Analyzing the Interaction Effects of Agency Type and Subsidy Receipt on Total Space Allocation

1. Introduction

In recent years, the Toronto government has been focused on expanding the availability of childcare spaces due to the challenges of affordability and scarcity. With only 25% of Toronto households able to secure a spot for their children in both licensed and non-licensed facilities, there's a pressing need to understand the allocation of these childcare spaces. This urgency underscores the importance of our study into the distribution mechanisms of childcare provisions in the city.

This analysis will focus on the problem how the total care spaces are allocated in Toronto, using one-way ANOVA and two-way ANOVA methodologies. To dig into the patterns, our study will use the dataset named 'INF2178_A2_data' provided by the government of Toronto. To be specific, not every column of variables will be studied, and our research will be focusing on three main research questions:

Research question1: Is there a significant difference in the total space occupied by different types of agencies? (different in 3 types of auspices, Commercial, Non-Profit, and public)

Research question2: Can we observe statistically significant disparities in space utilization between different subsidy types?

Research question3: Can we detect a significant interactive effect between agency types and subsidies on the occupation of space?

By addressing these questions, we will be able to see what factors can be possibly affect the total space allocation and get a further understanding of current Toronto childcare space situations.

2. Data Cleaning and Data Wrangling

In this dataset, we will perform a study on **1063** observations with **17** columns. By viewing the data information and format, we can draw a conclusion that not much data cleaning will be needed in this dataset. All numerical data are well recorded and only one categorical variable requires cleaning. Specifically, **BLDGNAME (Name of the building the childcare centre is located in)** shows 348 missing values, so I choose to fill the missing space with NAN in this cleaning process.

Next, with respect to the research questions there are a few variables of choice we will need in the future studies, which are:

Numerical:

- **IGSPACE:** Childcare spaces for infants 0-18 months
- **TGSPACE:** Childcare spaces for toddlers 18-30 months
- **PGSPACE:** Childcare spaces for preschoolers 30 months up until they enter grade one
- **KGSPACE:** Childcare spaces for children in full-day kindergarten
- **SGSPACE:** Childcare spaces for children grade one and up
- **TOTSPACE:** Childcare spaces for all age groups

Categorical:

- **AUSPICE:** Operating auspice (Commercial, Non-Profit, or Public)
- **subsidy:** Centre has a fee subsidy contract (Yes/No)

3. Data Visualization

We will then have a look into the variables of our interest in box plots.

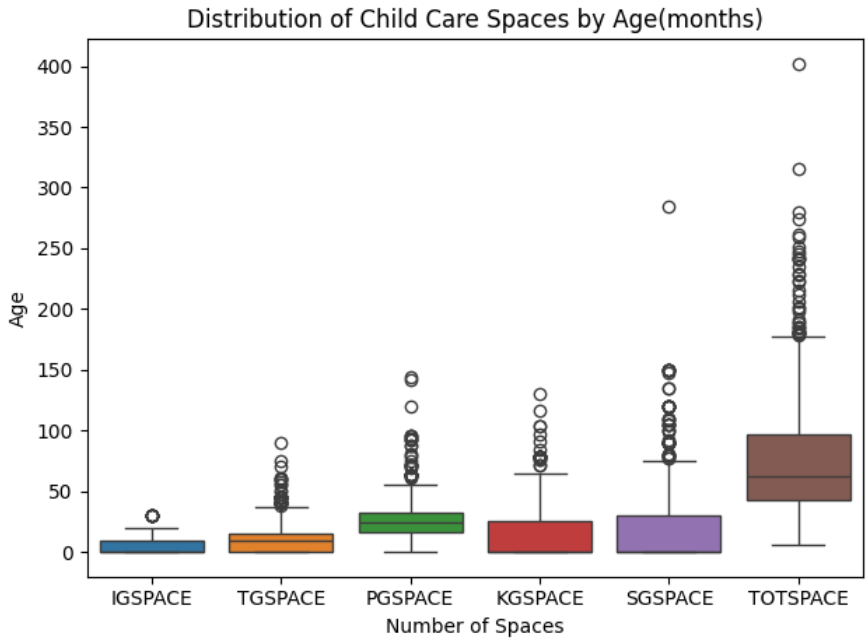

Figure 1: Distribution of Child Care Spaces by Ages(months)

In this figure, we use childcare for all age in comparison to the other variables representing child age. Specifically, no outlier has occurred below the lower whiskers for all variables. The median value appears to increase from infants to toddlers, but then decreases for up to grade one child, while full time children in full-day kindergarten and upper than grade one children's mean drop slightly. We can since have an overview of the central tendency and dispersion of the data.

