The Accomodation Revenue Problem

How to Increase the Revenue Generated Through Accommodation During Fests

Indian Institute of Technology Hyderabad

02-05-2024

Aim of the Project

Every year, during the college fest, our institute opens to the outsiders. People are invited to come and visit the campus and be a part of our fest. Some of these people chose to pay and stay in the institute guest house. This turns out to be a considerable source of revenue for the team organising the fest.

Our aim is to analyse the data and provide insights which can help increase the amount of people who stay in the campus. We are trying to achieve this through various hypothesis tests and analysing data plots.

Aim of the Project

The assumptions we have in mind before doing our analysis are:

- People from farther away stay more than those who come from nearby.
- People who are older generally tend to stay more as they are able to spend more and have more freedom than the younger ones
- People who study in private colleges visit more than those who study in government colleges.

Our goal is to debunk or confirm these assumptions and probably gain some more insights into the data.

Data Collection and Parameters

The data was collected from a randomly selected group of 158 people who applied for accommodation during Elan and nVision 2024, through a survey.

The data parameters collected from the group were:

- Residential Address
- College/Institute and whether it is public or private
- Age
- The number of days they were staying

Dataset

	Permanent address	distance from residence	Unnamed: 2	Age	Gender	Direction of college	Distance from college	Institution / Organization	private college or public	Category	Duration of stay (check the form description for check in and check out timings)	15th, 16th and 17th March	16th and 17th March
0	[redacted]	502.0	se	18	1	se	56.0	[redacted]	Private	Participant (in either of Workshops / Techy ev...	15th, 16th and 17th March	True	False
1		584.0	s	20	0	s	593.0		Public	Participant (in either of Workshops / Techy ev...	15th, 16th and 17th March	True	False
2		530.0	s	20	0	s	593.0		Public	Participant (in either of Workshops / Techy ev...	15th, 16th and 17th March	True	False
3		530.0	se	19	1	se	56.0		Private	Participant (in either of Workshops / Techy ev...	15th, 16th and 17th March	True	False
4		599.0	s	19	0	s	593.0		Public	Participant (in either of Workshops / Techy ev...	15th, 16th and 17th March	True	False

Figure: First 5 rows of the data

[redacted] Edited for security reasons

Central Tendencies

	distance from residence	Age	Gender	Distance from college
count	158.000000	158.000000	158.000000	151.000000
mean	371.158228	19.208861	0.525316	343.917881
std	312.991164	2.644677	0.500946	341.738894
min	8.000000	14.000000	0.000000	8.000000
25%	60.000000	18.000000	0.000000	49.000000
50%	350.000000	19.000000	1.000000	270.000000
75%	618.750000	20.000000	1.000000	600.000000
max	1954.000000	40.000000	1.000000	1822.000000

Figure: Central tendencies for data

Analysis of the effect of distance from college/residence

Introduction

Abstract

This section presents an analysis of the effect of distance on people visiting the college fest. We conduct a hypothesis test to determine the mean distance that is traveled by those who participate in the fest

