Student Information

Full Name : Mehmet Rüçhan Yavuzdemir
Id Number : 2522159

Answer 1

a)

Since the sample size is smaller than 30, which is 16, and also standard deviation of the sample is not given, we can use Student's t-distribution.

Since we want 98% confidence, our α value is 0.02, and it will be two-tailed because of the nature of the confidence intervals.

$\bar{X}$ = (8.4 + 7.8 + 6.4 + 6.7 + 6.6 + 6.6 + 7.2 + 4.1 + 5.4 + 6.9 + 7.0 + 6.9 + 7.4 + 6.5 + 6.5 + 8.5) / 16 = **6.81**

$$n = 16 \quad \mu_0 = 7.50 \quad \bar{X} = 6.81 \quad \alpha = 0.02 \quad v = n - 1 = 16 - 1 = 15$$

$\sum_{i=1}^{n}(X_i - \bar{X})^2 =$

$(8.40 - 6.81)^2 + (7.80 - 6.81)^2 + (6.40 - 6.81)^2 + (6.70 - 6.81)^2 + (6.60 - 6.81)^2 + (6.60 - 6.81)^2 + (7.20 - 6.81)^2 + (4.10 - 6.81)^2 + (5.40 - 6.81)^2 + (6.90 - 6.81)^2 + (7.00 - 6.81)^2 + (6.90 - 6.81)^2 + (7.40 - 6.81)^2 + (6.50 - 6.81)^2 + (6.50 - 6.81)^2 + (8.50 - 6.85)^2 = \mathbf{16.71}$

From MATLAB, we can use $tinv(\alpha, v)$ function to calculate $t_{0.01}$.

$t_{0.01} = -tinv(0.01, 15) = \mathbf{2.60}$

$s(X) = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n}(X_i - \bar{X})^2} = \sqrt{\frac{1}{15} \cdot 16.71} = \mathbf{1.06}$

$$\begin{aligned} X &= \bar{X} \pm t_{\alpha/2} \cdot s(\bar{X}) \\ &= \bar{X} \pm t_{\alpha/2} \cdot \frac{s(X)}{\sqrt{n}} \\ &= 6.81 \pm t_{0.01} \cdot \frac{s(X)}{\sqrt{16}} \\ &= 6.81 \pm 2.60 \cdot \frac{s(X)}{4} \end{aligned}$$

$$
\begin{aligned}
&= 6.81 \pm 2.60 \cdot \frac{1.06}{4} \\
&= 6.81 \pm 0.69 \\
&= [6.12, 7.50]
\end{aligned}
$$

b)

$$
H_0 : \text{Fuel consumption is similar, } F_i = F_f
$$
$$
H_A : \text{Fuel consumption is reduced, } F_i > F_f
$$

From part a),

$$
n = 16 \quad \mu_0 = 7.50 \quad \bar{X} = 6.81 \quad v = n - 1 = 16 - 1 = 15
$$

At a 5% significance level, $\alpha = 0.05$.

Since we are only interested in the cases where the fuel consumption is lesser, we can do a left-tailed T-test.

From MATLAB, we can use $tinv(\alpha, v)$ function to calculate $t_{0.05}$.

$t_{0.05} = -tinv(0.05, 15) = \mathbf{1.75}$

Acceptance interval for a t-score of a T-test is $[-t_{0.05}, \infty)$, $[-1.75, \infty)$

$$
\begin{aligned}
T &= \frac{\bar{X} - \mu_0}{s(\bar{X})} \\
&= \frac{\bar{X} - \mu_0}{\frac{s(X)}{\sqrt{n}}} \\
&= \frac{6.81 - 7.50}{\frac{1.06}{\sqrt{16}}} \\
&= 4 \cdot \frac{-0.69}{1.06} \\
&= \textbf{-2.60}
\end{aligned}
$$

Since -2.60 does not belong to the acceptance interval $[-1.75, \infty)$, we can say that we have enough evidence to reject H_0 in favor of H_A. Therefore, there is a significant reduction in fuel consumption, and the improvement is effective, with a significance level of 5%.

c)

$$T = \frac{\bar{X} - \mu_0}{s(\bar{X})}$$

From parts a) and b), $\bar{X} = 6.81$. In this question, we are given that μ_0 is 6.50.

Both $\bar{X} - \mu_0$ and $s(\bar{X})$ are greater than zero, meaning that our t-score is positive. Since we are doing a left-tailed test, our t-score should be negative and it should be smaller than $-t_\alpha$. Hence, there is no way to reject the null hypothesis H_0, we immediately accept it, and we do not need to calculate the test statistic.

Answer 2

a)

H_0 : Rent prices are similar to the last year, $R_i = R_f$, offered by Ali.
H_A : There is an increase in rent prices, $R_i < R_f$, offered by Ahmet.

Ali claimed the null hypothesis H_0.

b)

Since n = 100 > 30, we can use the right-tailed Z-test for this purpose.

$$n = 100 \quad \mu_0 = 5000 \quad \bar{X} = 5500 \quad \sigma = 2000$$

We have a 5% significance level, $\alpha = 0.05$. We should calculate $z_{0.05}$.

