

Student Information

Name : Murat Bolu

ID : 2521300

Answer 1

a)

Since the sample is small and the standard deviation of the population is unknown, Student's t distribution can be used. The critical value of t distribution is $t_{\alpha/2} = t_{0.01} = 2.602$ with $n - 1 = 15$ degrees of freedom.

$$\begin{aligned}\bar{X} &= \frac{8.4 + 7.8 + 6.4 + 6.7 + 6.6 + 6.6 + 7.2 + 4.1 + 5.4 + 6.9 + 7.0 + 6.9 + 7.4 + 6.5 + 6.5 + 8.5}{16} \\ &= 6.806\end{aligned}$$

$$\begin{aligned}s &= \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} \\ &= \sqrt{\frac{\sum_{i=1}^{16} (X_i - 6.806)^2}{15}} \\ &= 1.055\end{aligned}$$

The 98% confidence interval $\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$ becomes $6.806 \pm 2.602 \cdot \frac{1.055}{4} = [6.120, 7.493]$.

b)

Let μ_1 be the initial gasoline consumption per 100 kilometers and μ_2 be the improved gasoline consumption per 100 kilometers. Therefore, the null hypothesis H_0 is $\mu_1 = \mu_2$ and the alternative hypothesis H_A is $\mu_1 > \mu_2$. Since we don't know the population standard deviation, we can use T-statistic.

$$\begin{aligned}t &= \frac{\bar{X} - \mu}{s/\sqrt{n}} \\ &= \frac{6.806 - 7.5}{1.055/4} \\ &= -2.629\end{aligned}$$

The rejection region is $R = (-\infty, -t_\alpha] = (-\infty, -t_{0.05}] = (-\infty, -1.753]$ with 15 degrees of freedom, since we are using a left-tail alternative. Since $t \in R$, we can reject the null hypothesis. There is sufficient evidence that the improvement was effective.

c)

Since $\bar{X} = 6.806 > 6.5$, we can immediately accept the null hypothesis because the T-statistic is positive and the rejection region includes zero. Since the mean of the new sample is higher than the previous one, there is no evidence suggesting decrease in gasoline consumption.

Answer 2

a)

The null hypothesis H_0 should be considered to be that the rent prices stayed the same. The alternative hypothesis H_A then becomes the rent prices have increased.

b)

Since the population standard deviation is known, Z-statistic can be used.

$$\begin{aligned} Z &= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \\ &= \frac{5500 - 5000}{2000/10} \\ &= \frac{500}{200} \\ &= 2.5 \end{aligned}$$

The rejection region is $R = [z_\alpha, \infty) = [z_{0.05}, \infty) = [1.645, \infty)$, since we are using a right-tail alternative. Since $Z \in R$, we can reject the null hypothesis. There is sufficient evidence that the rent prices have increased. Ahmet can claim that there is an increase in the rent prices compared to the last year at 5% level of significance.

c)

The P -value is $P = \mathbf{P}\{Z \geq Z_{\text{obs}}\} = \mathbf{P}\{Z \geq 2.5\} = 1 - \Phi(2.5) = 0.0062$. This means Ahmet can claim that the rent prices have increased with 0.62% level of significance.

d)

The null hypothesis $H_0 : \mu_A = \mu_I$ is that the rent prices in Ankara, μ_A , are the same as the rent prices in Istanbul, μ_I . The alternative hypothesis $H_A : \mu_A < \mu_I$ is that the rent prices in

Ankara are lower than the rent prices in Istanbul. Since we know both of the standard deviations, we can use Z-statistic.

$$\begin{aligned}
Z &= \frac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \\
&= \frac{5500 - 6500 - 0}{\sqrt{\frac{2000^2}{100} + \frac{3000^2}{60}}} \\
&= \frac{-1000}{\sqrt{40,000 + 150,000}} \\
&= -\frac{1000}{435.890} \\
&= -2.294
\end{aligned}$$

We can calculate the P -value as $\Phi(-2.294) = 0.011$. Since the P -value is barely above 1%, the null hypothesis should be accepted. Therefore, they cannot claim that the prices in Ankara are lower than the prices in Istanbul.

Answer 3

Let's construct the table and calculate the sums for all rows and columns.

| Obs(i, j) | Winter | Spring | Summer | Autumn | $n_{i.}$ |
|---------------|--------|--------|--------|--------|----------|
| Rainy | 34 | 32 | 15 | 19 | 100 |
| Non-rainy | 56 | 58 | 75 | 71 | 260 |
| $n_{.j}$ | 90 | 90 | 90 | 90 | 360 |

Now, we can construct the table for expected number days being rainy and non-rainy if they were independent of season.

| $\widehat{\text{Exp}}(i, j)$ | Winter | Spring | Summer | Autumn | $n_{i.}$ |
|------------------------------|--------|--------|--------|--------|----------|
| Rainy | 25 | 25 | 25 | 25 | 100 |
| Non-rainy | 65 | 65 | 65 | 65 | 260 |
| $n_{.j}$ | 90 | 90 | 90 | 90 | 360 |

Now, we can calculate Chi-square statistic.

$$\begin{aligned}
\chi_{\text{obs}}^2 &= \frac{(34 - 25)^2}{25} + \frac{(32 - 25)^2}{25} + \frac{(15 - 25)^2}{25} + \frac{(19 - 25)^2}{25} \\
&\quad + \frac{(56 - 65)^2}{65} + \frac{(58 - 65)^2}{65} + \frac{(75 - 65)^2}{65} + \frac{(71 - 65)^2}{65} \\
&= \frac{81 + 49 + 100 + 36}{25} + \frac{81 + 49 + 100 + 36}{65} \\
&= \frac{266}{25} + \frac{266}{65} \\
&= \frac{3458 + 1330}{325} \\
&= 14.732
\end{aligned}$$

We have $(4 - 1) \cdot (2 - 1) = 3$ degrees of freedom. Accounting into three degrees of freedom, we can find the P -value by $P = \mathbf{P}\{\chi^2 \geq 14.732\} < 0.005$, since $\mathbf{P}\{\chi^2 \geq 12.8\} = 0.005$ as observed on the table. We have a significant evidence that the number of rainy days depends on the season, with a significance level above 99.5%, because P -value is less than 0.005.

Answer 4

The Octave code is as follows,

```
% Load statistics module for chi-square independence test
pkg load statistics;

% Define the matrix
X = [34 32 15 19; 56 58 75 71];

% Use library function
[pval, chisq, df] = chisquare_test_independence(X)
```

with a screenshot of some outputs,

```
>> the3_code
pval = 0.0020603
chisq = 14.732
df = 3
```

The findings agree with our calculations.