Analysis of the effect of distance from college/residence

Data visualisation

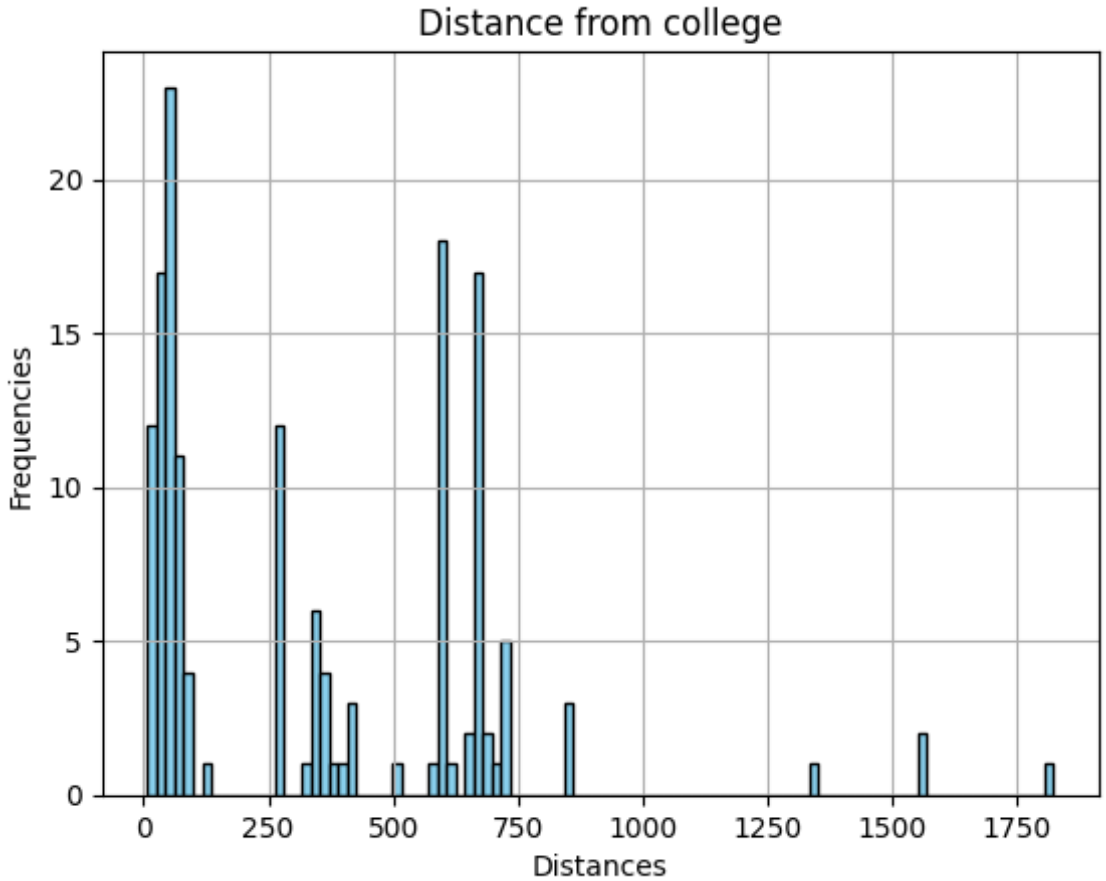


Figure: Sample distance from college

Can see that the sample is skewed

Verification of CLT for Sample

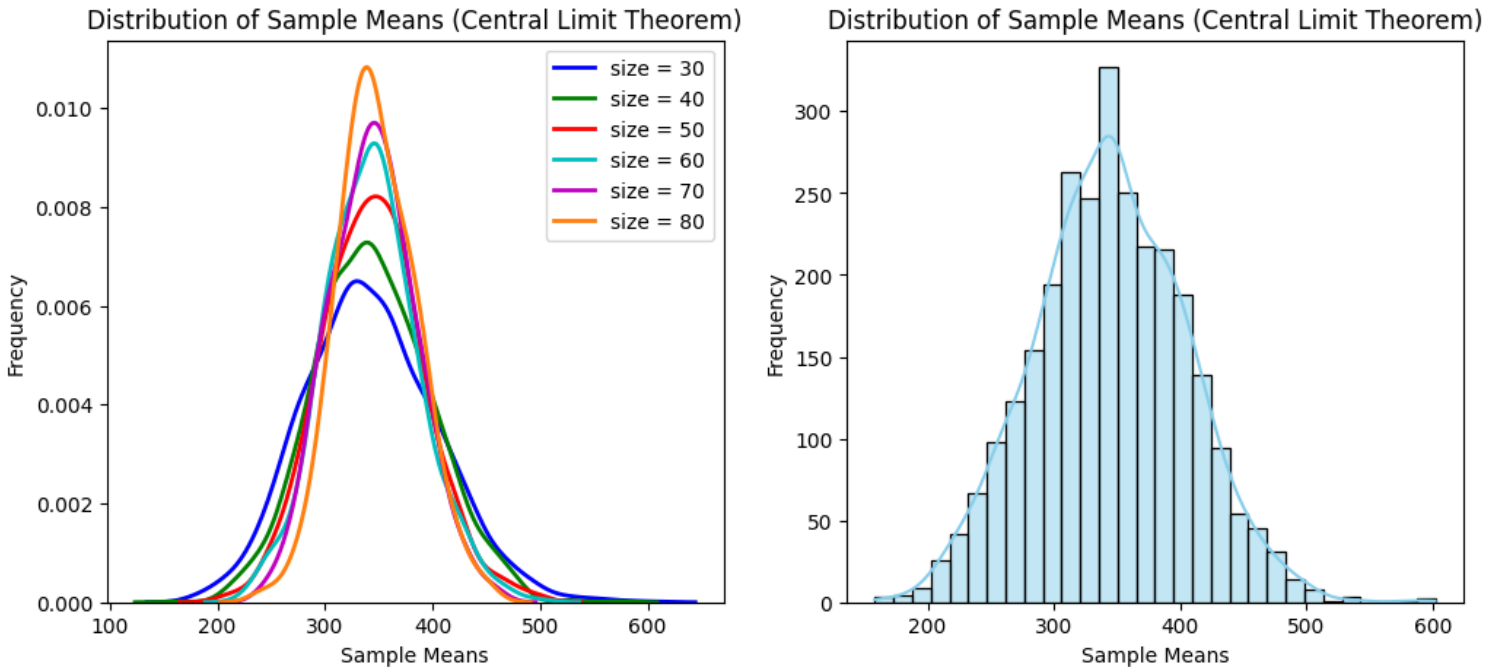


Figure: Visualisation of CLT for College Distance

Verification of CLT for Sample

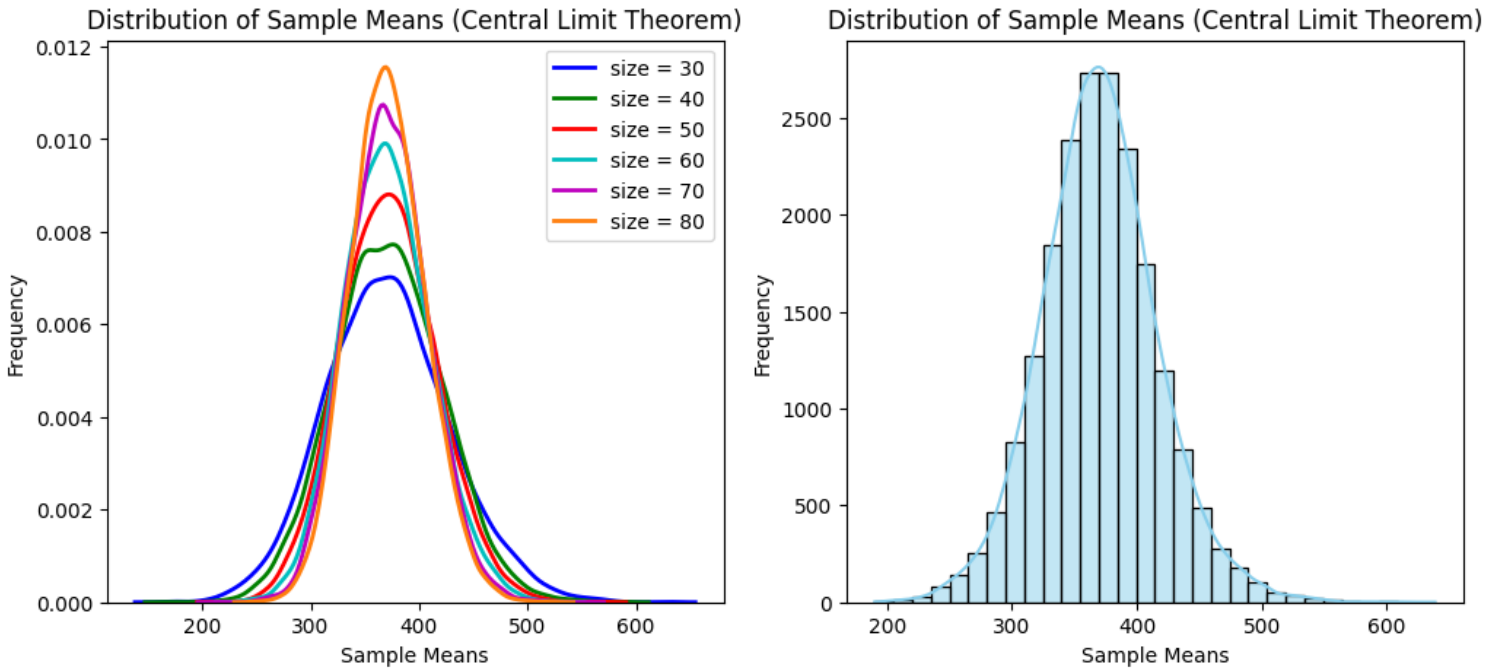


Figure: Visualisation of CLT for Resident Distance

Confidence Interval

95% Confidence Interval for Mean (μ) of Residential Distances:
$(321.9755, 420.3409)$

Confidence Interval

95% Confidence Interval for Mean (μ) of College Distances:
$(288.9673, 398.8684)$

Analysis of the effect of distance from college/residence

Hypothesis Testing

Hypothesis Test

Mean distance of attendees' college is not 300 km

$$\text{Null Hypothesis}(H_0) : \mu_0 = 300$$

$$\text{Alternate Hypothesis}(H_a) : \mu_0 \neq 300$$

To test the hypothesis, we can use the two-tailed t-test. The test statistic is given by:

$$t^* = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

where:

- $\bar{X} = 343.917$; the Sample Mean
