

17

HYPOTHESIS TESTING

SYLLABUS

- Concepts of null and alternative hypotheses, test statistic, critical region, critical value, level of significance and p -value
 - Understand the nature of a statistical test of hypothesis, the concepts and formulation of a null hypothesis (H_0) and an alternative hypothesis (H_1), the difference between 1-tail and 2-tail tests, and the concepts of significance level and test statistic;
 - Understand that the p -value is the smallest level of significance at which the null hypothesis can be rejected, calculate the p -value using a graphic calculator, and determine whether there is sufficient evidence to reject the null hypothesis
- Test for a population mean based on
 - A sample from a normal population of known variance;
 - A large sample from any population
- Formulate hypotheses and carry out a hypothesis test for a population mean in the following cases:
 - A sample from a normal population of known variance;
 - A large sample drawn from any population, using the Central Limit Theorem and an unbiased estimate of the population variance
- 1-tail and 2-tail tests
- Interpret the results of a hypothesis test in the context of the problem

CONTENTS

1	Hypothesis Tests.....	3
1.1	The Null and Alternative Hypotheses	6
1.2	Level of Significance of a Test	7
1.3	Test Statistic	8
1.4	The p -value	10
2	Z-Test.....	13
3	Critical Region and Critical Values.....	17
4	Using the Result of Hypothesis Testing to Find Unknowns	19

LECTURE 1

Lesson Outline

- Introduction to hypothesis testing and how it is done
 - Understand the meaning of null and alternative hypotheses, 1-tail and 2-tail tests, and significance level
-

1 HYPOTHESIS TESTS

In real life, we are often required to make decisions and draw conclusions about an entire population based on data from a random sample. For example, we may be asked whether eating a supplement decreases the risk of cancer based on a study of 80 randomly selected people.

To arrive at a conclusion, assumptions typically need to be made about the population. Such assumptions are called statistical hypotheses. They may or may not be true, and can be tested. Here are two other scenarios:

Scenario 1: The Fair Coin

We may hypothesise that a coin is fair. If we tossed that coin 100 times and obtained 52 heads, is the hypothesis that the coin is fair still valid?

What if 70 heads were obtained instead? More generally, under what conditions would we reject the hypothesis that the coin is fair?

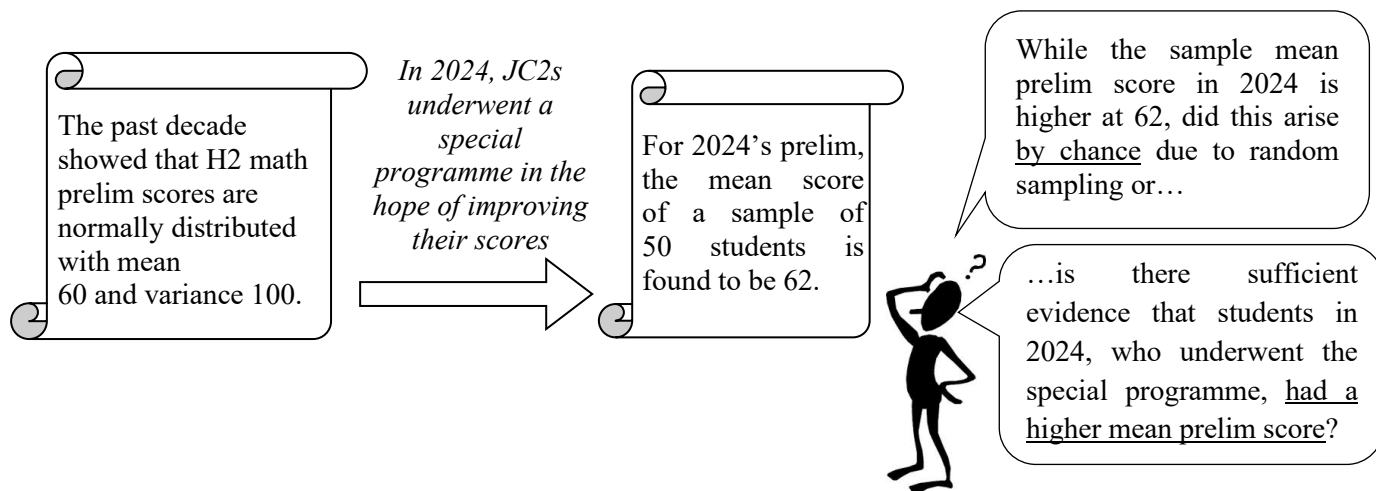
Scenario 2: Lifespan of Batteries

A manufacturer may claim that the mean lifespan of their batteries for pocket calculations is 600 hours. An independent party tests the manufacturer's claim using a sample of 50 batteries.

If the mean lifespan of the 50 batteries tested is 585 hours, how confident is the third party to conclude that the manufacturer overstated their battery lifespan?

