

Explore the Dynamics of Childcare Centre Capacities in Toronto via Statistical Analysis

Introduction

Embarking on a statistical journey to explore the complexities of licensed childcare centres in Toronto, I aimed to uncover the nuanced relationships between childcare centre capacities and various operational characteristics. Central to this investigation were the impacts of the operating auspice (**AUSPICE**) and the availability of subsidies (**subsidy**) on the total space available in these centres (**TOTSPACE**).

Research Questions

Guided by a keen interest in the operational dynamics of childcare centres, I formulated:

1. How does the type of operating auspice influence the total available space in childcare centres?
2. Is there an observable interaction effect between operating auspice and subsidy status on the total space available?

These questions steered the direction of my analysis, utilizing one-way and two-way ANOVAs to dissect the factors influencing childcare centre capacities.

Data Preparation

I removed non-essential variables to focus on **AUSPICE**, **subsidy**, and **TOTSPACE**. None of the above variables have missing values.

Examination of interested variables

Figure 1 and 2 states that the record of total space is not even across auspice, indicating variability. Public agency group has the largest central tendency. The other two groups have more or less outliers. This pattern shows the necessity of constructing ANOVA upon Types of agency and total spaces of care center.

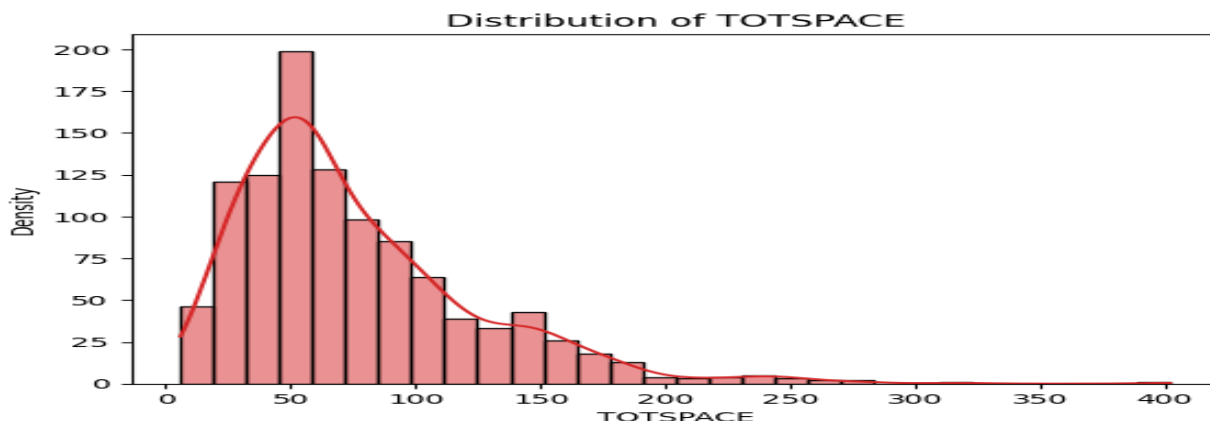


Figure-1 Distribution of total space

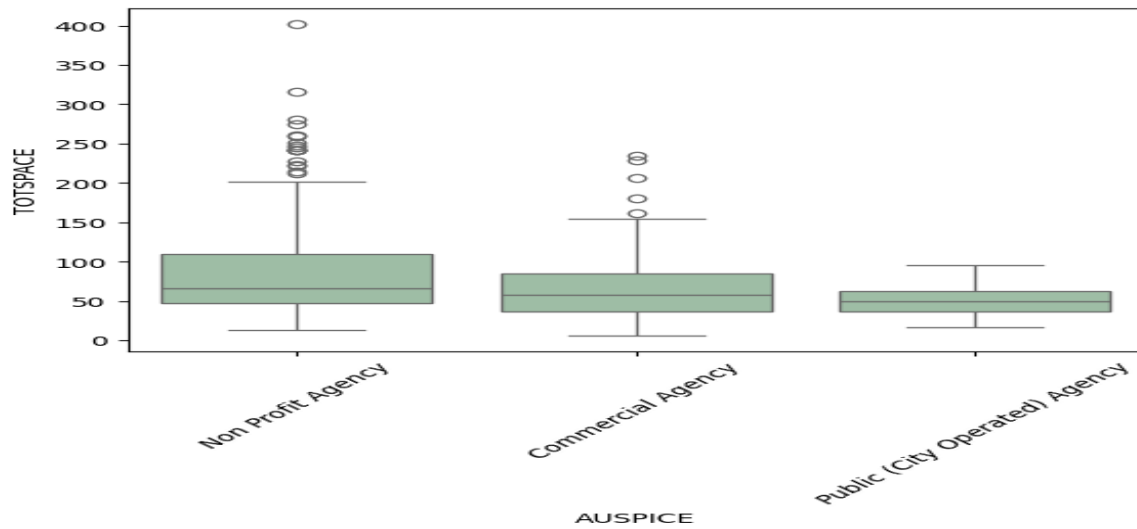


Figure-2 distribution of total space record over different auspices

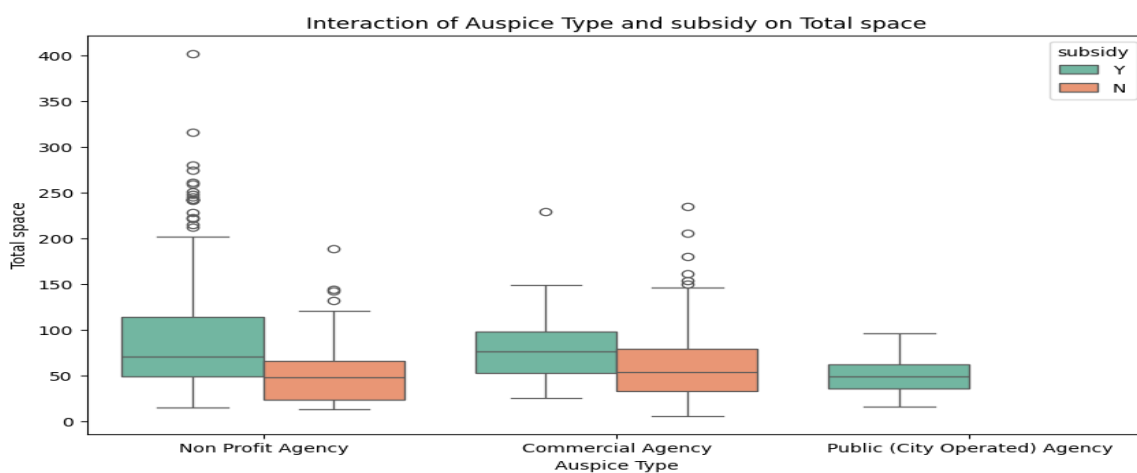


Figure-3 distribution of total space record across auspices and subsidization.

From figures 3, it is clear that receiving subsidy group has a larger variability than the non-subsidy group, but the advantages of subsidy are not the same for all agencies. Therefore, it is necessary to include question 2 in this topic.

One-way ANOVA Analysis(research question 1)

Assumption Testing

Table 1: Assumptions Test for One-way ANOVA

Test	Statistics	P-value	Interpretation
Shapiro-Wilk	0.902	< 0.001 *	Violation of normality
Levene's Test	17.927	< 0.001 *	Violation of homogeneity of variances

* p-values smaller than 0.001 are denoted as <0.001 for clarity.

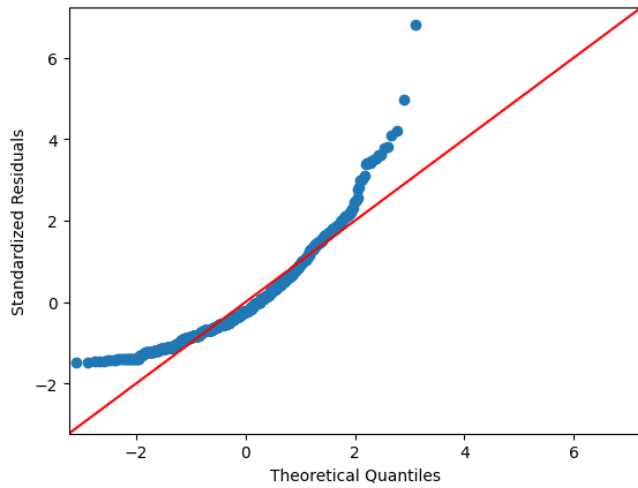


Figure-4 Q-Q plot of one-way ANOVA

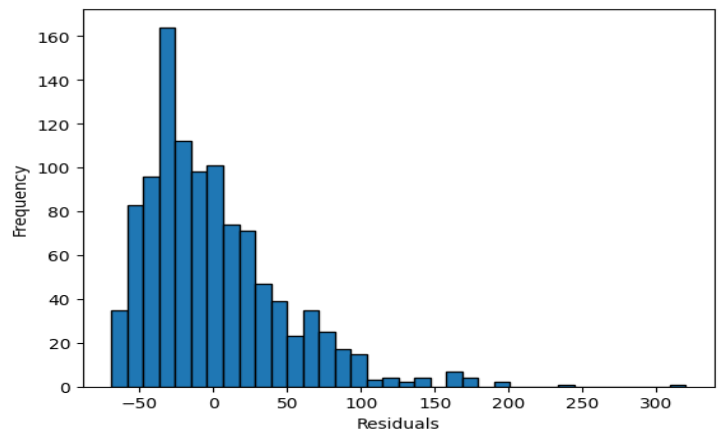


Figure-5 residual from one-way ANOVA

Explanation:

