

## Examination Of Toronto Childcare Centers

### 1. Introduction

Childcare is delivered by a mixture of non-profit, for-profit and publicly operated programs with non-profit services predominating (Friendly et al. (2013)). Three-quarters of families in Toronto can't afford the cost of licensed childcare. The issue of childcare has been on the agenda for both the federal and provincial governments lately, but rather than simply adding more childcare spaces governments need to also take a hard look at the affordability issue (Campbell, 2016).

This report explores Toronto licensed childcare centers through the dataset INF2178\_A2\_data.xlsx that outlines data points updated in February 2024 on the operation and the capacity of these centers for various age groups. The analysis is guided by the following questions:

- **Research question 1:** Is there a statistically significant difference in the total childcare spaces among different operating agencies in Toronto, considering the challenges associated with high fees and low availability of childcare spaces?
- **Research question 2.1.:** Is there a statistically significant joint effect of having a fee subsidy contract and being part of a CWELCC program on the total childcare spaces available for all age groups in childcare centers?
- **Research question 2.2:** Do fee subsidy contracts and CWELCC program interact in a way that significantly influences the variation in childcare spaces?

### 2. Data Cleaning and Wrangling

The dataframe INF2178\_A2\_data.xlsx has a total of 1053 rows and 16 columns: 8 integer columns and 8 object columns, with '\_id' as an index column. The dataframe requires some cleaning and treatment of the missing variables.

The following columns have been renamed respectively to make the dataset look consistent: '\_id' to 'ID'; 'ward' to 'WARD'; 'bldg\_type' to 'BLDGTYPE'; 'subsidy' to 'SUBCONTR'; 'cwelcc\_flag' to 'CWELCCFL'.

The values 'Y' and 'N' for 'CWELCCFL' variable has been renamed to "Part" for participation in the CWELCC program and "NOPart" to indicate that the center do not participate in the CWELCC program.

The missing variable under "BLDGTYPE" variable, i.e. building type, has 348 missing values that has been filled in with 'Undefined' value.

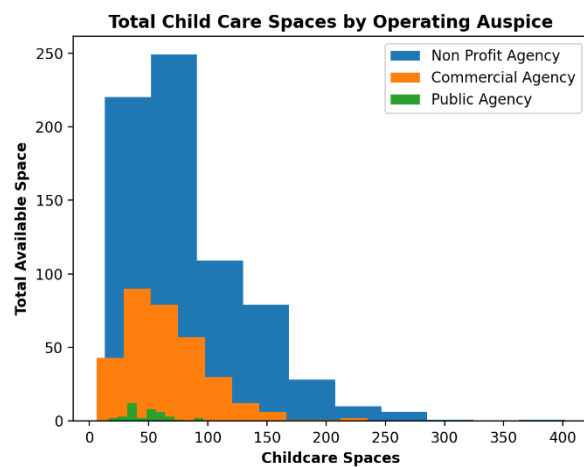
### 3. Exploratory Data Analysis

The analysis extensively explored the dataset, and it has been reflected in the Python notebook. Only key results have been included in this paper.

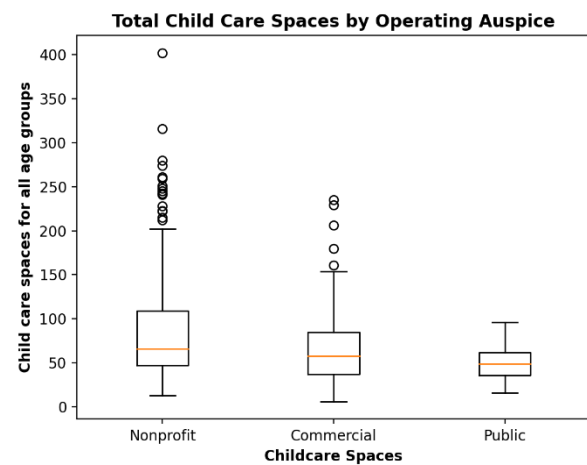
The majority of the childcare centers are operating as non-profits has greater number of centers (M=66, iqr=62) followed by commercial operating center that are moderate in number (M=58, iqr=48) and public ones are fewer and less dispersed (M=49, iqr=26). Thus, nonprofit childcare centers emerge as a prominent category, featuring a higher median, larger IQR, and substantial contribution to the total childcare spaces (Figure 1). Additionally, nonprofit childcare centers have considerably large number of outliers. In contrast, commercial centers, although demonstrating a moderate capacity and spread, fall behind nonprofit centers in terms of overall representation (Figure 2). The difference between operating agencies in

