

# 11 March Assignment

May 14, 2023

[ ]: 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.

ANS -

[ ]: The z test uses a normal distribution while the t test uses the Students t distribution. Degrees of freedom are needed in t testing, not in z testing. The z statistic is calculated with the standard error. The t statistic uses the estimated standard error. The z test is used for testing proportions when  $np > 10$  and  $n(1 - p) > 10$ .

The main difference between a t-test and a z-test is that 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 you want to compare the average height of two groups of people (say, men and women), you can use a t-test if you have a small sample size (less than 30) and don't know the population standard deviation. On the other hand, if you have a large sample size (more than 30) and know the population standard deviation, you can use a z-test<sup>2</sup>.

[ ]:

[ ]:

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

ANS -

[ ]: One and Two-Tailed Tests are ways to identify the relationship between the statistical variables. For checking the relationship between variables in a single direction (Left or Right direction), we use a one-tailed test. A two-tailed test is used for checking whether the relations between variables are in any direction or not.

One-Tailed Test :

A one-tailed test **is** based on a uni-directional hypothesis **↪** where the area of rejection **is** on only one side of the sampling distribution. It determines whether a particular population parameter **↪ is** larger **or** smaller than the predefined parameter. It uses one single critical value to test the data.

[ ]: Two-Tailed Test :

A two-tailed test **is** also called a nondirectional **↪** hypothesis. For checking whether the sample **is** greater **or** less than a **range** of values, we use the two-tailed. It **is** used **for** null hypothesis **↪** testing.

[ ]:

[ ]:

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

ANS -

[ ]: In hypothesis testing, a Type 1 error occurs when the null **↪** hypothesis **is** rejected even though it **is** true. A Type 2 error occurs when the null hypothesis **is not** rejected even though it **is** false.

For example, let's say a pharmaceutical company **is** testing a new drug that **↪** they claim can cure a disease **with** 95% confidence. The null hypothesis would be that the drug does **not** cure the disease. If the **↪** company rejects the null hypothesis **and** claims that the drug cures the disease when it actually does **not**, this would be a Type 1 **↪** error. On the other hand, **if** the company fails to reject the null hypothesis **and** claims that the drug does **not** cure the disease **↪** when it actually does, this would be a Type 2 error.

[ ]:

[ ]:

[ ]: Q4: Explain Bayess theorem **with** an example.

ANS -

[ ]: Bayes' theorem describes the probability of occurrence of an **↪** event related to **any** condition. It **is** also considered

for the case of conditional probability. 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. In  
→ this case, the probability of occurrence of an event  
is calculated depending on other conditions is known as conditional probability.

[ ]:

[ ]:

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

ANS -

[ ]: A confidence interval is a range of values that we can be confident  
→ contains the true population parameter with a certain  
degree of certainty<sup>1</sup>. It is calculated using a formula that takes into account  
→ the sample size, sample mean, and standard deviation.

The formula for calculating a confidence interval is as follows:

$$\text{Confidence interval (CI)} = \bar{X} \pm Z \left( S \div \sqrt{n} \right)$$

Where:

- X represents the sample mean
- Z represents the Z-value you get from the normal standard distribution
- S is the population standard deviation
- n represents the sample size you're surveying

example of how to calculate a confidence interval:

1. Find the sample mean by adding up all the values in your data set and  
→ dividing the sum by the number of values in the sample.
2. Calculate the standard deviation of your sample.
3. Determine your desired level of confidence (e.g., 95%).
4. Look up the corresponding Z-value for your desired level of confidence in a  
→ Z-table.
5. Plug in your values into the formula above.

[ ]:

[ ]:

[ ]: Q6. Use Bayes Theorem to calculate the probability of an event occurring given  
→ prior knowledge of the events probability and new evidence. Provide a sample problem and solution.

ANS -

[ ]: Bayes' Theorem is a mathematical formula for determining  
→ conditional probability. It describes the probability of an event based on prior knowledge of conditions that might be related to the event. The  
→ formula is named after Thomas Bayes, an 18th-century British mathematician.

Here's a sample problem and solution using Bayes' Theorem:

Suppose there are two factories that produce light bulbs: Factory A and Factory  
→ B. Factory A produces 60% of all light bulbs and  
Factory B produces 40%. Of the light bulbs produced by Factory A, 5% are  
→ defective, while 10% of those produced by Factory B are defective. If a light bulb is chosen at random and found to be defective, what  
→ is the probability that it was produced by Factory A?

