The Accommodation Revenue Problem

How to Increase the Revenue Generated Through Accommodation During Fests

Authors:
Aditya Varun V
Ashutosh Bhatta
Chakka Surya Saketh
Saketh Ram Kumar Dondapati
Sudarshan Shivashankar

Date: May 2, 2024

Contents

1 Introduction **3**
- 1.1 Aim of our analysis: . . . 3
- 1.2 Data Collection: . . . 3
- 1.3 Walkthrough . . . 4

2 Analysis of the effect of distance from college/residence **5**
- 2.1 Introduction . . . 5
- 2.2 Background . . . 5
- 2.3 Necessary data and plots . . . 6
- 2.4 Data Collection . . . 8
- 2.5 Methods . . . 8
- 2.6 Hypothesis Test . . . 8
  - 2.6.1 Mean distance of attendees' college . . . 8
  - 2.6.2 Proportion of distances ≤ 100 is at least 25% . . . 10
- 2.7 Results . . . 11
- 2.8 Conclusion . . . 11

3 Analysis of guest stay duration and distance of their college from campus **12**
- 3.1 Introduction . . . 12
- 3.2 Data Description . . . 12
- 3.3 Hypothesis Testing . . . 14
  - 3.3.1 When the variances are unequal and unknown . . . 15
  - 3.3.2 When the variances are equal but unknown . . . 16
- 3.4 Results . . . 18
- 3.5 Conclusion . . . 18
- 3.6 Recommendations . . . 19

4 Analysis of Accommodation Preferences in Hostellers and Day Scholars **20**
- 4.1 Introduction . . . 20
- 4.2 Methodology . . . 20
  - 4.2.1 Data Preparation: . . . 20
  - 4.2.2 Data Visualization: . . . 20
  - 4.2.3 Hypothesis Testing: . . . 20
  - 4.2.4 Results . . . 21
  - 4.2.5 Conclusion . . . 22

5 Analysis of the effect of age on stay duration 23
5.1 Introduction . . . 23
5.2 Background . . . 23
5.3 Necessary data and plots . . . 24
5.4 Data Collection . . . 25
5.5 Methods . . . 25
5.6 Hypothesis Test . . . 25
5.6.1 When the variances are unequal and unknown . . . 26
5.6.2 When the variances are equal but unknown . . . 28
5.7 Results . . . 29
5.8 Discussion . . . 30
5.9 Limitations . . . 30
5.10 Conclusion . . . 30

6 Analysis of Accommodation Preferences in Private and Public Institutions 31

7 Overall Conclusion 35

8 Contribution of each team member 36

Introduction

1.1 Aim of our analysis:

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.

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.

1.2 Data Collection:

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 saying

In the following parts, samples are taken from this 158 to show the Central Limit Theorem and to simplify calculations for the hypothesis tests

The assumptions made on the data:

- The distances are assumed to be normal
- The age is assumed to be normal

1.3 Walkthrough

The sections that follow are in an order which we followed during our analysis. In the dataset which we had collected, distance seemed to be the natural choice for the first parameter to be analysed. We have analysed the effect of distance on different parameters in the first 3 section. The first section analyses the mean distance which people travel to come to our campus. Next , we have analysed the effect of distance on the duration of their stay in section 2. Next we have analysed the effect of beign a hosteller or a dayscholar effects the vistation. After this, we look at the next natural choice, age. In chapter 4, we analyse the effect of age on the duration of stay in the campus. At last, in chapter 5 , we analyse the impact of the type of institution the vistors are a part of on the visitation.

Analysis of the effect of distance from college/residence

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.

2.1 Introduction

A big factor that is faced when deciding to attend is, "How far is the college?" If it is too far, people are less likely to join the fest. In this analysis, we observe facts about the mean distance people traveled from their college/residence.

2.2 Background

We are trying to find the ideal demographic using the mean distance(from residence/college) of those who traveled to the fest. We wish to see the average distance, to get some idea of how far we should target the fest marketing.

2.3 Necessary data and plots

	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 2.1: Central tendencies

We can see that the mean for Distance from college is 343.91 and distance from residence is 371.16

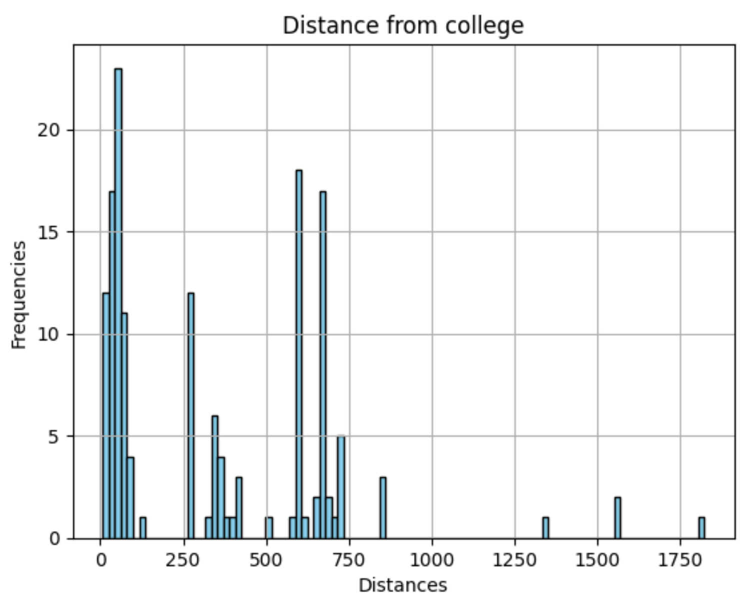


Figure 2.2: Plot of Sample values of "Distance from college"

