95% confidence interval

A 95% confidence interval is a range of values believed to include an unknown population parameter (μ) .You can be 95% confident that the population mean lies in the range $\bar{x} \mp 1.96 \frac{\sigma}{\sqrt{n}}$

Null Hypothesis H_0

A hypothesis is an assertion about one or more popular parameters. We are interested only in a population mean (μ) or a population proportion (P).In hypothesis testing you will have two hypothesises H_0 the null hypothesis and H_i the alternative hypothesis.

The null hypothesis is the hypothesis you hope to nullify and you must hold it to be true unless you have sufficient evidence to conclude otherwise.

P Value; The observed significance Level

When you perform a hypothesis test in statistics, a *p*-value helps you determine the significance of your results. **Hypothesis tests** are used to test the validity of a claim that is made about a population. This claim that's on trial, in essence, is called the ***null hypothesis***.

The *alternative hypothesis* is the one you would believe if the null hypothesis is concluded to be untrue. The evidence in the trial is your data and the statistics that go along with it. All hypothesis tests ultimately use a *p*-value to weigh the strength of the evidence (what the data are telling you about the population). The *p*-value is a number between 0 and 1 and interpreted in the following way:

- A small *p*-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so you reject the null hypothesis.
- A large *p*-value (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject the null hypothesis.

The **Central Limit Theorem** (**CLT**) is a **statistical** theory states that given a sufficiently large sample size from a population with a finite level of variance, the mean of all samples from the same population will be approximately equal to the mean of the population.

$\bar{x} = \mu$, and the standard deviation of the sample $= \frac{\sigma}{\sqrt{n}}$ (standard error)

www.tuition.ie for free leaving and junior cert maths notes

Inferential statistics Main points ©tuition.ie

Finding a probability from a Population

Example 1; suppose that score in the leaving cert are normally distributed with a mean μ = 500 and a standard deviation σ == 100.Find the probability that a randomly selected student will score at least 750?

Solution we are looking for P(x) > 750 we must now change to z we are looking for

P(z) > $\frac{x-\mu}{\sigma}$ = P(z) > $\frac{750-500}{100} \Rightarrow P(z) > 2.5$ = 1 – P(2.5) = 1 – 0.9938 = 0.0062

Finding a probility from a sample; we use $\frac{x-\mu}{\frac{\sigma}{\sqrt{n}}}$

Example 2: A random sample of size n = 36 is obtained from a population with μ =60 and a standard deviation σ =15 .What is the probability that the sample mean is at most 58.

We are looking for P (x) < 58 ⇒ P (z) <. $\frac{x-\mu}{\frac{\sigma}{\sqrt{n}}} \Rightarrow \frac{58-60}{\frac{15}{\sqrt{36}}} = \frac{-2}{\frac{15}{6}} = -.8$ P (z) < -0.8.

Which is the same as P (z) > 0.8 = 1 – P (0.8) = 1 -0.2119 = 0.7881

Example 3: The scores in the Junior Cert are normally distributed with a mean μ = 500 and a standard deviation σ = 100 .If a random sample of 200 junior certs is selected find the probability that the average score of the sample will be between 500 and 520.

WE are looking for $\frac{500-500}{\frac{100}{\sqrt{200}}}$ < P (z) < $\frac{520-500}{\frac{100}{\sqrt{200}}}$ ⇒ 0 < P (z) <2.828 = P (2.82) – P (0)

= 0.9977 – 0.500 = 0. 4977.

Example 4: 95% confidence interval of a large sample to estimate μ . We use $\bar{x} \mp 1.96\frac{\sigma}{\sqrt{n}}$;

A researcher wants to measure the average amount of fat in grams in a sandwich. A random sample of 40 sandwiches yielded a mean of $\bar{x} = 17, \sigma = 2.5$ .Construct a 95% confidence interval of the average amount of fat in grams in the sandwich.

$\bar{17} \mp 1.96\frac{2.5}{\sqrt{40}} = 17 - 1.96\frac{2.5}{\sqrt{40}} to 17 + 1.96\frac{2.5}{\sqrt{40}}$ = 16.23 to 17.77. The researcher can be 95% confident that the average amount of fat in the sandwich lies between 16.23g and 17.77g.

