null hypothesis to be rejected)?
- Is the result from the collected data in this set of values (the critical region)?

Using critical regions is efficient if you are going to carry out the same hypothesis test several times with different samples, as you can work out the critical region just once.

You can use the technique of critical values and critical regions to investigate some of the situations studied in the Notes and Examples in section 1.

Firstly, look at the situations from examples 3 and 4 in section 1.

Edexcel AS Maths Hyp testing 2 Notes and Examples

Example 1
A student takes a multiple-choice test. There are 10 questions with 4 possible answers. Unfortunately the student has attended very few lessons so has to guess. The student gets 5 questions correct.
The student claims to be an inspired guesser. Is this true?

Solution

Establish the null and alternative hypotheses

Let p be the probability the student guesses correctly
H_0: $p = \frac{1}{4}$ (equal chances of choosing each response)
H_1: $p > \frac{1}{4}$ (guesses correct in more than quarter of the questions)

Significance level set as 5%

Decide on the significance level

Let X be the number of correct guesses.
If H_0 is true then we are using $p = \frac{1}{4}$ and $X \sim B(10, \frac{1}{4})$

Find the critical region

Find the lowest value of a for which $P(X \geq a) < 0.05$
$\Rightarrow 1 - P(X \leq a - 1) < 0.05$
$\Rightarrow P(X \leq a - 1) > 0.95$
$P(X \leq 4) = 0.9219$
$P(X \leq 5) = 0.9803$

0.95 lies between 0.9219 and 0.9803

The lowest possible value of $a - 1$ is 5
so the lowest possible value of a is 6

So the critical value is $X = 6$
The critical region is $X \geq 6$

This is the region where H_0 is rejected

5 questions correct so $X = 5$
This does not lie in the critical region, so accept H_0

Look at the data

There is not sufficient evidence that the student is performing any better than through random guessing; he/she would be advised to attend lessons.

Interpret the results

Example 2
The probability that a student at a certain school passes an exam is 0.8.
A new teacher is appointed and takes the next group. Out of a group of 10 students only 6 students now pass.
Should the Head of Maths be concerned?

Solution

Establish the null and alternative hypotheses

Let p be the probability a student passes the exam
H_0: $p = 0.8$ (probability a student passes the exam is 0.8)
H_1: $p < 0.8$ (probability a student passes the exam has been lowered)

Significance level set as 5%

Decide on the significance level

Let X be the number of students passing.
If H_0 is true then we are using $p = 0.8$ and $X \sim B(10,0.8)$

Find the highest value of a where $P(X \leq a) < 0.05$

Find the critical region

$P(X \leq 5) = 0.0328 < 0.05$
$P(X \leq 6) = 0.1209 > 0.05$

So $a = 5$.
So the critical value is $X = 5$
The critical region is $X \leq 5$

This is the region where H_0 is rejected

6 students pass, so $X = 6$

Look at the data

This does not lie in the critical region, so accept H_0.

There is not sufficient evidence that the students are under performing with the new teacher, based on this sample of students.

Interpret the results

Now look at the coin situation (Example 1 in Section 1).
Why did we not start with working out the critical region for this situation?

Example 3
Andy and Bilal are playing a game with a coin. Andy wins if the coin shows Heads, and Bilal wins if the coin shows Tails. Bilal wins 4 consecutive times. Andy complains that the coin must be biased. Is his complaint justified?

Solution
Let p be the probability that the coin shows Heads

Establish the null and alternative hypotheses

H_0: $p = \frac{1}{2}$ (equal chances of showing Head or Tail)
H_1: $p < \frac{1}{2}$ (shows Heads in less than half of the times)

Significance level set as 5%

Decide on the significance level

Let X be the number of times Heads is shown.
If H_0 is true then we are using $p = \frac{1}{2}$ and $X \sim B(4, 0.5)$

Find the highest value of a where $P(X \leq a) < 0.05$
But $P(X = 0) = 0.0625 > 0.05$
So the value $X = 0$ does not lie in the critical region.

Find the critical region

Clearly there cannot be a critical region for this test, we will always accept H_0.

Heads is shown 0 times, so $X = 0$

Look at the data

This does not lie in the critical region, so accept H_0.

There is not sufficient evidence to show that the coin is biased.

Interpret the results

So we avoided using this example to illustrate a critical region or a critical value, as it does not have one!