Example 1

We use a fictional (but familiar) story to illustrate hypothesis testing.



We state the two possibilities as two statistical hypotheses, the null hypothesis against the alternative hypothesis.

Denote the population mean prelim score by μ .

The **null hypothesis**, denoted by H_0 , is that the mean prelim score of students in 2024 is still 60, i.e. the special programme may not have an effect on prelim scores.



We write

The **alternative hypothesis**, denoted by H_1 , is that the mean prelim score of students in 2024 is more than 60, i.e. the programme may have an effect on improving prelim scores.



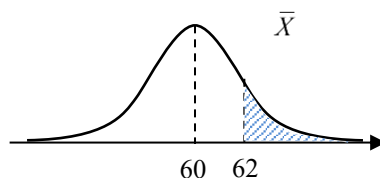
We write

Now, we **set up a model on the assumption that H_0 is true** to calculate the probability that a sample mean of 62 or more could arise by chance. We let X be the random variable denoting the “H2 math prelim score of a randomly chosen student”.

Under the assumption that H_0 is true, the mean prelim score is normally distributed with mean 60 and variance 100. That is, $X \sim N(60, 10^2)$.

For a random sample of 50 students, the sampling distribution is

$\bar{X} \sim N\left(60, \frac{10^2}{50}\right)$. Using the GC, $P(\bar{X} > 62) = 0.0786$ (to 3 s.f.).



So, **under the assumption that H_0 is true**, we have a probability of 0.0786 of seeing a sample mean score of at least 62, and we could either:

- **Not reject H_0** as 0.0786 is **large enough** that we feel a sample mean of at least 62 could arise by chance (insufficient evidence to reject H_0 that the mean prelim score in 2024 is still 60); or
- **Reject H_0 in favour of H_1** as 0.0786 is **too small** that we feel that the sample mean of 62 is unlikely to arise by chance (sufficient evidence to favour H_1 that the mean prelim score in 2024 is more than 60).

An equivalent way of interpreting the probability of 0.0786 is:

- **If H_0 were actually true, but we decided to reject H_0** whenever we see a sample mean score of at least 62, then we have a 7.86% chance of incorrectly rejecting H_0 .

So, we must set a threshold for when to reject H_0 , i.e. when it is considered too improbable for the sample data to arise by chance. But,

- The threshold cannot be too high, as we may incorrectly reject H_0 too often; and
- The threshold cannot be too low, as we may not reject H_0 even if the sample data were improbable and unlikely to arise by chance.

In practice, this threshold is decided before data is collected, and is called the **level of significance** of the test.

The significance level is typically taken as 5% (or 0.05), i.e. 5% chance of incorrectly rejecting H_0 given that H_0 is true.

In Example 1, we do not reject H_0 at a 5% level of significance, as the probability of 0.0786 is more than 0.05. So, it is considered probable enough for a sample mean score of 62 to arise by chance.

- Fields like medicine may use a lower level of significance at 1% (or 0.01) for it is vital to ensure that the effect seen is not due to chance;
- Some fields like psychology use a higher level of significance of 10% (or 0.1). In Example 1, we reject H_0 at a 10% level of significance. This is because the probability of 0.0786 is less than 0.1, which is now considered sufficiently improbable to arise by chance.

In Example 1, would we reject H_0 at a 1% level of significance?

Determining what to choose as the level of significance of a hypothesis test depends on the problem context and subject discipline.

1.1 The Null and Alternative Hypotheses

In general, when conducting a hypothesis test, we first state two statistical hypotheses to decide between, namely:

- The null hypothesis H_0 , which states that the population mean is of the claimed value; and
- The alternative hypothesis H_1 , which is used to contradict H_0 .

In Example 1, $H_0 : \mu = 60$ is tested against $H_1 : \mu > 60$. Notice that H_1 only considers if the population mean μ is more than 60, not whether it is less than 60. Such a hypothesis test is a 1-tail (right-tail) test. In general,

A hypothesis test is a **1-tail test** if H_1 is of the form

(i) $\mu > \mu_0$ (right-tail), or (ii) $\mu < \mu_0$ (left-tail),

for some number μ_0 .

Example 2

A student claims that the mean time spent solving a Vectors question is 20 min. He measured the time taken to solve the question for a random sample of 150 students, and found that the sample mean is 19.1 min with a standard deviation of 5.82 min. Did the student overstate the mean time spent to solve the question?

Hypotheses

Let μ be the population mean time spent solving the question in min.

To test H_0 : against H_1 :

Therefore, the hypothesis test in Example 2 is a 1-tail (left-tail) test.

A 1-tail test considers only an increase or a decrease in the parameter of interest, but not both. However, we may want H_1 which considers any change in the parameter, whether an increase or a decrease (i.e. not equal to a claimed value). We say that:

A hypothesis test is a **2-tail test** if H_1 has the form

$\mu \neq \mu_0$ for some number μ_0 .

