Cambridge (CIE) A Level Maths: Probability & Statistics 2

Your notes

Hypothesis Testing

Contents

* Hypothesis Testing
* Type I & Type II Errors

Language of Hypothesis Testing

What is a hypothesis test?

- A hypothesis test uses a **sample of data** in an experiment to test a **statement** made about the value of a **population parameter**
- A hypothesis test is used when the value of the assumed population parameter is questioned
- The hypothesis test will look at the which outcomes are unlikely to occur if assumed population parameter is true
- The probability found will be compared against a given **significance level** to determine whether there is **evidence to believe** that the assumed population parameter is not true

What are the key terms used in statistical hypothesis testing?

- Every hypothesis test **must** begin with a clear **null hypothesis** (what we believe to already be true) and **alternative hypothesis** (how we believe the data pattern or probability distribution might have changed)
- A **hypothesis** is an assumption that is made about a particular **population parameter**
  - A **population parameter** is a numerical characteristic which helps define a population
    - One example of a population parameter is the **probability, p** of an event occurring
    - Another example is the mean of a population
  - The **null hypothesis** is denoted $\mathbf{H_0}$ and sets out the assumed population parameter given that no change has happened
  - The **alternative hypothesis** is denoted $\mathbf{H_1}$ and sets out how we think the population parameter could have changed
  - When a hypothesis test is carried out, the null hypothesis is **assumed to be true** and this assumption will either be **accepted** or **rejected**
- A hypothesis test could be a **one-tailed** test or a **two-tailed** test
- The null hypothesis will always be $\mathbf{H_0 : \theta = \ldots}$
- The alternate hypothesis will depend on if it is a one-tailed or two-tailed test
  - A one-tailed test would test to see if the **population parameter, θ,** has **either increased or decreased**
    - The **alternative hypothesis, $\mathbf{H_1}$ will be $\mathbf{H_1 : \theta > \ldots}$ or $\mathbf{H_1 : \theta < \ldots}$**

- A two-tailed test would test to see if the **population parameter, θ**, has **changed**
    - The **alternative hypothesis, H_1 will be $H_1: \theta \neq \ldots$**
- It is important to read the wording of the question carefully to decide whether your hypothesis test should be one-tailed or two-tailed
- To carry out a hypothesis test an experiment will be carried out on a **sample of data**, the result of this experiment will be the **observed value**
    - A **sample of data** is a subset of data taken from the population
    - The **observed value** is a numerical value calculated from the of data
- A hypothesis test will always be carried out at an appropriate **significance level**
    - The significance level sets the **smallest probability** that an event could have occurred by chance.
        - Any probability smaller than the significance level would suggest that the event is unlikely to have happened by chance
    - The **significance level** must be set **before** the hypothesis test is carried out
    - The significance level will usually be 1%, 5% or 10%, however it may vary

Your notes

Worked Example

A hypothesis test is carried out at the 5% level of significance to test if a normal coin is fair or not.

(i) Describe what the population parameter could be for the hypothesis test.

(ii) State whether the hypothesis test should be a one-tailed test or a two-tailed test, give a reason for your answer.

(iii) Clearly defining your population parameter, state suitable null and alternative hypotheses for the test.

Your notes

(i) The population parameter, p, could be the probability that the coin lands on heads

Equally, this could have been the probability that the coin lands on tails.

(ii) The test should be a two–tailed test because we are testing to see if the probability that the coin lands on heads has changed

If tails had been chosen in Part (i), then tails should be used here.

(iii) The population parameter, p, is the probability the coin lands on heads

The null hypothesis is what we believe should be true:

$H_0: p = \frac{1}{2}$

Probability the coin lands on heads

The alternative hypothesis is how we believe the population parameter could have changed.

$H_1: p \neq \frac{1}{2}$

The probability the coin lands on heads is not $\frac{1}{2}$

Alternative hypothesis

$H_0: p = \frac{1}{2}$

$H_1: p \neq \frac{1}{2}$

Examiner Tips and Tricks

- Make sure you read the question carefully to determine whether the test you are carrying out is for a one-tailed or a two-tailed test.

Critical Regions

How do we decide whether to reject or accept the null hypothesis?

