INF2178 Techincal Assignment one

University of Toronto
Yingliang Ding
1009324492

Introducation

In Toronto's vibrant cityscape, a disturbing trend has emerged as a result of urban expansion and socioeconomic change: The city's homeless population has grown significantly. By stretching the capacity of the City of Toronto's shelter support system, this increase is challenging the very fabric of the city's social welfare system. Although the system is designed to provide overnight shelter and support services to the unsheltered, it is struggling with a critical bottleneck: there are not enough shelter spaces to accommodate all those in need. The inadequacy of the current infrastructure to meet the growing demand becomes apparent as individuals and families seeking shelter face the disheartening reality of being turned away.

This paper examines the systemic strains on shelter availability and explores the broader societal implications of this urban phenomenon, delving into the nuances of Toronto's homelessness crisis. The study aims to uncover the underlying patterns and inequalities within the shelter system through statistical analysis of shelter occupancy rates using methods such as one-sample, two-sample, and Welch's t-tests. By examining occupancy rates in different sectors, including "families," "mixed adults," and varying shelter capacity types, this research seeks to illuminate the complexities of service delivery and policy effectiveness in addressing homelessness. The findings present a compelling story of a city at a crossroads, and highlight the urgent need for comprehensive policy interventions and innovative solutions to ensure that no one is left without shelter in Toronto's urban landscape.

One-Sample T-Test

Hypothesis for 1-Sample T-Test: (assume the population mean as 0.8 (or 80%))

- H0: The mean occupancy rate in the 'Families' sector is equal to 0.8.
- H1: The mean occupancy rate in the 'Families' sector is different from 0.8.

```
print("t-statistic = " + str(t_stat_1sample))
print("p-value = " + str(p_value_1sample))
```

```
t-statistic = 73.91863704056344
p-value = 0.0
```

The results from the one-sample t-test suggest a very significant difference between the actual observed occupancy rate for the 'Families' sector in Toronto shelters and the hypothesized occupancy rate of 80%. A t-statistic as high as approximately 73.92 coupled with a p-value of 0 indicates that the observed occupancy rate is significantly different from the hypothesized rate, and the probability that this difference is due to random chance is extremely low.

From a data analysis perspective, this result might indicate that the 'Families' sector is experiencing either a much higher or much lower occupancy rate than what was expected (the hypothesized value). Given the context of an increase in homelessness in Toronto and the fact that individuals seeking shelter are often turned away due to a lack of space, this result could

suggest that shelters are operating at a different capacity than what is considered standard or targeted.

Two-Sample T-Test(1)

Hypothesis for 2-Sample T-Test

- H0 (Null Hypothesis): The mean occupancy rates between the 'Families' and 'Mixed Adult' sectors are equal.
- H1 (Alternative Hypothesis): The mean occupancy rates between the 'Families' and 'Mixed Adult' sectors are not equal.

To provide a fuller picture and a better understanding of the operation of the shelter system, We compare the mean occupancy rates between the 'Families' and 'Mixed Adult' sectors.

```
]: print("t-statistic = " + str(t_stat_2sample))
   print("p-value = " + str(p_value_2sample))

   t-statistic = 8.660754065466053
   p-value = 5.290740247388472e-18
```

A t-statistic of approximately 8.6607 and a p-value of approximately 5.29e-18 were obtained from the two-sample t-test. Therefore, we reject the null hypothesis and conclude that there is a significant difference in the average occupancy rate between the "Families" sector and the "Mixed Adults" sector. Which indicating a statistically significant difference between the average occupancy rates of two different groups within the shelter system. Given the increase in homelessness in Toronto and the strain on the shelter support system, the significant t-statistic suggests that the difference in occupancy rates between the two groups is greater than would be expected by chance. This could mean that one group has consistently higher or lower occupancy rates than the other, which could reflect different levels of service delivery, demand, or policy effectiveness in different parts of the shelter system.

In addition, given the growing number of homeless people in the city and the lack of shelter space that results in some people being turned away, the results of the test may indicate that some groups may be disproportionately affected. For example, if the test compares "families" to "mixed adults," it may indicate that families are more or less likely to find shelter than mixed adults, suggesting a potential policy focus.

Welch's T-Test

Hypothesis for Welch's T-Test

- H0 (Null Hypothesis): The mean occupancy rates between the 'Families' and 'Mixed Adult' sectors are equal, without assuming equal variances.
- H1 (Alternative Hypothesis): The mean occupancy rates between the 'Families' and 'Mixed Adult' sectors are not equal, without assuming equal variances.

This is similar to the two-sample t-test(1) but does not assume equal variances between the two groups.

```
print("t-statistic = " + str(t_stat_welch))
print("p-value = " + str(p_value_welch))
```

```
t-statistic = 9.17540822190609
p-value = 5.3308370032743536e-20
```

The result of the Welch's t-test, with a t-statistic of about 9.175 and a p-value effectively equal to zero, again rejects the null hypothesis and confirms that there is a significant difference in the mean occupancy rates between the "Family" and "Mixed Adult" sectors when not assuming equal variances.

