INF2178H (Winter 2024)
Michelle Lee (#1001448169)

Introduction

Child Care in Ontario has been a consistent problem for many years, with the City of Toronto being no exception to this trend. Toronto is not on track to meet its 2026 goal for additional child care spaces and is currently experiencing a deficit in qualified daycare staff for the existing spots (Hasham, 2023). In particular, auspice types can greatly impact whether or not certain centers receive sufficient funding, as well as the quality of care offered. This report will therefore aim to investigate this relationship by performing statistical techniques to quantitatively analyze information on Child Care centers in Toronto City.

Our main research questions will be:

Q: How do auspice types (public, commercial, and non-profit) influence the availability and distribution of child care spaces in Toronto?
Q: What is the impact of building type on the total group space available in child care centres across Toronto?
Q: Are there any specific combinations of auspice type and building type that are significantly beneficial or detrimental to the amount of space in child care centers available for age groups specific to toddlers (18-30 months)?

Data and Methodology

Data Source

Our dataset was sourced from Open Toronto's online repository, and consists of 1063 entries, indicating child care centers, with various attributes across 17 columns.

We isolated our variables of concern from the dataset to focus primarily on the following:

- **AUSPICE:** Operating auspice type, divided by Commercial, Non Profit or Public
- **SUBSIDY:** Whether or not the center has a fee subsidy contract, divided by Y/N (Yes/No)
- **TGSPACE:** Child care spaces for toddlers 18-30 months

We chose AUSPICE as a categorical variable to assess its impact across different age groups. It helps in understanding the variety of service delivery models and their potential impacts on quality, accessibility, and affordability of child care.

We also chose to focus on SUBSIDY as another categorical variable of interest to assess whether or not existing government fee subsidy contracts and related policies have had any statistically significant impact on supply and quality of child care centers.

TGSPACE in particular was selected as the continuous dependent variable from all age groups because children of toddler age (1-3 years) are considered to be a significant age group due to significant leaps in language development, motor skills, as well social and cognitive skill developments at this particular age for children, hence the importance of this demographic receiving childcare. (Howes et al., 1986)

Data Preparation

Data preparation involved melting the DataFrame to restructure it for analysis. We transformed the data from a wide format, where each age group capacity (IGSPACE, TGSPACE, PGSPACE, KGSPACE, SGSPACE) was in its own column, into a long format with two new columns: 'AgeGroup' for the age categories and 'Capacity' for

their respective values. This was essential for conducting the ANOVAs to help make the comparison across different age groups and auspice types. We also converted relevant columns to categorical types and removed missing values to ensure the data's cleanliness for subsequent analysis.

Statistical Analysis Methodology

Our methodology will primarily involve performing ANOVA, as well as multiple tests to check if assumptions for ANOVA are met, and post-hoc tests on results. ANOVAs are useful because they analyze the influence of one or more categorical independent variables on a continuous dependent variable, helping investigate underlying relationships between different variables. A one-way ANOVA will help us check the impact of a single factor on a continuous outcome, while a Two-way ANOVA was also chosen because it helps us to explore the effects of two independent categorical variables on a continuous outcome variable while also investigating the interaction between these factors.

Descriptive Analysis

This section will help examine, summarize, and visualize the main characteristics of our dataset. It will help provide a brief overview of some trends and patterns through descriptive and graphical visualizations.