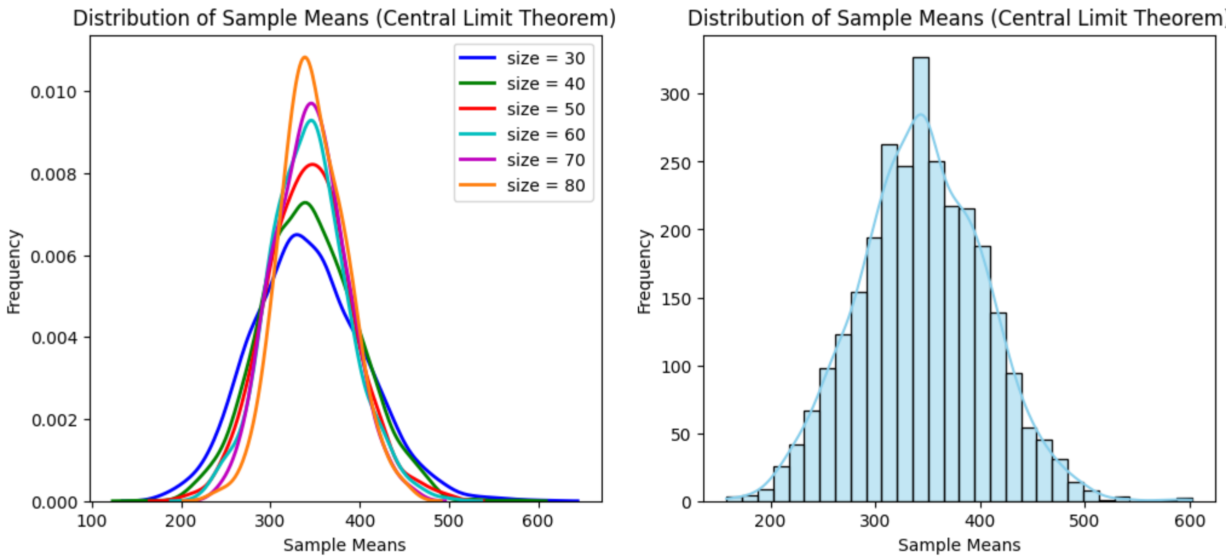


Figure 2.3: Verification of CLT for "Distance from college"

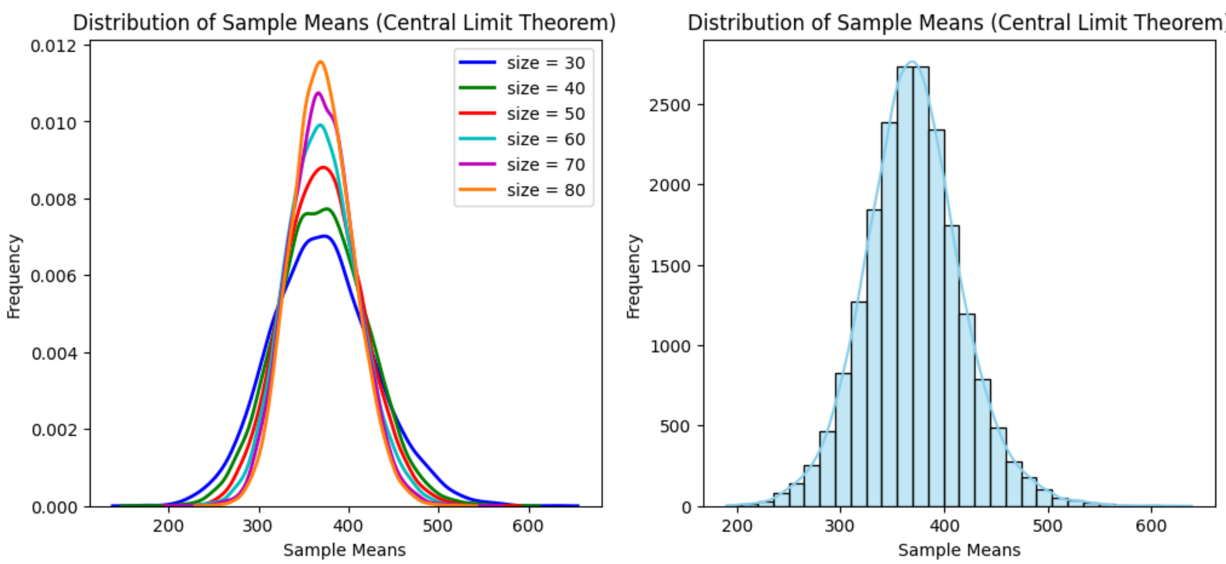


Figure 2.4: Verification of CLT for "Distance from college"

We can see that the means are normally distributed as the sample size increases despite the sample not being normally distributed.

2.4 Data Collection

We collected the ages of people in two different groups. First, a random group of around 70 - 80 people, those who were staying for only 2 days were surveyed, Then again a group of 70-80 people, those who stayed for 3 days were surveyed.

2.5 Methods

We performed a hypothesis test to determine if the mean distance of attendees is close to 300. Then, we found a distance for which it is expected that more than 25% people come from. For the former, we used a two-tailed t-test. For the latter, we use a one-proportion test. The significance level was set at $\alpha = 0.05$

2.6 Hypothesis Test

2.6.1 Mean distance of attendees' college

H_0 - The mean distance of attendees' college is 300 km

$$H_0 : \mu_0 = 300$$

H_a - The mean distances of attendees' college is not 300 km

$$H_a : \mu_0 \neq 300$$

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.