- $S = 341.739$; is the Sample Standard Deviations
- $\mu_0 = 300$; the hypothesized mean
- $df = 150$; the Degree of Freedom
- $\alpha = 0.05$; the confidence coefficient

Using the above formula, we get :

$$t^* = 1.579$$

Also $|t_{\alpha/2,df}| = 1.975$
Since $|t^*| < |t_{\alpha/2,df}|$,

We fail to reject the null hypothesis.

Therefore, there isn't enough statistical evidence to show that the mean distance is not 300. Therefore, the mean distance is not too different from 300.
For p-value test,

$$p = 2^* P(t \geq t^*) = 0.116$$

$$\alpha = 0.05$$

Since $p > \alpha$, we fail to reject the null hypothesis.

Analysis of the effect of distance from college/residence

Hypothesis Testing

Hypothesis Test

Proportion of distances travelled by people being ≤ 100 is at least 25%

$$\text{Null Hypothesis}(H_0) : p < p_0$$

$$\text{Alternate Hypothesis}(H_a) : p \geq p_0$$

To test the hypothesis, we can use a one-proportion test.

$$Z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where:

- $n = 158$; the number of samples
- $\hat{p} = 0.335$ is the sample proportion
- $p_0 = 0.25$ is the hypothesized proportion
- $\alpha = 0.05$; the confidence coefficient

Using the above formula, we get :

$$Z* = 2.480$$

$$Z_\alpha = 1.645$$

Since

$$|Z^*| < |Z_\alpha|$$

,

We reject the null hypothesis.

For p-value test,

$$p = P(t \geq t^*) = 0.007$$

$$\alpha = 0.05$$

Since $p < \alpha$, we reject the null hypothesis.

Analysis of the effect of distance from college/residence

Result and Conclusion

Result

The mean distance of attendees' college is not significantly different from 300.
We can say with 95% confidence that the proportion of people traveling less than 100 km is at least 25%

Conclusion

The result of the hypothesis test indicates that at least a fourth of the population is within 100 km. Thus this is a good radius to market the fest in.

Analysis of guest stay duration and college distance

Abstract

In this section, we conduct an analysis to investigate the relationship between the duration of guest stays during our college fest and the distance they travel from. Our primary objective is to explore whether guests staying for 3 days come from farther away on average compared to those staying for 2 days.