Example 3

The optimal brightness of a lightbulb in school is 110 lux. In a random sample of 250 lightbulbs in a school, the mean brightness is 108.8 lux with a standard deviation of 11.9 lux. Assuming the sample is from a normal distribution, would the sample provide sufficient evidence that the population mean brightness of lightbulbs in the school is 110 lux?

Hypotheses

Let μ be the population mean brightness of lightbulbs in lux.

To test H_0 : against H_1 :

Therefore, the hypothesis test in Example 3 is a 2-tail test.

In summary,

	Example 1: Did the special programme improve prelim exam scores?	Example 2: Did the student overstate the mean solving time?	Example 3: Is the mean brightness of lightbulbs different from 110 lux?
Null hypothesis	$H_0 : \mu = 60$	$H_0 : \mu = 20$	$H_0 : \mu = 110$
Alternative hypothesis	$H_1 : \mu > 60$	$H_1 : \mu < 20$	$H_1 : \mu \neq 110$
Type of test	1-tail (right-tail)	1-tail (left-tail)	2-tail

1.2 Level of Significance of a Test

After writing the hypotheses and before collecting data, we decide on the level of significance of the test. Recall that:

An $\alpha\%$ **level of significance** means that “there is a probability of $\frac{\alpha}{100}$ of rejecting H_0 when H_0 is true”.

The significance level is usually stated in the question. If it is not stated, it must be chosen sensibly. In practice, a 5% significance level is typically used, i.e. “a probability of 0.05 of rejecting H_0 when H_0 is true”.

If asked what the level of significance means, the phrase “rejecting H_0 when H_0 is true” should be written in the context of the question.

LECTURE 2

Lesson Outline

- Understand the meaning of a test statistic and the p -value
- Apply the z -test in contexts with known and unknown population variance

1.3 Test Statistic

Now that we have set up the hypotheses and significance level of our test, we can collect data from our sample. After collecting data, we summarise the data into a single number – called a test statistic – which allows us to determine whether our findings were likely have arisen by chance.

In the subsequent discussion, suppose $H_0 : \mu = \mu_0$ for a fixed number μ_0 .

Case 1: For **any sample size n** from a **normal distribution** with **known population variance σ^2**

Under H_0 , the test statistic has distribution $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$ *exactly*, and the sample mean $\bar{x}$ is the value of the test statistic.

We use the standardised test statistic if there is an unknown. For example, if the sample size n is unknown.

Under H_0 , the standardised test statistic is $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ *exactly*, and $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ is the value of the standardised test statistic.

We use Example 1 to illustrate how the test statistic is a number that can be evaluated. The context of Example 1 falls into Case 1 above.

Recall Example 1:

The past decade showed that H2 math prelim scores are normally distributed with mean 60 and variance 100.

For 2024's prelim, the mean score of a sample of 50 students is found to be 62.

Did the special programme improve prelim exam scores?

Example 1 (cont'd from Pg 4)

Recall that $H_0 : \mu = 60$, and so, $\mu_0 = 60$.

From the sample data, $\bar{x} = 62$, $\sigma = 10$, $n = 50$.

The sample mean is the test statistic, i.e. $\bar{x} = 62$, and we can find how likely it was for the value of 62 to arise from a $N\left(60, \frac{10^2}{50}\right)$ distribution.

Alternatively, we can standardise, $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{62 - 60}{10/\sqrt{50}} = 1.4142$ (5 sf).

We can then find how likely it was for the value of 1.4142 to arise from a $N(0,1)$ distribution.

Case 2: For a large sample ($n \geq 30$) from a non-normal distribution with known population variance σ^2

Under H_0 , the test statistic $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$ approximately by the Central

Limit Theorem (CLT) for sufficiently large sample size n , and the sample mean $\bar{x}$ is value of the test statistic.

The standardised test statistic $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$ approximately by

CLT, and $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ is the value of the standardised test statistic.

In many real-life situations, the population variance σ^2 is unknown. That said, we can calculate an unbiased estimate of the population variance s^2 from the sample. Provided that the sample size is sufficiently large, we can reasonably replace σ^2 by its unbiased estimate s^2 .

Case 3: For a large sample ($n \geq 30$) from a normal distribution with unknown population variance

Under H_0 , the test statistic $\bar{X} \sim N\left(\mu_0, \frac{s^2}{n}\right)$ approximately, and $\bar{x}$ is value of the test statistic.

In Case 3, why is it not necessary to apply CLT?

The standardised test statistic $Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim N(0,1)$ approximately, and

$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ is the value of the standardised test statistic.

Case 4: For a large sample ($n \geq 30$) from a non-normal distribution with unknown population variance

Under H_0 , the test statistic $\bar{X} \sim N\left(\mu_0, \frac{s^2}{n}\right)$ approximately by the Central

Limit Theorem (CLT) for sufficiently large sample size n , and $\bar{x}$ is the value of the test statistic.

