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

Your notes

## Hypothesis Testing (Normal Distribution)

### Contents


* Hypothesis Testing for the Mean of a Normal Distribution

# Normal Hypothesis Testing

## What steps should I follow when carrying out a hypothesis test for the mean of a normal distribution?

- Following these steps will help when carrying out a hypothesis test for the mean of a normal distribution:
- **Step 1.** Define the distribution of the **population mean** usually $X \sim N(\mu, \sigma^2)$
- **Step 2.** Write the null and alternative hypotheses clearly
- **Step 3.** Assuming the null hypothesis to be true, define the **statistic**
- **Step 4.** Calculate either the **critical value(s)** or the probability of the observed value for the test
- **Step 5.** Compare the observed value of the test statistic with the critical value(s) or the probability with the significance level
  - Or compare the z-value corresponding to the observed value with the z-value corresponding to the critical value
- **Step 6.** Decide whether there is enough evidence to reject $H_0$ or whether it has to be accepted
- **Step 7.** Write a conclusion in context

## How should I define the distribution of the population mean and the statistic?

- The **population parameter** being tested will be the **population mean, $\mu$** in a **normally distributed random variable** $N(\mu, \sigma^2)$

## How should I define the hypotheses?

- A hypothesis test is used when the value of the assumed population mean is questioned
- The **null hypothesis**, $H_0$ and **alternative hypothesis**, $H_1$ will always be given in terms of $\mu$
  - Make sure you clearly define $\mu$ before writing the hypotheses, if it has not been defined in the question
  - The null hypothesis will always be $\mathbf{H_0 : \mu = \ldots}$
  - The alternative hypothesis will depend on if it is a **one-tailed** or **two-tailed** test
  - A one-tailed test would test to see if the **value of $\mu$** has **either increased or decreased**
    - The **alternative hypothesis, $H_1$ will be $H_1 : \mu > \ldots$ or $H_1 : \mu < \ldots$**
  - A two-tailed test would test to see if the **value of $\mu$** has **changed**

- The **alternative hypothesis, $H_1$ will be $H_1: \mu \neq ..$**

## How should I define the statistic?

- The population mean is tested by looking at the mean of a sample taken from the population
  - The sample mean is denoted $\bar{x}$
  - For a random variable $X \sim N(\mu, \sigma^2)$ the distribution of the sample mean would be $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$
- To carry out a hypothesis test with the normal distribution, the **statistic used to carry out the test** will be the **sample mean,** $\bar{X}$
  - Remember that the **variance** of the sample mean distribution will be the variance of the population distribution **divided by *n***
  - the **mean** of the sample mean distribution will be the same as the mean of the population distribution

## How should I carry out the test?

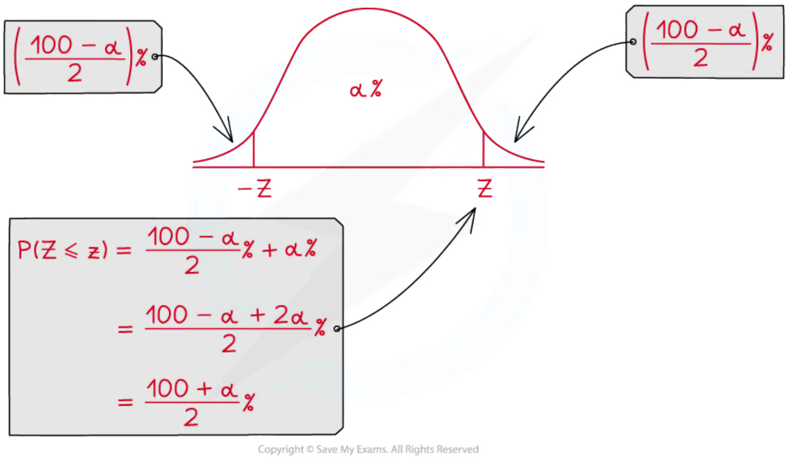
- The hypothesis test can be carried out by
  - either calculating the probability of a value taking the observed or a more extreme value and comparing this with the significance level
    - The normal distribution will be used to **calculate the probability** of a value of the random variable $\bar{X}$ taking the observed value **or a more extreme value**
  - or by finding the critical region and seeing whether the observed value lies within it
    - Finding the critical region can be more useful for considering more than one observed value or for further testing
- A third method is to compare the z-values of your observed value with the z-values at the boundaries of the critical region(s)
  - Find the z-value for your sample mean using $z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
    - This is sometimes known as your **test statistic**
  - Use the table of critical values to find the z-value for the significance level
  - If the z-value for your test statistic is further away from 0 than the critical z-value then reject $H_0$