Furthermore, we will then generate a boxplot to see the data distribution by treatments. Using this boxplot in figure 2, we can easily detect the differences between different treatments.

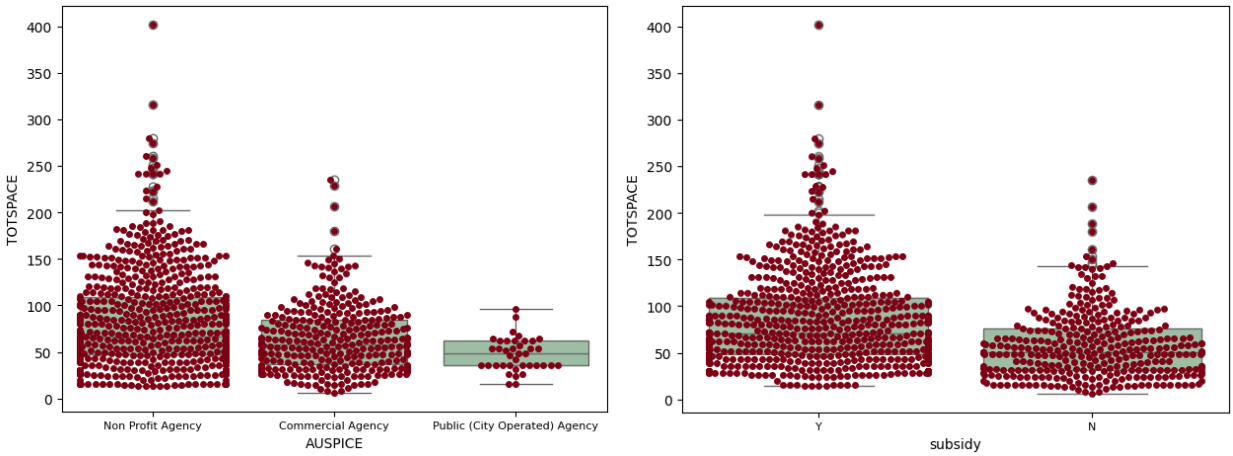

Figure 2: Box plot of childcare of all age groups by different treatments

Based on the research questions of our interest, according to the graphic, we will see in different auspice, non-profit agency has the widest interquartile range, and the spread of the data is very large, while public agency has the narrowest interquartile range and shows decent crowded data without outliers. For commercial agency, the median is positioned closer to the center of the box, which suggests a more symmetric distribution of data around the median. For another thing, the case in subsidy is more similar for either center has subsidy contract or not.

4. One-Way ANOVA

We will then move to our one-way ANOVA step to see if there are any significant differences, the figures can be seen for auspice in table 1 and subsidy in table 2.

	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	9.61e+04	48056.06	21.84	5.06e-10
Residual	1060.0	2.33e+06	2200.06	\	\

Table 1: ANOVA table for auspice

	df	sum_sq	mean_sq	F	PR(>F)
C(subsidy)	1.0	1.61e+05	160765.39	75.23	1.55e-18
Residual	1061.0	2.27e+06	2137.05	\	\

Table 2: ANOVA table for subsidy

The results of Table 1 and 2 suggest that there are statistically significant differences in childcare space for all age groups when comparing different types of auspices and subsidies. As we can see the P-values are extremely small (<0.05), we can reject the null hypothesis that there is no

difference in these factors. In addition, the subsidy variable has a higher F-statistic, which could suggest a stronger association than the auspice.