Let D be the event that a bulb is defective and A be the event that it was  
→ produced by Factory A. We want to find  $P(A|D)$ , which is the probability that a bulb was produced by Factory A given that it was found  
→ to be defective. Using Bayes' Theorem, we have:

$$P(A|D) = P(D|A) * P(A) / P(D)$$

where  $P(D|A) = 0.05$  (the probability that a bulb produced by Factory A is  
→ defective)  $P(A) = 0.6$  (the probability that a bulb was produced by Factory A)  $P(D) = P(D|A) * P(A) + P(D|B) * P(B) = 0.05 * 0.6 + 0.1$   
→  $* 0.4 = 0.07$

Therefore,

$$P(A|D) = (0.05 * 0.6) / 0.07 = 0.43$$

So there's a 43% chance that the defective bulb was produced by Factory A.

[ ]:

[ ]:

[ ]: 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.

ANS -

[ ]: The formula for calculating a 95% confidence interval for a sample of data with a mean of 50 and a standard deviation of 5 is:

[ ]:  $95\% \text{ confidence interval} = \text{mean} \pm (1.96 * (\text{standard deviation} / \sqrt{\text{sample size}}))$

[ ]:  $50 \pm (1.96 * (5 / \sqrt{\text{sample size}}))$

[ ]: The interpretation of the results is that if we were to take many samples of the same size from this population and calculate a 95% confidence interval for each sample, then approximately 95% of those 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.

ANS -

[ ]: The margin of error (MOE) is a measure of the accuracy of an estimate based on a sample of data. It is defined as the range of values above and below the sample statistic in a confidence interval (CI) that is expected to contain the population parameter with a certain degree of confidence. The MOE decreases as the sample size increases because larger samples provide more information about the population and reduce sampling error.

For example, suppose we want to estimate the proportion of people who support a particular political candidate in an election.

We take a random sample of 500 voters and find that 60% of them support the candidate. We construct a 95% confidence interval for the true proportion using the formula  $CI = \bar{x} \pm z^* (s / \sqrt{n})$ , where  $\bar{x}$  is the sample mean,  $s$  is the sample standard deviation,  $n$  is the sample size, and  $z^*$  is the z-score corresponding to the desired level of confidence (1.96 for 95%).

The MOE for this interval is given by  $z^* (s / \sqrt{n})$ , which equals 0.04 or 4%. If we had taken a larger sample of 1000 voters instead, then our MOE would have been smaller at 0.03 or 3%.

[ ]:

[ ]:

[ ]: 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.

ANS -

[ ]: The formula for calculating a z-score is  $z = (x - \mu) / \sigma$ , where x is the raw score,  $\mu$  is the population mean, and  $\sigma$  is the population standard deviation<sup>1</sup>.

Using this formula, we can 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 as follows:

[13]:  $z = (75 - 70) / 5$

[13]: 1.0

[ ]: The z-score of this data point is 1.

Interpreting the results: A z-score of 1 indicates that this data point is one standard deviation above the mean.

[ ]:

[ ]:

[ ]: 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.

ANS -

[ ]: To conduct a hypothesis test to determine if the drug is significantly effective at a 95% confidence level using a t-test, we can use the following steps:

State the null hypothesis  $H_0$  and alternative hypothesis  $H_a$ .

$H_0$ : The drug is not significantly effective.

$H_a$ : The drug is significantly effective.

Determine the level of significance  $\alpha = 0.05$ .

Calculate the test statistic  $t$  using the formula:

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

where  $\bar{x}$  is the sample mean (6 pounds),  $\mu$  is the population mean (unknown),  $s$  is the sample standard deviation (2.5 pounds), and  $n$  is the sample size (50).

Determine the degrees of freedom  $df = n - 1 = 49$ .

Determine the critical value  $t/2$  with  $df = 49$  and  $\alpha = 0.05/2 = 0.025$  using a  $t$ -table or calculator.

Compare the calculated test statistic  $t$  with the critical value  $t/2$ .

If  $|t| > t/2$ , reject  $H_0$  and conclude that  $H_a$  is true.

If  $|t| \leq t/2$ , fail to reject  $H_0$ .

Interpret the result in context.

Using this method, we can calculate that  $t = (6 - \mu) / (2.5 / \sqrt{50}) = 8.94431$ . With  $df = 49$  and  $\alpha = 0.025$ , we can find that  $t/2 = \pm 2.00962$ . Since  $|t| > t/2$ , we reject  $H_0$  and conclude that  $H_a$  is true at a 95% confidence level. Therefore, we can say that there is evidence that the drug is significantly effective in weight loss.

[ ]:

[ ]:

[ ]: 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.