So although we may be surprised with the number of questions the student got right, the number of students who passed the exam and the number of times the coin showed Tails, in examples 1, 2 and 3 respectively, the results turned out to be not significant. We have now shown this by demonstrating that they were not in the critical region for the test.

However, in all these cases the sample sizes were relatively small. What happens if we investigate results from larger samples?

Examples using larger samples

Consider these three new examples which are the same as Examples 1, 2 and 3.

Example 4
Andy and Bilal are playing a game with a coin. Andy wins if the coin shows Heads, and Bilal wins if the coin shows Tails. Bilal wins 10 consecutive times. Andy complains that the coin must be biased. Is his complaint justified?

Example 5
A student takes a multiple-choice test. There are 20 questions with 4 possible answers. Unfortunately the student has attended very few lessons so has to guess. The student gets 10 questions correct.
The student claims to be an inspired guesser. Is this true?

Example 6
The probability that a student at a certain school passes an exam is 0.8.
A new teacher is appointed and takes the next group. Out of a group of 20 students only 12 students now pass.
Should the Head of Maths be concerned?

Note the proportions of success in example 5 and 6 have stayed the same as in examples 1 and 2, i.e. 50% and 60% respectively. Do we come to the same conclusions?

Example 4
Andy and Bilal are playing a game with a coin. Andy wins if the coin shows Heads, and Bilal wins if the coin shows Tails. Bilal wins 4 consecutive times. Andy complains that the coin must be biased. Is his complaint justified?

Solution

Let p be the probability that the coin shows Heads

H_0: $p = \frac{1}{2}$ (equal chances of showing Heads or Tails)

H_1: $p < \frac{1}{2}$ (shows Heads in less than half of the times)

Establish the null and alternative hypotheses

Significance level set as 5%

Decide on the significance level

Let X be the number of times Heads is shown.

If H_0 is true then we are using $p = \frac{1}{2}$ and $X \sim B(10, 0.5)$

Find the highest value of a where $P(X \leq a) < 0.05$

Find the critical region

$P(X \leq 1) = 0.0107 < 0.05$

$P(X \leq 2) = 0.0547 > 0.05$

So $a = 1$.

So the critical value is $X = 1$

The critical region is $X \leq 1$.

This is the region where H_0 is rejected

0 calls correct, so $X = 0$

This does lie in the critical region, so reject H_0.

Look at the data

There is evidence that to suggest that the coin is biased.

Interpret the results

Example 5

A student takes a multiple-choice test. There are 20 questions with 4 possible answers. Unfortunately the student has attended very few lessons so has to guess. The student gets 10 questions correct.

The student claims to be an inspired guesser. Is this true?

Solution

Let p be the probability the student guesses correctly

H_0: $p = \frac{1}{4}$ (equal chances of choosing each response)

H_1: $p > \frac{1}{4}$ (guesses correct in more than quarter of the questions)

Establish the null and alternative hypotheses

Significance level set as 5%

Decide on the significance level

Let X be the number of correct guesses.

If H_0 is true then we are using $p = \frac{1}{4}$ and $X \sim B(20, \frac{1}{4})$

Find the lowest value of a where $P(X \geq a) < 0.05$

$\Rightarrow 1 - P(X \leq a - 1) < 0.05$

$\Rightarrow P(X \leq a - 1) > 0.95$

Find the critical region

$P(X \leq 7) = 0.8982$

$P(X \leq 8) = 0.9591$

The lowest possible value of $a - 1$ is 8
So the lowest possible value of $a = 9$

So the critical value is $X = 9$
The critical region is $X \geq 9$

This is the region where H_0 is rejected

10 questions correct so $X = 10$
This does lie in the critical region, so reject H_0.

Look at the data

There is evidence to suggest that the student is an inspired guesser.

Interpret the results

Example 6
The probability that a student at a certain school passes the S1 module is 0.8.
A new teacher is appointed and takes the next group. Out of a group of 20 students only 12 students now pass.
Should the Head of Maths be concerned?

Solution
Let p be the probability a student passes the exam
H_0: $p = 0.8$ (probability a student passes the exam is 0.8)
H_1: $p < 0.8$ (probability a student passes the exam has been lowered)

Establish the null and alternative hypotheses

Significance level set as 5%

Decide on the significance level

Let X be the number of students passing.
If H_0 is true then we are using $p = 0.8$ and $X \sim B(20, 0.8)$

Find the highest value of a where $P(X \leq a) < 0.05$

Find the critical region

$P(X \leq 12) = 0.0321 < 0.05$
$P(X \leq 13) = 0.0867 > 0.05$

So $a = 12$.
So the critical value is $X = 12$
The critical region is $X \leq 12$

This is the region where H_0 is rejected

12 students pass so $X = 12$
This does lie in the critical region, so reject H_0.

