

Statistics Advance Assignment - 4

January 24, 2024

0.1 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 t-test and z-test are statistical tests used to compare the means of two groups or samples. However, a t-test is used when the sample size is small, and the population standard deviation is unknown, while a z-test is used when the sample size is large, and the population standard deviation is known.

For example, if a researcher wants to compare the mean height of two groups of people, where one group is small ($n=20$) and the other group is large ($n=2000$), the researcher would use a t-test for the smaller group and a z-test for the larger group.

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

A one-tailed test is a statistical test that tests the hypothesis in only one direction. It is used when the researcher has a specific hypothesis about the direction of the effect. For example, a researcher may want to test whether a new drug is better than an existing drug, so the one-tailed test would be used to test if the new drug is significantly better.

A two-tailed test is a statistical test that tests the hypothesis in both directions. It is used when the researcher does not have a specific hypothesis about the direction of the effect. For example, a researcher may want to test whether a new drug has any effect, so the two-tailed test would be used to test if the new drug has a significant effect.

0.3 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 is when a researcher rejects a null hypothesis when it is actually true. It means that the researcher finds a significant effect when there is none. For example, a researcher may conclude that a new drug is effective when it is actually not, resulting in patients being given unnecessary treatment.

Type 2 error is when a researcher accepts a null hypothesis when it is actually false. It means that the researcher fails to find a significant effect when there is one. For example, a researcher may conclude that a new drug is not effective when it is actually effective, resulting in patients missing out on a potentially beneficial treatment.

0.4 Q4: Explain Bayes's theorem with an example.

Bayes's theorem is a mathematical formula used to calculate the probability of an event occurring given prior knowledge of the event's probability and new evidence. It is used to update the prior

probability based on new information.

For example, a doctor may use Bayes's theorem to calculate the probability of a patient having a disease based on the patient's symptoms and test results. The doctor would start with the prior probability of the patient having the disease based on the prevalence of the disease in the population. The doctor would then update the probability based on the patient's symptoms and test results to get a more accurate estimate of the probability of the patient having the disease.

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

A confidence interval is a range of values that is likely to contain the true value of a population parameter with a certain level of confidence. It is calculated using the sample statistics and the standard error.

For example, if a researcher wants to estimate the average height of a population with a 95% confidence interval, the researcher would take a sample of the population and calculate the mean and standard deviation. The confidence interval would then be calculated by adding and subtracting the margin of error from the mean, where the margin of error is calculated using the standard error and the critical value for the confidence level.

0.6 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.

Bayes' Theorem is a formula used to calculate the probability of an event occurring based on prior knowledge and new evidence. It is expressed as:

$$P(A|B) = P(B|A) \times P(A) / P(B)$$

where $P(A|B)$ is the probability of A given B, $P(B|A)$ is the probability of B given A, $P(A)$ is the prior probability of A, and $P(B)$ is the probability of B.

For example, let's say that a certain disease affects 1 in 1000 people, and a new test has been developed that correctly identifies the disease in 99% of cases. If a person tests positive for the disease, what is the probability that they actually have the disease?

Solution: $P(\text{Disease}) = 0.001$ (prior probability) $P(\text{Positive Test} | \text{Disease}) = 0.99$ (probability of testing positive given the person has the disease) $P(\text{Positive Test} | \text{No Disease}) = 0.01$ (probability of testing positive given the person does not have the disease) $P(\text{No Disease}) = 0.999$ (1 - prior probability)

Using Bayes' Theorem:

$$P(\text{Disease} | \text{Positive Test}) = P(\text{Positive Test} | \text{Disease}) \times P(\text{Disease}) / (P(\text{Positive Test} | \text{Disease}) \times P(\text{Disease}) + P(\text{Positive Test} | \text{No Disease}) \times P(\text{No Disease}))$$
$$P(\text{Disease} | \text{Positive Test}) = 0.99 \times 0.001 / ((0.99 \times 0.001) + (0.01 \times 0.999))$$

$P(\text{Disease} | \text{Positive Test}) = 0.09$ or 9%

So, even with a highly accurate test, the probability of a positive result indicating the presence of the disease is still relatively low.

0.7 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.

