# Student Information

Name : Tahsin ELMAS

ID : 2476844

# Answer 1

## a)

To construct a 98% confidence interval for the mean per 100 km gasoline first need to compute the mean ($\mu$) and the standard deviation (s) of the sample.

The t-value for a 98% confidence interval with 15 degrees of freedom (which is 2.947) comes from a t-distribution table. We use the t-distribution because the population standard deviation is not known. So let's calculate:

The mean $\mu$ of the sample

$$\mu = \frac{1}{16}(8.4 + 7.8 + 6.4 + 6.7 + 6.6 + 6.6 + 7.2 + 4.1 + 5.4 + 6.9 + 7.0 + 6.9 + 7.4 + 6.5 + 6.5 + 8.5)$$

$$\mu = 6.80$$

Standard deviation (s) of the sample:

$$s = \sqrt{\frac{1}{16-1}\left((8.4-6.8)^2 + (7.8-6.8)^2 + \ldots + (6.5-6.8)^2 + (8.5-6.8)^2\right)} \approx 1.05$$

Finally, the 98% confidence interval is:

$$ConfidenceInterval = 6.80 \pm 2.947 \cdot \frac{1.05}{\sqrt{16}} = 6.80 \pm 0.77$$

So, the 98% confidence interval is approximately $(6.03, 7.57)$.

## b)

In order to test the effectiveness of the improvement, we can perform a one-sample t-test. We're testing if the mean gasoline consumption has decreased after the improvement. Therefore, we set up our null and alternative hypotheses as follows:

- Null Hypothesis, $H_0 : \mu = 7.5$.
- Alternative Hypothesis, $H_a : \mu < 7.5$. Now, we conduct the t-test. The $t$-statistic is given by:

$$t = \frac{6.80 - 7.5}{1.05/\sqrt{16}} = -2.66$$

With 15 degrees of freedom, the critical value for a one-sided test at the 5% level of significance from a t-distribution table or calculator is approximately -1.753 . The observed t-statistic $(-2.66)$ is less than the critical $t$-value $(-1.753)$, thus falling in the critical region. Therefore, we reject the null hypothesis. This indicates that there's significant statistical evidence at the 5% level of significance to support the claim that the engine improvement led to a reduction in gasoline consumption.

### c)

Without any calculations, we cannot immediately accept or reject the null hypothesis $H_0$ that the gasoline consumption before the improvement was 6.5 liters per 100 km. The sample data and calculations are necessary to assess the evidence and make a conclusion. Hypothesis testing requires statistical analysis to determine the significance of the difference between the observed data and the hypothesized value. Therefore, we need to perform the necessary calculations and hypothesis test to evaluate $H_0$.

## Answer 2

### a)

The choice of the null hypothesis $(H_0)$ often goes with the assumption of no change or no effect. So, in this question, Ali's claim that the rent prices are similar to those of the last year should be considered as the null hypothesis $(H_0)$. Ahmet's claim that there is an increase in the prices represents the alternative hypothesis $(H_a)$.

### b)

To test Ahmet's claim,

$$Z = (5500 - 5000)/(2000/\sqrt{100}) = 500/(2000/10) = 500/200 = 2.5$$

Our calculated $Z$ score of 2.5 is greater than the critical $Z$ score of 1.645 . This means that we would reject the null hypothesis at a 5% level of significance. Therefore, at a 5% level of significance, Ahmet can claim that there is an increase in the rent prices compared to the last year.

### c)

Given our $Z$ score of 2.5 , we find that the cumulative probability is approximately 0.9938 . This represents the probability of observing a value less than or equal to our test statistic. Since we're conducting a one-tailed test, we want the probability of observing a value greater than our test statistic,

$$P - value = 1 - 0.9938 = 0.0062$$

The $P$-value of 0.0062 is less than the significance level of 0.05 . This is further evidence against the null hypothesis. If the null hypothesis were true (the rents haven't increased), it would be very unlikely 0.62% to see an average rent as extreme as 5500TL in a sample of 100 houses. This strengthens Ahmet's claim that the average rent has indeed increased.