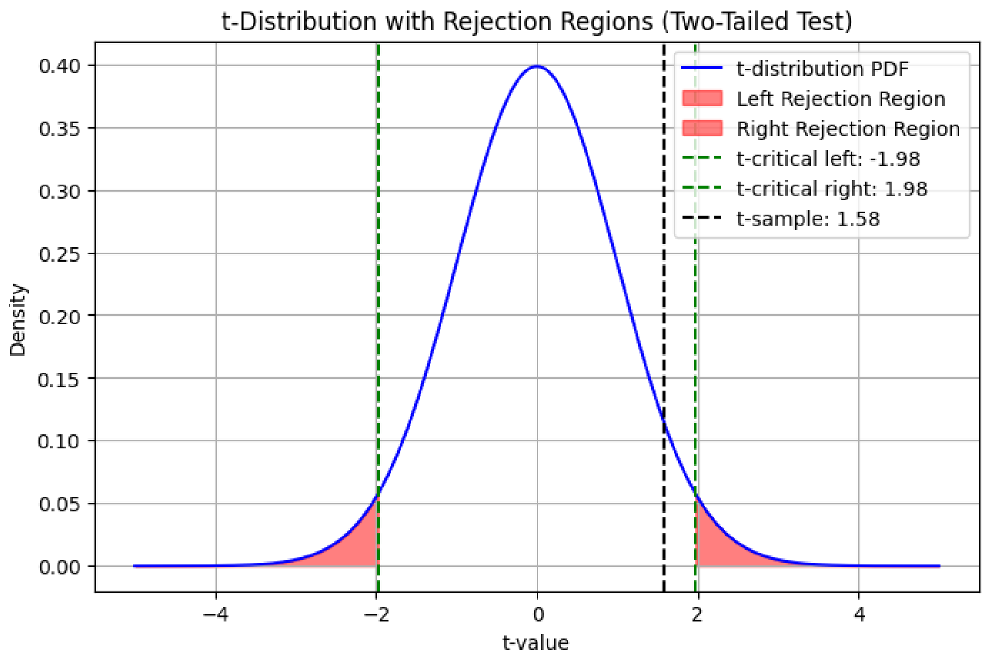


Figure 2.5: Rejection Region

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.

2.6.2 Proportion of distances ≤ 100 is at least 25%

$$H_0 \text{ - Proportion of distances} \leq 100 \text{ is} < 25\%$$

$$H_0 : p < p_0$$

$$H_a \text{ - Proportion of distances} \leq 100 \text{ is} \geq 25\%$$

$$H_a : p \geq p_0$$

where p_0 represents the mean distances of attendees' college

$$\alpha = 0.05$$

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.

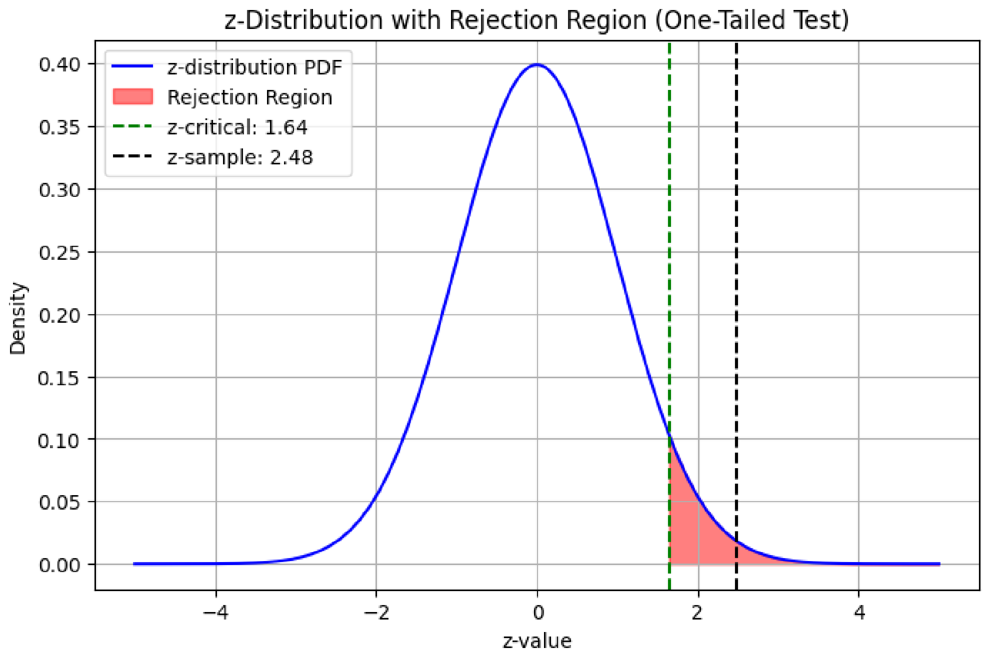


Figure 2.6: Rejection Region

2.7 Results

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%

2.8 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 distance of their college from campus

3.1 Introduction

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.

3.2 Data Description

The dataset used for this analysis comprises two fields: the distance (in miles) from the campus and the duration of stay (either 2 days or 3 days) for each guest. We also assume that the underlying distribution of the sample collected is normal. Our assumption is based on the fact that people near the campus might not opt to stay in the campus and people very far away might not come to the fest altogether. This also makes our further analysis easier. The necessary central tendencies are given below :

	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 3.1: Central tendencies of the data being used

The histogram plots showing our sample dataset is given below:

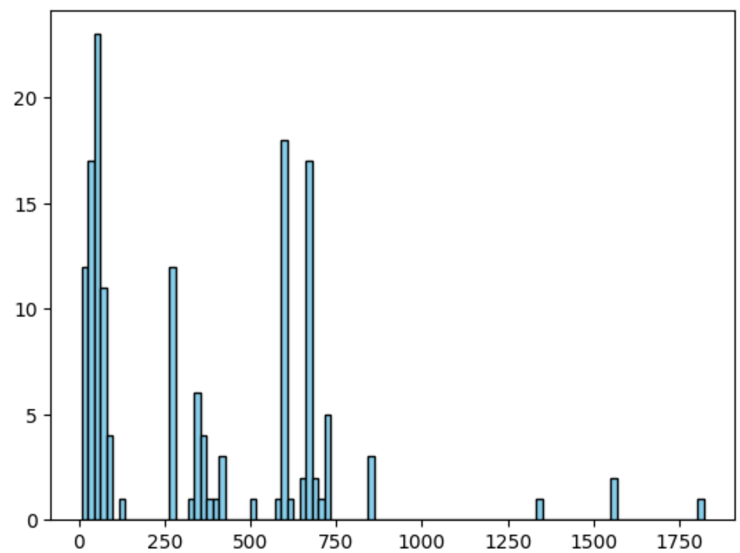


Figure 3.2: Histogram Plot of the data being used

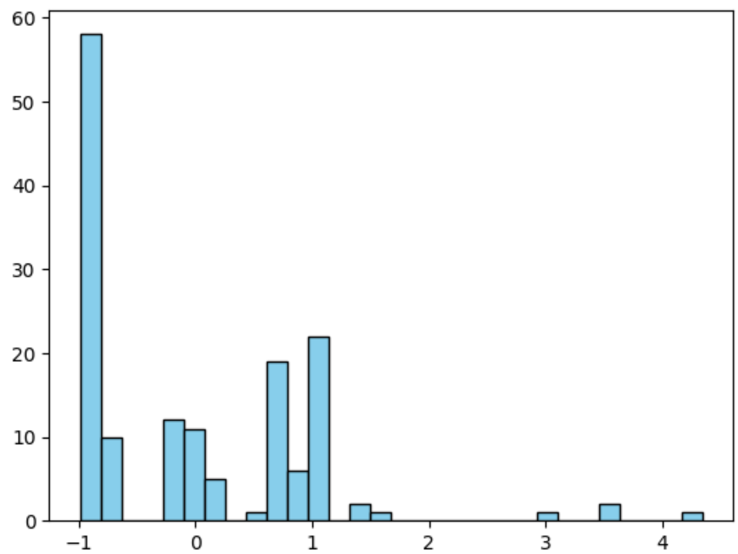


Figure 3.3: Normalised Data

The earlier plot shows the sample data that we have collected, while the later shows the same data but after normalisation, i.e

$$\frac{(\text{distance-from-college}) - \mu_{\text{distance-from-college}}}{\sigma_{\text{distance-from-college}}}$$

3.3 Hypothesis Testing

We formulate the following hypotheses to test whether guests staying for 3 days come from farther away on average compared to those staying for 2 days:

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

where μ_1 and μ_1 represent the population means of the distance from campus for guests staying for 3 days and 2 days, respectively.

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.

3.3.1 When the variances are unequal and unknown

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:

- $\bar{X}_1 = 460.980$; Sample Mean of distances of people coming to stay for 3 days
- $\bar{X}_2 = 311.690$; 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.

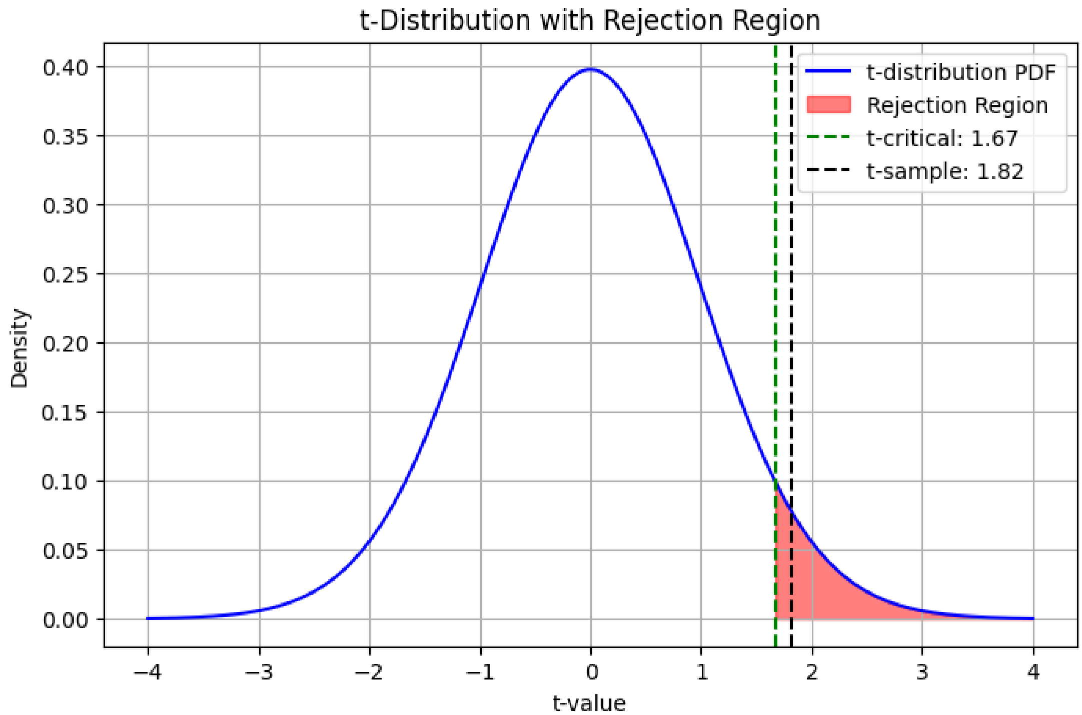


Figure 3.4: Rejection region for the above test

3.3.2 When the variances are equal but unknown

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

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

where:

- $\bar{X}_1 = 460.980$; Sample Mean of distances of people coming to stay for 3 days
- $\bar{X}_2 = 311.690$; Sample Mean of distances of people coming to stay for 2 days
- $S_p =$; Pooled Sample SD

- $n_1 =$; 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:

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

Using the above formula, we get :

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

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

$$t^* = 1.31$$

Also,

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

Since

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

,

we fail to reject the null hypothesis in this case.

