BT307 Quiz ■

Total Marks: 10 Time: 45 min

Name: **Roll No.:**

Q1. A factory produces light bulbs, and the lifespan of these bulbs is known to be normally distributed, with a mean of 800 h and a standard deviation of 100 h. We repeatedly sample 25 bulbs randomly and calculate the sample means. What will be the standard deviation of the sample mean?

Ans 1:

Given, $\mu = 800$ h and $\sigma = 100$ h. The population follows Normal distribution. Sample size $n = 25$

Following the Central Limit Theorem,

the std. dev of the sample mean, $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{100}{\sqrt{25}} = 20$ h

Q2. The heights of adult males in a certain country are normally distributed with a mean of 70 inch and a standard deviation of 3 inch. Estimate the range of heights that encompasses approximately 68% of adult males. Mention the range's lower and upper values.

Ans 2:

Given, $\mu = 70$ inch and $\sigma = 3$ inch. The population follows Normal distribution.

Using the 68-95-99.7 rule for normal distribution, we know that approximately 68% of population will be in the range $[\mu - \sigma, \mu + \sigma]$.

Therefore,

the lower value of the range = $\mu - \sigma = 70 - 3 = 67$ inch

the upper value of the range = $\mu + \sigma = 70 + 3 = 73$ inch

Q3. A researcher is testing a new drug that is believed to lower blood pressure. The null hypothesis (H_0) states that the drug has no effect on blood pressure, while the alternative hypothesis (H_1) states that the drug affects blood pressure. If the researcher sets the significance level (α) at 0.01, what is the probability of making a Type I error?

Ans 3:

As the chosen significance level (α) is 0.01, the probability of making a Type I error is 0.01.

Q4. We are creating a scatter plot using the following R code. Edit the code to add a trend line overlayed on the scatter plot. Do not add any confidence interval around the trend line. Highlight/mark the edited section of your code so that it can be recognized easily.

```
library(ggplot2)
sales_data <- read.csv("sales_data.csv", header = TRUE)
plot_sales <- ggplot(sales_data, aes(x = Spend, y = Revenue)) +
                geom_point(size = 3, alpha = 0.7) +
                labs(x = "Spend (in $)",
                     y = "Revenue (in $)")
plot_sales
```

Ans 4:

```
library(ggplot2)
sales_data <- read.csv("sales_data.csv", header = TRUE)
plot_sales <- ggplot(sales_data, aes(x = Spend, y = Revenue)) +
                geom_point(size = 3, alpha = 0.7) +
                geom_smooth(se = FALSE, span = 5) +
                labs(x = "Spend (in $)",
                     y = "Revenue (in $)")
plot_sales
```

Note: You should write the whole code so that one can understand where you are adding the additional script. As the question asks, you should highlight (use pencil/colour pen/underline etc) to show your edits.

● **BT307 Quiz**

Total Marks: 10 Time: 45 min

Roll No. **Name:**

Q1. A t-test is performed using the following R script:

```
t.test(h.data$Treated,h.data$Untreated, var.equal =FALSE)
```

What are the assumptions made for this test?

Ans 1:

The assumptions are:

1) The population distribution of both groups is Normal distribution.

2) The population variances of the groups are not the same.

Note: The first assumption is a key assumption for t-test in general. The second assumption is evident from the script (`var.equal =FALSE`)

Q2. Scores for five students are given in the table. If you standardize this data (i.e. scores), what will be the standardized score of the second student?

Student	1	2	3	4	5
Score	80	85	90	75	95

Ans 2:

Standardized score of a student $x_i = \frac{X_i - \bar{X}}{s}$

Here, X_i is the score of the student. $\bar{X}$ is the sample mean of scores. s is the sample standard deviation.

For the given data, $\bar{X} = 85$

Therefore, the standardized score of the 2nd student,

$$x_2 = \frac{X_2 - \bar{X}}{s} = \frac{85-85}{s} = 0$$

Q3. Suppose the average height of adults in a community is normally distributed with a mean of 175 cm and a standard deviation of 5 cm. What percentage of adults (approximately) in this group would you expect to be between 165 cm and 185 cm tall?

Ans 3:

Given, $\mu = 175$ cm and $\sigma = 5$ cm. The population follows Normal distribution.

$\therefore \mu - 2\sigma = 165$ cm and $\therefore \mu + 2\sigma = 185$ cm

Using the 68-95-99.7 rule for normal distribution, we know that approximately 95% of the population will be in the range $[\mu - 2\sigma, \mu + 2\sigma]$.

So, we can expect that approximately 95% of the population will be between 165 cm and 185 cm.

Q4. Suppose you want to compare the average scores of two groups of students who took different versions of a test. Group A has a mean score of 85 with a standard deviation of 10, and Group B has a mean score of 90 with a standard deviation of 12. Group A consists of 20 students, and Group B consists of 25 students. You are performing a two-sample t-test. What will be the degree of freedom for the t-distribution in this test?

Ans 4:

Given $n_A = 20$ and $n_B = 25$.

