Hypothesis testing

Problems and Solutions

Question 1:

The 95% confidence for the difference of proportions of students who uses their own vehicle and who use the school bus to reach their school is (0.03, 0.08). Test the claim that the proportion of students who uses their own vehicle is equal to the proportion of students who use school bus.

Solution 1:

From the given information,
Let S_1 represents the proportion of students who use their own vehicle.
Let S_2 represents the proportion of students who use school bus.

The claim is to test whether the proportion of students who uses their own vehicles is equal to the proportion of students who use school bus.

Null and alternative hypothesis:
H_0 : There is no significant difference between proportion of students who use their own vehicle and students who use school bus.

$$H_0 : P_1 - P_2 = 0$$

H_1 : There is significant difference between proportion of students who use their own vehicle and students who use school bus.

$$H_1 : P_1 - P_2 \neq 0$$

The confidence level under study is 0.95 .
Hence, the level of significance is,

$$1 - \alpha = 0.95$$
$$\alpha = 1 - 0.95$$
$$\alpha = 0.05$$

It is clearly stated that the 95% confidence interval for the difference of proportions of students who uses their own vehicles and who use the school bus to reach school is (0.03,0.08) and the hypothesized value is 0.

Here,

The lower limit of the 95% confidence interval is **0.03**.
The upper limit of the 95% confidence interval is **0.08**.

Since, the hypothesized value of 0 does not **lie** within the confidence interval.

Therefore, the null hypothesis is rejected.
Hence, there is no sufficient evidence to conclude that the proportion of students who uses their own vehicles is equal to the proportion of students who use school bus at 0.05 level of significance.

Question 2:

The 99% confidence interval for proportion of students who quit their education after school level because of financial problems is (0.1, 0.18). Test the claim that the percentage of student who quit their education is equal to 12%.

Solution 2:

From the given information,

The objective is to test whether the percentage of students who quit their education after school level because of financial problems is 20% or not.

Null hypothesis and alternative hypothesis:

H_0: The percentage of students who quit their education after school level is 20%

$$H_0 : P = 20$$

H_1: the percentage of students who quit their education after school level is not equal to 20%

$$H_0 : P \neq 20$$

From the given information,

The confidence level under study is 0.99
Hence, the level of significance is calculated as follows:

$$1 - \alpha = 0.99$$
$$\alpha = 1 - 0.99$$
$$\alpha = 0.01$$

From the given information,
Let P = 0.2 represents the hypothesized proportion under study.
It is clearly stated that the 99% confidence interval for proportions of students who quit their education because of financial problems is (0.1, 0.18)
.
Therefore, the lower limit of the 99% confidence interval is 0.1.
The upper limit of the 99% confidence interval is 0.18.

Since, the hypothesized value for the testing the proportion 0.2 does not lies within the confidence interval.
Therefore, the null hypothesis is rejected.

Hence, there is no sufficient evidence to say that 20% of students who quit their education after school level because of financial problems at 0.01 level of significance.

Question 3:

The 90% confidence interval for the average height of the female students in the university of the California is (149, 167). Test the claim that the average height of female students is 155 cm.

Solution 3:

From the given information, observe that the 90% confidence interval for the average height of the female students in the university of the California is (149, 167).

Therefore, the objective is check whether the average height of female students is 155cm.

Null hypothesis and alternative hypothesis:

H_0 : Average height of the female students is 155cm.

$$H_0 : \mu = 155cm$$

H_1 : Average height is not equal to 155cm.

$$H_1 : \mu \neq 155cm$$

It is given that the 90% confidence interval for the average height of the female students in the university of the California is (149, 167) and the hypothesized value is 155.
As the hypothesized value of 155 lies between the lower limit 149 and upper limit 167, so we can not reject the null hypothesis.

But there is insufficient evidence to say that the average height of female students is 155 cm 0.1 level of significance.

Question 4:

Eight students have join the mathematics tuition to improve the score in mathematics. The following data was recorded before joining the tuition and after joining the tuition relating to their score in mathematics tests (out of 50). Can we conclude that the tuitions has made significant improvement impact? Test at 0.1 level of significance.

The data is given below:

Students	1	2	3	4	5	6	7	8
Score before joining the tuition	23	38	29	45	39	38	19	37

Score after joining the tuition	36	40	44	43	48	49	32	33

Solution 4:

From the given information,

Null hypothesis and alternative hypothesis:

H_0 : There is no significant difference between scores before joining the tuition and after joining the tuition.

$$H_0 : \mu_d = 0$$

H_1 : There is significant difference between scores before joining the tuition and after joining the tuition.

$$H_1 : \mu_d \neq 0$$

Sample size, n = 8.