This result, which is statistically significant, indicates that there is a profound difference between the average occupancy rates of the two groups under consideration. This difference may reflect an inequitable distribution of resources or different levels of need within the homeless community in the context of Toronto's growing homeless population and strains on the shelter system. For example, certain demographics may be over-represented in shelters due to various socioeconomic factors, or some shelters may be more heavily utilized due to where they're located, the services they provide, or the demographics they serve.

A number of narratives could be built from this preliminary analysis. It could suggest that certain policies or programs may be more effective for one group than another, resulting in varying utilization. Alternatively, it could suggest that shelter capacity is not aligned with the demographics of the homeless population. In this case, targeted interventions would be needed.

Two-Sample T-Test(2)

Hypothesis for 2-Sample T-Test(2)

- H0 (Null Hypothesis): There is no significant difference in capacity rates between 'Room Based Capacity' and 'Bed Based Capacity'.
- H1 (Alternative Hypothesis): There is a significant difference in capacity rates between 'Room Based Capacity' and 'Bed Based Capacity'.

```
print("t-statistic = " + str(t_stat))
print("p-value = " + str(p_value))
```

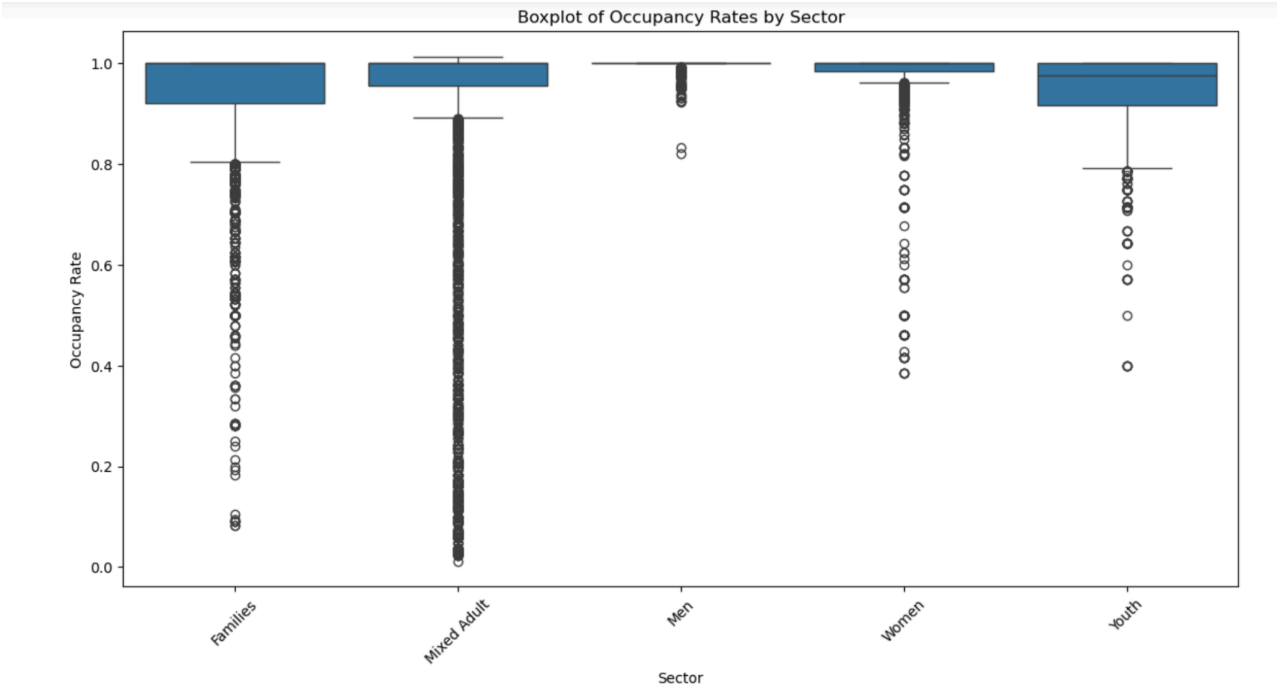
```
t-statistic = 4.498751771925636
p-value = 6.860477551487939e-06
```

The two-sample t-test results yielded a t-statistic of approximately 4.498 and a p-value of approximately 6.86e-06, indicating a statistically significant difference between two shelter capacity types: room-based and bed-based.

