Student Information

Full Name : Yarkın Özcan
Id Number : 2580835

Answer 1

a)

Our sample is of size 9, which is not large; however, since the measurements follow a normal distribution we can construct a confidence interval. Also, in this case, standard deviation is known, hence we can use the formula for confidence interval for the mean when standard deviation σ is known;

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

First we need to calculate the sample mean $\bar{X}$;

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{6.4 + 9.5 + 8.2 + 10.2 + 7.6 + 11.1 + 8.7 + 7.3 + 9.1}{9} = \frac{78.1}{9} \approx 8.678$$

Now we need to calculate the standard normal quantile z_α. The question wants a %95 confidence interval, (1 - α) = 0.95. So $\alpha = 0.05$ and $\frac{\alpha}{2} = 0.025$. After checking the standard normal distribution table, we can find that $z_{0.025} = 1.960$.
Also in the question it is given that the measurement device guarantees standard deviation of 2.7, hence $\sigma = 2.7$.
Now, we have all the information we need in order to use the formula.

$$8.678 \pm 1.960 \cdot \frac{2.7}{\sqrt{9}}$$
$$[6.914, 10.442]$$

b)

We can describe the confidence interval as;

$$center \pm margin$$

Where center is the mean $\bar{X}$ and margin is $z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$

$$\Delta \geq z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$n \geq (z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\Delta})^2$$

$$n \geq (1.96 \cdot \frac{2.7}{1.25})^2 = 17.923$$

Now, it is crucial to round up the n to 18 because it cannot be less then 17.923. Hence we need a sample size of 18.

Answer 2

a)

A null hypothesis is always an equality, hence for this study, null hypothesis should be the equality between the average monthly revenue of last year and average monthly revenue of this year. More formally;

$$H_0 : \mu = 20.000TL$$

In this case, Mecnun's claim should considered as the null hypothesis because Mecnun claimed that the average revenue has remained same.
For this study, the alternative hypothesis is one sided because of Leyla's claim. We are only interested in if the average revenue has increased or not. Therefore, the alternative hypothesis is;

$$H_A : \mu > 20.000TL$$

b)

Since Leyla and Mecnun collected data from a random sample of 50 customers which can be considered as a large sample (by rule of thumb), we can use Z-Test in order to test Mecnun's claim.

Step 1: Test statistic. We are given $\sigma = 3000$, n $= 50$, $\alpha = 0.05$, $\mu_0 = 20.000$. Also from the sample, we have $\mu = \bar{X} = 22.000$. The test statistic is;

$$Z = \frac{\bar{X} - \mu_0}{\sqrt{Var(\bar{X})}} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$Z = \frac{22.000 - 20.000}{\frac{3000}{\sqrt{50}}} \approx 4.714$$

Step 2: Acceptance and rejection regions. We have %5 level of significance, hence $\alpha = 0.05$. Also since our Z-Test is one-sided right tail test, we don't divide α by 2. So the critical value is;

$$z_\alpha = z_{0.05} = 1.645$$

With the right tail alternative, we;

$$\text{reject } H_0 \text{ if } Z \geq 1.645$$
$$\text{accept } H_0 \text{ if } Z < 1.645$$

Step 3: Result. Our test statistic Z = 4.714 belongs to the rejection region; therefore, we reject the null hypothesis which is Mecnun's claim. We can also say that the data provides sufficient evidence in favor of alternative hypothesis.

c)

In part b, we found that our $Z_{obs} = 4.714$. And since our alternative hypothesis was one-sided right-tail, our P value must be P = $\alpha = P\{Z \geq z_\alpha\} = P\{Z \geq Z_{obs}\}$ = 1 - $\Phi(Z_{obs})$. So we want to calculate $\Phi(4.714)$; however, when we look at the standard normal distribution table, we see that after Z is 3.9, probabilities gets closer to 1. So $\Phi(4.714)$ is almost 1 hence 1 - $\Phi(4.714)$ is approximately 0. At the end, we have P-value close to 0 which is smaller than 0.01. From this value, we can reject the null hypothesis H_0 since for smaller values of P, rejection areas get larger.

d)

