Student Information

Full Name : Berk Ulutaş
Id Number : 2522084

Answer 1

a)

We can use formula 9.9 from our textbook page 259 since we have small sample and we don't know population standard deviation.

$$CI = \bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

Assuming Normal distribution of these consumptions.

- The sample size $n = 16$
- Confidence $1 - \alpha = 0.98$ then $\alpha = 0.02$
- The sample mean consumption is $\bar{X} = \frac{8.4+7.8+\cdots+8.5}{16} \approx 6.81$
- The sample standard deviation is $s = \sqrt{\frac{\Sigma(x_i-\bar{x})^2}{n-1}} = \frac{(8.4-6.81)^2+(7.8-6.81)^2+\cdots+(8.5-6.81)^2}{16-1} \approx 1.06$
- The critical value of t distribution with $n - 1 = 15$ degrees of freedom is $t_{\alpha/2} = t_{0.01} = 2.602$ (using appendix table A5 from textbook)

$$CI = 6.81 \pm 2.602 \cdot \frac{1.06}{\sqrt{16}}$$
$$= [6.12, 7.50]$$

b)

To determine if the improvement in the engine is effective, we need to set up the null and alternative hypotheses and conduct a hypothesis test.

- H_0 : The improvement in the engine did not result in a significant reduction in the gasoline consumption. ($\mu = 7.5$)
- H_A : The improvement in the engine resulted in a significant reduction in the gasoline consumption. ($\mu < 7.5$)
- We will conduct a one-tailed hypothesis test using the given significance level $\alpha = 0.05$

To perform hypothesis test we can use t test formula from the textbook page 276, since we have the sample mean, sample standard deviation, and sample size.

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

We have:

- $\bar{X} = 6.81$ (sample mean)
- $\mu_0 = 7.5$ (population mean under null hypothesis)
- $s = 1.06$ (sample standard deviation)
- $n = 16$ (sample size)

Compute the T-statistic:

$$t = \frac{6.81 - 7.5}{1.06/\sqrt{16}} = -2.60$$

The rejection region $\mathcal{R} = (-\infty, -1.75]$, where we used T-distribution with $16 - 1 = 15$ degree of freedom and $\alpha = 0.05$ since we used left tail test ($t_\alpha = \text{tinv}(0.05, 15) = -1.75$ computed with octave online)

Since $t \in \mathcal{R}$, we reject the null hypothesis and conclude that there is significant evidence of the improvement resulted reduction in the gasoline consumption.

c)

If we assume car was consuming 6.5 liters of gasoline per 100 km. Our null and alternative hypothesis are followings:

- $H_0 : \mu = 6.5$
- $H_A : \mu < 6.5$

Alternative $H_A : \mu < 6.5$ covering the region to the left of H_0 is one-sided, left-tail. Since it is left tail, the rejection are in negative side.

Let's analyze the the test statistic

$$t = \frac{6.81 - 6.5}{1.06/\sqrt{16}}$$

- Numerator ($6.81 - 6.5 > 0$) and Denominator ($1.06/\sqrt{16} > 0$)
- Without any statistic test calculation we can easily see that $t > 0$
- We said that our rejection area in negative side and we know that $t > 0$. So, $t \notin \mathcal{R}$
- Since we know our test statistic is not in rejection region we can immediately accept H_0 without any statistic test calculation.

Answer 2

a)

- $H_0 : \mu = 5000$ (the prices are similar to those of the last year.)
- $H_A : \mu > 5000$ (there is an increase in the prices.)

Ali's claim should be considered as null hypothesis.

b)

We test the null hypothesis $H_0 : \mu = 5000$ against a one sided right-tail alternative $H_A : \mu > 5000$, since we only interested to know if the mean number of rent prices has increased with significance level $\alpha = 0.05$

To perform hypothesis test we can use z test formula from the textbook page 273, since we have one sample and we know population mean and standard deviation.

$$z = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

We have:

- $\bar{X} = 5500$ (sample mean)
- $\mu_0 = 5000$ (mean)
- $s = 2000$ (standard deviation)
- $n = 100$ (sample size)

$$z = \frac{5500 - 5000}{2000/\sqrt{100}} = 2.5$$

The critical value is (norminv(0.05) computed with octave online and normal distribution is symmetric):

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

With right-tail alternative the rejection region $\mathcal{R} = [1.645, \infty)$.

Since our test statistic $z = 2.5 \in \mathcal{R}$, we reject the null hypothesis. The data provided sufficient evidence in favor of the alternative hypothesis. So, Ahmet can claim that there is an increase in the rent prices compared to the last year at a 5% level of significance.

c)

In part b we have found at 5% level of significance Ahmet can claim that increase in the rent prices compared to the last year. Let us compute the P-value.