The standardised test statistic $Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim N(0,1)$ approximately by the

CLT, and $z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ is the value of the standardised test statistic.

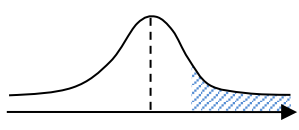
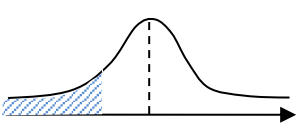
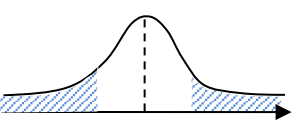
1.4 The p -value

We have summarised our sample data into test statistic, and we also know the distribution of the test statistic under H_0 . This allows us to calculate how likely it was for test statistic to have arisen by chance. Formally,

The **p -value** is the probability of observing a test statistic at least as extreme as the one calculated from the sample, given that H_0 is true.

If asked what the p -value means, it should be stated in the context of the question.

What is “as extreme” depends on whether the test is a (i) right-tail; (ii) left-tail; or (iii) 2-tail test. “Extreme values” are shaded in each case below.

Right-tail test	Left-tail test	2-tail test
 Extreme values are those <i>larger</i> than the observed test statistic	 Extreme values are those <i>less</i> than the observed test statistic	 Extreme values <i>differ more</i> than the observed test statistic

The p -value, p , is compared with the level of significance, $\alpha\%$, to arrive at a conclusion for the test, i.e.

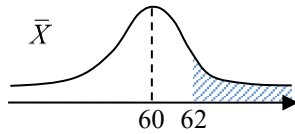
- If $p < \frac{\alpha}{100}$, “we reject H_0 at $\alpha\%$ level of significance.”
- If $p > \frac{\alpha}{100}$, “we do not reject H_0 at $\alpha\%$ level of significance.”

So, we can also define the p -value as the smallest level of significance at which H_0 can be rejected.

Example 1 (cont'd from Pg 8)

As $H_1 : \mu > 60$, this is a right-tail test.

Under H_0 , $\bar{X} \sim N\left(60, \frac{10^2}{50}\right)$ exactly.



The value of the test statistic is $\bar{x} = 62$.

The p -value is $p = P(\bar{X} > 62) = 0.0786$.

Since $p = 0.0786 > 0.05$, we do not reject H_0 at a 5% level of significance. There is insufficient evidence at a 5% level of significance to conclude that the mean prelim score of 2024 students is more than 60.

Let us apply the concepts learnt to Examples 2 and 3. In both examples, we use a 5% level of significance.

Example 2 (cont'd from Pg 6)

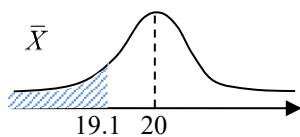
We test $H_0 : \mu = 20$ against $H_1 : \mu < 20$ at a 5% significance level.

Under H_0 , $\bar{X} \sim N\left(\mu_0, \frac{s^2}{n}\right)$ *approximately by the Central Limit Theorem*

(as $n = 150 \geq 30$ is large), where $\mu_0 = 20, n = 150, \bar{x} = 19.1$, and, as

$$s^2 = \frac{n}{n-1} \times \text{sample variance}, \text{ so } s^2 = \frac{150}{149} (5.82)^2 \Rightarrow s = 5.8395.$$

The value of the test statistic is $\bar{x} = 19.1$.



The p -value is $p = P(\bar{X} < 19.1) = 0.0295$.

As $p = 0.0295 < 0.05$, we reject H_0 at the 5% level of significance, and conclude there is sufficient evidence at the 5% level of significance that the student overstated the mean solving time of 20 minutes.

Recall Example 1:

The past decade showed that H2 math prelim scores are normally distributed with mean 60 and variance 100.

For 2024's prelim, the mean score of a sample of 50 students is found to be 62.

Did the special programme improve prelim exam scores?

Recall Example 2:

A student claims that the mean time spent solving a Vectors question is 20 min.

He measured the time taken to solve the question for a random sample of 150 students, and found that the sample mean is 19.1 min with a standard deviation of 5.82 min.

Did the student overstate the mean time spent to solve the question?

Example 2's context falls into Case 4 (Pg 9), as

- sample size is large;
- the distribution is not given to be normal; and
- population variance is unknown.

Recall Example 3:

The optimal brightness of a lightbulb in school is 110 lux. In a random sample of 250 lightbulbs in a school, the mean brightness is 108.8 lux with a standard deviation of 11.9 lux.

Assuming the sample is from a normal distribution, would the sample provide sufficient evidence that the population mean brightness of lightbulbs in the school is 110 lux?

Example 3's context falls into Case 3 (Pg 9). Why so?