## d)

The test statistic is:

$$t = \frac{5500 - 6500}{\sqrt{\frac{2000^2}{100} + \frac{3000^2}{60}}} \approx -2.29$$

We compare the $Z$ score to the critical $Z$ score for a 1% level of significance. We look up the critical $Z$ score for a one-tailed test at the 1% level, which is approximately $-2.33$ .
$Z$ score of $-2.29$ is less than the critical $Z$ score of $-2.33$. This means that we would fail to reject the null hypothesis at a 1% level of significance. Therefore, at a 1% level of significance, they cannot claim that the rent prices in Ankara are lower than the prices in Istanbul.

# Answer 3

The null hypothesis $H_O$ is that there's no association between the number of rainy days and the season - in other words, the weather does not depend on the season. The alternative hypothesis $H_1$ is that there is an association - that the weather does depend on the season. Here are the observed frequencies from your data:

| | Winter | Spring | Summer | Autumn | Total |
|---|---|---|---|---|---|
| Rainy Days | 34 | 32 | 15 | 19 | 100 |
| Non-Rainy Days | 56 | 58 | 75 | 71 | 260 |
| Total | 90 | 90 | 90 | 90 | 360 |

The expected frequencies, assuming the null hypothesis is true. The expected frequency for each cell in the table is (row total * column total) / grand total.

| | Winter | Spring | Summer | Autumn |
|---|---|---|---|---|
| Rainy Days | 25 | 25 | 25 | 25 |
| Non-Rainy Days | 65 | 65 | 65 | 65 |

Next, we compute the chi-square test statistic, which is

$$\Sigma[(Observed - Expected)^2/Expected]$$

.

| | Winter | Spring | Summer | Autumn |
|---|---|---|---|---|
| Rainy Days | $(34-25)^2/25$ | $(32-25)^2/25$ | $(15-25)^2/25$ | $(19-25)^2/25$ |
| Non-Rainy Days | $(56-65)^2/65$ | $(58-65)^2/65$ | $(75-65)^2/65$ | $(71-65)^2/65$ |

The degrees of freedom for this test are (number of rows -1 ) * (number of columns -1)=$(2-1)*(4-1)=3$ A low P-value means that the probability of obtaining these results by chance alone is low. In this case, it would mean that the number of rainy days in Ankara is dependent on the season.

# Answer 4

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


total_observed = sum(observed(:));
row_totals = sum(observed, 2);
col_totals = sum(observed, 1);
expected = (row_totals * col_totals) / total_observed;


chi_square = sum((observed(:) - expected(:)).^2 ./ expected(:));

n_rows = size(observed, 1);
n_cols = size(observed, 2);
df = (n_rows - 1) * (n_cols - 1);


p = 1 - chi2cdf(chi_square, df);

disp(['Chi-square statistic X^2: ', num2str(chi_square)]);
disp(['P-value: ', num2str(p)]);
```

```
octave:27> pkg load statistics
% Input
observed = [34, 32, 15, 19; 56, 58, 75, 71]; % observed data

% Chi-Square Test
total_observed = sum(observed(:));
row_totals = sum(observed, 2);
col_totals = sum(observed, 1);
expected = (row_totals * col_totals) / total_observed;

% Calculate chi-square statistic
chi_square = sum((observed(:) - expected(:)).^2 ./ expected(:));

% Degrees of freedom
n_rows = size(observed, 1);
n_cols = size(observed, 2);
df = (n_rows - 1) * (n_cols - 1);

% P-value
p = 1 - chi2cdf(chi_square, df);

% Display Results
disp(['Chi-square statistic X^2: ', num2str(chi_square)]);
disp(['P-value: ', num2str(p)]);

Chi-square statistic X^2: 14.7323
P-value: 0.0020603
```

Q4 - The octave code and the result from octave online.