Analyzing Capacity in Toronto's Child Care Centres

Songjia Liu 1005137503

1. Introduction

The accessibility and capacity of childcare services are critical issues Toronto is facing, where the demand for childcare centres is often greater than the supply. This assignment applies statistical analysis to investigate the operational capacities of childcare centres in Toronto, exploring how various factors, including operating auspices and subsidy availability, impact their ability to serve the community. By conducting one-way and two-way ANOVA, the study aims to identify potential disparities and inform policy decisions, ultimately contributing to the enhancement of childcare services in the city.

2. Data cleaning and warning

The following steps are doing data cleaning and preprocessing to ensure the availability and quality of the Childcare dataset:

a. Select the necessary columns

The raw dataset has a total of 17 columns and 1064 rows, the columns I need are:

_id: Unique row identifier for the database
AUSPICE: Operating auspice (Commercial, Non-Profit or Public)
IGSPACE: Spaces for infants 0-18 months
TGSPACE: Spaces for toddlers 18-30 months
PGSPACE: Spaces for preschoolers 30 months up until they enter grade one
KGSPACE: Spaces for children in full-day kindergarten
SGSPACE: Spaces for children grade one and up
TOTSPACE: Spaces for all age groups
subsidy: The Centre has a fee subsidy contract (Yes/No)

b. Check whether there are missing values or outliers

There is no missing value because the count of nulls in all columns is zero.

c. Check if TOTSPACE equals the overall age groups' childcare spaces

The TOTSPACE equals the sum of IGSPACE, TGSPACE, PGSPACE, KGSPACE and SGSPACE

3. Research questions

Research Question 1 (using One-Way ANOVA):

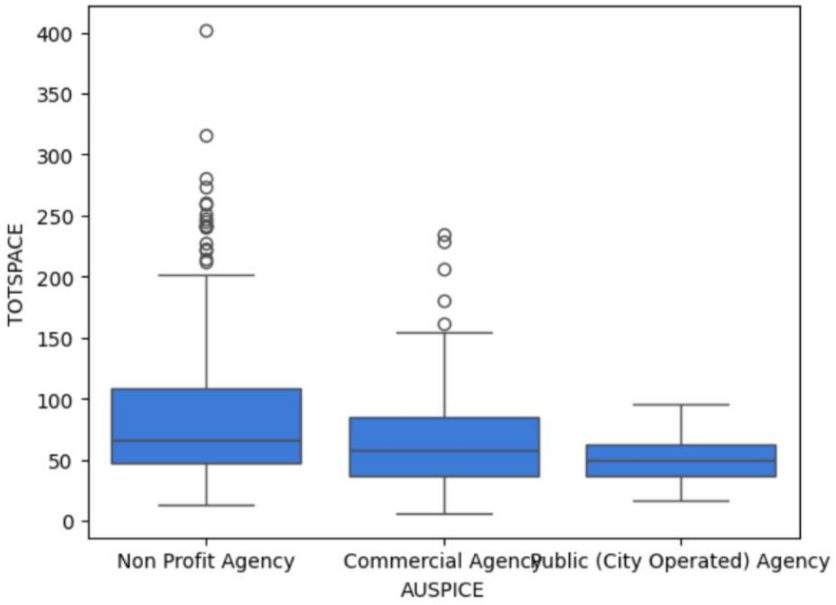
Is there a significant difference in the total space of childcare available in Toronto among different types of operating facilities (Commercial, Non-Profit, Public)?

Research Question 2 (using Two-Way ANOVA):

How do both the auspice of childcare and subsidies collectively impact the total number of childcare spaces available in Toronto?

4. One-Way ANOVA Analysis: Auspice Influence on Child Care Spaces

a. Box plot



The boxplot visual analysis suggests that there are differences in total space available among Non-Profit, Commercial, and Public childcare facilities in Toronto:

Non-Profit Agency: Has a wider IQR than the other two, indicating more variability in data. Lower median space, with some high-space outliers.

Commercial Agency: Higher median space than Non-Profit, wider range of spaces, and some high-space outliers.

Public Agency: Highest median space, similar variability to Commercial, with high-space outliers.