Analysis of guest stay duration and college distance

	3 Day Distances	2 Day Distances
count	30.000000	30.000000
mean	460.980000	311.690000
std	370.448375	267.177005
min	28.000000	34.000000
25%	277.750000	56.000000
50%	399.000000	266.500000
75%	605.000000	544.250000
max	1954.000000	837.000000

Figure: Central tendencies

Analysis of guest stay duration and college distance

Hypothesis Test

Guests staying for 3 days come from farther away on average compared to those that stay for 2 days.

Analysis of guest stay duration and college distance

Null and Alternate Hypothesis

$$\text{Null Hypothesis}(H_0) : \mu_1 - \mu_2 \leq 0$$

$$\text{Alternate Hypothesis}(H_a) : \mu_1 - \mu_2 > 0$$

To test the hypothesis, we can use Welch's t-test, which is appropriate when the variances of the two populations are unknown and unequal. The test statistic is given by :

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Analysis of guest stay duration and college distance

Null and Alternate Hypothesis

where:

- $\bar{X}_1$: Sample Mean of distances of people coming to stay for 3 days
- $\bar{X}_2$: Sample Mean of distances of people coming to stay for 2 days
- S_1 : Sample SD for 3 days
- S_2 : Sample SD for 2 days
- n_1 : Number of samples for 3 days
- n_2 : Number of samples for 2 days
- $\alpha = 0.05$; the confidence coefficient

The degrees of freedom for the above t-distribution is given by:

$$\text{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Using the above formula, we get :

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t^* = 1.82$$

Also,

$$t_{\alpha,df} = 1.67$$

Since

$$t^* > t_{\alpha,df}$$

,

we have enough evidence to reject the null hypothesis.

Result

After conducting the hypothesis test, we find that in the case where the variances are assumed to be unequal the calculated value of the test statistic is $t = 1.83$, and the corresponding critical value t_α is 1.67. Since $t > t_\alpha$, we reject the null hypothesis. This suggests that there is evidence to support the idea that guests staying for 3 days come from farther away on average compared to those staying for 2 days.

Conclusion

Based on the results of our analysis, we conclude that there is a significant difference in the distances traveled by guests staying for different durations during our fest. Understanding this relationship can inform our fest planning and accommodation arrangements to better cater to the needs of our guests.

Preferences in Hostellers and Day Scholars

Abstract

This section aims to investigate whether people who are hostellers are more likely to opt for accommodation than day scholars.

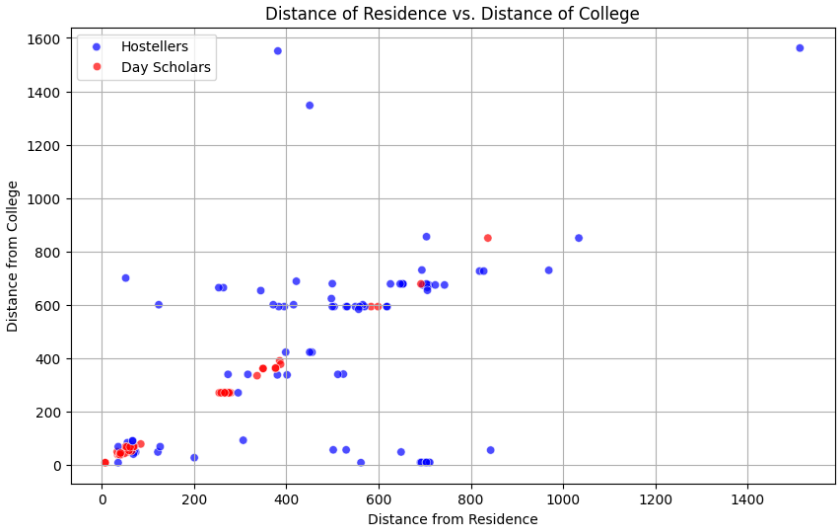
Preferences in Hostellers and Day Scholars