	Group1	Group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	17.12	9.70	24.53	7.66	0.001
1	Non Profit Agency	Public Agency	34.33	16.22	52.45	6.29	0.001
2	Commercial Agency	Public Agency	18.22	-1.45	35.88	3.06	0.078

Table 3: Post-hoc test for auspice

	Group1	Group2	Diff	Lower	Upper	q-value	p-value
0	Y	N	26.27	20.32	32.21	12.27	0.001

Table 4: Post-hoc test for subsidy

Referring to Table 3, the p-value in Non-Profit Agencies and Commercial Agencies, it is lower than 0.05 which suggests that there is a statistically significant difference in the total space between Non-Profit Agencies and Commercial Agencies. A same result will be observed for the case between Non-Profit Agencies and Public Agencies, with subsidy or without subsidy. However, when we move to the result of Commercial Agencies and Public Agencies, the p-value 0.078 is greater than 0.05, where we can conclude that there is no significant difference lies between these two groups.

5. Assumption Check

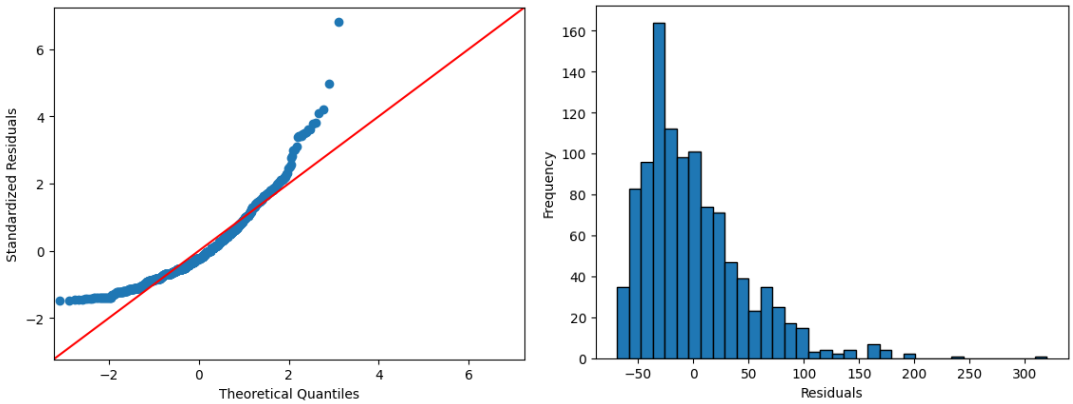

Figure 3: Q-Q & Distribution plot of residual (auspice)

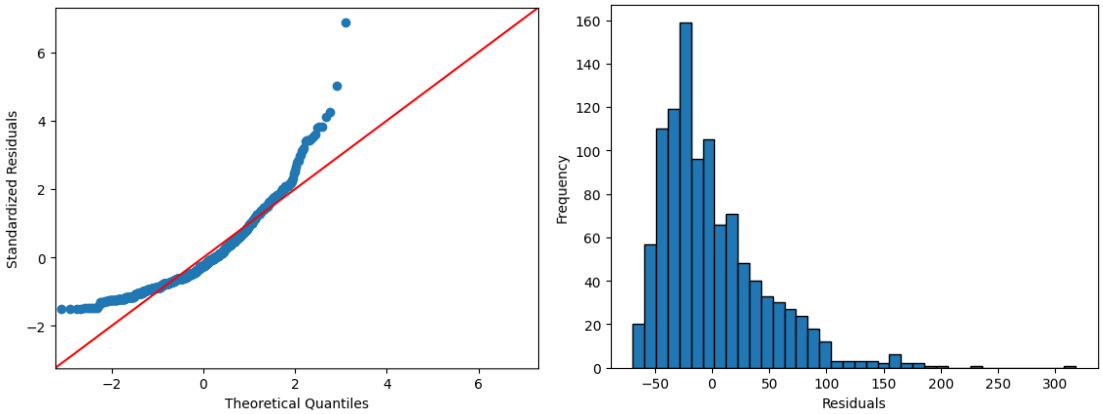

Figure 4: Q-Q & Distribution plot of residual (subsidy)

In both factors, from Q-Q plots we can see the residuals from the regression analysis do not perfectly follow a normal distribution. Histograms of residuals show similar patterns of right skewed, which indicates potential outliers. The result could fail to meet assumption 1: Normality, which could affect the validity of statistical tests that assume normality of residuals.