In this question, we want to compare the average monthly revenue of Leyla and Mecnun μ_{LM} and the average monthly revenue of competitor μ_C. To do that, we can use two-sample Z-test for means. Firstly, we need to formally write the null and alternative hypothesis;

$$H_0 : \mu_{LM} = \mu_C \text{ OR } H_0 : \mu_{LM} - \mu_C = 0$$

$$H_A : \mu_{LM} > \mu_C$$

This is a one sided-right tail test because we are interested in whether Leyla and Mecnun's average monthly revenue is higher than the competitors.
Step 1: Test statistic. We are given $\hat{\mu}_{LM} = \bar{X}$ = 22.000TL from a sample of size n = 50 with a standard deviation σ_{LM} = 3000TL. Also we are given $\hat{\mu}_C = \bar{Y}$ = 24.000TL from a sample of size m = 40 with a standard deviation σ_Y = 4000TL. The tested value E$(\mu_{LM} - \mu_C)$ = D = 0TL. Now that we know these values, we can use the two-sample Z-test formula;

$$Z = \frac{(\hat{\mu}_{LM} - \hat{\mu}_C) - E(\mu_{LM} - \mu_C)}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = \frac{(22.000 - 24.000) - 0}{\sqrt{\frac{(3000)^2}{50} + \frac{(4000)^2}{40}}} = \frac{-2000}{\sqrt{180.000 + 400.000}} = \frac{-2000}{761.6} \approx -2.626$$

Step 2: Acceptance and rejection regions. The claim in the question is made with a %1 level of significance. Also since this is a one sided-right tail test, we take $\alpha = 0.01$ and the critical value becomes $z_\alpha = z_{0.01} = 2.326$ (Looking at the standard distribution table). With the right-tail alternative, we;

$$\text{reject } H_0 \text{ if } Z \geq 2.326$$
$$\text{accept } H_0 \text{ if } Z < 2.326$$

Step 3: Result. Our test statistic Z = -2.626 is smaller than $z_{0.01}$ = 2.326, therefore belongs to acceptance region which means that the data we sampled is not sufficient to claim that the average monthly revenue of Leyla and Mecnun's store is higher than that of the competitors. Thus we accept the null hypothesis.

Answer 3

To determine if there is a significant association between gender and coffee preference, we can perform a Chi-square test of independence. Firstly we need to state the hypothesis.

H_0: There is no association between gender and coffee preference (they are independent).

H_A: There is an association between gender and coffee preference (they are dependent).

Now we need to create the observed frequency table with the given data;

Observed Counts	Male	Female	Row Total
Black Coffee	52	17	69
Coffee with Milk	16	63	79
Coffee with Sugar	32	20	52
Column Total	100	100	200

Now we can calculate the expected frequencies for each entry using the formula;

$$E_{ij} = \frac{(Row\,Total_i \cdot Column\,Total_j)}{Grand\,Total}$$

Now let's create the expected frequency table;

Expected Counts	Male	Female
Black Coffee	$\frac{69 \cdot 100}{200} = 34.5$	$\frac{69 \cdot 100}{200} = 34.5$
Coffee with Milk	$\frac{79 \cdot 100}{200} = 39.5$	$\frac{79 \cdot 100}{200} = 39.5$
Coffee with Sugar	$\frac{52 \cdot 100}{200} = 26$	$\frac{52 \cdot 100}{200} = 26$

Now, we can compute Chi-Square test using the formula;

$$X^2 = \sum \frac{(Obs_{ij} - Exp_{ij})^2}{Exp_{ij}}$$

Where Obs_{ij} are the observed frequencies and Exp_{ij} are the expected frequencies. Let's calculate each component;

For Black Coffee (Male): $\frac{(52-34.5)^2}{34.5} = \frac{17.5^2}{34.5} = \frac{306.25}{34.5} \approx 8.88$
For Black Coffee (Female): $\frac{(17-34.5)^2}{34.5} = \frac{17.5^2}{34.5} = \frac{306.25}{34.5} \approx 8.88$
For Coffee with Milk (Male): $\frac{(16-39.5)^2}{39.5} = \frac{23.5^2}{39.5} = \frac{552.25}{39.5} \approx 13.98$
For Coffee with Milk (Female): $\frac{(63-39.5)^2}{39.5} = \frac{23.5^2}{39.5} = \frac{552.25}{39.5} \approx 13.98$
For Coffee with Sugar (Male): $\frac{(32-26)^2}{26} = \frac{6^2}{26} = \frac{36}{26} \approx 1.38$
For Coffee with Sugar (Female): $\frac{(20-26)^2}{26} = \frac{6^2}{26} = \frac{36}{26} \approx 1.38$

Summing these we get;

$$X^2 = 8.88 + 8.88 + 13.98 + 1.38 + 1.38 = 48.48$$

Now we need to determine the degrees of freedom. Degrees of freedom (df) for a Chi-square test of independence is given by;

$$df = (rows - 1) \cdot (columns - 1) = (2 - 1) \cdot (3 - 1) = 1 \cdot 2 = 2$$

Using the Chi-square distribution table, we can see that for $df = 2$, $X^2 = 48.48$ exceeds the value at the table even for the least significance level of 0.001, thus we can reject the null hypothesis. So we can conclude that there is a significant association between gender and coffee preference.