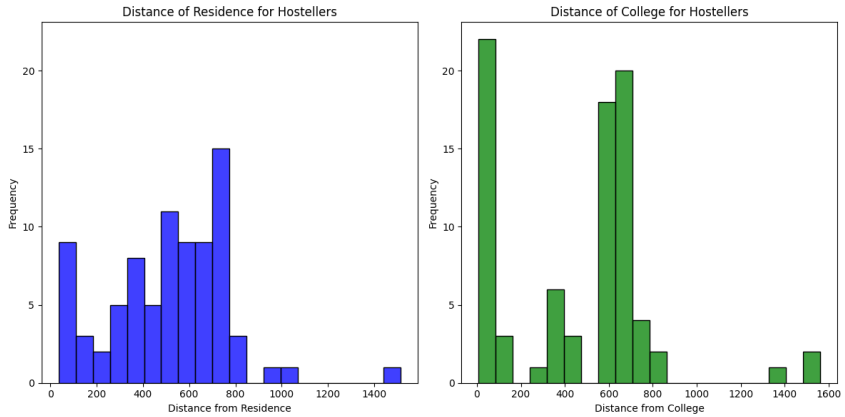
Preparation of data

Assumption

Found distance between residential address and college address. Created binary variable 'hosteller'. If the distance between the residential address and college address is more than 20, then assigned 1 (hosteller), otherwise 0 (non-hosteller).

Preferences in Hostellers and Day Scholars





Preferences in Hostellers and Day Scholars

Hypothesis Test

People who are hostellers are more likely to opt for accommodation than day scholars

$$\text{Null Hypothesis}(H_0) : p \leq p_0$$

$$\text{Alternate Hypothesis}(H_a) : p > p_0$$

To test the hypothesis, we can use a one-proportion test.

$$Z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where:

- n : the number of samples
- $\hat{p}$: the sample proportion
- $p_0 = 0.5$: the hypothesized proportion
- $\alpha = 0.05$: the confidence coefficient

We use $p_0 = 0.5$ since if the Hypothesis is true, then hostellers must be a majority.

Using the above formula, we get :

$$Z* = 2.307$$

$$Z_\alpha = 1.645$$

Since

$$|Z^*| > |Z_\alpha|$$

,

We reject the null hypothesis.

For p-value test,

$$p = P(t \geq t^*) = 0.011$$

$$\alpha = 0.05$$

Since $p < \alpha$, we reject the null hypothesis.

Preferences in Hostellers and Day Scholars

Result and Conclusion

Result

As the test statistic is greater than the critical value and the p-value is less than the significance level α, we reject the null hypothesis.

Conclusion

Therefore, we conclude that people who are hostellers are more likely to opt for accommodation than day scholars. It may be beneficial to direct marketing towards hostellers.

Analysis of the effect of age on stay duration

Abstract

This section presents an analysis of the effect of age on the duration of stay. We conducted hypothesis testing to determine if there is a significant difference in hotel stay duration between different age groups.

Analysis of the effect of age on stay duration

Hypothesis Test

Determine if there is a significant difference in hotel stay duration between Group 1 and Group 2

Analysis of the effect of age on stay duration

Null and Alternate Hypothesis

$$\text{Null Hypothesis}(H_0) : \mu_1 - \mu_2 = 0$$

$$\text{Alternate Hypothesis}(H_a) : \mu_1 - \mu_2 \neq 0$$

Where μ_1 represents the mean of the age group staying for 3 days and μ_2 represents the mean of the age group staying for 2 days where:

Now, we need to test the hypothesis in two conditions,

- when both the variances are unequal and unknown.
- when both the variances are equal but unknown.

Analysis of the effect of age on stay duration

Unequal and Unknown Variances

To test the hypothesis, we can use Welch's t-test, which is appropriate when the variances of the two populations are unknown and unequal. The test statistic is given by :

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where:

- $\bar{X}_1 = 18.7$: Sample Mean of distances of people coming to stay for 3 days
- $\bar{X}_2 = 19.03$: Sample Mean of distances of people coming to stay for 2 days
- $S_1 = 1.71$: Sample SD for 3 days
- $S_2 = 1.49$: Sample SD for 2 days
- $n_1 = 30$: Number of samples for 3 days
- $n_2 = 30$: Number of samples for 2 days
- $\alpha = 0.05$; the confidence coefficient

The degrees of freedom for the above t-distribution is given by:

$$\text{df} = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Using the above formula, we get :

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$|t^*| = 0.903$$

$$|t_{\alpha/2,df}| = 2.003$$

Therefore, $|t^*| < |t_{\alpha/2,df}$ We fail to reject the null hypothesis.