For assumption 2: Homogeneity, we will perform Bartlett's test and Levene's test

	Parameter	Value
0	T	89.59
1	Df	2.00
2	p-value	<0.001

Table 5: Bartlett's test for auspice

	Parameter	Value
0	T	49.08
1	Df	1.00
2	p-value	<0.001

Table 6: Bartlett's test for subsidy

	Parameter	Value
0	T	17.93
1	Df	2.00
2	p-value	<0.001

Table 7: Levene*'s test for auspice*

	Parameter	Value
0	T	22.99
1	Df	2.00
2	p-value	<0.001

Table 8: Levene*'s test for subsidy*

According to these tables, where all p-values are significantly small, we will conclude that we reject the null hypothesis of equal variances across the groups, which means that we fail to meet assumption 2: Homogeneity.

6. Two-Way ANOVA

To see what impact the subsidy variable brings to childcare space for all age groups in auspice, we can further perform this on a boxplot and interaction effect plot:

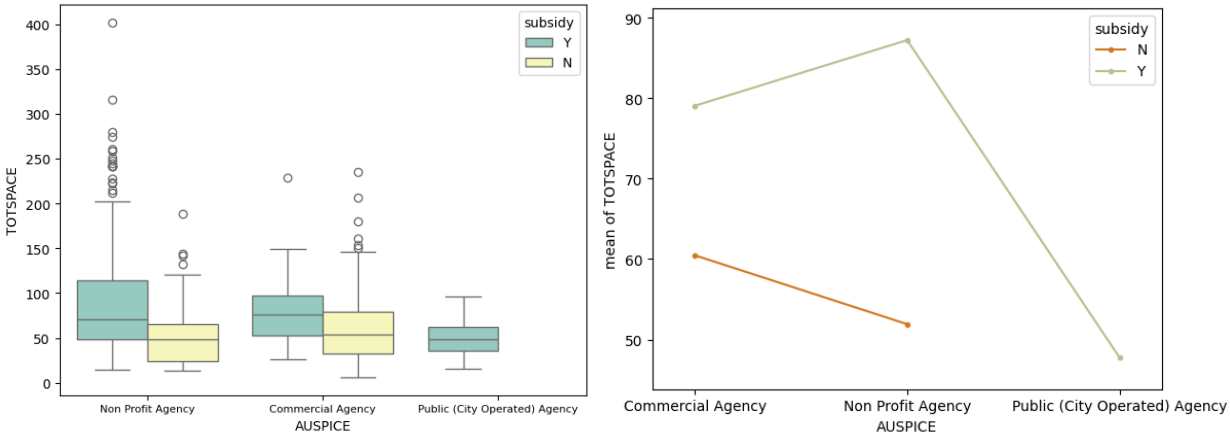

Figure 5: Effect of subsidy on different auspice

Specifically in the box plot, Public Agencies exhibit only cases with subsidy. For the right graph, as the lines are not parallel, there is an interaction effect present, meaning the effect of different auspice on total space depends on the whether a subsidy is received.

	Group1	Group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	17.12	9.91	24.33	7.87	0.001
1	Non Profit Agency	Public Agency	34.33	16.72	51.95	6.49	0.001
2	Commercial Agency	Public Agency	17.22	-0.95	35.88	3.15	0.068

Table 9: Post-hoc test for auspice considering interacting effect

	Group1	Group2	Diff	Lower	Upper	q-value	p-value
0	Y	N	26.27	20.40	32.13	12.43	0.001

Table 10: Post-hoc test for subsidy considering interacting effect

7. Conclusion

We will finally reach to the answer for **research questions**, for question 1, there is a significant difference in the total space occupied by different types of agencies, which was supported by the results of one-way ANOVA and post-hoc tests. Affirmative answer will also be drawn from the results of table 2 & 4 for question 2 like question 1.

Given the result of box plot comparison, interaction effect plot, and two-way ANOVA result, similar to the first two questions we will reach a conclusion that there is significant interactive effect between agency types and subsidies on the occupation of space in question 3.