This significant finding may suggest that existing shelter capacity types may not be aligned with the needs or preferences of homeless people, given the context of Toronto's shelter system and the challenges of housing the growing homeless population. For instance, an increase in the number of families seeking shelter or individuals seeking more privacy may

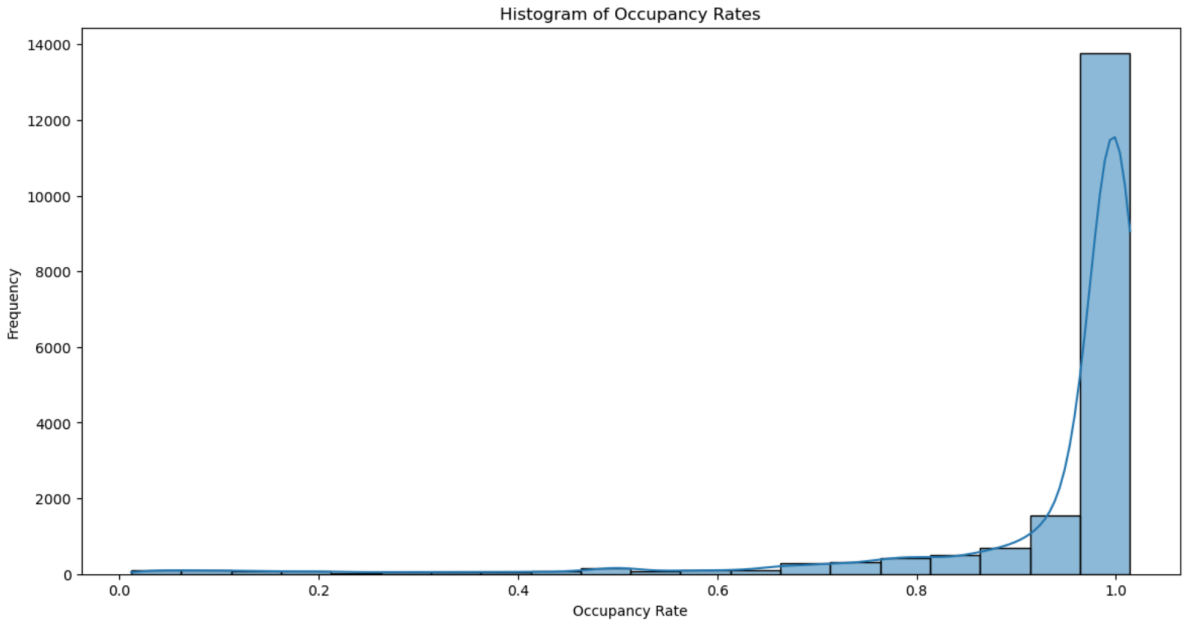
indicate a higher demand for private or family rooms if room-based shelter capacity is more heavily utilized than bed-based shelter capacity.
Conversely, a higher utilization rate of "bed-based capacity" shelters may reflect a larger single-person population in need of shelter services, and may also indicate that the distribution of bed-based shelters better matches the locations where homeless people seek services.
In order to assess the adequacy of the current shelter system and make decisions about future investments, these findings are critical for city planners and policymakers. To better serve the homeless, cities may need to consider adjusting the balance between room and bed capacities, or examine other factors that may affect these occupancy levels, such as the location and accessibility of sheltering facilities, the types of services provided, and even the sense of safety and comfort in different sheltering environments.

Boxplot



The box plot of the occupancy rates by sector shows a considerable range in the distribution of the rates for different groups. In particular, the categories "Mixed Adults" and "Men" show considerable variance. The variation can be seen in the spread of the interquartile range and the number of outliers, which suggests an uneven use of these refuges. Some of the shelters that serve these sectors are operating at occupancy rates that are much lower than the median, while others are operating at or near full capacity. The Family and Youth sectors show less variability and generally higher median occupancy rates, indicating a more consistent demand for shelter in these sectors. This may be due to the unique needs and challenges faced by these groups.

Histogram



The histogram of occupancy rates shows a bimodal distribution. There are peaks at the low and high ends of the occupancy rate spectrum. This suggests that there are a significant number of shelters with much lower occupancy rates, while many shelters are operating at or near full capacity. The concentration of high-occupancy shelters supports the narrative that there is a pressing need for shelter services in Toronto, which often leads to strained capacity where individuals seeking shelter may be turned away.

Further Analysis

Building on the analysis done previously, the next step in exploring the dataset to understand the work of the Toronto shelter system will include several key approaches. The first is that it will be instructive to delve into temporal trends. Analyzing occupancy rates over time can reveal seasonal patterns or responses to specific policy implementation. If daily, weekly, or monthly data are available, the impact of certain events or policy changes on shelter utilization can be assessed. It is also useful to secondly relate shelter occupancy rates to other socio-economic indicators such as unemployment rates, housing prices, or weather conditions. This can help identify external factors that affect the demand for shelter services. Regression analysis can reveal the extent to which these factors predict changes in occupancy rates.

Finally, an examination of shelter capacity and resource allocation can reveal if there are inefficiencies in the current distribution of services. An analysis of staff turnover in the shelter system may identify bottlenecks or opportunities to improve shelter space management.