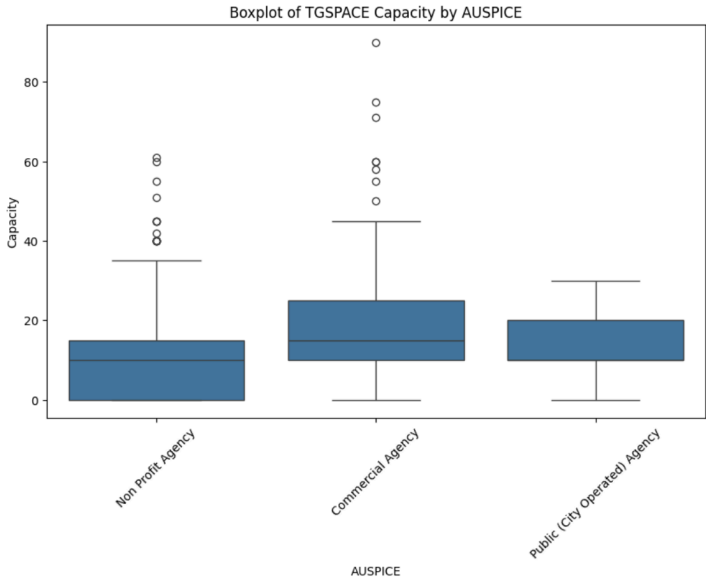


Figure 1: Boxplot of Auspice types vs Capacity

The boxplot shows that there are some slight differences in capacity variability and central tendency among the three auspice types. In general: the Commercial Agencies category has a higher median capacity than others, while Non-Profit Agencies has fewer outliers and a slightly smaller range, and Public (City Operated) Agencies tend to be more consistent, with the fewest outliers and the least variability. Commercial Agencies show the largest range as shown by its large whiskers, indicating the greatest variation in capacity.

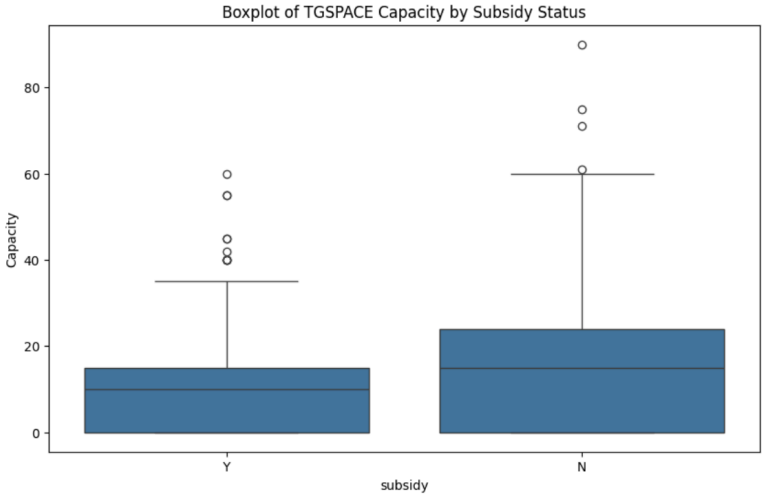


Figure 2: Boxplot of Subsidy Types vs Capacity

Figure 2 meanwhile indicates that while central tendency for capacity is overall very similar regardless of subsidy type, variability seems relatively lower for subsidized centers (shown by a narrower IQR). Both Y and N groups have outliers, which means that there are some centers with capacities much larger than typical for both subsidized and unsubsidized centers. This could be due to differences in facility size, efficiency in space utilization, or other reasons. Both categories have medians approximately at the same level, suggesting a similar central tendency for capacity regardless of subsidy status.The IQR for centers with Y subsidies seems slightly narrower than N subsidies, possibly indicating less variability in capacity for Y subsidy centers.

