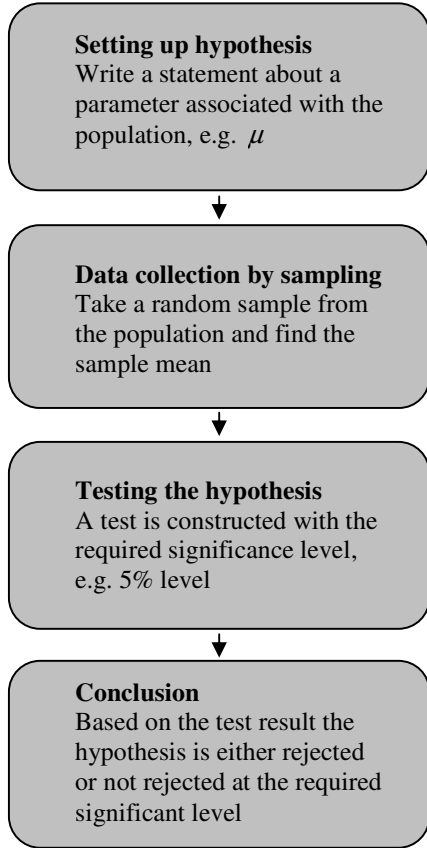




Statistical Testing

Here we concentrate on testing of the population mean.
The sequence in statistical testing:



Setting up hypothesis

Over time a population may change.

A hypothesis: *The population mean remains the same.*

A 'no change' hypothesis is called a null hypothesis.

It is denoted by H_0 .

Alternative hypothesis: *The population mean is less than before.*

The alternative hypothesis is denoted by H_1 .

For example, suppose $\mu = c$ originally,

then $H_0: \mu = c$, $H_1: \mu < c$

Data collection by taking a random sample of size n

The sample mean can be calculated by $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$ if it is

not already known.

The sample mean random variable has a normal distribution (or a distribution approximately normal),

$$\therefore \Pr(\bar{X} \leq \bar{x} | \mu = c) = \Pr\left(Z \leq \frac{\bar{x} - c}{\text{sd}(\bar{X})}\right)$$

If the alternative hypothesis is: *The population mean is greater than before*, then $H_0: \mu = c$, $H_1: \mu > c$

$$\text{and } \Pr(\bar{X} \geq \bar{x} | \mu = c) = \Pr\left(Z \geq \frac{\bar{x} - c}{\text{sd}(\bar{X})}\right)$$

Testing the hypothesis

The value of the probability above is called the p -value of the test,

$$\text{i.e. } p\text{-value} = \Pr\left(Z \leq \frac{\bar{x} - c}{\text{sd}(\bar{X})}\right) \text{ or } p\text{-value} = \Pr\left(Z \geq \frac{\bar{x} - c}{\text{sd}(\bar{X})}\right)$$

It is used to *decide* whether the null hypothesis H_0 is to be rejected or not. The test itself is called a z -test.

By convention,

$p\text{-value} > 0.05$ means insufficient evidence against H_0

$p\text{-value} < 0.05$ means good evidence against H_0

$p\text{-value} < 0.01$ means strong evidence against H_0

$p\text{-value} < 0.001$ means very strong evidence against H_0

Conclusion

If the null hypothesis H_0 is rejected in favour of the alternative hypothesis H_1 when the $p\text{-value} < 0.05$, the condition for rejection, we say H_0 is rejected at the 0.05 level of significance.

0.05 (5%) is the most commonly used value for the significance level.

Other levels like 0.01 (1%) and 0.001 (0.1%) are sometimes used.

One-tail test and two-tail test

For $H_0: \mu = c$, $H_1: \mu < c$ OR $H_0: \mu = c$, $H_1: \mu > c$

a one-tail test is applied to test the hypothesis as discussed above

For $H_0: \mu = c$, $H_1: \mu \neq c$, a two-tail test is applied to test the hypothesis.

The difference between the two tests is in the p -values.

$$p\text{-value (two-tail test)} = 2 \times p\text{-value (one-tail test)}$$

When the $p\text{-value (two-tail test)} < 0.05$, the condition for rejection, we say H_0 is rejected at the 0.05 level of significance.

A 95% confidence interval for the population mean μ , given by

$$\left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right), \text{ is also used to test the hypothesis.}$$

If $\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < c < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$, H_0 is *not* rejected, otherwise it would be rejected at the 5% level of significance.

Types of error in statistical testing

Type I error: H_0 is rejected when H_0 is true

Type II error: H_0 is not rejected when H_0 is not true

	H_0 is true	H_0 is not true
H_0 is rejected	Type I error	Correct decision
H_0 is not rejected	Correct decision	Type II error

Note 1: A type I error is viewed to be more serious than a type II error.