- Shapiro-Wilk Test: The statistic is around 0.902, and the p-value is significantly small (<0.001), indicating the data doesn't follow a normal distribution.
- In the Q-Q plot, the pronounced curve away from the line, especially at the ends, supports the Shapiro-Wilk test's suggestion of non-normality. The histogram provides another perspective on the distribution of residuals. The distribution appears to be skewed, with a tail extending to the right, which further indicates that the normality assumption may not hold for our data.
- Levene's Test: With a statistic of approximately 17.927 and a p-value also significantly small (<0.001), there is evidence against the homogeneity of variance assumption.
- Table-1 indicates that the assumptions necessary for conducting a traditional one-way ANOVA are violated. While ANOVA is known to be robust, the results of the ANOVA should be interpreted with caution.

ANOVA Results and Interpretation

The one-way ANOVA unveiled significant differences in **TOTSPACE** across the **AUSPICE** categories (Table 2).

Table-2 one-way ANOVA Results

	Sum of square	Degree of freedom	F-stats	p-value PR(>F)
C(AUSPICE)	96112.11	2.0	21.843	<0.001
Residual	2332065.00	1060.0	NaN	NaN

Explanation:

- A p-value of <0.001 for types of agency(AUSPICE) means there is a statistically significant difference in total spaces(TOTSPACE) across different auspices, strong enough to reject the null hypothesis that there is no difference.
- The F-statistic indicates that the variance between the different types of auspices is significantly greater than the variance within each auspice type.

- This discovery warranted a deeper investigation through post-hoc testing to elucidate the pairwise differences.

Post-hoc testing Results and Interpretation

Table-3 Post-hoc test using Tukey's HSD(one-way ANOVA)

group1	group2	Diff	Lower	Upper	q-value	p-value
Non-Profit Agency	Commercial Agency	17.119	9.704	24.535	7.662	0.001
Non-Profit Agency	Public Agency	34.335	16.224	52.445	6.293	0.001
Commercial Agency	Public Agency	17.215	-1.453	35.884	3.061	0.078

Interpretations from the Post-hoc Analysis:

- **Non-Profit Agency vs. Commercial Agency:** A significant difference in mean **TOTSPACE** of 17.119, with a tight confidence interval that does not include 0, indicating that Non-Profit Agencies tend to have significantly more spaces than Commercial Agencies.
- **Non-Profit Agency vs. Public (City Operated) Agency:** The largest observed difference in means at 34.335, again significantly different with a p-value of 0.001. Non-Profit Agencies provide substantially more spaces than Public (City Operated) Agencies.
- **Commercial Agency vs. Public (City Operated) Agency:** Though the mean difference of 17.215 suggests a disparity, the confidence interval crosses 0, and the p-value of 0.078 exceeds the 0.05 threshold, indicating that this difference is not statistically significant at the conventional level.
- These results underscore the variability in **TOTSPACE** among the different categories of childcare centres, specifically highlighting the capacity advantages of Non-Profit Agencies.

Two-way ANOVA Analysis(research question 2)

Assumption Testing

Table 4: Assumptions Test for two-way ANOVA

Test	Statistics	P-value	Interpretation
Shapiro-Wilk	0.902	< 0.001 *	Violation of normality
Levene's Test	12.985	< 0.001 *	Violation of homogeneity of variances

* p-values smaller than 0.001 are denoted as <0.001 for clarity.

Explanation:

- **Shapiro-Wilk Test:** The test statistic is approximately 0.902, and the p-value is significantly less than 0.001, which suggests that the residuals from the two-way ANOVA model do not follow a normal distribution.
- **QQ plot and residual histogram show the same pattern in one-way ANOVA.**
- **Levene's Test:** With a statistic of approximately 12.985 and a p-value of less than 0.001, there is a significant indication that variances across different groups defined by the combination of **AUSPICE** and **subsidy** are not homogenous.

- These test results imply that the assumptions required for a traditional two-way ANOVA may not be fully met in the dataset. While ANOVA is robust to some violations, the two-way ANOVA results need to be interpreted with caution.