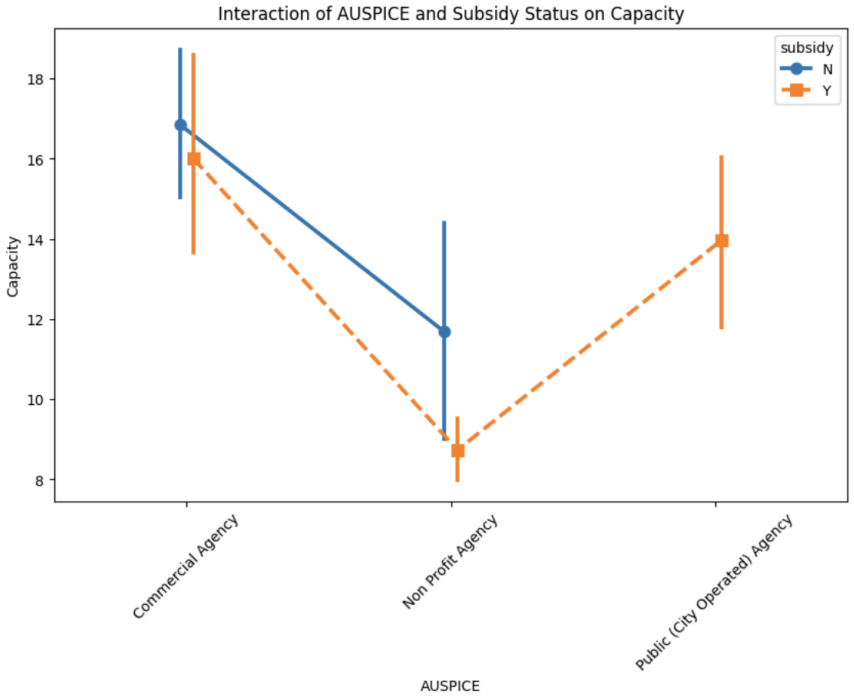


Figure 3: Interaction Plot

Figure 3 compares the interaction between capacity of child care centers across different auspice types with centers that have a subsidy (Y) and centers that do not (N). It shows how the effect of subsidies differs significantly based on whether a child care center is commercially operated, non-profit, or publicly operated.

The non-parallel lines indicate some interaction effect between auspice and subsidy based on capacity. Specifically, capacity trends based on subsidy types are not consistent across auspices, whereby capacity decreases with subsidies in Commercial Agencies but increases in Non-Profit and Public Agencies. This might mean in reality, subsidies vary the ability for some centers to increase spaces available - subsidies might help Non-Profit and Public Agencies to provide more spaces, whereas Commercial Agencies may not rely on subsidies to the same extent to offer more spaces.

Results

One-Way ANOVA: Auspice and TGSPACE

	Sum of Squares	Degrees of Freedom	F-Value	P-Value
C(Auspice)	**12587.669972**	**2.0**	**46.785974**	**3.376175e-20**
Residual	**142595.40999**	**1060.0**	**NaN**	**NaN**

The one-way ANOVA results show a statistically significant difference in the capacity among the different types of child care auspices, with an F-value of 46.79 and a p-value approximately zero (3.38e-20). The sum of squares for the auspice effect is 12,587.67, meaning that the type of auspice accounts for some variability in capacity. The residual sum of squares is 142,595.41, meaning that while auspice type is significant, there is also considerable variation within auspice categories that it does not account for. Given the highly significant p-value, we can confidently reject the null hypothesis that all auspices have the same average capacity.

Two-Way ANOVA: Auspice, Subsidy and TGSPACE

	Sum of Squares	Degrees of Freedom	F-Value	P-Value
C(Auspice)	**6100.351067**	**2.0**	**22.759087**	**2.101687e-10**
C(Subsidy)	**858.516307**	**1.0**	**6.405876**	**1.151831e-02**
C(Auspice):C(Subsidy)	**192.767266**	**2.0**	**0.719173**	**4.873930e-01**
Residual	**141793.285497**	**1058.0**	**NaN**	**NaN**

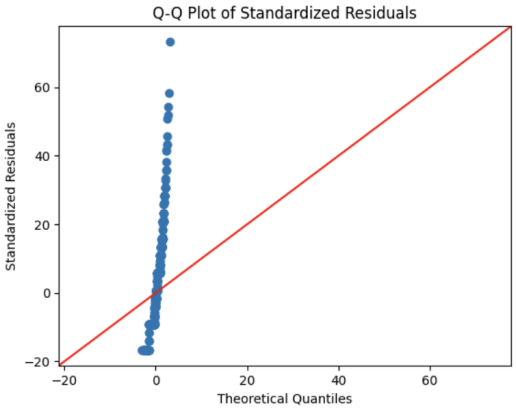
The two-way ANOVA meanwhile shows significant effects on 'TGSPACE' capacity due to 'AUSPICE' with an F-value of 22.759 and a p-value of approximately 2.10e-10. This means that different auspice types have different average capacities. The effect of 'subsidy' is also significant with a p-value of 0.0115, meaning that the availability of subsidies does affect capacity. However, the interaction between 'AUSPICE' and 'subsidy' is not significant given an F-value of 0.719 and a p-value of 0.487. This means that subsidy does not vary by auspice type in a statistically significant way. Overall, while both the type of child care center and subsidy status have their own effects on capacity, there is no compounded interaction effect between auspice and subsidy factors.