To statistically confirm whether these differences

b. OLS model and ANOVA table

index	sum_sq	df	F	PR(>F)
C(AUSPICE)	96112.11	2.0	21.84	5.06e-10
Residual	2332065.26	1060.0	NaN	NaN

LS model result

index	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	96112.11	48056.06	21.84	5.06e-10
Residual	1060.0	2332065.26	2200.06	NaN	NaN

ANOVA table result

- Notice that AUPICE is the independent variable, TOTSPACE is the dependent variable. The ANOVA analysis results show that there is a statistically significant difference in total space among the different types of childcare facilities (Commercial, Non-Profit, Public).
- The Sum of Squares for the grouping variable AUSPICE is 961121.11, which quantifies the variation attributable to the differences among the group levels. Conversely, the Residual Sum of Squares is 2332065.26, representing the variation within the groups themselves.
- The F-statistic is 21.84 with a very small p-value (5.06e-10), which is far below the usual significance level of 0.05. This indicates that the observed differences in total space across these categories are highly unlikely to have occurred by chance.

c. The post hoc test using Tukey's HSD

index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	17.12	9.70	24.54	7.66	0.001
1	Non Profit Agency	Public (City Operated) Agency	34.33	16.22	52.45	6.29	0.001
2	Commercial Agency	Public (City Operated) Agency	17.22	-1.45	35.88	3.06	0.078

Interpretation:
Non Profit Agency vs Commercial Agency:
The mean difference (Diff) is approximately 17.12.
The 95% confidence interval of this difference ranges from about 9.70 to 24.54.
The p-value is 0.001, which is statistically significant ($p < 0.05$).

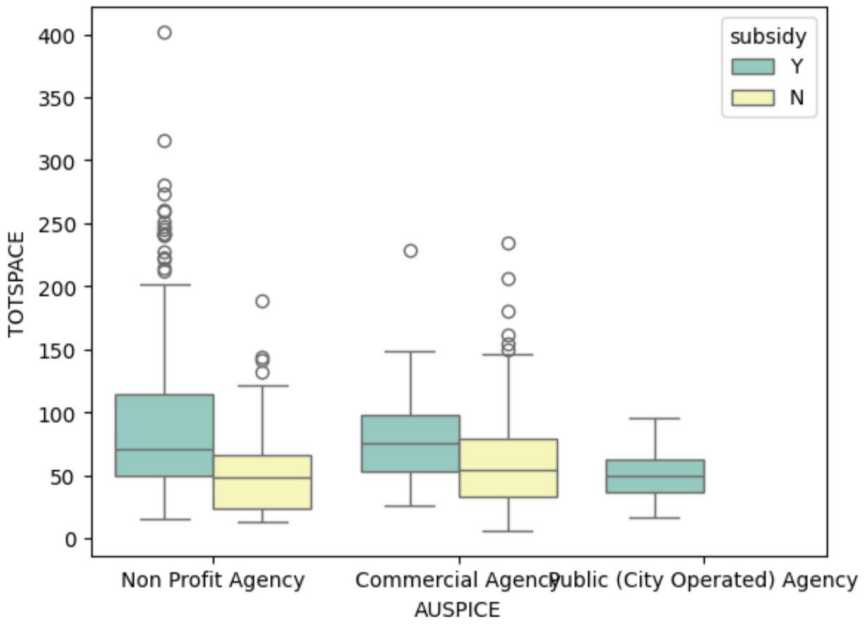
Non Profit Agency vs Public (City Operated) Agency:
The mean difference is approximately 34.33.
The 95% confidence interval ranges from about 16.22 to 52.45.
The p-value is 0.001, indicating a statistically significant difference.

Commercial Agency vs Public (City Operated) Agency:
The mean difference is approximately -1.45, indicating that the Commercial Agency has slightly less space on average compared to the Public Agency, but this is not statistically significant ($p > 0.05$) as indicated by the p-value of 0.078.

Conclusion: The p-vlue and q-values suggest that the differences between Non Profit and the other two categories are significant, while the difference between Commercial and Public agencies is not.

5. Two-Way ANOVA Analysis: The Combined Impact of Childcare Auspice and Subsidies on Availability of Childcare Spaces in Toronto

a. Box plot



The boxplot analysis:
Central Tendencies: Subsidized facilities, especially public ones, tend to offer more childcare spaces than non-subsidized. All public facilities are subsidized and show higher median spaces than non-profit and commercial ones.
Data Spread: There's a broader spread of total spaces in subsidized non-profit agencies, while commercial ones have a more consistent space offering. Public agencies show the least variability.
Outliers: Numerous non-profit agencies receiving subsidies are outliers with exceptionally large spaces. Subsidized commercial agencies have some outliers, but less than non-profit ones. Public agencies show the fewest outliers.

b. OLS model and ANOVA table

index	sum_sq	df	F	PR(>F)
C(AUSPICE)	8567.996	2.0	2.06	0.13
C(subsidy)	83527.44	1.0	40.12	3.53e-10
C(AUSPICE):C(subsidy)	56034.45	2.0	13.46	1.69e-06
Residual	2202809.39	1058.0	NaN	NaN