- The null hypothesis would be rejected if the **observed value** falls **within the critical region**
  - The **critical region** is the range of values that the observed value could take which will lead to the **null hypothesis** being **rejected**
- The **critical value** is the boundary of the critical region
  - It is the least extreme value that would lead to the rejection of the null hypothesis

Your notes

- The **critical value** is determined by the **significance level**
- In a **two-tailed test** the significance level is halved and both the upper and the lower tails are tested
- For **discrete distributions** the critical value is the first value that falls within the critical region and so the probability of the observed value falling within the critical region may be lower than the given significance level
  - This probability will be known as the **actual significance level**
  - The actual significance level is the probability of incorrectly rejecting the null hypothesis
- Finding the critical region will be different for a **two-tailed test** than it is for a **one-tailed test**

- For an $\alpha\%$ significance level
  - In a one-tailed test the critical region will consist of $\alpha\%$ in the tail that is being tested for
  - In a two-tailed test the critical region will consist of $\frac{\alpha}{2}\%$ in each tail

Do we always need to find the critical region?

- In most cases the best method of conducting a hypothesis test is to find the critical region
  - It allows you to see how far the observed value is from the critical value and make decisions about whether further testing is necessary
- In some cases a hypothesis test can be carried out without finding the critical region
- The null hypothesis would be rejected if the probability of a value being **at least as extreme** as the observed value, assuming that the null hypothesis is true, is **less than the significance level**
  - If the test is looking for a decrease then extreme values are smaller than the observed value, so find the probability of less than or equal to the observed value
  - If the test is looking for an increase then extreme values are bigger than the observed value, so find the probability of greater than or equal to the observed value
  - This probability is called the "**p-value**"
- In a **two-tailed test** it is common to half the significance level and compare this with the probability found in one of the tails

Worked Example

For the following situations, state at the 1% and 5% significance levels whether the null hypothesis should be rejected or not.

(i) The critical region is $X \leq 3$ and the observed value is 4.

(ii) Assuming the null hypothesis is true, the probability of a value being at least as extreme as the test statistic in a one-tailed hypothesis test is 0.0429.

(iii) Assuming the null hypothesis is true, the probability of a value being at least as extreme as the test statistic in a two-tailed hypothesis test is 0.00705.

(i) The observed value is outside of the critical region so do not reject Ho

(ii) For a one-tailed test the significance level is all in one tail

0.0429 < 5% so reject Ho at the 5% level of significance

0.0429 > 1% so do not reject Ho at the 1% level of significance

(iii) For a two-tailed test use half of the significance level in each tail

5% significance level → 2.5% either side

1% significance level → 0.5% either side

0.00705 < 2.5% so reject Ho at the 5% level of significance

0.00705 > 0.5% so do not reject Ho at the 1% level of significance

Conclusions of Hypothesis Testing

How is a hypothesis test carried out?

- There are a number of ways that a hypothesis test can be carried out for different models, however the following steps should form the base for your test:
- **Step 1.** Define the test statistic and population parameter
- **Step 2.** Write the null and alternative hypotheses clearly
- **Step 3.** Calculate the **critical value(s)** or the necessary probability for the test
- **Step 4.** Compare the observed value with the critical value(s) or the probability with the significance level
- **Step 5.** Decide whether there is enough evidence to reject H_0 or whether it has to be accepted
- **Step 6.** Write a conclusion in context

How should a conclusion be written for a hypothesis test?

- Your conclusion **must** be written in the context of the question
- Use the wording in the question to help you write your conclusion
  - If rejecting the null hypothesis your conclusion should state that there is **sufficient evidence to suggest** the alternative hypothesis is true at this level of significance
  - If accepting the null hypothesis your conclusion should state that there is **not enough evidence** to suggest the alternative hypothesis is true at this level of

Your notes

significance

- Your conclusion **must not** be definitive
  - There is a chance that the test has led to an incorrect conclusion
  - The outcome is dependent on the sample, a different sample might lead to a different outcome
- The conclusion of a two-tailed test can state if there is evidence of a change
  - You should not state whether this change is an increase or decrease

Your notes

Worked Example

A teacher carried out a hypothesis test at the 10% significance level to test if her students perform better in exams after using a new revision technique. Under the null hypothesis she calculates the probability that a value will be at least as extreme as the observed value to be 0.09142. Write a conclusion for her hypothesis test.