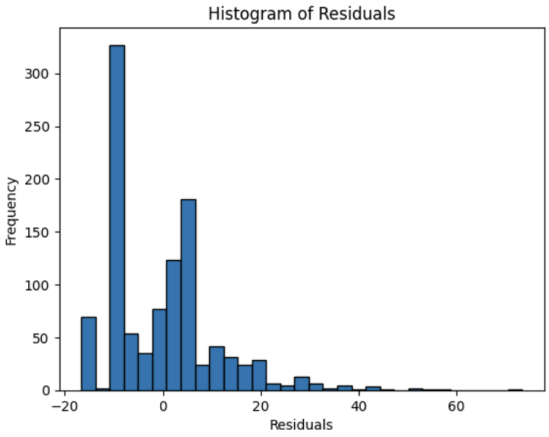
Post-Hoc Analysis

This section will evaluate using statistical tests whether the conditions for ANOVA were met based on the characteristics of our dataset.

INF2178H (Winter 2024)
Michelle Lee (#1001448169)

One-Way ANOVA Post-Hoc:



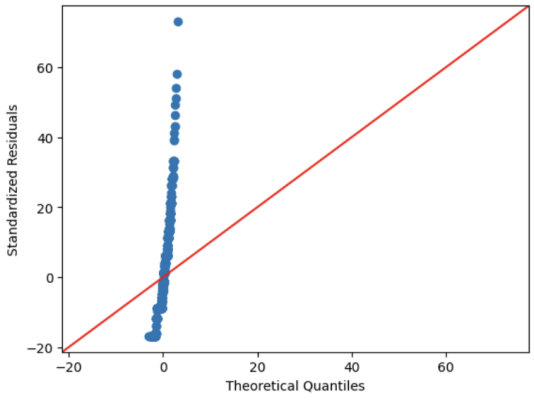


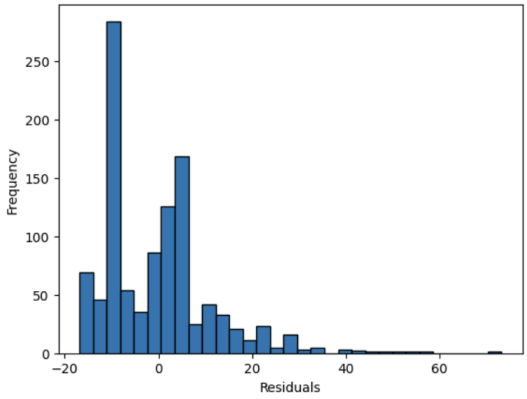
The Q-Q plot and histogram of residuals from the ANOVA model indicate a violation of the normality assumption: the plot has an S-shaped pattern, which diverges away from the theoretical line, which suggests a non-normal distribution with pronounced tails. Additionally, the histogram displays a rightward skew rather than the bell curve typical of normal distributions. Although ANOVA has a robustness to normality violations in cases of large sample sizes, the significant skewness and non-normality observed may mean it's necessary to transform the data or the use of non-parametric tests.

Shapiro-Wilk test	W Statistic = **0.886004**	P-Value = **2.908532e-27**
Levene's test	Test Statistic = **7.515801**	P-Value = **0.0005739**

Furthermore, based on Shapiro-Wilk and Levene's tests performed on the residuals, two key assumptions of ANOVA may have been violated. The Shapiro-Wilk test yields a W statistic substantially less than 1 and a very small p-value, indicating a significant deviation from normality in the residuals. Additionally, Levene's test shows a small p-value, implying that the variances across groups are not equal.