## How is the critical value found in a hypothesis test for the mean of a normal distribution?

- The **critical value(s)** will be the boundary of the **critical region**

Your notes

- The probability of the observed value being within the critical region, given a true null hypothesis will be the same as the **significance level**
- 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




- To find the critical value(s) use the standard normal distribution:
  - Step 1. Find the distribution of the sample means, assuming $H_0$ is true
  - Step 2. Use the coding $Z = \dfrac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$ to standardise to Z
  - Step 3. Use the table to find the z - value for which the probability of Z being equal to or more extreme than the value is equal to the significance level
    - You can often find this in the table of the critical values
  - Step 4. Equate this value to your expression found in step 2
  - Step 5. Solve to find the corresponding value of $\bar{x}$
- If using this method for a two-tailed test be aware of the following:
  - The symmetry of the normal distribution means that the z - values will have the same absolute value
  - You can solve the equation for both the positive and negative z – value to find the two critical values
    - Check that the two critical values are the same distance from the mean

Your notes

Your notes

### Worked Example

The time, minutes, that it takes Amelea to complete a 1000-piece puzzle can be modelled using $X \sim \mathrm{N}(204, 81)$. Amelea gets prescribed a new pair of glasses and claims that the time it takes her to complete a 1000-piece puzzle has decreased. Wearing her new glasses, Amelea completes 12 separate 1000-piece puzzles and calculates her mean time on these puzzles to be 201 minutes. Use these 12 puzzles as a sample to test, at the 5% level of significance, whether there is evidence to support Amelea's claim. You may assume the variance is unchanged.

Step 1: Define the distribution of the population mean:

For $X \sim N(\mu, 81)$

Step 2: Write the hypotheses

$H_0 : \mu = 204$

$H_1 : \mu < 204$

Step 3: Define the statistic

Watch out, $\sigma^2 = 81$ so $\sigma = 9$

Assuming $H_0$ true, $\bar{X} \sim N\left(204, \frac{9^2}{12}\right)$

Step 4: Calculate the z-value for the test-statistic

Assuming $H_0$ is true, then

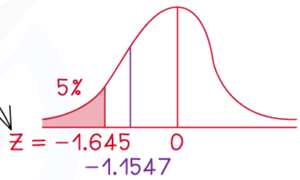
$$P(X \leqslant 201) = P\left(Z \leqslant \frac{201 - 204}{\frac{9}{\sqrt{12}}}\right) = P(Z \leqslant -1.1547...)$$

At least as extreme as the observed value

Alternatively, you could find this probability and compare with the significance level ($P(X < 201) = 0.1241$)

Step 5: Compare the test-statistic with the critical z-value

The z-value for a 5% significance level in the lower tail is $Z = -1.645$ (table of critical values)




$-1.1547 > -1.645$

Step 6: Decide whether there is enough evidence to reject $H_0$

The test-statistic is outside of the critical region so there is not enough evidence to reject $H_0$

Step 7: Write a conclusion in context

There is insufficient evidence to reject the null hypothesis at the 5% level of significance

The test does not support Ameleas' claim that the amount of time it takes her to complete a 1000-piece puzzle has decreased

* Note: If finding the critical region, steps 4 and 5 would be:

Step 4: Calculate the critical value

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{n}} \Rightarrow z = \frac{\bar{x} - 204}{\frac{9}{\sqrt{12}}}$$

$$-1.645 = \frac{\sqrt{12}}{9}(\bar{x} - 204)$$

$$\bar{x} = 199.72...$$

Step 5: Compare the observed value of the test-statistics with the critical region:

$201 > 199.72...$ so 201 is outside of the critical region



## Examiner Tips and Tricks

- Use a diagram to help, especially if looking for the critical value and comparing this with an observed value of a test statistic or if working with z-values.

Your notes