Analysis of the effect of age on stay duration

Unknown but Equal Variances

To test the hypothesis, we can use pooled t-test, which is appropriate when the variances of the two populations are unknown and equal. The test statistic is given by:

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where:

- $\bar{X}_1 = 18.7$; Sample Mean of Sample 1
- $\bar{X}_2 = 19.03$; Sample Mean of Sample 2
- $S_p = 1.61$; Pooled Sample SD
- $n_1 = 30$; Number of samples in Sample 1
- $n_2 = 30$; Number of samples in Sample 2
- $\alpha = 0.05$; the confidence coefficient

Analysis of the effect of age on stay duration

Unknown but Equal Variances

The degrees of freedom for the above t-distribution is given by:

$$\text{df} = n_1 + n_2 - 2$$

Using the above formula, we get :

$$t^* = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t^* = -0.8$$

$$|t^*| = 0.8$$

$$|t_{\alpha/2,df}| = 2.001$$

Since $|t^*| < |t_{\alpha/2,df}|$,
We fail to reject the null hypothesis.

Analysis of the effect of age on stay duration

Result and Conclusion

Result

The result of the hypothesis test indicates that there is no significant difference in the mean of age group that stays for 3 days and the one that stays for 2 days. Although we have a narrow interval of ages that visit the campus, there is not much distinction as to who stay longer.

Conclusion

Our analysis indicates that age has no significant effect on stay duration.

Analysis of Accommodation of Private/Public Institutions

Abstract

This section aims to investigate whether students from private institutions have a higher likelihood of opting for accommodation compared to students from public institutions in Telangana.

Analysis of Accommodation of Private/Public Institutions

```
Public Institutions:
count     40.000000
mean     405.447500
std      168.986648
min       52.400000
25%      267.000000
50%      390.000000
75%      557.500000
max      706.000000
Name: distance from residence, dtype: float64
```

Figure: Central tendencies for public institutions

```
Private Institutions:
count     117.000000
mean      357.103419
std       348.914287
min         8.000000
25%        51.000000
50%       317.000000
75%       653.000000
max      1954.000000
Name: distance from residence, dtype: float64
```