Look at the data

There is evidence to suggest that that the students are under performing with the new teacher, based on this sample of students.

Interpret the results

However, there could be several other factors: did the students also perform badly in other subjects, was it a very difficult exam paper etc?

The outcomes in examples 4, 5 and 6 are now all significant.

As sample sizes get larger, the conclusion can change even if the ratio of success stays the same.

The chance of getting 10 or more questions correct out of 20 is much smaller than the chance of getting 5 or more questions correct out of 10.

Examination style question

Example 7
Using recent data provided by the low-cost airline Brianair, the probability of a flight arriving on time is estimated to be 0.8.
After some rescheduling, Brianair state that their performance has improved. In a recent survey 19 out of 20 flights arrived on time.
Construct a critical region, using a significance level of 5%. Is Brianair's conclusion correct?

Solution
Let p be the probability a flight is on time.
H_0: $p = 0.8$ (probability a flight is on time is 0.8)
H_1: $p > 0.8$ (probability a flight is on time has increased)
Significance level set as 5%

Let X be the number of times a flight is on time.
If H_0 is true then we are using $p = 0.8$ and $X \sim B(20, 0.8)$.

Find the lowest value of a where $P(X \geq a) < 0.05$
$\Rightarrow 1 - P(X \leq a - 1) < 0.05$
$\Rightarrow P(X \leq a - 1) > 0.95$

$P(X \leq 18) = 0.9308$
$P(X \leq 19) = 0.9885$

The lowest possible value of $a - 1$ is 19
So the lowest possible value of $a = 20$

So the critical value is $X = 20$
The critical region is $X \geq 20$ which of course with $n = 20$ is just $X = 20$

19 flights on time so $X = 19$
This does not lie in the critical region, so accept H_0

There is not sufficient evidence to suggest that the flights have improved. Brianair's conclusion is not correct. They need to take a larger sample size to investigate further, especially as the data value was close to the critical value.

Two-tailed tests

The technique you have just used is an essential part of doing an asymmetrical two-tailed test. This is 'asymmetrical' because $p \neq 0.5$, so that the lower 'tail' has 5 elements; 0, 1, 2, 3 and 4, and the upper tail has 7 elements; 14, 15, 16, 17, 18, 19 and 20.

When you are testing for a **change**, without specifying the direction of the change, you are dealing with a situation that requires a two-tailed test. You are examining the probabilities at both the upper and lower ends of the distribution at the same time.

There are two methods of dealing with two-tailed tests.

- Using critical regions. In two-tailed tests, the critical region has two parts, one at each tail. If you are asked to give the critical region, you must give both tails.
- Using probabilities. As in section 1, you work out the probability of the observed result or a more extreme result. Although you will only look at one tail, you have to allow for the possibility that the observed result could have been in the other tail. So the *p*-value is the probability of this result or a more extreme result plus the probability of an equivalent result or more extreme in the other direction. You therefore need to double the probability that you have found, before comparing it with the significance level.

The examples below show how both of these methods can be used to investigate some new situations involving two-tailed tests.

Example 8

An anti-smoking campaign is held in a city. Before the campaign, health workers estimated that one-third smoked cigarettes. Some time after the end of the campaign, a survey is conducted to find out if it has had any impact on the number of smokers, positive or negative.

A random sample of 16 students is selected from Year 11. When the data is analysed it is found that the sample contains exactly 2 smokers. At the 5% significance level, is there sufficient evidence to suggest that there has been any change in the number of smokers?

Note: you are not looking at whether the number of smokers has increased or decreased, just whether it has changed.

Solution 1 (using critical regions)

Let p be the probability the student smokes.

H_0: $p = \frac{1}{3}$ (probability of being a smoker stays the same)

H_1: $p \neq \frac{1}{3}$ (probability of being a smoker changes)

2-tail test

Establish the null and alternative hypotheses

Significance level set as 5%

Decide on the significance level

Let X be the number of smokers.
If H_0 is true then we are using $p = \frac{1}{3}$ and $X \sim B(16, \frac{1}{3})$

$5\% \div 2 = 2.5\% = 0.025$

5% is 2.5% in each 'tail'.