OLS model result

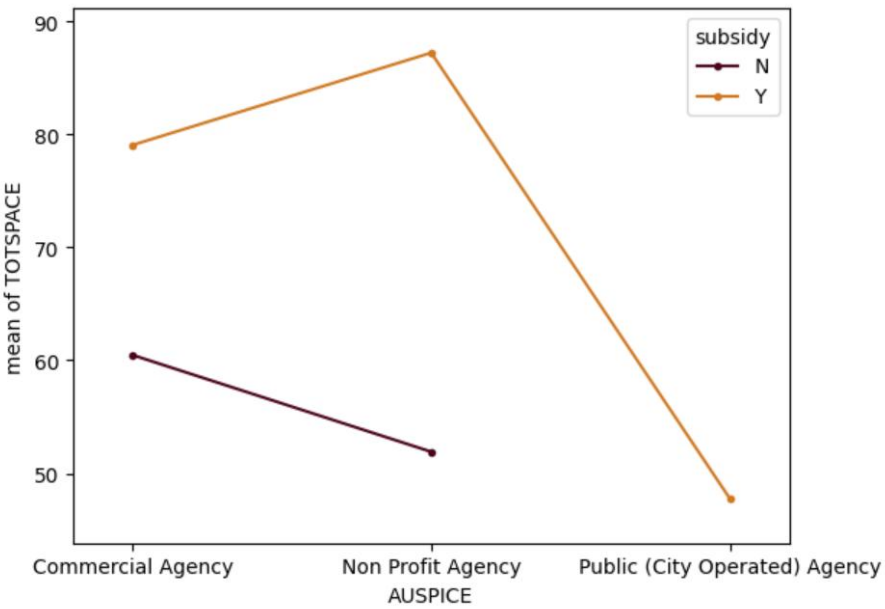
index	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	8567.996	4283.998	2.06	0.13
C(subsidy)	1.0	83527.44	83527.44	40.12	3.53e-10
C(AUSPICE):C(subsidy)	2.0	56034.45	28017.23	13.46	1.69e-06
Residual	1058.0	2202809.39	2082.05	NaN	NaN

ANOVA table result

- The effect of AUSPICE on total spaces is not statistically significant, with an F-value of 2.057586 and a p-value of 0.128273 (larger than 0.05).
- The effect of subsidy on total spaces is statistically significant, with a high F-value of 40.11876 and a very low p-value of 3.529094e-10 (less than 0.05).
- There is a significant interaction between the type of childcare agency and subsidy on total spaces, indicated by an F-value of 13.456555 and a p-value of 0.00000169 (less than 0.05).

In summary, while the type of agency alone does not significantly affect the number of spaces, the presence of a subsidy does. Moreover, the interaction effect suggests that the impact of subsidies on total spaces varies by the type of childcare agency.

c. The Interaction Plot



The interaction plot shows the mean total space of childcare facilities by auspice type and whether they receive subsidies:

Subsidized non-profit agencies have the highest mean total space, significantly higher than subsidized commercial and public agencies.
Non-subsidized facilities have relatively consistent mean total spaces across all auspice types, with only slight variation.

Summary: The interaction effect is noticeable: the subsidy appears to benefit non-profit agencies the most in terms of increasing total space, whereas commercial and public agencies see less of an impact from subsidies.

d. The post hoc test using Tukey's HSD

index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	17.12	9.91	24.33	7.88	0.001
1	Non Profit Agency	Public (City Operated) Agency	34.33	16.72	51.95	6.47	0.001
2	Commercial Agency	Public (City Operated) Agency	17.22	-0.95	35.38	3.15	0.068

Interpretaton: Significant disparities exist in 'TOTSPACE' when comparing Non Profit Agencies with Commercial Agencies and Public Agencies, indicated by a p-value less than 0.05. On the other hand, the difference in 'TOTSPACE' between Commercial Agencies and Public Agencies is not statistically significant, as the p-value is greater than 0.05.

index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Y	N	26.27	20.40	32.13	12.43	0.001