Figure: Central tendencies for private institutions

Analysis of Accommodation of Private/Public Institutions

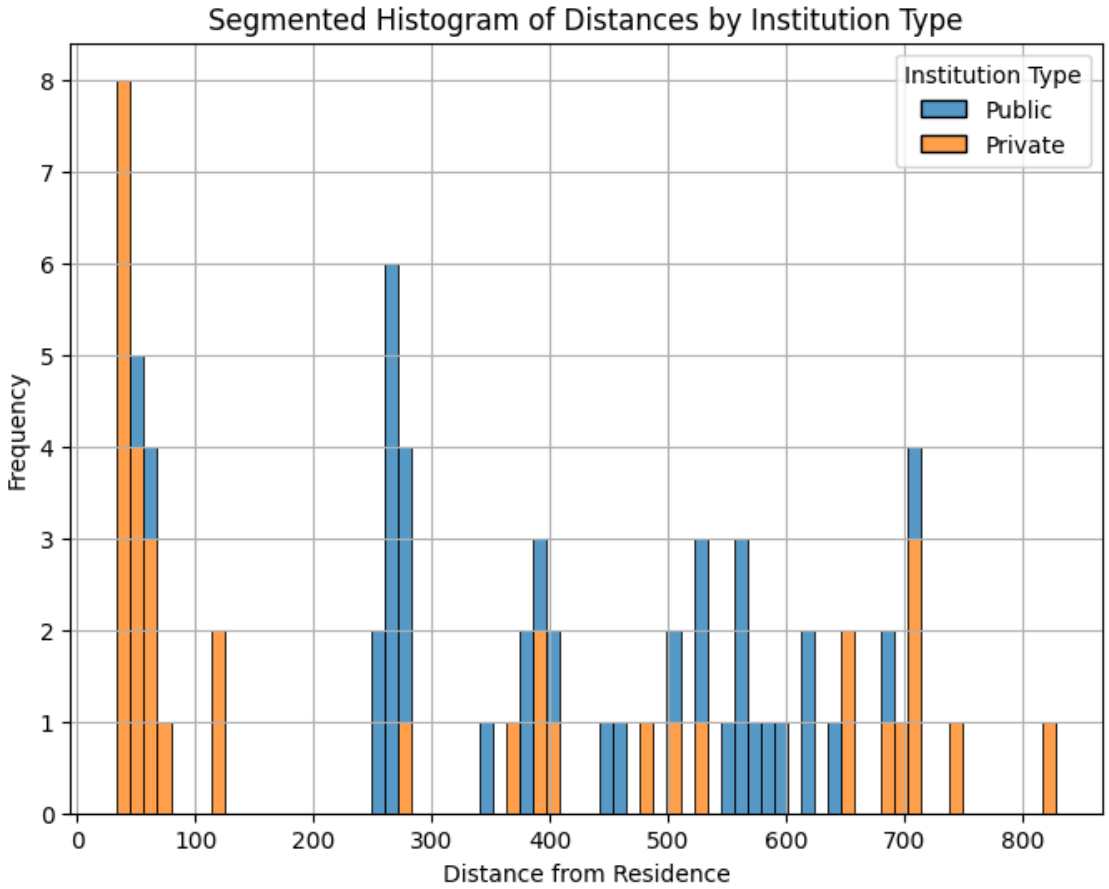


Figure: Visual Representation of data for private and public institutions

Analysis of Accommodation of Private/Public Institutions

Hypothesis Test

Private Institutions have more chance of their students from the short range (in and around Telangana) to take accommodation than public institutions

$$\text{Null Hypothesis}(H_0) : p_1 - p_2 \leq 0$$

$$\text{Alternate Hypothesis}(H_a) : p_1 - p_2 > 0$$

To test the hypothesis, we can use the two-proportions right tailed test. The test statistic is given by:

$$Z^* = \frac{\hat{p}_1 - \hat{p}_2 - p_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

where

- $\hat{p}_1 = 0.56410$; Sample Proportion 1
- $\hat{p}_2 = 0.05$; Sample Proportion 2
- $n_1 = 117$; Number of samples in Sample 1
- $n_2 = 40$; Number of samples in Sample 2
- $\alpha = 0.05$; the confidence coefficient

Analysis of Accommodation of Private/Public Institutions

Result and Conclusion

Result

The p-value obtained from both proportion tests was below the significance level, indicating a rejection of the null hypothesis.

Conclusion

Based on the analysis, it can be concluded that private institutions indeed have a significantly higher proportion of students from the short range opting for accommodation compared to public institutions in Hyderabad.

Final Conclusion

Based on the conclusions drawn from the above analyses:

- **Accommodation Preference**: People who are hostellers are more likely to opt for accommodation than day scholars, as indicated by the rejection of the null hypothesis in the statistical analysis.
- **Distance Traveled**: There is a significant difference in the distances traveled by guests staying for different durations during the fest. This suggests that understanding the geographical distribution of guests can inform fest planning and accommodation arrangements.
- **Geographical Distribution**: Considering the geographical distribution of guests when planning fest activities and accommodations is recommended. Providing transportation options for guests traveling longer distances can enhance their fest experience and participation.
- **Marketing Strategy**: A radius of 100 km is identified as effective for marketing the fest, as at least a fourth of the population falls within this distance.

Final Conclusion

- **Institutional Differences**: Private institutions have a significantly higher proportion of students from the short range opting for accommodation compared to public institutions in Hyderabad.

Overall, based on these conclusions, it can be concluded that fest organizers should tailor their planning, accommodation arrangements, and marketing strategies to accommodate the preferences and geographical distribution of guests. Additionally, recognizing the differences between private and public institutions can help in targeted outreach and accommodation planning.