Using 95% confidence interval to estimate P (the population proportion) $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}}$

We use $\hat{p}$ and $\hat{q}$ where $\hat{p}$ = the proportion of the random sample, $\hat{q} = 1 - \hat{p}$

Example 5: A decorator needs to know the proportion of residents who prefer blue as a colour. A random survey of 300 residents revealed that 240 preferred blue .Construct a 95% confidence interval for the proportion of residents who prefer blue. Also give the margin of error

Given $\hat{p} = 0.8$ (240/300), $\hat{q} = 1 - \hat{p} = 1 - 0.8 = 0.2$ .The 95% confidence interval for the population proportion who prefer blue is $\hat{p} \pm 1.96\sqrt{\frac{\hat{p}\hat{q}}{n}} = \hat{0.8} \pm 1.96\sqrt{\frac{0.8 x 0.2}{300}} = 0.7547 \text{ to } 0.8453$

Margin of error $1.96\sqrt{\frac{0.8 x 0.2}{300}} = 0.0453$

Hypothesis testing

Example 6: A battery manufacturer claims that the average life of its batteries is 45 months. A random sample of 80 batteries has a mean life $\bar{x} = 39, \sigma = 8$ months Does the data dispute the manufacturers claim?

The null hypothesis $H_0 = 45, H_1 \neq 45$. Find the Test statistic $\frac{x-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{39-45}{\frac{8}{\sqrt{80}}} = -3\sqrt{5} = -6.71$

Since – 6.71 does not lie in the interval $-1.96 < z < 1.96$ we reject $H_0 = 45$, and reject the manufacturers claim

Example 7: A health professional claims that the average calorie intake of the citizens of Dublin is 2300 calories per day .A Survey of 500 people found their mean intake per day was 2400 calories with a standard deviation of 930 calories. Use a hypothesis test to see if the claim is true .There are 3 ways to do this. $H_0 = 2300, H_1 \neq 2300$

Method (i)) Find the test statistic $\frac{x-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{2400-2300}{\frac{930}{\sqrt{500}}} = 2.404$ since 2.404 does not lie in the interval $-1.96 < z < 1.96$ we reject $H_0 = 2300$.

Method (ii) Find the 95% confidence interval for the mean of the population

$$2300 \mp 1.96\frac{930}{\sqrt{500}} = 2218.48 \text{ to } 2381.5$$

Since 2400 does not lie in the 95% confidence interval we reject H_0

Method (iii) P Value.

Calculate the P value as follows 2(1-Probility of test statistic) =2 (1- P (2.404) = 2(1-0.9927) =0.0146

Since The P Value 0.0146 < 0.05 we reject H_0."If the P is low the null must go".

Example 8: A radio station manager claims that they have 40% listeners in a local area. A survey of 100 listeners revealed 34 listened Perform a hypothesis test at the 5% level to test the manager's claim.

$H_0 = 0.4, H_1 \neq 0.4 \quad \hat{p} = 0.34, \hat{q} = 1 - \hat{p} = 1 - 0.34 = 0.66 . p_o = 0.4, q_o = 0.6$

The test statistic is $\dfrac{\hat{p} - p_o}{\sqrt{\dfrac{p_o q_0}{n}}} = \dfrac{0.34 - 0.40}{\sqrt{\dfrac{(0.4)(0.6)}{100}}} = -1.225$

Since -1.225 lies in the interval $-1.96 < z < 1.96$ we fail to reject H_0 and accept the manager's claim. Again there are two more ways to answer this question

Method 2;

95% confidence interval $\hat{p} \pm 1.96\sqrt{\dfrac{\hat{p}\hat{q}}{n}} = 0.34 \pm 1.96\sqrt{\dfrac{0.34 x 0.66}{100}} = 0.2472 to 0..4328$

Since 0.40 lies in the 95% confidence interval we fail to reject H_0.

Method 3; Find the P value =2(1 – 0.8907) = 0.2186 since 0.2186 > .05 we fail to reject H_0.

"If the P is high the null must fly"

2018 Q8 Paper2 (i) Find the maximum value of p(1-p). Given

$y = p - p^2, \dfrac{dy}{dp} = 1 - 2p = 0 \Rightarrow p = \dfrac{1}{2} \Rightarrow \max = 0.25$

Hence find the largest possible value of the radius of the 95% confidence interval =

$1.96\sqrt{\dfrac{.5x.5}{800}} = 0.346$.

To calculate the correct sample size for a given margin of error E.

$1.96\sqrt{\dfrac{\hat{p}\hat{q}}{n}} = E \Rightarrow (1.96)^2 \dfrac{\hat{p}\hat{q}}{n} = E^2 \Rightarrow (1.96)^2 \dfrac{\hat{p}\hat{q}}{E^2} = n$ Or $1.96\dfrac{\sigma}{\sqrt{n}} = E \Rightarrow (1.96)^2 \dfrac{\sigma^2}{E^2} = n$

Definitions