$\therefore df = n_B + n_B - 2 = 43$

BT307 Quiz

Total Marks: 10 Time: 45 min

Name: **Roll No.:**

Q1. We are creating a scatter plot using the following R code. Edit the code to add a trend line overlayed on the scatter plot. Do not add any confidence interval around the trend line. Highlight/mark the edited section of your code so that it can be recognized easily.

```
library(ggplot2)
sales_data <- read.csv("sales_data.csv", header = TRUE)
plot_sales <- ggplot(sales_data, aes(x = Spend, y = Revenue)) +
                geom_point(size = 3, alpha = 0.7) +
                labs(x = "Spend (in $)",
                     y = "Revenue (in $)")
plot_sales
```

Ans 1:

```
library(ggplot2)
sales_data <- read.csv("sales_data.csv", header = TRUE)
plot_sales <- ggplot(sales_data, aes(x = Spend, y = Revenue)) +
                geom_point(size = 3, alpha = 0.7) +
                geom_smooth(se = FALSE, span = 5) +
                labs(x = "Spend (in $)",
                     y = "Revenue (in $)")
plot_sales
```

Note: You should write the whole code so that one can understand where you are adding the additional script. As the question asks, you should highlight (use pencil/colour pen/underline etc) to show your edits.

Q2. The diameters of RBCs for a particular mammalian species are found to be normally distributed with a mean of 7.8 μm and a standard deviation of 0.4 μm. Estimate the range of diameters that encompasses approximately 68% of these RBCs. Write the range's lower and upper values.

Ans 2:

Given, $\mu = 7.8$ μm and $\sigma = 0.4$ μm. The population follows Normal distribution.

Using the 68-95-99.7 rule for normal distribution, we know that approximately 68% of the population will be in the range $[\mu - \sigma, \mu + \sigma]$.

Therefore,

the lower value of the range = $\mu - \sigma = 7.8 - 0.4 = 7.4$ μm

the upper value of the range = $\mu + \sigma = 7.8 + 0.4 = 8.2$ μm

Q3. It is known that the weight of the seeds of a particular plant is normally distributed with a mean of 50 mg and a standard deviation of 5 mg. A botanist repeatedly takes random samples of 36 seeds and calculates the sample mean each time. What is the standard deviation of the sample mean for these 36-seed samples?

Ans 3:

Given, $\mu = 50$ mg and $\sigma = 5$ mg. The population follows Normal distribution. Sample size $n = 36$

Following the Central Limit Theorem,

the std. dev of the sample mean, $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{36}} = 0.833$ mg

Q4. A plant biologist is studying a new growth supplement that is hypothesized to increase the height of a particular crop variety. The null hypothesis (H_0) states that the supplement has no effect on the average height of the plants, while the alternative hypothesis (H_1) states that the supplement affects the average height. The biologist chooses a significance level (α) of 0.05 for the experiment. What is the probability of making a Type I error in this study?

Ans 4:

As the chosen significance level (α) is 0.05, the probability of making a Type I error is 0.05.

BT307 Quiz

Time: 45 min | Total Marks: 10

Roll No. **Name:**

Q1. A marine biologist is studying a fish species whose body length (from snout to tail) is found to be normally distributed with a mean of 60 cm and a standard deviation of 5 cm. What percentage of these fish (approximately) would you expect to have lengths between 55 cm and 65 cm?

Ans 1:

Given, $\mu = 60$ cm and $\sigma = 5$ cm. The population follows Normal distribution.

$\therefore \mu - \sigma = 55$ cm and $\therefore \mu + \sigma = 65$ cm

Using the 68-95-99.7 rule for normal distribution, we know that approximately 68% of the population will be in the range $[\mu - \sigma, \mu + \sigma]$.

So, we can expect that approximately 68% of the population will be between 55 cm and 65 cm.

Q2. A t-test is performed using the following R script:

```
t.test(h.data$Treated,h.data$Untreated, var.equal =TRUE)
```

What are the assumptions made for this test?

Ans 2:

The assumptions are:

1) The population distribution of both groups is Normal distribution.

2) The population variances of the groups are the same.

Note: The first assumption is a key assumption for t-test in general. The second assumption is often used in t-test and is evident from the script (`var.equal =TRUE`)

Q3. Scores for six students are given in the table. If you standardize this data (i.e. scores), what will be the standardized score of the fourth student?

Student	1	2	3	4	5	6
Score	10	30	50	50	90	70

Ans 3:

Standardized score of a student $x_i = \dfrac{X_i - \bar{X}}{s}$

Here, X_i is the score of the student. $\bar{X}$ is the sample mean of scores. s is the sample standard deviation.

For the given data, $\bar{X} = 50$

Therefore, the standardized score of the 2nd student,

$$x_4 = \frac{X_4 - \bar{X}}{s} = \frac{50 - 50}{s} = 0$$

Q4. A researcher is studying the effect of a new dietary supplement on laboratory mice. The control group (Group A) has 12 mice and shows a mean weight of 25 grams with a standard deviation of 3 grams. The treatment group (Group B) has 15 mice and shows a mean weight of 28 grams with a standard deviation of 2 grams. The researcher plans to compare the two groups using a two-sample t-test. What will be the degree of freedom for the t-distribution in this test?

Ans 4:

Given $n_A = 12$ and $n_B = 15$.

$\therefore df = n_B + n_B - 2 = 25$