- We have found Z-statistic in part b
$$Z_{obs} = 2.5$$
- To compute P-value we can use formula from Table 9.3 page 283 in our textbook.
- We find that the P-value for the right-tail alternative is: ($\Phi(2.5) = 0.9938$ taken from textbook appendix table A4)

$$P = \mathbf{P}\{Z \geq Z_{obs}\} = \mathbf{P}\{Z \geq 2.5\} = 1 - \Phi(2.5) = 0.0062$$

The P-value is very low, a low P-value indicates that such an extreme test statistic is unlikely if H_0 is true. Therefore, P-value supports that Ahmet can reject the null hypothesis.

d)

Let's start with determining null and alternative hypothesis. Subscript x represents Ankara a and subscript y represents Istanbul.

- $H_0 : \mu_x = \mu_y$ (the prices in Ankara is equal to the prices in Istanbul)
- $H_A : \mu_x < \mu_y$ (the prices in Ankara is less than the prices in Istanbul)

We test the null hypothesis $H_0 : \mu_x = \mu_y$ against one sided left-tail alternative $H_A : \mu_x < \mu_y$ with significance level $\alpha = 0.01$

We have two-sample and to perform hypothesis test we can use z test formula from the textbook page 273.

$$z = \frac{\bar{X} - \bar{Y} - D}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

We have:

- $\bar{X} = 5500$
- $\sigma_X = 2000$
- $n = 100$
- $\bar{Y} = 6500$
- $\sigma_Y = 3000$
- $m = 60$

The tested value is $D = 0$. The test statistic the equals

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

Let's calculate acceptance and rejection regions. This is a left tail test. The critical value is:

$$z_\alpha = z_{0.01} = -2.326$$

With left-tail alternative the rejection region $\mathcal{R} = (-\infty, -2.326]$

Since our test statistic $z = -2.29 \notin \mathcal{R}$, we cannot reject H_0. The evidence against H_0 is insufficient. They cannot claim that the prices in Ankara is lower than the prices in Istanbul.

Answer 3

Let's start with determining null and alternative hypothesis

- H_0 : number of rainy days in Ankara is independent of the season
- H_A : number of rainy days in Ankara is dependent of the season

Here is the obtained data:

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

Testing independence, we compute the estimated expected counts with this formula which is given in our textbook page 312:

$$\widehat{Exp}(i,j) = \frac{(n_{i\cdot})(n_{\cdot j})}{n}$$

- $\widehat{Exp}(1,j) = \frac{90 \cdot 100}{360} = 25$ for $j = 1,2,3,4$
- $\widehat{Exp}(2,j) = \frac{90 \cdot 260}{360} = 65$ for $j = 1,2,3,4$

Expected	Winter	Spring	Summer	Autumn	Total
Rainy	25	25	25	25	100
Non-rainy	65	65	65	65	260
Total	90	90	90	90	360

Then we can use the formula from our textbook page 312 for calculating test statistic:

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

The test statistic is:

$$\chi^2_{obs} = \frac{(34-25)^2}{25} + \cdots + \frac{(71-65)^2}{65} = 14.73$$

and it has $(2-1)\cdot(4-1) = 3$ degrees of freedom.

We can calculate p value using octave online:

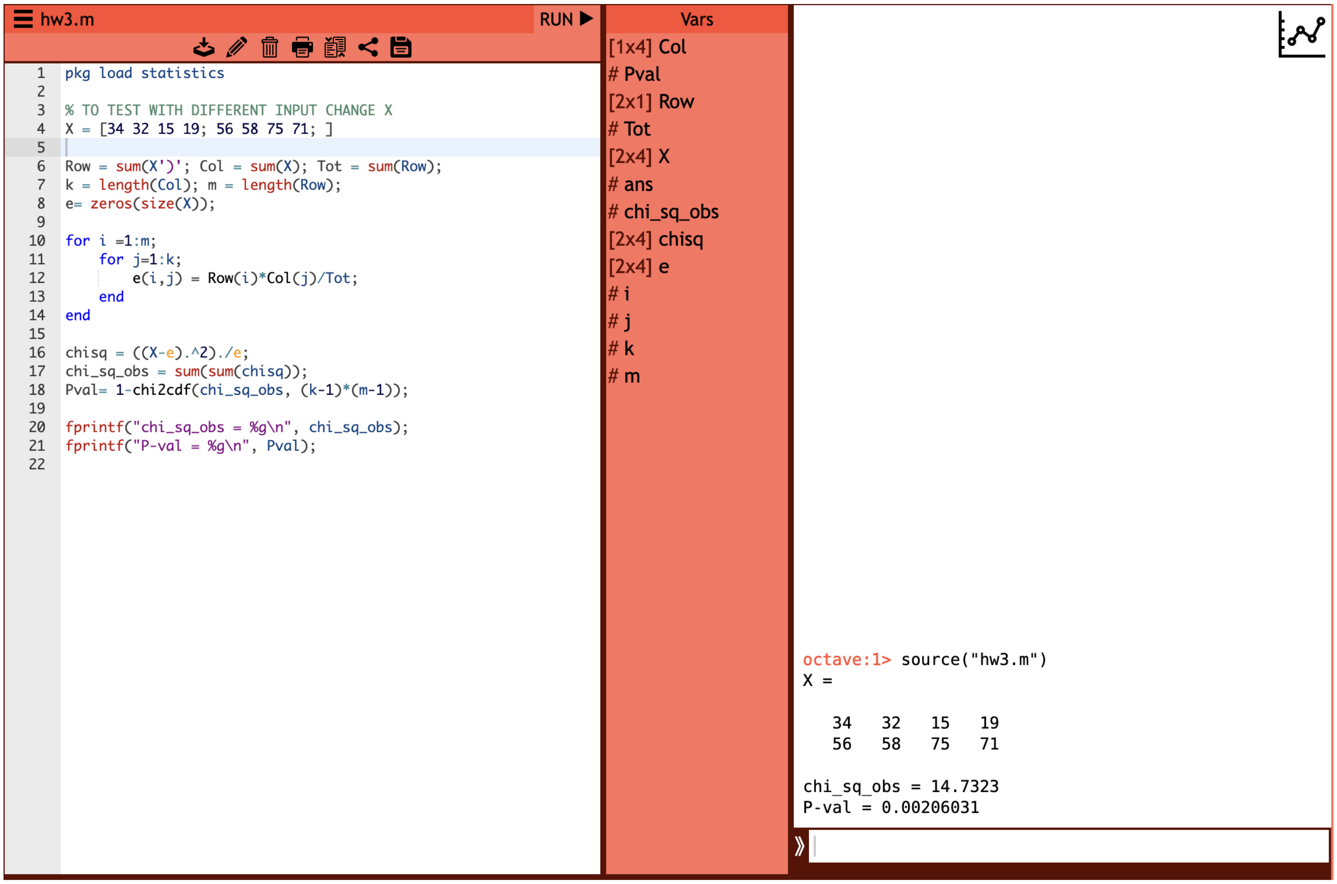
$$\mathrm{P} = 1 - \mathit{chi2cdf}(14.73, 3) \approx 0.002$$

According to table in our textbook page 282. Practically we can reject H_0 if $P < 0.01$.

Based on our calculation, the obtained p-value of 0.002 is indeed less than 0.01. Therefore, we have significant evidence to reject the null hypothesis, which suggests that there is no relationship between the number of rainy days in Ankara and the season. Instead, the evidence supports the alternative hypothesis, indicating that the number of rainy days in Ankara is dependent on the season

Answer 4

```
pkg load statistics

% TO TEST WITH DIFFERENT INPUT CHANGE X
X = [34 32 15 19; 56 58 75 71; ]

Row = sum(X')'; Col = sum(X); Tot = sum(Row);
k = length(Col); m = length(Row);
e= zeros(size(X));

for i =1:m;
    for j=1:k;
        e(i,j) = Row(i)*Col(j)/Tot;
    end
end

chisq = ((X-e).^2)./e;
chi_sq_obs = sum(sum(chisq));
Pval= 1-chi2cdf(chi_sq_obs, (k-1)*(m-1));

fprintf("chi_sq_obs = %g\n", chi_sq_obs);
fprintf("P-val = %g\n", Pval);
```


hw3.m
RUN
pkg load statistics

% TO TEST WITH DIFFERENT INPUT CHANGE X
X = [34 32 15 19; 56 58 75 71;]

Row = sum(X')'; Col = sum(X); Tot = sum(Row);
k = length(Col); m = length(Row);
e= zeros(size(X));

for i =1:m;
    for j=1:k;
        e(i,j) = Row(i)*Col(j)/Tot;
    end
end

chisq = ((X-e).^2)./e;
chi_sq_obs = sum(sum(chisq));
Pval= 1-chi2cdf(chi_sq_obs, (k-1)*(m-1));

fprintf("chi_sq_obs = %g\n", chi_sq_obs);
fprintf("P-val = %g\n", Pval);
Vars
[1x4] Col
Pval
[2x1] Row
Tot
[2x4] X
ans
chi_sq_obs
[2x4] chisq
[2x4] e
i
j
k
m
octave:1> source("hw3.m")
X =
34 32 15 19
56 58 75 71
chi_sq_obs = 14.7323
P-val = 0.00206031