$z_{0.05} = 1.64$, therefore our acceptance interval is $(-\infty, 1.64]$.

$$\begin{aligned} Z &= \frac{\bar{X} - \mu_0}{\sigma(\bar{X})} \\ &= \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \\ &= \frac{5500 - 5000}{\frac{2000}{\sqrt{100}}} \\ &= \frac{500}{200} = \mathbf{2.50} \end{aligned}$$

Since 2.50 are not in the acceptance interval $(-\infty, 1.64]$, we reject H_0, saying that house prices are similar to the last year. There is a significant increase in rent prices. Therefore, Ahmet can claim and provide mathematical evidence for supporting that increase.

c)

Since this is a right-tailed test, the p-value means the probability of more extreme cases, or the area under the curve after the z-score of 2.50.

We can use MATLAB to calculate our p-value for a Z-test:

1 - *normcdf*(2.5) = $6.21 \cdot 10^{-3}$ = **0.00621**.

With a 99.38% confidence level, we can say that there is an increase in rent prices; in addition, as a rule of thumb, p-values around 0.05 is considered to be small enough to reject H_0. 0.00621 is much smaller than 0.05; therefore, we reject H_0 in favor of H_A, Ahmet is right.

d)

$$H_0 : \text{Rent prices in Ankara and Istanbul are similar, } X = Y$$
$$H_A : \text{Rent prices in Ankara are lower than in Istanbul, } X < Y$$

Ankara:

$$n = 100 \quad \bar{X} = 5500 \quad \sigma = 2000$$

Istanbul:

$$n = 60 \quad \bar{X} = 6500 \quad \sigma = 3000$$

where X is rent prices in Ankara and Y is rent prices in Istanbul.

$\hat{\theta} = \hat{X} - \hat{Y}$

$\bar{\theta} = \bar{X} - \bar{Y}$ = 5500 - 6500 = -1000

From MATLAB, we can use $norminv(\alpha)$ function to calculate $z_{0.01}$.

$z_{0.01} = -norminv(0.01) = \mathbf{2.33}$

Since 100 and 60 are large numbers, we can conduct a left-tailed Z-test. Because we are given that the significance level is 1%, $\alpha = 0.01$. So our acceptance interval is $[-z_{0.01}, \infty)$, $[-2.33, \infty)$.

To find the z-score, we need $s(\bar{\theta})$.

As a general rule, we know that $Var(\gamma) = \frac{\sigma^2}{n}$.

$s(\bar{\theta}) = \sqrt{Var(\bar{X}) + Var(\bar{Y})} = \sqrt{\frac{2000^2}{100} + \frac{3000^2}{60}} = \mathbf{435.89}$

$$\begin{aligned} Z &= \frac{\bar{\theta} - \mu_0}{s(\bar{\theta})} \\ &= \frac{-1000 - 0}{435.89} \\ &= \frac{-1000}{435.89} \\ &= \mathbf{-2.29} \end{aligned}$$

Since -2.29 is in the acceptance region, we accept H_0. Hence, there is no significant difference between the two cities. We cannot claim that rent prices in Ankara are lower than in Istanbul with a significance level of 1%.

Answer 3

H_0 : Number of rainy days is independent of the season
H_A : Number of rainy days is dependant on the season

Observed:

Weather	Winter	Spring	Summer	Autumn
Rainy	34	32	15	19
Non-Rainy	56	58	75	71

Expected:

Weather	Winter	Spring	Summer	Autumn
Rainy	25	25	25	25
Non-Rainy	65	65	65	65

Since all expected values Exp(k) are greater than 5, we do not need to join rows.

v = (2 - 1) · (4 - 1) = 3.

$$\chi^2_{obs} = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\{Obs(i,j) - \hat{Exp}(i,j)\}^2}{\hat{Exp}(i,j)}$$

$\chi^2_{obs} = \frac{(34-25)^2}{25} + \frac{(32-25)^2}{25} + \frac{(15-25)^2}{25} + \frac{(19-25)^2}{25} + \frac{(56-65)^2}{65} + \frac{(58-65)^2}{65} + \frac{(75-65)^2}{65} + \frac{(71-65)^2}{65}$

$= (\frac{1}{25} + \frac{1}{65}) \cdot (6^2 + 7^2 + 9^2 + 10^2) = \mathbf{14.73}$

We can use $chi2cdf(\chi^2, v)$ from MATLAB to calculate our p-value for Chi-square test:

$P(\chi^2 > 14.73) = 1 - chi2cdf(14.73, 3) = 2.06 \cdot 10^{-3} = \mathbf{0.00206}$.

Our p-value is extremely small, which means we can definitely reject H_0 with 99.79% confidence. We have significant evidence that the number of rainy days is dependent on the season.

Answer 4

```matlab
observed = [34 32 15 19; 56 58 75 71];

row = sum(observed')';
column = sum(observed);
total = sum(row);

x = length(row);
y = length(column);

expected = zeros(size(observed));

for i = 1:x
    for j = 1:y
        expected(i,j) = row(i) * column(j) / total;
    end
end

formula = (observed - expected).^2./ expected;

degree_of_freedom = (x-1) * (y-1);

chi_square = sum(sum(formula))

p_value = 1 - chi2cdf(chi_square, degree_of_freedom)
```

```
observed = [34 32 15 19; 56 58 75 71];

row = sum(observed')';
column = sum(observed);
total = sum(row);

x = length(row);
y = length(column);

expected = zeros(size(observed));

for i = 1:x
    for j = 1:y
        expected(i,j) = row(i) * column(j) / total;
    end
end

formula = (observed - expected).^2./ expected;

degree_of_freedom = (x-1) * (y-1);

chi_square = sum(sum(formula))

p_value = 1 - chi2cdf(chi_square, degree_of_freedom)
```

```
octave:1> source("Q4.m")
chi_square = 14.732
p_value = 2.0603e-03
```