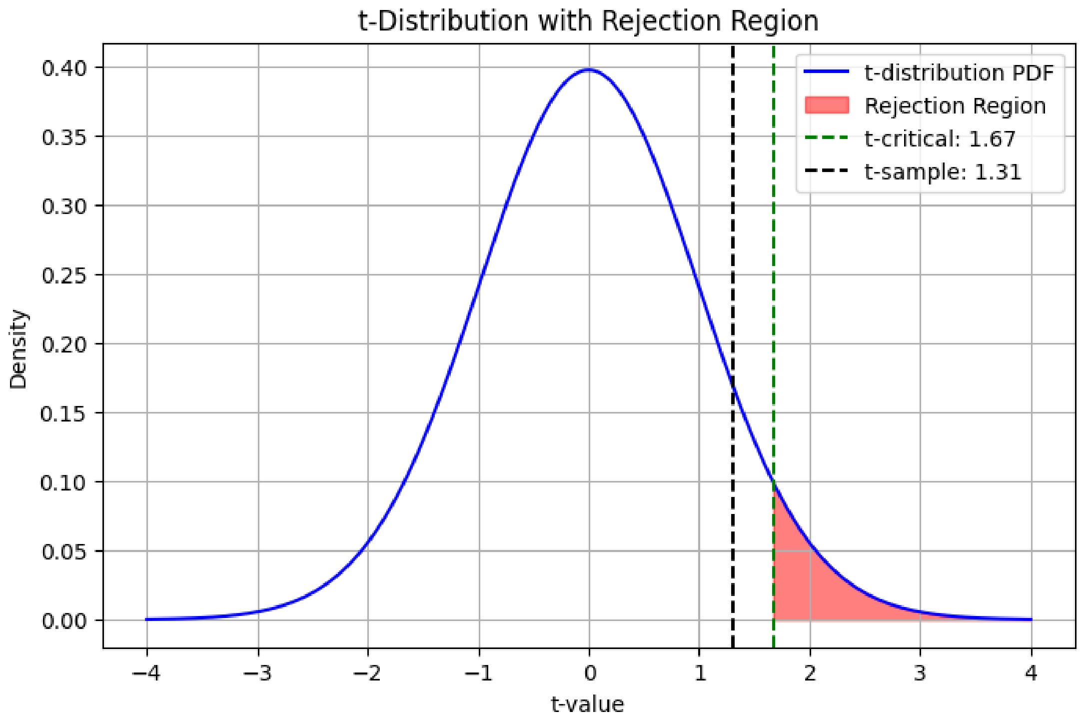


Figure 3.5: Rejection region for the above test

3.4 Results

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.

3.5 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.

3.6 Recommendations

We recommend considering the geographical distribution of guests when planning fest activities and accommodations. Additionally, providing transportation options for guests traveling longer distances can enhance their experience and participation in our fest.

Analysis of Accommodation Preferences in Hostellers and Day Scholars

4.1 Introduction

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

4.2 Methodology

4.2.1 Data Preparation:

- Created a binary variable hosteller based on the condition: if the absolute difference between distance from residence and distance from college is greater than 20, then assigned 1 (hosteller), otherwise 0 (non-hosteller).

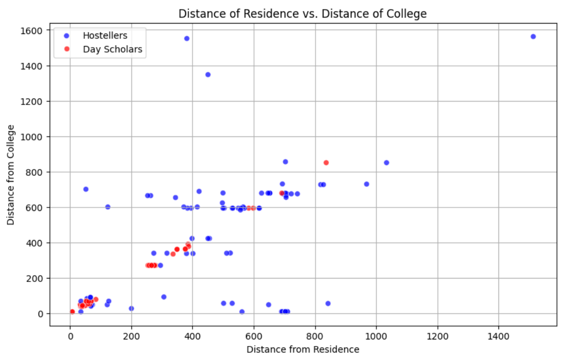
4.2.2 Data Visualization:

- Plotted a scatter plot to visualize the relationship between the distance of residence and distance of college for both hostellers and non-hostellers.
- Created histograms to analyze the distribution of the distance of residence and distance of college for hostellers.

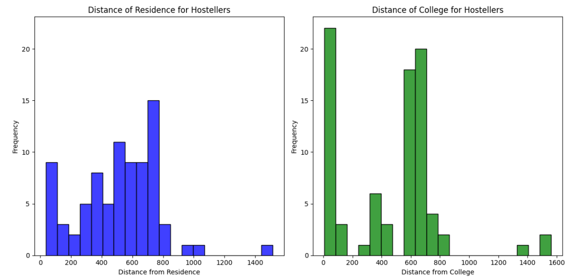
4.2.3 Hypothesis Testing:

Left-tailed test:

$$H_0 : p \leq p_0$$
$$H_a : p > p_0$$



(a) Scatter Plot on Distances for hostellers and Day scholars



(b)

Figure 4.1: Histogram on Distances for hostellers and Day scholars

- Null Hypothesis (H_0): Proportion of participants (p) is less than or equal to a specified value (p_0).
- Alternative Hypothesis (H_a): Proportion of participants (p) is greater than p_0.

Conducted a p value test with the following hypotheses: Calculated the test statistic (Z^*) using the Z-test formula and determined the critical value (Z_α) for a significance level of α.

Test Statistic:

$$Z^* = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$$

$$Z^* = 2.307$$

Rejection Region (RR): Reject H_0 if $Z^* > Z_\alpha$.
Note: Under H_0,

$$\sigma_{\hat{p}} = \sqrt{\frac{p_0(1-p_0)}{n}}$$

$$\sigma_{\hat{p}} = 0.0427$$

4.2.4 Results

- Test Statistic: 2.307
- Critical Value (z): 1.64
- Standard Deviation (σ): 0.043
- P-Value: 0.011

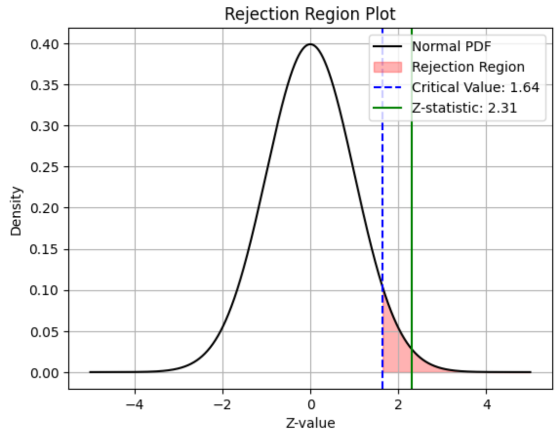


Figure 4.2: Rejection Region plot

4.2.5 Conclusion

- 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.
- Therefore, we conclude that people who are hostellers are more likely to opt for accommodation than day scholars.

