Student Information

Full Name : Anıl Eren Göçer
Id Number : 2448397

Answer 1

a)

Let's propose an estimator $\hat{\theta}$ for the population mean such that:

$$\hat{\theta} = \overline{X}$$

where $X = \{20.1, 12.8, 18.9, 16.4, 20.3, 10.1, 15.4, 12.4, 24.7, 18.5\}$ is the given sample.

$$\hat{\theta} = \overline{X} = \frac{1}{n}.\sum_{i=1}^{10} = \frac{1}{10}.\sum_{i=1}^{10} = 16.96$$

Note that;

$$E(\overline{X}) = E(\frac{1}{n}.\sum_{i=1}^{10}) = \frac{1}{n}.\sum_{i=1}^{10} E(X_i) = \frac{1}{n}.n.\mu = \mu$$

Thus, $\hat{\theta} = \overline{X}$ is an unbiased estimator for population mean μ.

The population standard deviation is given as $\sigma = 3$

i) We want to obtain a 90% confidence interval:

$$1 - \alpha = 0.90 \longrightarrow \alpha = 0.10 \longrightarrow \alpha/2 = 0.05$$

Confidence interval calculated as:

$$\hat{\theta} \pm z_{\alpha/2}.\sigma(\hat{\theta}) = 16.96 \pm \Phi(0.05).\frac{1}{\sqrt{10}}.3 = 16.96 \pm 1.560584$$
$$= [\mathbf{15.39341,\ 18.520574}]$$

ii) We want to obtain a 99% confidence interval:

$$1 - \alpha = 0.99 \longrightarrow \alpha = 0.01 \longrightarrow \alpha/2 = 0.005$$

Confidence interval calculated as:

$$\hat{\theta} \pm z_{\alpha/2}.\sigma(\hat{\theta}) = 16.96 \pm \Phi(0.005).\frac{1}{\sqrt{10}}.3 = 16.96 \pm 2.443808176$$
$$= [\mathbf{14.51619182,\ 19.403808}]$$

b)

We want a confidence level of 0.98 :

$$1 - \alpha = 0.98 \longrightarrow \alpha = 0.02 \longrightarrow \alpha/2 = 0.01$$
$$\text{margin} = z_{\alpha/2}.\sigma(\hat{\theta}) = \Phi(0.01).\frac{3}{\sqrt{n}}$$
$$\Phi(0.01).\frac{3}{\sqrt{n}} \leq 1.55 \longrightarrow \frac{\Phi(0.01).3}{1.55} \leq \sqrt{n}$$
$$\longrightarrow n \geq (\frac{3.\Phi(0.01)}{1.55})^2 \longrightarrow n \geq 20.2674231$$

By rounding it up, $n_{min} = \mathbf{21}$.

Hence, we need a sample of at least size **21**.

Answer 2

a)

No, they are not enough. To calculate test statistic and test hypotheses, we need to also know the population standard deviation σ . Moreover, intuitively, knowing only average and size of the rating does not give much clue about the restaurant, how the ratings are distributed is also important. For example, consider two restaurant, namely A and B. Both A and B have the average 7 and sample size 5. In addition, sample for A is $\{10,2,10,3,10\}$ and sample for B is $\{7,7,7,6,8\}$. Which one would you prefer ? Of course, you prefer restaturant A because you do not want to take the risk that you may get an order that is worth for 2 rating while you get an order which is worth for 6,7 or 8 rating which are more or less acceptable.

b)

We test the null hypotheses $H_0 : \mu = 7.5$ against a one sided left-tail alternative $H_A : \mu < 7.5$.

Test statistic: We are given n = 256, $\sigma = 0.8$, $\alpha = 0.05$, $\mu_0 = 7.5$ and $\overline{X} = 7.4$. The test statistic is:

$$\mathrm{Z} = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{7.4 - 7.5}{0.8/\sqrt{256}} = -2$$

Acceptance and rejection regions:

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

With the left tail alternative, we

$$\begin{cases} \text{reject } H_0 & \text{if Z} \leq -1.645 \\ \text{accept } H_A & \text{if Z} > -1.645 \end{cases}$$

Result: Our test statistic Z = -2 belongs to the rejection region; therefore we reject the null hypotheses. The sample data provided sufficient evidence in favor of the alternative that $H_A : \mu < 7.5$.