The differences between the scores before joining the tuition and after joining the tuition are obtained as follows:

Students	Score before joining the tuition	Score after joining the tuitions	Difference (d_i)	Difference ^{2}
1	23	36	-13	169
2	38	40	-2	4
3	29	44	-15	225
4	45	43	2	4
5	39	48	-9	81
6	38	49	-11	121

7	19	32	-13	169
8	37	33	4	16
Sum			-57	789

The mean of the difference is,

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$= -\frac{57}{8}$$

$$= -7.125$$

The standard deviation of the errors is,

$$s_d = \sqrt{\left(\frac{1}{n-1}\right)\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2}$$

$$= \sqrt{\frac{(-13--7.125)^2+(-2--7.125)^2+\ldots+(4--7.125)^2}{8-1}}$$

$$= \sqrt{163.872}$$

$$= 12.801$$

The standard deviation of the errors is 12.801.

The degrees of freedom,

$$
\begin{aligned}
df &= n - 1 \\
&= 8 - 1 \\
&= 7
\end{aligned}
$$

From the t table critical value of the test is,

$$
\begin{aligned}
t_{0.1,n-1} &= t_{0.1,7} \\
&= -1.415
\end{aligned}
$$

The t test statistics value is,

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

$$= \frac{-7.125 - 0}{\frac{12.801}{\sqrt{8}}}$$

$$= -1.574$$

The test statistic value lies in the critical region. So reject the null hypothesis. There is insufficient evidence to conclude that the tuitions has made significant improvement impact.

Question 5:

The department of transportation of UK considers sample of 6 metro cities in UK and collect the data on average number of road accidents in two year, before and after pandemic. Transportation ministry believes that there is significant difference in average number of road accidents in year 2019 and 2020 after lockdown in pandemic situation . Assume that the difference between two samples follow normal distribution.

Cities	1	2	3	4	5	6
Average of road accidents in month of 2018 (μ_1)	120	142	105	167	109	94

Average of road accidents in month of 2019 (μ_2)	139	123	89	144	97	68

Can we conclude that there is effect of lockdown on road accidents? Test $\alpha = 0.05$

Solution 5:

From the given information,
Null hypothesis and alternative hypothesis:
H_0 : There is no significant difference between average number of road accidents in year 2018 and 2019.

$$H_0 : \mu_1 - \mu_2 = 0$$

H_1 : There is significant difference between average number of road accidents in year 2018 and 2019.

$$H_1 : \mu_1 - \mu_2 \neq 0$$

Sample size, n = 6.

The differences between the average number of accidents in the year 2018 and 2019 as follows:

Cities	Average of road accidents in month of 2018 (μ_1)	Average of road accidents in month of 2019 (μ_2)	Difference (d_i)	Difference ^{2} $(d_i)^2$
1	120	139	-19	361
2	142	123	19	361
3	89	105	-16	256
4	167	144	23	529
5	97	109	-12	144
6	94	68	26	676
Sum			21	2327

The mean of the difference is,

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$= \frac{21}{6} = 3.5$$

The standard deviation is,

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}}$$

$$= \sqrt{\frac{2327 - \left(\frac{441}{6}\right)}{5}}$$

$$= \sqrt{\frac{2327-73.5}{5}}$$

$$= \sqrt{450.7}$$

$$= 21.23$$

Degree of freedom is,

$$df = n - 1$$

$$= 6 - 1$$

$$= 5$$

From the t table critical value of the test is,

$$\begin{aligned} t_{0.05,n-1} &= t_{0.05,5} \\ &= 2.57 \end{aligned}$$

The t test statistics is,

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

$$= \frac{3.5 - 0}{\frac{21.23}{\sqrt{6}}}$$

$$= \frac{3.5}{8.668}$$

$$= 0.404$$

The calculated t statistics value is 0.404.

Compare the absolute calculated value with the critical value.

Here, the absolute calculated value of the test statistic (tcal=0.404) is less than the critical value of the test statistic ($t_{0.05/2,\,5}$ = 2.57), so fail to reject the null hypothesis.

Hence, we can conclude that there is an insufficient evidence that there is significant difference on average number of road accidents before and after lockdown due to pandemic situation.