ANOVA Results and Interpretation

Table-5 two-way ANOVA Results

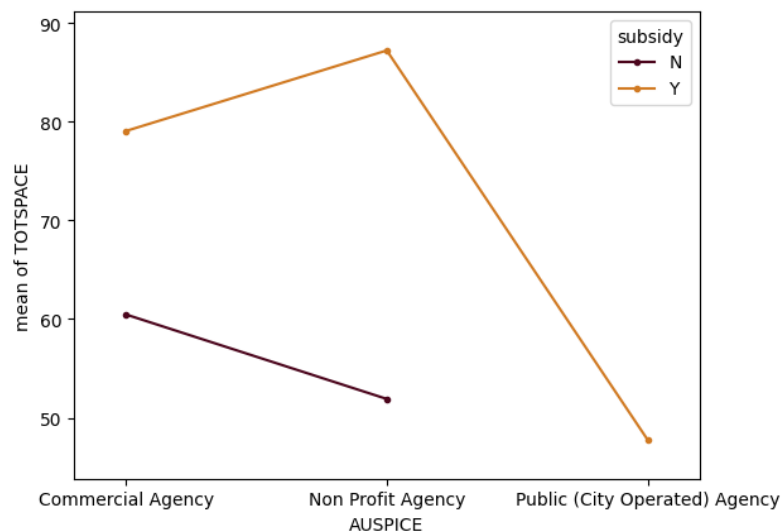
	sum_sq	df	F	PR(>F)
C(AUSPICE)	8568	2.0	2.058	0.128
C(subsidy)	83527.44	1.0	40.118	< 0.001*
C(AUSPICE):C(subsidy)	56034.45	2.0	13.457	< 0.001*
Residual	2202809	1058.0	NaN	NaN

* p-values smaller than 0.001 are denoted as <0.001 for clarity.

Explanation:

- The significant p-values for subsidy and its interaction with AUSPICE indicate that these factors and their interplay significantly affect TOTSPACE. The non-significant p-value for AUSPICE alone suggests that it does not independently influence TOTSPACE when not considering subsidy.

Interaction plot



Explanation:

- The orange line (subsidy "Y") generally sits above the brown line (subsidy "N"), suggesting that agencies with subsidies tend to have more TOTSPACE on average than those without.
- There's a clear interaction effect visible as the lines are not parallel. The difference between subsidized and non-subsidized centres varies depending on the type of AUSPICE. For example, the disparity in TOTSPACE between subsidy statuses for Non-Profit Agencies.
- The steep drop for Public (City Operated) Agencies indicates that subsidies make a significant difference in TOTSPACE for these types of agencies compared to others.

Post-hoc testing Results and Interpretation

Table-6 Post-hoc test using Tukey's HSD(two-way ANOVA)

group1	group2	Diff	Lower	Upper	q-value	p-value
(Non Profit Agency, Y)	(Non Profit Agency, N)	35.328	21.377	49.278	10.225	0.001
(Non Profit Agency, Y)	(Commercial, Y)	8.166	-7.512	23.843	2.103	0.650
(Non Profit Agency, Y)	(Commercial , N)	26.765	16.862	36.668	10.912	0.001
(Non Profit Agency, Y)	(Public Agency, Y)	39.460	17.935	60.986	7.401	0.001
(Non Profit Agency, Y)	(Public Agency, N)	0.000	-inf	inf	0.000	0.900
(Non Profit Agency, N)	(Commercial, Y)	27.162	7.568	46.756	5.597	0.001
(Non Profit Agency, N)	(Commercial , N)	8.563	-6.806	23.932	2.249	0.590
(Non Profit Agency, N)	(Public Agency, Y)	4.133	-20.393	28.658	0.680	0.900
(Non Profit Agency, N)	(Public Agency, N)	0.000	-inf	inf	0.000	0.900
(Commercial, Y)	(Commercial , N)	18.599	1.646	35.552	4.430	0.022
(Commercial ,Y)	(Public Agency, Y)	31.295	5.747	56.843	4.946	0.006
(Commercial, Y)	(Public Agency, N)	0.000	-inf	inf	0.000	0.900
(Commercial , N)	(Public Agency, Y)	12.696	-9.776	35.167	2.281	0.578
(Commercial, N)	(Public Agency, N)	0.000	-inf	inf	0.000	0.900
(Public Agency, Y)	(Public Agency, N)	0.000	-inf	inf	0.000	0.900

Explanation:

The results show significant differences in TOTSPACE when comparing Non-Profit Agencies with and without subsidies, as well as when comparing Non-Profit Agencies with subsidies to Commercial Agencies without subsidies, and to Public (City Operated) Agencies with subsidies. No significant difference was observed between Public (City Operated) Agencies regardless of subsidy status.

These findings highlight the significant influence of subsidy status on the capacity differences within and between different types of childcare agencies. The lack of significant differences in some pairwise comparisons suggests that for certain groupings, operating auspice and subsidy status do not result in substantial changes in total childcare space.

Conclusion

The investigation into Toronto's childcare capacities concluded that both the type of operating auspice and the presence of subsidies significantly affect total available space, with Non-Profit Agencies generally offering more space. Despite violations of normality and homogeneity assumptions, the robust ANOVA framework revealed an interaction between auspice type and subsidy status, suggesting that subsidies' impact on space varies across agency types. This conclusion underscores the need for nuanced policy approaches that consider the multifaceted influences on childcare capacity.