The teacher is testing for an increase in exam scores so the test is a one-tailed test.

For a one-tailed test the significance level is all in one tail.

$0.09142 < 10\%$ so reject H_0 at the 10% level of significance

There is sufficient evidence to suggest at the 10% level of significance that the students did perform better after using the new revision technique.

Examiner Tips and Tricks

- It is best to use the exact wording from the question when writing your conclusion for the hypothesis test, do not be afraid to sound repetitive.

Type I & Type II Errors

Any hypothesis test will only provide evidence about whether a parameter has changed or not. A conclusion can not claim with certainty whether to accept or reject the null hypothesis as the test is based on probability, and therefore errors are possible.

What is a Type I error?

- A Type I error occurs when the **null hypothesis is rejected incorrectly**
  - In order for a Type I error to happen, the null hypothesis **must have been rejected**
- If a Type I error has been made, the hypothesis test has provided evidence that there is a change when in fact there is **not a change**
  - Think about the impact of this in some scenarios
    - For example a test saying that a student had cheated in an exam when in fact they had not
- The probability of a Type I error occurring in any hypothesis test is the same as the probability of **rejecting a true null hypothesis**
  - This is the probability of the observed value being at least as extreme as the critical value(s)
  - It is the same or a little bit less than **the significance level**
- In a true hypothesis test you would not need to calculate the probability of a Type I error as it would be the same as the **actual significance level**

What is a Type II error?

- A Type II error occurs when the **null hypothesis is accepted incorrectly**
  - In order for a Type II error to happen, the null hypothesis **must not** have been rejected
- If a Type II error has been made, the hypothesis test has provided evidence that there is no change when in fact there **was** a change
  - Think about the impact of this in some scenarios
    - For example a test saying that a car's brakes have not worn down, when in fact they have
- To find probability of a Type II error occurring in any hypothesis test you would need to be given the true value of the population parameter being tested
  - For example, you would be given the true probability of the event occurring or the true population mean

- The probability of a Type II error would be the probability of the observed value being **outside of the rejection region**, given the true value of the population parameter

		Conclusion	
		Reject H_0	Accept H_0
Reality	H_0 True	Type I	No error
	H_0 False	No error	Type II

Can the probabilities of making the errors be manipulated?

- It is possible to reduce the probability of making a Type I error by reducing the significance level before carrying out the test
  - However, this would decrease the size of the rejection region and therefore could increase the probability of a Type II error
- It is possible to reduce the probability of making a Type II error by increasing the significance level before carrying out the test
  - This would increase the size of the rejection region, making it easier to reject the null hypothesis
  - As the probability of rejecting the null hypothesis has increased, this would increase the probability of making a Type I error
- Before setting the significance level a researcher could consider which error they would want to reduce the likelihood of
  - For example, if the test is for a company advertising that their product works 90% of the time, but customers believe it may be less than this:
    - the company would want to reduce the probability of a Type I error (incorrectly declaring a change)
    - the customers would want to reduce the probability of a Type II error (incorrectly declaring no change)

Worked Example

In the following scenarios, decide whether a Type I error or Type II error could have occurred

Your notes

(i) A farmer is testing for a change in crop growth after trying a new fertiliser. The test concludes that there is no evidence of change at the 5% significance level.

(ii) A dentist's receptionist believes that the waiting times have been reduced due to a new scheduling system. They conduct a hypothesis test and will reject the null hypothesis if no more than two customers wait more than ten minutes. Exactly two customers have to wait more than ten minutes.

(i) No evidence of change $\rightarrow$ the null hypothesis has not been rejected

A Type II error could have occurred

(ii) 2 is inside the critical region $\rightarrow$ the null hypothesis has been rejected

A Type I error could have occurred

Your notes

Examiner Tips and Tricks

- Here are two tips if you cannot remember which error is which but are asked to calculate one on the exam:
  - Look to see if you are given a new population parameter, this will be a Type II error.
  - Check the number of marks, a Type I error is normally only 1 mark whilst a Type II error needs to be calculated and so will be more.