Example 3 (cont'd from Pg 7)

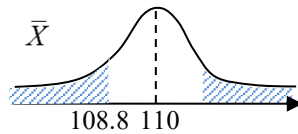
We test $H_0 : \mu = 110$ against $H_1 : \mu \neq 110$ at a 5% significance level.

Under H_0 , $\bar{X} \sim N\left(\mu_0, \frac{s^2}{n}\right)$ *approximately* where $\mu_0 = 110, n = 250$,

$\bar{x} = 108.8$, and

$$s^2 = \frac{n}{n-1} \times \text{sample variance}, \text{ so } s^2 = \frac{250}{249}(11.9)^2 \Rightarrow s = 11.924.$$

The value of the test statistic is $\bar{x} = 108.8$.



The p -value is $p = 2 \times P(\bar{X} < 108.8) = 0.112$.

As $p = 0.112 > 0.05$, we do not reject H_0 at the 5% level of significance, and conclude there is insufficient evidence at the 5% level of significance that the mean brightness of lightbulbs in school is different from 110 lux.

2 Z-TEST

The hypothesis tests we conducted in the previous section are all known as the **z-test**. Here we summarise the process before discussing examples.

Six Step Process for Hypothesis Testing Using the z-test

Step 1: State the null and alternative hypotheses

To test $H_0 : \mu = \mu_0$ against (i) $H_1 : \mu > \mu_0$; (ii) $H_1 : \mu < \mu_0$ or (iii) $H_1 : \mu \neq \mu_0$ depending on the question context.

μ_0 is a fixed number.

Step 2: State the level of significance at which the test is conducted

... at $\alpha\%$ level of significance.

Step 3: State distribution of $\bar{X}$ (or Z) under H_0

σ^2 is known	σ^2 is unknown
Under H_0 , $\bar{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right)$ or $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$	Under H_0 , $\bar{X} \sim N\left(\mu_0, \frac{s^2}{n}\right)$ or $Z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \sim N(0, 1)$ approximately when n large

The value of μ_0, n, σ and/or s must be stated. Central Limit Theorem must be applied for non-normal populations.

Step 4: State value of the test statistic used and the p -value

Value of test statistic: $\bar{x} = \underline{\hspace{2cm}}$ (or $z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \underline{\hspace{2cm}}$)

p -value, $p = \underline{\hspace{2cm}}$

Step 5: Compare the p -value with the level of significance

“As $p < \frac{\alpha}{100}$, reject H_0 at $\alpha\%$ level of significance.” OR

“As $p > \frac{\alpha}{100}$, do not reject H_0 at $\alpha\%$ level of significance.”

Step 6: Conclude in the context of the problem

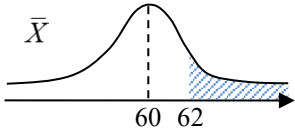
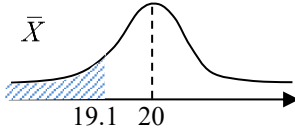
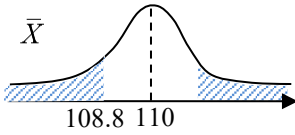
In the case of rejection of H_0 :

“There is sufficient evidence at the $\alpha\%$ level of significance that H_1 is true (in the context of the question).”

In the case of non-rejection of H_0 :

“There is insufficient evidence at the $\alpha\%$ level of significance that H_1 is true (in the context of the question).”

The following table summarises the z-test for Examples 1 to 3.

	Example 1: Did the special programme improve prelim exam scores?	Example 2: Did the student overstate the mean solving time?	Example 3: Is the mean brightness of lightbulbs different from 110 lux?
Null hypo.	$H_0 : \mu = 60$	$H_0 : \mu = 20$	$H_0 : \mu = 110$
Alt. hypo.	$H_1 : \mu > 60$	$H_1 : \mu < 20$	$H_1 : \mu \neq 110$
Sig. level	5%	5%	5%
Test statistic under H_0	Under H_0 , $\bar{X} \sim N\left(60, \frac{10^2}{50}\right)$	Under H_0 , $\bar{X} \sim N\left(20, \frac{s^2}{150}\right)$ approximately by CLT, where $s^2 = \frac{150}{149}(5.82)^2$	Under H_0 , $\bar{X} \sim N\left(110, \frac{s^2}{250}\right)$ approximately, where $s^2 = \frac{250}{249}(11.9)^2$
Value of test statistic	$\bar{x} = 62$	$\bar{x} = 19.1$	$\bar{x} = 108.8$
			
p-value	$P(\bar{X} > 62)$ $= 0.0786$ (3 s.f.)	$P(\bar{X} < 19.1)$ $= 0.0295$ (3 s.f.)	$2P(\bar{X} < 108.8)$ $= 0.112$ (3 s.f.)
Compare p-value with $\alpha\%$	$p\text{-value} > 0.05$ (<u>Do not reject</u> H_0 at the 5% level of significance)	$p\text{-value} < 0.05$ (<u>Reject</u> H_0 at the 5% level of significance)	$p\text{-value} > 0.05$ (<u>Do not reject</u> H_0 at the 5% level of significance)
Conclude	There is <u>insufficient evidence</u> at a 5% level of significance to conclude that the mean prelim score of 2024 students is more than 60 marks.	There is <u>sufficient evidence</u> at the 5% level of significance that the student overstated the mean solving time of 20 minutes.	There is <u>insufficient evidence</u> at the 5% level of significance that the mean brightness of lightbulbs differs from 110 lux.