A confidence interval is a range of values that is likely to contain the true population parameter with a certain level of confidence. The formula to calculate a 95% confidence interval is:

$$CI = \bar{x} \pm z(\alpha/2) * (s / \sqrt{n})$$

where \bar{x} is the sample mean, $z(\alpha/2)$ is the z-score associated with a 95% confidence level (1.96), s is the population standard deviation, and n is the sample size.

Plugging in the values from the problem:

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

To calculate the interval, we need to know the sample size. Let's assume it's $n = 100$:

$$CI = 50 \pm 1.96 * (5 / \sqrt{100}) \quad CI = 50 \pm 0.98 \quad CI = (49.02, 50.98)$$

Interpretation: With a 95% level of confidence, the true population mean is likely to fall between 49.02 and 50.98.

0.8 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 range of values above and below the point estimate that is likely to contain the true population parameter with a certain level of confidence. Sample size affects the margin of error inversely; as the sample size increases, the margin of error decreases. For example, a survey of 100 people may have a margin of error of $\pm 10\%$, whereas a survey of 1000 people may have a margin of error of $\pm 3\%$.

0.9 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.

The z-score is calculated as $(75-70)/5 = 1$. The z-score of 1 indicates that the data point is one standard deviation above the population mean. This means that the data point is relatively far from the mean and is considered unusual.

0.10 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.

The null hypothesis is that the mean weight loss for the population is equal to 0. The alternative hypothesis is that the mean weight loss for the population is greater than 0. With a sample size of 50 and a sample mean of 6 pounds, the t-value is calculated as $(6-0)/(2.5/\sqrt{50}) = 12$. The critical t-value at a 95% confidence level with 49 degrees of freedom is 1.677. Since the calculated t-value is greater than the critical t-value, we reject the null hypothesis and conclude that the drug is significantly effective at a 95% confidence level.

0.11 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.

The standard error of the proportion is calculated as $\sqrt{(0.65(1-0.65))/500} = 0.025$. The 95% confidence interval is calculated as $0.65 \pm 1.96(0.025)$, which results in the interval of (0.601, 0.699). Therefore, we can say with 95% confidence that the true proportion of people who are satisfied with their job is between 60.1% and 69.9%.

0.12 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.

To 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, we can use the following formula:

$$t = (\bar{x}_1 - \bar{x}_2) / \sqrt{(s_1^2/n_1 + s_2^2/n_2)}$$

where $\bar{x}_1 = 85$, $s_1 = 6$, $n_1 = ?$, $\bar{x}_2 = 82$, $s_2 = 5$, $n_2 = ?$, and $\alpha = 0.01$.

To find the degrees of freedom (df), we use the following formula:

$$df = (s_1^2/n_1 + s_2^2/n_2)^2 / [(s_1^2/n_1)^2/(n_1-1) + (s_2^2/n_2)^2/(n_2-1)]$$

Once we have calculated t and df, we can use a t-distribution table to find the critical value of t for $\alpha = 0.01$ and df.

If our calculated t-value is greater than the critical value of t, we reject the null hypothesis that there is no significant difference between the two teaching methods.

0.13 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.

To calculate the 90% confidence interval for the true population mean, we can use the following formula:

$$CI = \bar{X} \pm Z_{\alpha/2} * (s / \sqrt{n})$$

where $\bar{X} = 65$, $s = 8$, $n = 50$, and $\alpha = 0.10$ (since we want a 90% confidence interval).

Using a standard normal distribution table, we can find $Z_{\alpha/2}$ for $\alpha = 0.10$.

Once we have calculated CI, we can say with 90% confidence that the true population mean falls within this interval.

0.14 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.

To conduct a hypothesis test to determine if caffeine has a significant effect on reaction time at a 90% confidence level using a t-test, we can use the following formula:

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

where $\bar{x} = 0.25$ seconds, $\mu = ?$, $s = 0.05$ seconds, $n = 30$, and $\alpha = 0.10$ (since we want a 90% confidence level).

To find the degrees of freedom (df), we use the following formula:

$$df = n - 1$$

Once we have calculated t and df , we can use a t-distribution table to find the critical value of t for $\alpha = 0.10$ and df .

If our calculated t -value is greater than the critical value of t , we reject the null hypothesis that caffeine has no significant effect on reaction time at a 90% confidence level.