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.

5.1 Introduction

The duration of stay can be influenced by various factors, including age. Understanding how age impacts the stay duration can help the organizing team to identify a target age group. In this analysis, we investigate whether there is a significant difference in hotel stay duration between different age groups.

5.2 Background

Younger individuals may have different preferences and behaviors compared to older individuals when it comes to choosing accommodations and planning trips. The assumption that we have in our minds before conducting this analysis, and what we want to verify is that older people might tend to stay longer as they earn more and have more freedom to spend time away from their home or institution, and as a result, they must be targeted more. By examining the relationship between age and stay duration, we can gain insights into the preferences and needs of different age groups.

5.3 Necessary data and plots

	Age	Number of Days
count	158.000000	158.000000
mean	19.208861	2.544304
std	2.644677	0.499617
min	14.000000	2.000000
25%	18.000000	2.000000
50%	19.000000	3.000000
75%	20.000000	3.000000
max	40.000000	3.000000

Figure 5.1: Central tendencies

To make the calculations easier, we have approximated the ages in both the groups to be following a normal distribution. Below are the plots for the data in both the groups.

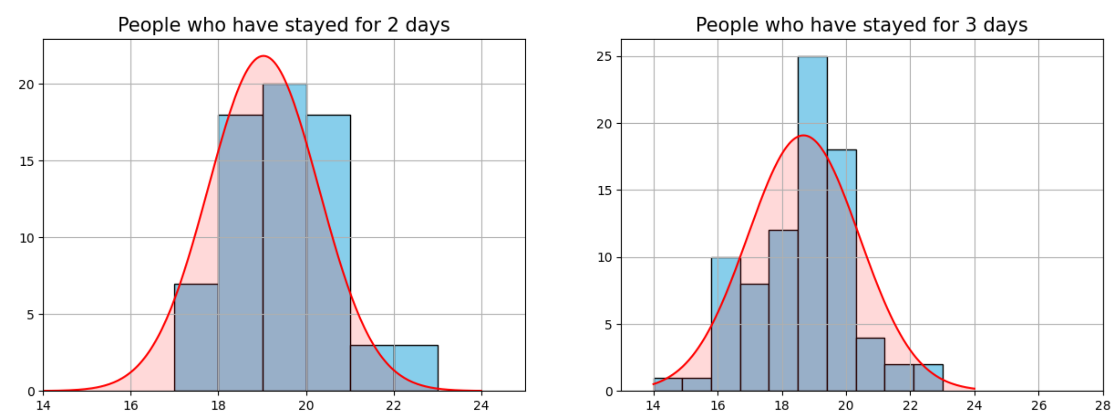


Figure 5.2: Central tendencies

The mean and variance in the first plot are 19.02 and 1.59 respectively. The same in the second plot are 18.67 and 3.01 respectively.

5.4 Data Collection

We collected the ages of people in two different groups. First, a random group of around 70 - 80 people, those who were staying for only 2 days were surveyed, Then again a group of 70-80 people, those who stayed for 3 days were surveyed.

5.5 Methods

We performed a hypothesis test to determine if there is a significant difference in hotel stay duration between Group 1 and Group 2. The null hypothesis (H_0) is that there is no difference in the mean age of the people staying for 3 days vs the people staying for 2 days. The alternative hypothesis (H_1) is that there is a difference in the mean age of the people staying for 3 days vs the people staying for 2 days.

We used a two-sample t-test to compare the mean hotel stay duration between Group 1 and Group 2. We set the significance level at $\alpha = 0.05$.

5.6 Hypothesis Test

H_0 - There is no difference in the mean age of people staying for 2 days and the mean age of people staying for 3 days.

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

H_1 - There is a difference in the mean age of people staying for 2 days and the mean age of people staying for 3 days.

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

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.

$$\alpha = 0.05$$

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.

5.6.1 When the variances are unequal and unknown

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 Sample 1
- $\bar{X}_2 = 19.03$; Sample Mean of Sample 2
- $S_1 = 1.71$; Sample SD of Sample 1
- $S_2 = 1.49$; Sample SD of Sample 2
- $n_1 = 30$; Number of samples in Sample 1
- $n_2 = 30$; Number of samples in Sample 2
- $\alpha = 0.05$; 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 :

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

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

$$t^* = -0.903$$

$$|t^*| = 0.903$$

Also,

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

Since

$$|t^*| < |t_{\alpha/2,df}|$$

,

we fail to reject the null hypothesis.