Example 4

The mean lifetime of the fluorescent light bulbs produced by a company is μ and the standard deviation is 120 hours. If the mean lifetime of a random sample of 100 fluorescent light bulbs produced by the company is found to be 1570 hours, use a 1% level of significance to test if the mean lifetime of the fluorescent light bulbs is less than 1600 hours.

State the meaning of this p -value and level of significance in the context of the question.

Solution

To test _____ against _____ at _____ level of significance.

Under H_0 , as $n = 100$, _____ approximately by Central

Limit Theorem.

Value of test statistic:

The p -value,

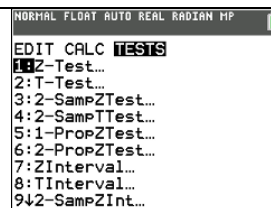
Since $p = 0.00621 < 0.01$, we reject H_0 , and conclude there is sufficient evidence at the 1% level of significance that the mean lifetime of the fluorescent light bulbs is less than 1600 hours.

The p -value of 0.00621 is the probability that the mean lifetime of a random sample of 100 light bulbs is at most 1570 hours (the sample mean observed), given that the mean lifetime is indeed 1600 hours.

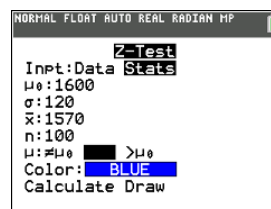
A 1% level of significance indicates a probability of 0.01 to conclude that the mean lifetime of light bulbs is less than 1600 hours, given that it is in fact 1600 hours.

Using GC to calculate p-value and value of test statistic

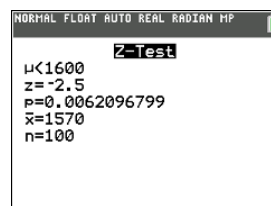
- 1) Press **[STATS]** and select the **TESTS** menu. Then select **1: Z-Test...**



- 2) Select **Stats** and enter as shown on the screenshot on the right. Note that σ here refers to the population standard deviation.
Select the appropriate test. Then select **Calculate** and press **[ENTER]**



- 3) From GC, $p = 0.00621$ and $z = -2.5$.



We could also standardise.

Under H_0 :

$$Z = \frac{\bar{X} - 1600}{120/\sqrt{100}} \sim N(0,1)$$

approximately by CLT.

Value of standardised test statistic:

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{1570 - 1600}{120/10} = -2.5$$

p -value:

$$p = P(Z < -2.5) = 0.00621$$

Example 5

The variance of a population is unknown, and in order to test the hypothesis $H_0 : \mu = 5.35$, a random sample of 1000 is taken. The results of the sample are as follows:

$$\sum x = 5384, \quad \sum x^2 = 29210$$

Test this hypothesis against the alternative hypothesis $H_1 : \mu \neq 5.35$ at the 5% level of significance.

Write down the smallest level of significance at which the null hypothesis can be rejected.

Solution

$$\text{Sample mean, } \bar{x} = \frac{\sum x}{n} = \frac{5384}{1000} = 5.384$$

$$s^2 = \frac{1}{999} \left(29210 - \frac{(5384)^2}{1000} \right) = 0.22277 = 0.223 \text{ (to 3 s.f.)}$$

To test $H_0 : \mu = 5.35$ against $H_1 : \mu \neq 5.35$ at a 5% level of significance

Under H_0 , approximately

by the Central Limit Theorem.

$$p\text{-value} = 0.0227 \text{ (from GC)}$$

Since

Let $\alpha\%$ be the smallest level of significance at which the null hypothesis can be rejected.

Therefore, the smallest level of significance at which the null hypothesis can be rejected is 2.28%.

LECTURE 3

Lesson Outline

- Critical Region and Critical Values
- Using given conclusion of a hypothesis test to find unknowns

3 CRITICAL REGION AND CRITICAL VALUES

There is another approach to hypothesis testing related to the z -test. Suppose we test H_0 against H_1 at an $\alpha\%$ level of significance, then:

- The set of all values of $\bar{X}$ or Z (depending on the test statistic used) for which H_0 is rejected is called the **critical region**; and
- The boundaries of the critical region are called the **critical values**.