ANS -

[ ]: To calculate the 95% confidence interval for the true proportion of people who are satisfied with their job, we can use the following formula<sup>1</sup>:

$$\text{Confidence Interval} = p \pm z \sqrt{p(1-p) / n}$$

where:

- $p$ : sample proportion
- $z$ : the chosen  $z$ -value
- $n$ : sample size

In this case, we have a sample size of 500 and a sample proportion of 0.65 (65%). To find the z-value for a 95% confidence interval, we can use a standard normal distribution table or calculator and find that  $z = 1.96$ .

Confidence Interval =  $0.65 \pm 1.96 \sqrt{0.65(1-0.65) / 500}$

Confidence Interval = [0.60, 0.70]

Therefore, we can say with 95% confidence that the true proportion of people who are satisfied with their job is between 60% and 70%.

[ ]:

[ ]:

[ ]: 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.

[ ]: ANS -

[ ]: 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 a two-sample t-test<sup>1</sup>.

The null hypothesis for a two-sample t-test is that the means of the two populations are equal, while the alternative hypothesis can be either two-tailed, left-tailed, or right-tailed<sup>1</sup>.

In this case, we can use a two-tailed alternative hypothesis since we want to determine if there is a significant difference between the two teaching methods<sup>1</sup>.

The formula for the two-sample t-test (a.k.a. the Student's t-test) is shown below<sup>1</sup>:

$$t = (x_1 - x_2) / \sqrt{s_1^2/n_1 + s_2^2/n_2}$$

where

- t is the t value



- $x_1$  and  $x_2$  are the means of the two groups being compared
- $s_1$  and  $s_2$  are the standard deviations of each group
- $n_1$  and  $n_2$  are the sample sizes

Using this formula, we can calculate the t value for our samples A and B as follows:

$$t = (85 - 82) / \sqrt{6^2/100 + 5^2/100} = 3 / \sqrt{0.36 + 0.25} = 3 / \sqrt{0.61} \approx 3 / 0.78 \approx 3.85$$

The degrees of freedom for a two-sample t-test with equal variances can be calculated as follows:

$$df = n_1 + n_2 - 2 = 100 + 100 - 2 = 198$$

Using a t-table or calculator with a significance level of 0.01 and degrees of freedom of 198, we can find that the critical t-value is approximately  $\pm 2.58$ . Since our calculated t-value of 3.85 is greater than this critical value, we can reject the null hypothesis and conclude that there is a significant difference between the two teaching methods at a significance level of 0.01.

Therefore, we can conclude that there is evidence to suggest that there is a significant difference in student performance between sample A and sample B using these two teaching methods.

[ ]:

[ ]:

[ ]: 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.

ANS -

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

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

where:

CI is the confidence interval

$\bar{X}$  is the sample mean

$Z/2$  is the z-score for a 90% confidence level (1.645)

$s$  is the population standard deviation

n is the sample size  
Plugging in the values given in your question, we get:

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

$$CI = [63.02, 66.98]$$

Therefore, we can say with 90% confidence that the true population mean lies between 63.02 and 66.98.

[ ]:

[ ]:

[ ]: 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.

ANS -

[ ]: 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 need to follow these steps:

State the null hypothesis  $H_0$  and alternative hypothesis  $H_a$ .

Determine the level of significance .

Calculate the test statistic t using the formula:  $t = (\bar{x} - \mu) / (s / \sqrt{n})$  where  $\bar{x}$  is the sample mean,  $\mu$  is the population mean under

$H_0$ , s is the sample standard deviation, and n is the sample size.

Determine the critical value  $t_c$  using a t-distribution table with  $n - 1$  degrees of freedom and  $\alpha/2$  significance level.

Compare t and  $t_c$  values and make a decision.

In this case, we have:

$H_0$ : Caffeine has no significant effect on reaction time

$H_a$ : Caffeine has a significant effect on reaction time  
= 0.1 (90% confidence level)

$\bar{x}$  = 0.25 seconds

= ?

s = 0.05 seconds

n = 30

We don't know what  $\mu$  is, so we assume it's equal to  $\bar{x}$  under  $H_0$ .

$$t = (\bar{x} - \mu) / (s / \sqrt{n}) = (0.25 - 0) / (0.05 / \sqrt{30}) \approx 8.66$$

The critical value  $t_c$  with 29 degrees of freedom and  $\alpha/2 = 0.05$  is approximately  $\pm 1.699$ .

Since  $|t| > t_c$ , we reject  $H_0$  and conclude that caffeine has a significant effect on reaction time at a 90% confidence level.