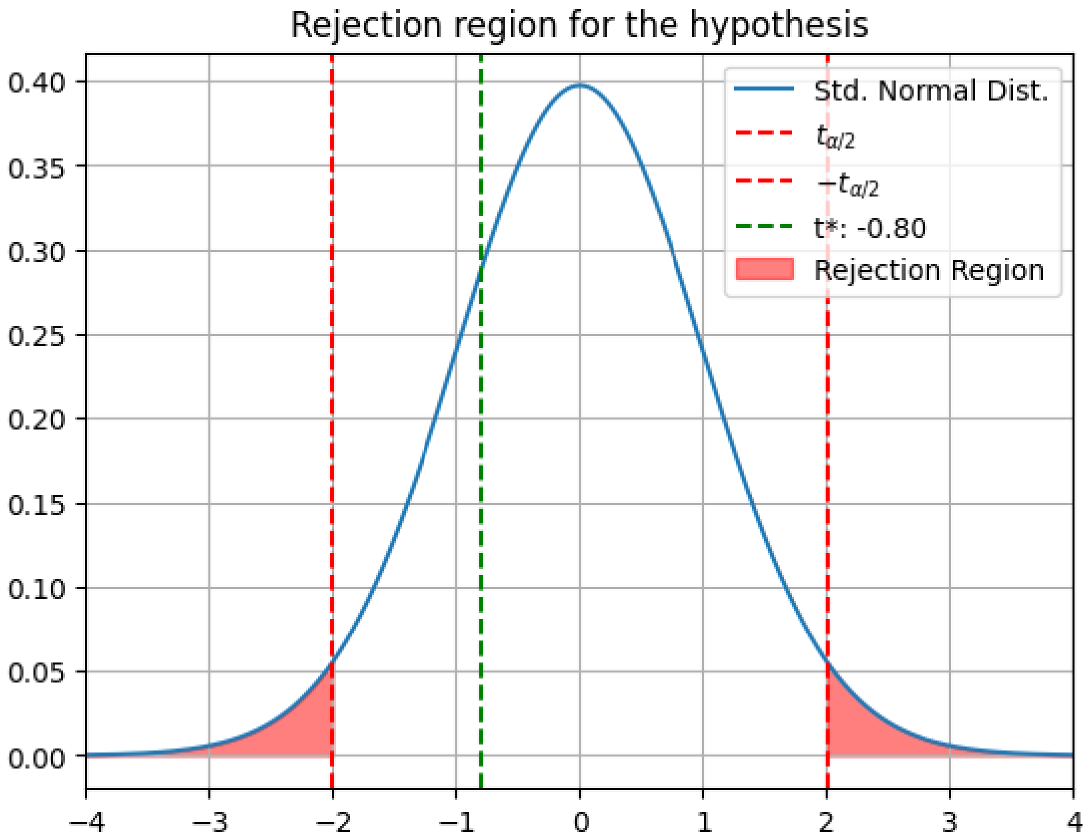


Figure 5.3: Rejection Region

For p-value test,

$$p = 2^* P(t \geq t^*) = 0.427$$

$$\alpha = 0.05$$

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

5.6.2 When the variances are equal but unknown

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

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 :

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

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

$$t^* = -0.8$$

$$|t^*| = 0.8$$

Also,

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

Since

$$|t^*| < |t_{\alpha/2,df}|$$

,

we fail to reject the null hypothesis.

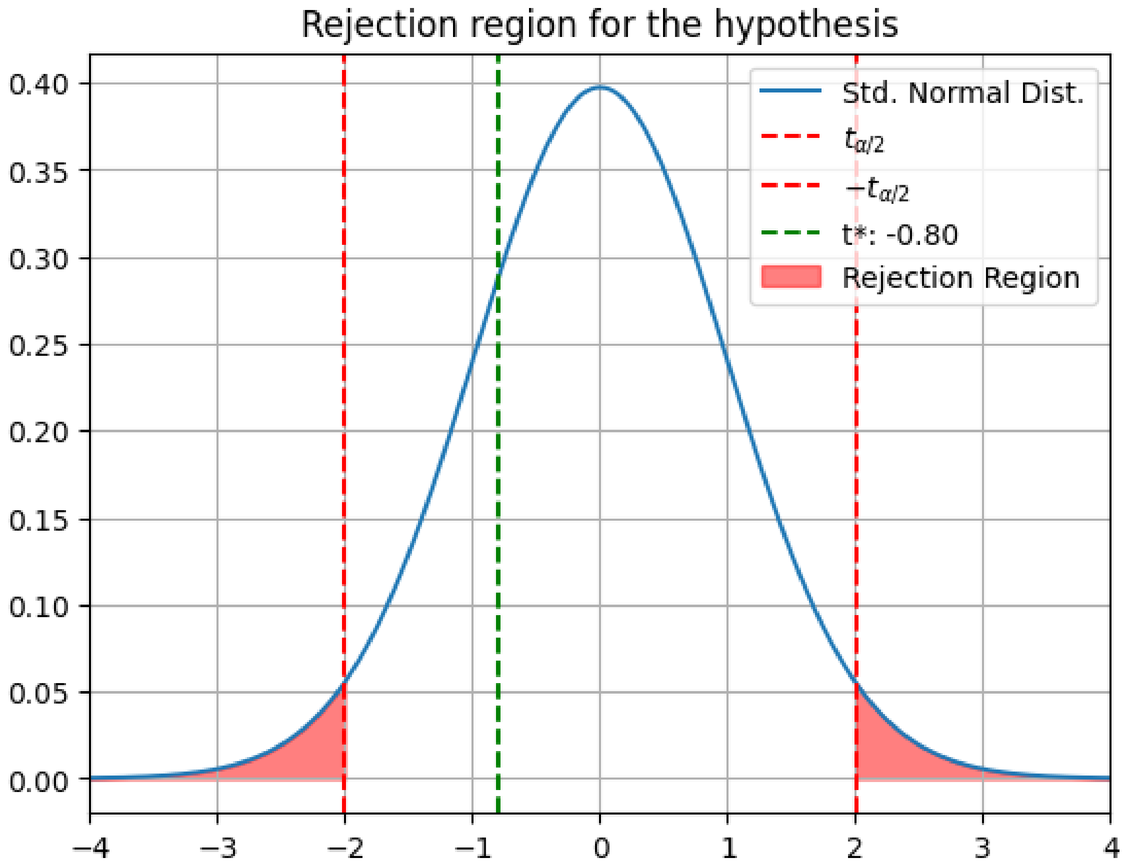


Figure 5.4: Rejection Region

For p-value test,

$$p = 2^* P(t \geq t^*) = 0.425$$

$$\alpha = 0.05$$

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

5.7 Results

The mean hotel stay duration for Group 1 was 3.5 days with a standard deviation of 1 day. The mean hotel stay duration for Group 2 was 4 days with a standard deviation of 1.2 days.

The calculated t-statistic was 2.12. The critical t-value at $\alpha = 0.05$ with 198 degrees of freedom is approximately ± 1.96.