Find the highest value of a for which $P(X \le a) < 0.025$ and the lowest value of b for which $P(X \ge b) < 0.025$

Find the critical regions

$P(X \le 1) = 0.0137 < 0.025$
$P(X \le 2) = 0.0594 > 0.025$

The highest possible value of a is 1
So the critical value is $X = 1$
The critical region is $X \le 1$

$P(X \ge b) < 0.025$
$1 - P(X \le b - 1) < 0.025$
$P(X \le b - 1) > 0.975$

0.975 lies between 0.9500 and 0.9841

$P(X \le 8) = 0.9500$
$P(X \le 9) = 0.9841$

So the smallest possible value of $b - 1$ is 9.
So the smallest possible value of b is 10.
So the critical value is $X = 10$
The critical region is $X \ge 10$

This is the region where H_0 is rejected

The critical region is $X \le 1$ and $X \ge 10$

Look at the data

There are 2 smokers, so $X = 2$
This does not lie in the critical region, so accept H_0.

Interpret the results

There is not sufficient evidence to suggest that there has been any change in the proportion of smokers.

Solution 2 (using probabilities)
Let p be the probability the student smokes.
H_0: $p = \frac{1}{3}$ (probability of being a smoker stays the same)
H_1: $p \ne \frac{1}{3}$ (probability of being a smoker changes)
2-tail test

Establish the null and alternative hypotheses

Significance level set as 5%

Decide on the significance level

Let X be the number of smokers.
If H_0 is true then we are using $p = \frac{1}{3}$ and $X \sim B(16, \frac{1}{3})$

$X = 2$ lies in the lower tail.
$P(X \le 2) = 0.0594$
p-value $= 2 \times 0.1188$
p-value > 0.5, so accept H_0.

Find the p-value

There is not sufficient evidence that there has been any change in the number of smokers.

Interpret the results

Note this is a two-tailed test. The wording in the question clearly indicates this. Do not be put off by the data, which suggests a reduction in smokers.

In practice you should be setting up the hypothesis test **before** we have collected the data.

Make sure you do not get fooled by the question! Read it carefully to determine whether it is a 1-tail or 2-tail test.

Example 9
A new postal delivery system is being trialled in a city.
The postal service watchdog wants to find out if this new system **has changed** the next day delivery rate from the previous value of 80%.
A survey is conducted on 20 letters.
The survey showed that 18 letters were delivered the next day.
Test at the 10% significance level, whether the new system has changed the next day delivery rate.

Solution 1 (using critical regions)

Establish the null and alternative hypotheses

Let p be the probability a letter is delivered the next day.
H_0: $p = 0.8$ (probability of a letter being delivered next day stays the same)
H_1: $p \ne 0.8$ (probability of a letter being delivered next day changes)
Hence we have a 2 tail-test.

Significance level set as 10%

Decide on the significance level

Let X be the number of letters delivered the next day.
If H_0 is true then we are using $p = 0.8$ and $X \sim B(20, 0.8)$
$10\% \div 2 = 5\% = 0.05$

Find the highest value of a for which $P(X \le a) < 0.05$ and the lowest value of b for which $P(X \ge b) < 0.05$

Find the critical regions

$P(X \le 12) = 0.0321 < 0.05$
$P(X \le 13) = 0.0867 > 0.05$

So the highest possible value a is 12.
So the critical value is $X = 12$
The critical region is $X \le 12$

$P(X \geq b) < 0.05$
$1 - P(X \leq b - 1) < 0.05$
$P(X \leq b - 1) > 0.95$

$P(X \leq 18) = 0.9308$
$P(X \leq 19) = 0.9885$

0.95 lies between 0.9308 and 0.9885

So the lowest possible value of $b - 1$ is 19.
So the lowest possible value of b is 20.
So the critical value is $X = 20$
The critical region is $X \geq 20$

This is the region where H_0 is rejected

The critical region is $X \leq 12$, and $X \geq 20$.

18 delivered the next day so $X = 18$.
This does not lie in the critical region, so accept H_0.

Look at the data

There is insufficient evidence to suggest that the new system has changed the next day delivery rate.

Interpret the results

Solution 2 (using probabilities)
Let p be the probability a letter is delivered the next day.
H_0: $p = 0.8$ (probability of a letter being delivered next day stays the same)
H_1: $p \neq 0.8$ (probability of a letter being delivered next day changes)
Hence we have a 2 tail-test.