Thus, using a “critical region” approach, we:

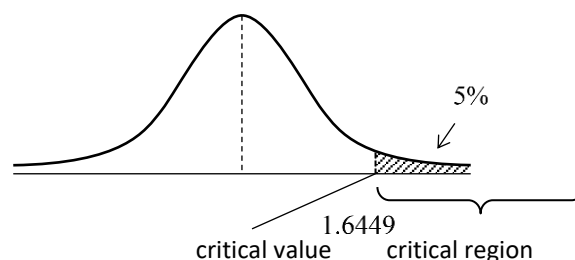
- Reject $H_0 \Leftrightarrow$ the value of the test statistic lies inside the critical region; and
- Do not reject $H_0 \Leftrightarrow$ the value of the test statistic lies outside the critical region.

The “critical region” approach is useful when we are given the result of a hypothesis test, but we must find unknown parameters in a distribution. So, we may construct the critical region with z as the standardised test statistic, as it has a standard normal distribution $N(0,1)$ under H_0 .

For ease of discussion, we use a 5% significance level with z as the test statistic. However, the approach generalises easily for $\alpha\%$ significance level and when using the test statistic $\bar{x}$.

Constructing the Critical Region for a Right-Tail Test ($H_1 : \mu > \mu_0$)

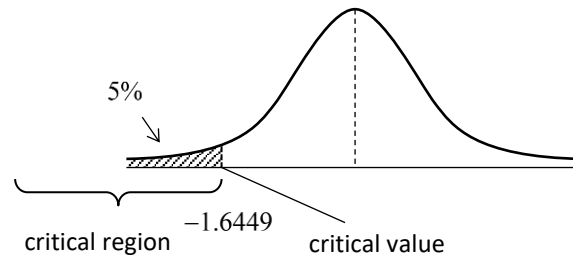
We seek z such that $P(Z > z) = 0.05$, i.e. area of 0.05 in the right-tail of $N(0,1)$. Using the GC, we find that $z = 1.6449$.



The critical value is 1.6449 and $z > 1.6449$ is the critical region.

Constructing the Critical Region for a Left-Tail Test ($H_1 : \mu < \mu_0$)

We seek z such that $P(Z < z) = 0.05$, i.e. area of 0.05 in the left-tail of $N(0,1)$. Using the GC, we find that $z = -1.6449$.

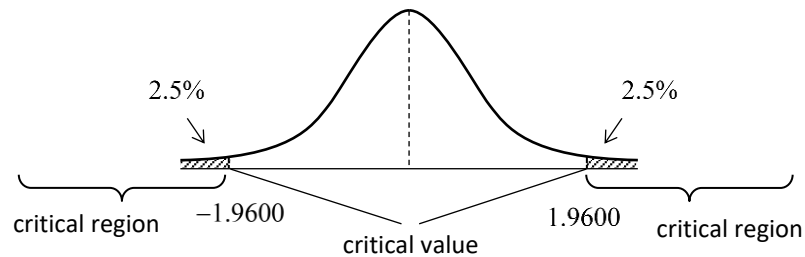


The critical value is -1.6449 and $z < -1.6449$ is the critical region.

Constructing the Critical Region for a Two-Tail Test ($H_1 : \mu \neq \mu_0$)

We seek z such that $P(Z < -z) + P(Z > z) = 0.05$, i.e. area of 0.05 in total from both tails of $N(0,1)$. By symmetry, $P(Z > z) = 0.025$.

Using the GC, we find that $z = 1.9600$.



The critical values are -1.9600 and 1.9600 ; and the critical region is given by $z < -1.9600$ or $z > 1.9600$.

4 USING THE RESULT OF HYPOTHESIS TESTING TO FIND UNKNOWNNS

Example 6

A manufacturer claims that a particular model of mobile phone has a mean standby-time of 36 hours. A random sample of n batteries of this model is tested and the mean 'standby-time' is found to be 33.8 hours. Given that the standard deviation of the standby-time is 12.1 hours and that n is large, find the least value of n which will result in the manufacturer's claim being overstated at the 8% level of significance.

It is important to read and identify the correct information:

Popn mean $\mu_0 = 36$

Sample mean $\bar{x} = 33.8$

Popn variance $\sigma^2 = 12.1^2$

State, with a reason, if it is necessary to assume that the standby-time of a mobile phone battery is normally distributed.

Solution

Let μ be the population mean stand-by time of a mobile phone in hours.

To test _____ against _____ at an _____ level of significance.

Under H_0 , since n is large, $\bar{X} \sim N\left(36, \frac{12.1^2}{n}\right)$ approximately by Central Limit Theorem.

The standardised test statistic $Z = \frac{\bar{X} - 36}{12.1/\sqrt{n}} \sim N(0, 1)$ approximately by Central Limit Theorem.

We use Z as there is an unknown constant in the distribution of $\bar{X}$.

