

Q1: What is the difference between a t-test and a z-test? Provide an example scenario where you would use each type of test.

Q1: What is the difference between a t-test and a z-test? Provide an example scenario where you would use each type of test. A z-test is used to test a Null Hypothesis if the population variance is known, or if the sample size is larger than 30, for an unknown population variance. A t-test is used when the sample size is less than 30 and the population variance is unknown.

Z-test is the statistical hypothesis used to determine whether the two samples' means calculated are different if the standard deviation is available and the sample is large. In contrast, the T-test determines how averages of different data sets differ in case the standard deviation or the variance is unknown.

If the population standard deviation is known and the sample size is greater than 30, Z-test is recommended to be used. If the population standard deviation is known, and the size of the sample is less than or equal to 30, T-test is recommended. If the population standard deviation is unknown, T-test is recommended

Q2: Differentiate between one-tailed and two-tailed tests.

One-tailed tests allow for the possibility of an effect in one direction. Two-tailed tests test for the possibility of an effect in two directions—positive and negative.

A statistical hypothesis test in which alternative hypothesis has only one end, is known as one tailed test. A significance test in which alternative hypothesis has two ends, is called two-tailed test.

Q3: Explain the concept of Type 1 and Type 2 errors in hypothesis testing. Provide an example scenario for each type of error.

Type 1 error (FP) :- If , Jury sent him jail but in real he is not done crime. Type 2 error (FN) :- If, Jury not sent him jail but in real he is done the crime.

There are two errors that could potentially occur: Type I error (false positive): the test result says you have coronavirus, but you actually don't. Type II error (false negative): the test result says you don't have coronavirus, but you actually do.

Q4: Explain Bayes's theorem with an example.

Bayes theorem gives the probability of an "event" with the given information on "tests". There is a difference between "events" and "tests". For example there is a test for liver disease, which is different from actually having the liver disease, i.e. an event. Rare events might be having a higher false positive rate.

Bayes theorem is also known as the formula for the probability of "causes". For example: if we have to calculate the probability of taking a blue ball from the second bag out of three different bags of balls, where each bag contains three different colour balls viz. red, blue, black

Q5: What is a confidence interval? How to calculate the confidence interval, explain with an example.

A confidence interval is the mean of our estimate plus and minus the variation in that estimate. This is the range of values we expect our estimate to fall between if we redo our test, within a certain level of confidence. Confidence, in statistics, is another way to describe probability.

Once the standard error is calculated, the confidence interval is determined by multiplying the standard error by a constant that reflects the level of significance desired, based on the normal distribution. The constant for 95 percent confidence intervals is 1.96.

For example, in 2005 the statewide estimated percentage of adults currently smoking was 20.7%. The 95%

Q6. Use Bayes' Theorem to calculate the probability of an event occurring given prior knowledge of the event's probability and new evidence. Provide a sample problem and solution.

Lets assume there are 5 balls in bucket, 3-yello, 2-red

$$\Pr(Y) = 3/5$$

$$\Pr(R) = 2/4 = 1/2$$

$$\Pr(R \text{ and } Y) = \Pr(R) * \Pr(Y/R) = 1/2 * 3/5 = 3/10$$

$$\Pr(A \& B) = \Pr(B \& A)$$

$$\Pr(A) * \Pr(B/A) = \Pr(B) * \Pr(A/B)$$

$$\Pr(B/A) = [\Pr(B) * \Pr(A/B)] / \Pr(A) \quad \text{.....Bayes' Theorem}$$

$$\Pr(A/B) \quad \text{.....Conditional Probability}$$

Q7. Calculate the 95% confidence interval for a sample of data with a mean of 50 and a standard deviation of 5. Interpret the results.

$$s_mean = 50$$

$$s_std = 5$$

To find the critical value, we can look it up in a standard normal distribution table or use a calculator. For a 95% confidence level, the critical value is 1.96.

$$CI = 50 \pm 1.96 * (5/\sqrt{n})$$

$$CI = 50 \pm 1.96 * (5/\sqrt{100}) = 50 \pm 0.98$$

Therefore, the 95% confidence interval would be (49.02, 50.98). This means that we can be 95% confident that the true population mean falls between 49.02 and 50.98. If we took repeated samples and calculated the confidence interval for each sample, 95% of the intervals would contain the true population mean.

Q8. What is the margin of error in a confidence interval? How does sample size affect the margin of error? Provide an example of a scenario where a larger sample size would result in a smaller margin of error.