Establish the null and alternative hypotheses

Significance level set as 10%

Decide on the significance level

Let X be the number of letters delivered the next day.
If H_0 is true then we are using $p = 0.8$ and $X \sim B(20, 0.8)$

18 delivered the next day so $X = 18$, which is in the upper tail.
$P(X \geq 18) = 1 - P(X \leq 17) = 1 - 0.7939 = 0.2061$

Work out the probability

p-value $= 2 \times 0.2061 = 0.4122$
p-value > 0.1, so accept H_0.

There is insufficient evidence to suggest that the new system has changed the next day delivery rate.

Interpret the results

Note that this is again a 2-tail test. The wording in the question clearly indicates this. Do not be put off by the data, which indicates an increase in the delivery rate.

In practice we will be setting up the hypothesis test **before** we have collected the data.

Make sure you do not get fooled by the question! Read it carefully to determine whether it is a 1-tail or 2-tail test.

Example 10

People entering an exhibition have to decide whether they turn right or left. The people organising the exhibition want to know whether there will be a preference for one of the directions.

A trial run is done using 20 people.
It is then found that 15 people went to the left.
At the 5% significance level, does this mean that people have a preference for a particular direction?

Establish the null and alternative hypotheses

Solution 1 (using critical regions)

Let p be the probability a person turns to the left.
H_0: $p = 0.5$ (a person is equally likely to choose left or right)
H_1: $p \neq 0.5$ (there will be a preference in one direction)
Hence we have a 2 tail-test.

Note: you could equally have chosen the probability that a person turns to the right.

Note, we are **not** investigating $p > 0.5$, as we do not know which is the preferred direction at this stage.

Significance level set as 5%

Decide on the significance level

Let X be the number of people who turn left.
If H_0 is true then we are using $p = 0.5$ and $X \sim B(20, 0.5)$
$5\% \div 2 = 2.5\% = 0.025$

Find the highest value of a for which $P(X \leq a) < 0.025$ and the lowest value of b for which $P(X \geq b) < 0.025$

Find the critical regions

$P(X \leq 5) = 0.0207 < 0.025$
$P(X \leq 6) = 0.0577 > 0.025$

So the highest possible value of a is 5.
So the critical value is $X = 5$
The critical region is $X \leq 5$

Now find the lowest value of b where $P(X \geq b) < 0.025$

$P(X \geq b) < 0.025$
$1 - P(X \leq b - 1) < 0.025$
$P(X \leq b - 1) > 0.975$

$P(X \leq 13) = 0.9423$
$P(X \leq 14) = 0.9793$

0.975 lies between 0.9423 and 0.9793

So the lowest possible value of $b - 1$ is 14.
So the lowest possible value of b is 15.
So the critical value is $X = 15$.
The critical region is $X \geq 15$.
The critical region is $X \leq 5$ and $X \geq 15$.

This is the region where H_0 is rejected

15 went to the left, so $X = 15$.
This does lie in the critical region, so reject H_0.

Look at the data

There is evidence to suggest that people have a preference for a particular direction.

Interpret the results

Solution 2 (using probabilities)
Let p be the probability a person turns to the left.
H_0: $p = 0.5$ (a person is equally likely to choose left or right)
H_1: $p \neq 0.5$ (there will be a preference in one direction)
Hence we have a 2 tail-test.

Note, we are **not** investigating $p > 0.5$, as we do not know which is the preferred direction at this stage.

Significance level set as 5%

Decide on the significance level

Let X be the number of people who turn left.
If H_0 is true then we are using $p = 0.5$ and $X \sim B(20, 0.5)$

15 went to the left, so $X = 15$, which is in the upper tail.
$P(X \geq 15) = 1 - P(X \leq 14) = 1 - 0.9793 = 0.0207$.

Work out the probability

p-value $= 2 \times 0.0207 = 0.0414$
p-value < 0.05, so reject H_0.

There is evidence to suggest that people have a preference for a particular direction.

Interpret the results

When you are using a two-tailed test, it is usually obvious whether the observation you have is in the upper or lower tail. However, if the value of p is close to 0 or 1, it may not be clear.
For example, if the distribution is $X \sim B(20, 0.9)$, and the observed value is $X = 14$, it may not be obvious which tail this value of 14 is in, since the probabilities of low values of X are very small.

In cases like these, it is helpful to know that the mean of a binomial distribution is given by np. Then any observation below the mean is in the lower tail and any observation above the mean is in the upper tail. In the case above, the mean is $20 \times 0.9 = 18$ so $X = 14$ is in the lower tail.