Value of standardised test statistic: $z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} =$

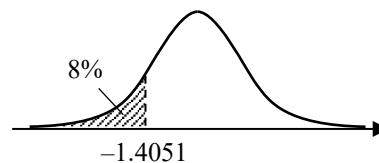
Given the manufacturer's claim is being overstated at 8% significance level (i.e. H_0 is rejected), the critical region is $z < -1.4051$.

Construct the critical region for a left-tail test at an 8% significance level by finding z such that $P(Z < z) = 0.08$.
From GC, $z = -1.4051$.

$$\sqrt{n} > 7.72805$$

$$n > 59.723$$

$$\text{Least } n = 60$$



It is not necessary to assume that the standby-time of a mobile phone battery is normally distributed, because sample size n is large,

$\bar{X} \sim N\left(36, \frac{12.1^2}{n}\right)$ approx. by Central Limit Theorem. ■

Example 7

A factory produces bolts and nuts. It is claimed that the mass of nuts produced is normally distributed with a mean mass of 6 g and standard deviation of 0.8 g. This claim is not rejected at 5% level of significance, on the basis that the mean mass obtained from a random sample of 50 nuts. Find the range of values of the mean mass of the 50 nuts in the sample.

Solution

Let μ denote the population mean mass of a nut produced in grams.

To test $H_0 : \mu = 6$ against $H_1 : \mu \neq 6$ at a 5% level of significance.

Construct the critical region for a 2-tail test at a 5% significance level by finding $\bar{x}$ such that

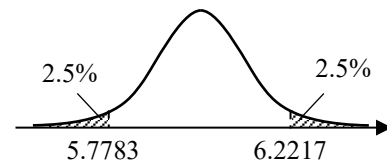
$$P(\bar{X} < \bar{x}) = 0.025$$

and

$$P(\bar{X} > \bar{x}) = 0.025$$

These values can be found using the GC.

Under H_0 ,



Given the result that the claim is not rejected at 5% significance level, i.e. do not reject H_0 , this means that

Example 8

In a factory, the time in minutes for an employee to install an electronic component is a normally distributed continuous random variable T . The standard deviation of T is 5.0 and under ordinary conditions the expected value of T is 38.0. After background music is introduced into the factory, a sample of n components is taken and the mean time taken for randomly chosen employees to install them is found to be $\bar{t}$ minutes. A test is carried out, at the 5% significance level, to determine whether the mean time taken to install a component has been reduced.

- (i) State appropriate hypotheses for the test, defining any symbols you use.
- (ii) Given that $n = 50$, state the set of values of $\bar{t}$ for which the result of the test would be to reject the null hypothesis.
- (iii) It is given instead that $\bar{t} = 37.1$ and the result of the test is that the null hypothesis is not rejected. Obtain an inequality involving n , and hence find the set of values that n can take.

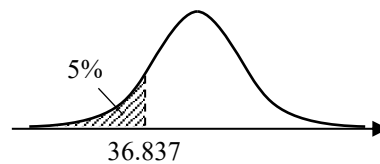
[9740 N2011/II/10]

Solution

(i)

- (ii) Let T denote the time taken for an employee to install an electronic component.

Given the that the null hypothesis would be rejected at 5% significance level,



Construct the critical region for a left-tail test at a 5% significance level by finding $\bar{t}$ such that $p(\bar{T} < \bar{t}) = 0.05$.
From the GC, $\bar{t} = 36.837$.

Therefore, the set of values is $\{\bar{t} \in \mathbb{R} : 0 < \bar{t} < 36.8\}$.

(iii) To test $H_0 : \mu = 38.0$ against $H_1 : \mu < 38.0$

$$\text{Under } H_0, \bar{T} \sim N\left(38.0, \frac{5.0^2}{n}\right)$$

$$\text{Standardised test statistic: } Z = \frac{\bar{T} - 38}{5/\sqrt{n}} \sim N(0, 1)$$

$$\text{Value of standardised test statistic: } z = \frac{\bar{t} - 38}{5/\sqrt{n}}$$

Construct the critical region for a left-tail test at a 5% significance level by finding z such that $P(Z < z) = 0.05$.
From GC, $z = -1.6449$.

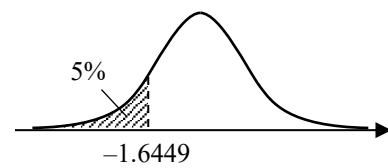
Given the result that null hypothesis is not rejected at 5% significance level,

$$z > -1.6449$$

$$\Rightarrow \frac{37.1 - 38}{5/\sqrt{n}} > -1.6449$$

$$\Rightarrow \frac{-0.9}{5} \sqrt{n} > -1.6449$$

$$\Rightarrow n < \left((-1.6449) \left(\frac{5}{-0.9} \right) \right)^2 = 83.504$$



Therefore, the set of values is $\{n \in \mathbb{Z}^+ : n \leq 83\}$, or equivalently, $\{n \in \mathbb{Z} : 0 < n \leq 83\}$.