The margin of error in a confidence interval is the measure of the precision of the estimate. It is the amount of error that we allow for in the confidence interval and is usually expressed as a percentage. The margin of error depends on the level of confidence desired, the sample size, and the standard deviation of the population. As the level of confidence increases, the margin of error also increases. Conversely, as the sample size increases, the margin of error decreases.

Sample size affects the margin of error because a larger sample size results in a smaller margin of error. This is because it reduces the amount of sampling error, which is the error that occurs due to the sample not being a perfectly representative subset of the population. As the sample size increases, the sample statistic (such as the sample mean) is likely to be closer to the true population parameter, resulting in a smaller margin of error.

For example, let's say we want to estimate the average height of all adults in a city. We take a sample of 100 people and calculate the sample mean height to be 170 cm with a standard deviation of 10 cm. If we want to construct a 95% confidence interval for the population mean height, we would use the formula Margin of

Q9. Calculate the z-score for a data point with a value of 75, a population mean of 70, and a population standard deviation of 5. Interpret the results.

$$z = (x - \mu) / \sigma$$

$$z = (75 - 70) / 5$$

$$z = 1$$

Q10. In a study of the effectiveness of a new weight loss drug, a sample of 50 participants lost an average of 6 pounds with a standard deviation of 2.5 pounds. Conduct a hypothesis test to determine if the drug is significantly effective at a 95% confidence level using a t-test.

Null hypothesis: The weight loss drug is not significantly effective, and the true population mean weight loss is equal to or less than 0 pounds.

Alternative hypothesis: The weight loss drug is significantly effective, and the true population mean weight loss is greater than 0 pounds.

$$t = (\bar{x} - \mu) / (s / \sqrt{n})$$

$$t = (6 - 0) / (2.5 / \sqrt{50})$$

$$t = 15.48$$

The critical value for a one-tailed t-test with 49 degrees of freedom and a significance level of 0.05 is 1.676. Since our test statistic of 15.48 is greater than the critical value of 1.676, we reject the null hypothesis.

Q11. In a survey of 500 people, 65% reported being satisfied with their current job. Calculate the 95% confidence interval for the true proportion of people who are satisfied with their job.

confidence interval = sample population \pm z-score * standard error

First, we can calculate the sample population using the formula: sample population = number of people satisfied / sample size = 0.65

standard error = $\sqrt{(\text{sample population} * (1 - \text{sample population})) / \text{sample size}}$ = $\sqrt{(0.65 * 0.35) / 500}$ = 0.032

confidence interval = $0.65 \pm 1.96 * 0.032 = (0.588, 0.712)$

Q12. A researcher is testing the effectiveness of two different teaching methods on student performance. Sample A has a mean score of 85 with a standard deviation of 6, while sample B has a mean score of 82 with a standard deviation of 5. Conduct a hypothesis test to determine if the two

teaching methods have a significant difference in student performance using a t-test with a significance level of 0.01.

meanA = 85 sA = 6 nA = ? meanB = 82 sB = 5 nB = ? $\alpha = 0.01$ Degrees of freedom = (nA + nB - 2)

Unfortunately, we cannot conduct the hypothesis test because the sample sizes of the two groups are not given. The t-test formula requires the sample sizes to calculate the degrees of freedom and the t-value.

Q13. A population has a mean of 60 and a standard deviation of 8. A sample of 50 observations has a mean of 65. Calculate the 90% confidence interval for the true population mean.

p_mean = 60 p_std = 8 n = 50 s_mean = 65

confidence interval = sample mean \pm z-score * (standard deviation / sqrt(sample size))

confidence interval = $65 \pm 1.645 * (8 / \sqrt{50})$

confidence interval = (62.18, 67.82)

Q14. In a study of the effects of caffeine on reaction time, a sample of 30 participants had an average reaction time of 0.25 seconds with a standard deviation of 0.05 seconds. Conduct a hypothesis test to determine if the caffeine has a significant effect on reaction time at a 90% confidence level using a t-test.

$t = (\text{sample mean} - \text{hypothesized mean}) / (\text{sample standard deviation} / \sqrt{\text{sample size}})$

sample mean = 0.25

hypothesized mean = 0

sample standard deviation = 0.05

sample size = 30

$\alpha = 0.1$

Degrees of freedom = (sample size - 1)

he t-value can be calculated using the t-distribution table or a statistical software, which is approximately 5.06 for a one-tailed test with 29 degrees of freedom and a significance level of 0.1.

confidence interval = sample mean \pm t-value * (sample standard deviation / sqrt(sample size))

confidence interval = $0.25 \pm 5.06 * (0.05 / \sqrt{30})$

confidence interval = (0.21, 0.29)