Note 2: For a good statistical test, we would want the probabilities of making type I and type II errors to be small.

Note 3: The probability of making type I error can be reduced by increasing the significance level. However, it will increase the probability of making type II error.

Using 0.05 as the significance level is a compromise.



Example 1 (2016 VCAA SM Sample Exam 1 Q3b, modified)

A coffee machine dispenses volumes of coffee that are normally distributed. The owner has been told that the mean volume is 240 mL. The owner is concerned that the coffee machine is, on average, dispensing less coffee than the mean 240 mL.

A sample of 16 cups of coffee (with no milk) is dispensed and it is found that the mean volume of all coffees served in this sample is 235 mL.

Assume that the population standard deviation of 8 mL.

i. State appropriate null and alternative hypotheses for the volume V in this situation. 1 mark

ii. The p value for this test is given by the expression $\Pr(Z \leq a)$, where Z has the standard normal distribution. Find the value of a and hence determine whether the null hypothesis should be rejected at the 0.05 level of significance. 2 marks

b i Null hypothesis: The second machine is, on average, dispensing **not** less coffee than the first.

OR $H_0: \mu = 240$ mL

Alternative hypothesis: The second machine is, on average, dispensing less coffee than the first.

OR $H_1: \mu < 240$ mL

b ii $\text{sd}(\bar{X}) = \frac{8}{\sqrt{16}} = 2$, $a = \frac{235 - 240}{2} = -2.5$,

$p = \Pr(Z \leq -2.5) \approx 0.0062$

Since $p < 0.05$, the null hypothesis should be rejected at the 0.05 level of significance.

Example 2 (2016 VCAA SM Sample Exam 2 SECTION B Q6)

A certain type of computer, once fully charged, is claimed by the manufacturer to have $\mu = 10$ hours lifetime before a recharge is needed. When checked, a random sample of $n = 25$ such computers is found to have an average lifetime of $x = 9.7$ hours and a standard deviation of $s = 1$ hour. To decide whether the information gained from the sample is consistent with the claim $\mu = 10$, a statistical test is to be carried out. Assume that the distribution of lifetimes is normal and that s is a sufficiently accurate estimate of the population (of lifetimes) standard deviation σ .

a. Write down suitable hypotheses H_0 and H_1 to test whether the mean lifetime is less than that claimed by the manufacturer. 2 marks

b. Find the p value for this test, correct to three decimal places. 2 marks

c. State with a reason whether H_0 should be rejected or not rejected at the 5% level of significance. 1 mark

Let the random variable X denote the mean lifetime of a random sample of 25 computers, assuming $\mu = 10$.

d. Find C^* such that $\Pr(\bar{X} < C^* | \mu = 10) = 0.05$. Give your answer correct to three decimal places. 2 marks

e. i. If the mean lifetime of all computers is in fact $\mu = 9.5$ hours, find $\Pr(\bar{X} > C^* | \mu = 9.5)$, giving your answer correct to three decimal places, where C^* is your answer to part d. 2 marks

e ii. Does the result in part e.i. indicate a type I or type II error? Explain your answer. 1 mark

a H_0 : The mean lifetime is that claimed by the manufacturer.

($\mu = 10$)

H_1 : The mean lifetime is less than that claimed by the manufacturer.

($\mu < 10$)

b $\sigma \approx s = 1$, $\text{sd}(\bar{X}) = \frac{\sigma}{\sqrt{n}} \approx \frac{1}{\sqrt{25}} = 0.2$, $E(\bar{X}) = \mu = 10$

p -value = $\Pr(\bar{X} \leq 9.7 | \mu = 10) = \Pr\left(Z \leq \frac{9.7 - 10}{0.2}\right) \approx 0.067$

c Since p -value > 0.05 , $\therefore H_0$ should not be rejected at the 5% level of significance.

d $\Pr(\bar{X} < C^* | \mu = 10) = 0.05$, $\Pr\left(Z < \frac{C^* - 10}{0.2}\right) = 0.05$

$\therefore \frac{C^* - 10}{0.2} \approx -1.64485$, $C^* \approx 9.671$

e i $\Pr(\bar{X} > 9.671 | \mu = 9.5) = \Pr\left(Z > \frac{9.671 - 9.5}{0.2}\right)$

$\approx \Pr(Z > 0.855) \approx 0.196$

e ii Type II error:

p -value $\approx 0.196 > 0.05$, it supports that the actual mean lifetime is $\mu = 9.5$, $\therefore H_0$ is not true but it is not rejected.