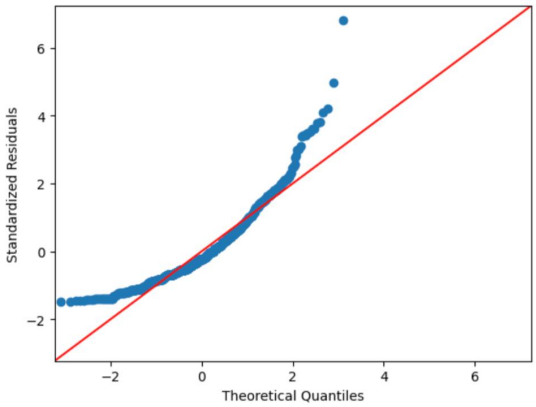
Interpretaton: A statistically significant variation exists in the average 'TOTSPACE' between groups receiving subsidies and those that do not, as evidenced by a p-value below 0.05. Specifically, groups with subsidies ('Y') exhibit a higher average 'TOTSPACE' in comparison to those without subsidies ('N')

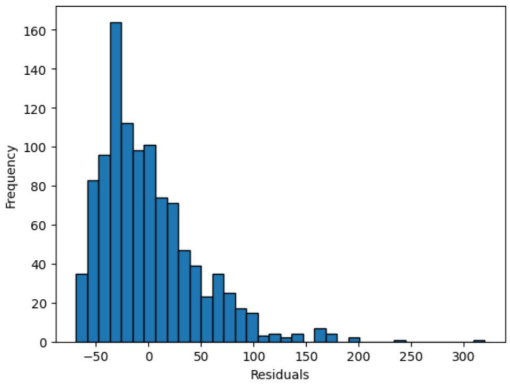
index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency,Y	Non Profit Agency,N	35.33	21.38	49.28	10.22	0.001
1	Non Profit Agency,Y	Commercial Agency,Y	8.17	-7.51	23.84	2.10	0.65
2	Non Profit Agency,Y	Commercial Agency,N	26.76	16.86	36.67	10.91	0.001
3	Non Profit Agency,Y	Public (City Operated) Agency,Y	39.46	17.93	60.99	7.40	0.001
4	Non Profit Agency,Y	Public (City Operated) Agency,N	0.0	-Infinity	Infinity	0.0	0.9

Interpretaton: There is a significant difference in average 'TOTSPACE' when comparing subsidized Non-Profit Agencies to non-subsidized Non-Profit Agencies, with the former having more space on average (p-value = 0.001). However, there is no statistically significant difference in 'TOTSPACE' between subsidized Non-Profit Agencies and subsidized Commercial Agencies, as indicated by a p-value of 0.9. A significant difference is noted between subsidized Non-Profit Agencies and non-subsidized Commercial Agencies (p-value = 0.001). Additionally, when subsidized Non-Profit Agencies are compared to subsidized Public Agencies, a significant difference in 'TOTSPACE' exists (p-value = 0.004757). The comparison between subsidized Non-Profit Agencies and non-subsidized Public Agencies shows no difference, as demonstrated by a p-value of 0.9, signifying uniformity between these groups in terms of 'TOTSPACE'.

6. ANOVA assumptions check

Assumption1: res.anova_std_residuals are standardized residuals obtained from ANOVA





Q-Q Plot: The points significantly deviate from the red reference line at both ends of the distribution, indicating that the residuals do not follow a normal distribution.

Histogram: The histogram shows that the residuals are skewed to the right, which also suggests a departure from normality.
These suggest that the assumptions of normality for the ANOVA test are violated. Non-normality of residuals can impact the validity of ANOVA's results, making it less reliable for determining whether group means are significantly different.

Assumption1: Residuals are normally distributed: Shapiro Wilk test
The Shapiro-Wilk test results in a W statistic of 0.90 and a p-value of approximately 1.5e-25. This extremely low p-value suggests that the residuals from the ANOVA model are not normally distributed, which is a deviation from one of the key assumptions of ANOVA.

Assumption2: Variances are homogenous: Levene's test when the sample is not normally distributed.

index	Parameter	Value
0	Test statistics (W)	9.1994
1	Degrees of freedom (Df)	2.0
2	p value	0.0001

Levene's test results in a low p-value, indicating unequal variances in 'TOTSPACE' across 'AUSPICE' groups, violating ANOVA assumption.

7. Conclusion

The analysis reveals that the total availability of childcare spaces in Toronto significantly differs across various types of operating auspices, including commercial, non-profit, and public entities. Government subsidies contribute notably to the availability of these spaces, with effects that are contingent on the agency type. Furthermore, the interaction between the operating auspice and subsidy presence has a substantial effect on the number of available childcare spaces.
Although the study faced deviations from the normality and homogeneity of variance assumptions, which could affect the ANOVA results' reliability, the outcomes provide valuable insights for decision-making processes, such as increasing subsidies, supporting for non-profit agencies and consider agency type in policy making.