Two-Way ANOVA Post Hoc:





Very similar to the previous one-way ANOVA, the histogram and Q-Q plot for the residuals from the two-way ANOVA model reveal violations and deviations from normality. The Q-Q plot shows an S-shaped curve which indicates heavier tails in the distribution of residuals. In addition, the histogram is skewed rightwards, showing a greater frequency of smaller residuals and a few larger outliers, rather than the expected symmetric bell curve for normally distributed data. Overall, the normally distributed assumption for residuals has not been satisfied based on these results.

Shapiro-Wilk test	W Statistic = **0.899658**	P-Value = **8.582114 e-26**
Levene's test	Test Statistic = **25.285676**	P-Value = **8.228154 e-21**

Similar to above, the Shapiro-Wilk and Levene's tests for the two-way ANOVA reveal that two key assumptions, normality of residuals and homogeneity of variances, are not met. The Shapiro-Wilk test shows a significant deviation from normality with a W statistic of 0.8997 and a very low p-value. Levene's test also shows lack of equal variances across groups with a large test statistic and a very small p-value. These violations imply that the standard ANOVA results may not be reliable, overall.

Discussion

Our findings indicate that the data does not satisfy the normality and homogeneity of variance assumptions, casting doubt on the validity of the ANOVA results and potentially necessitating alternative analysis methods, especially for small or unbalanced sample sizes. This could potentially compromise the validity of the ANOVA's results, especially if the sample size is not sufficiently large to invoke the Central Limit Theorem and there may be risks regarding the sample size being too small or if equal variances across groups are also in question.

Overall, we can say from our findings that significant disparities in childcare space availability across different auspice types do exist in Toronto, underlining the influence of organizational structure on childcare provision. The variance attributed to subsidies further emphasized the role of financial assistance in shaping the landscape of childcare services. However, the assumption checks for both one-way and two-way ANOVA exposed non-normality in the data distribution and unequal variances across groups, which puts the integrity of the ANOVA findings into question, particularly for smaller sample sizes. This insight aligns with Cleveland's research (2018) on the affordability of childcare and supports ongoing discussions, such as those put forth by the Government of Canada (2022) and The Toronto Star (Hasham, 2023), about the need for policy reforms. The study's outcomes suggest potential directions for policy amendments and further empirical inquiries to substantiate these findings and foster improved childcare models that accommodate the developmental needs of toddlers, as emphasized by Howes and Olenick (1986).

Conclusion

This study highlights the need for further research assessing the factors influencing childcare space availability in Toronto, where auspice type and subsidy status have been identified as significant contributors. The implications of these findings include how there is an urgent need for interventions aimed at equalizing the distribution of childcare spaces - such as subsidy allocations to ensure equitable access. Future research should focus on replicating this study with larger sample sizes and exploring statistical methods that could provide more reliable insights into the complex dynamics of childcare space distribution. In conclusion, the insights from this study support the need to address existing disparities within childcare spaces across Toronto.

References

Cleveland, Gordon. "Affordable for all: Making licensed child care affordable in Ontario." *Toronto: Cleveland Consulting* (2018). https://www.childcarepolicy.net/wp-content/uploads/2018/05/Presentation-at-OISE-on-Affordable-for-All-report.pdf

Government of Canada. 2022. “Reducing Fees for Licensed Child Care for Ontario Families by 50% on Average.” Canada.ca. Government of Canada. December 19, 2022. https://www.canada.ca/en/employment-social-development/news/2022/12/reducing-fees-for-licensed-child-care-for-ontario-families-by-50-on-average.html .

Hasham, Alyshah. 2023. “The Toronto Star.” Toronto Star. November 28, 2023. https://www.thestar.com/news/gta/we-are-in-a-perfect-storm-toronto-faces-child-care-space-crisis-advocates-say/article_5764ba8c-7411-58fe-994d-b84251948506.html#:~:text=Toronto%20is%20set%20to%20fall.report%20and%20child%20care%20advocates.

Howes, Carollee, and Michael Olenick. “Family and Child Care Influences on Toddler’s Compliance.” *Child Development* 57, no. 1 (1986): 202–16. https://doi.org/10.2307/1130652