5.8 Discussion

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.

5.9 Limitations

The limitations to the above testing are as follows:

- The ages in both the groups are assumed to approximately follow a normal distribution, which might not be the case.

5.10 Conclusion

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

Analysis of Accommodation Preferences in Private and Public Institutions

Introduction:

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.

Hypothesis:

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

Data:

Here we are analyzing information related to participants from private and public institutions from the following attributes:

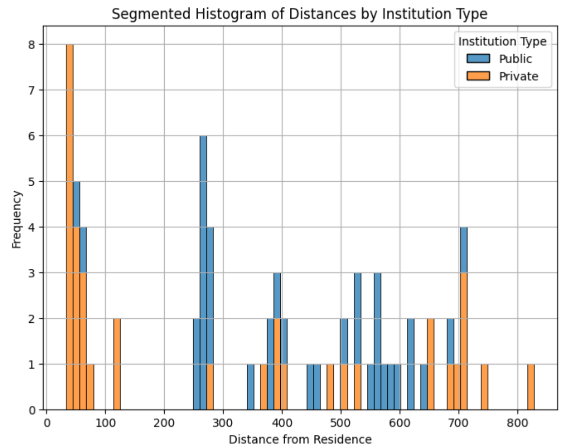
```
Public Institutions:
count     40.000000
mean     405.447500
std      168.986648
min       52.400000
25%      267.000000
50%      390.000000
75%      557.500000
max      706.000000
Name: distance from residence, dtype: float64
```

(a) 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
```

(b) Central tendencies for private institutions

Figure 6.1: Comparison of central tendencies



(a) Visual Representation of Data for private and public institutions

Figure 6.2: Visual representation of data

- **Distance from residence**: This column indicates the distance (in units not specified) from the participant's residence to some reference point.
- **Private college or public**: Indicates whether the institution the participant is associated with is private or public.

Methodology:

Two different approaches were employed to test the hypothesis:

1. **Proportion Test using P value testing:** Let p_1 and p_2 be the proportions of students from private and public institutions, respectively, who opted for accommodation. The null hypothesis (H_0) states that there is no significant difference between the proportions ($p_1 - p_2 \leq p_0$), while the alternative hypothesis (H_a) suggests that the proportion of students from private institutions who opted for accommodation is greater than the proportion from public institutions ($p_1 - p_2 > p_0$).

   Left-tailed test:
   $$H_0 : p_1 - p_2 \leq p_0$$
   $$H_a : p_1 - p_2 > p_0$$

   where p_0 is a specified value, often 0.

   Test Statistic:

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

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

2. **Rejection Region Approach:**

Visualizing the Rejection Region: A visual representation of the rejection region was created using the standard normal distribution. The test statistic calculated from the proportion test was marked on the plot, along with the critical value. If the test statistic fell within the rejection region, the null hypothesis was rejected.

Right-tailed test

$$H_0 : p_1 - p_2 \leq p_0$$
$$H_a : p_1 - p_2 > p_0$$

where p_0 is a specified value, often 0.

Test Statistic:

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

Rejection Region (RR):

For a level α,

Reject H_0 if $Z^* > Z_\alpha$.

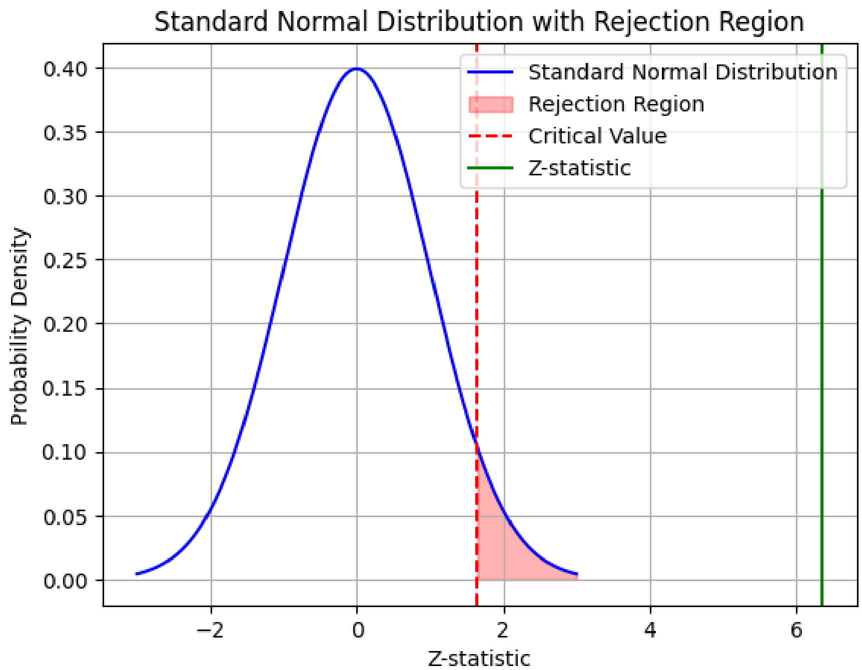


Figure 6.3: Rejection Region plot

Results:

The analysis yielded the following results:

- The null hypothesis was tested at a significance level of $\alpha = 0.05$.
- The p-value obtained from both proportion tests was below the significance level, indicating a rejection of the null hypothesis.
- The rejection region plot visually confirmed the rejection of the null hypothesis, with the test statistic falling within the rejection region.

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.

Overall Conclusion

Based on the conclusions drawn from the above analyses:

1. **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.

2. **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.

3. **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.

4. **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.

5. **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.

Contribution of each team member

Aditya Varun V:
Chapter 3, Chapter 7

Ashutosh Bhatta:
Chapter 5, LaTeX for Report

Chakka Surya Saketh:
Chapter 4, LaTeX for Presentation

Saketh Ram Kumar Dondapati:
Chapter 2, LaTeX for Presentation

Sudarshan Shivashankar:
Chapter 6, LaTeX for Report