Hence, restaurant A would **NOT** be in my list of candidate restaurants to order from.

c)

In this case, a change in sample standard deviation only changes test statistic. **Acceptance and rejection regions remain same:**

$$\begin{cases} \text{reject } H_0 & \text{if Z} \leq -1.645 \\ \text{accept } H_A & \text{if Z} > -1.645 \end{cases}$$

Test statistics is now:

$$\text{Z} = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{7.4 - 7.5}{1/\sqrt{256}} = -1.6$$

Result: The evidence against H_0 which is in favor of H_A is insuffficient. Since -1.6 > -1.645

Hence, I would include the restaurant in may list of candidate restaurants now.

d)

If the mean of user ratings for restaurant A is $X_0 = 7.6$ which is greater than $\mu_0 = 7.5$. Our test statistic $\text{Z} = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ will always be greater than zero, since $\overline{X} = 7.6 > \mu_0 = 7.5$ and σ and $\sqrt{n}$ are always positive. Moreover, our alternative hypothesis is left-tailed. Therefore, we can make sure that our test statistic will never fall into the rejection region which is in somewhere to the left of zero. Hence, we can include restaurant A in our list without applying a statistical test completely.

Answer 3

a)

In the question, we are given that

* The mean run time on computer A: $\overline{X_A} = 211$
* The mean run time on computer B: $\overline{X_B} = 133$

* The sample standard deviation for run times on computer A: $s_A = 5.2$
* The sample standard deviation for run times on computer B: $s_B = 22.8$

* The number of runs on computer A: $n_A = 20$
* The number of runs on computer B: $n_B = 32$

We will test the null hypothesis H_0: $\overline{X_A} - \overline{X_B} = 90$ against the alternative hypothesis H_A: $\overline{X_A} - \overline{X_B} < 90$ at a %1 level of significance.

So, $\alpha = 0.01$

The pooled standard deviation can be calculated as:

$$s_p^2 = \frac{(n_A - 1).s_A^2 + (n_B - 1).s_B^2}{n_A + n_B - 2} = 332.576 \longrightarrow s_p = 18.23666636 \approx 18.2367$$

Test statistic can be calculated as follows:

$$t = \frac{\overline{X_A} - \overline{X_B} - D}{s_p.\sqrt{\dfrac{1}{n_A} + \dfrac{1}{n_B}}} = -2.308465069 \text{ where } D = 90$$

Degree of freedom is: $n_A + n_B - 2 = 50$

$t_\alpha = t_{0.01}$ with degree of freedom 50 is equal to 2.403.

The rejection region for this left-tail test is $(-\infty,-t_\alpha] = (-\infty,-2.403]$. Since t = -2.308465069 $\notin (-\infty,-2.403]$, we **accept** H_0 concluding that the researcher can claim that the computer B provides a 90-minute or better performance.

Answer: Yes, the researcher can claim that the computer B provides a 90-minute or better performance.

b)

In the question, we are given that

* The mean run time on computer A: $\overline{X_A} = 211$
* The mean run time on computer B: $\overline{X_B} = 133$

* The sample standard deviation for run times on computer A: $s_A = 5.2$
* The sample standard deviation for run times on computer B: $s_B = 22.8$

* The number of runs on computer A: $n_A = 20$
* The number of runs on computer B: $n_B = 32$

We will test the null hypothesis H_0: $\overline{X_A} - \overline{X_B} = 90$ against the alternative hypothesis H_A: $\overline{X_A} - \overline{X_B} < 90$ at a %1 level of significance.

So, $\alpha = 0.01$

Test statistic can be calculated as follows:

$$t = \frac{\overline{X_A} - \overline{X_B} - D}{\sqrt{\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B}}} = -2.860631582 \text{ where } D = 90.$$

By Satterthwaite Approximation, the degree of freedom can be calculated as follows:

$$d.f = \frac{(\dfrac{s_A^2}{n_A} + \dfrac{s_B^2}{n_B})^2}{\dfrac{s_A^4}{n_A^2.(n_A - 1)} + \dfrac{s_B^4}{n_B^2.(n_B - 1)}} = 35.97$$

By rounding d.f = 35.97 to the nearest integer value, degree of freedom is 36.
$\alpha = 0.01, d.f = 36 \longrightarrow t_{0.01} = 2.434$

The rejection region for this left-tail test is (-∞, -t_α] = (-∞, -2.434].
Since t = -2.860631582 $\in$ ($-\infty$, -2.434], we **reject** H_0 concluding that the researcher cannot claim that the computer B provides a 90-minute or better performance.

Answer: No, the researcher cannot claim that the computer B provides a 90-minute or better performance.