terms of the total space available for childcare has driven the first research question to identify whether that difference is statistically significant.

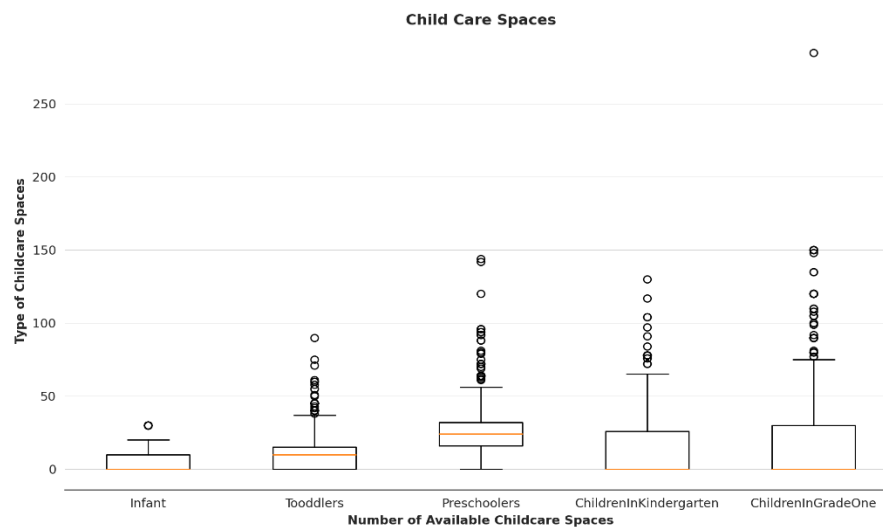
**Figure 1**



**Figure 2**



Childcare spaces by the age groups have also been examined (Figure 3). Based on the summary statistics, childcare spaces for preschoolers 30 months up until they enter grade one has the highest median ( $M=24$ ,  $IQR=16$ ) and the highest spread has been observed for childcare spaces for grade one and up ( $M=0$ ,  $IQR=30$ ). The data for all types of childcare centers has been skewed.



**Figure 3**

The total number of childcare spaces for all age groups have been examined from the position of participation in the fee subsidy program and Canada-Wide Early Learning & Childcare (CWELCC). Both programs help families with the cost of licensed childcare programs. The majority of the total spaces are available for both programs with only a few not participating in the CWELCC (Figure 4, 5). These variables have been selected to examine joint effect of the programs on the total childcare spaces.

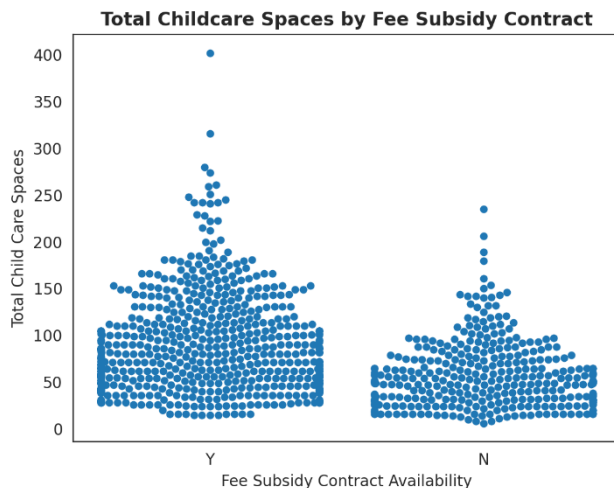


Figure 4

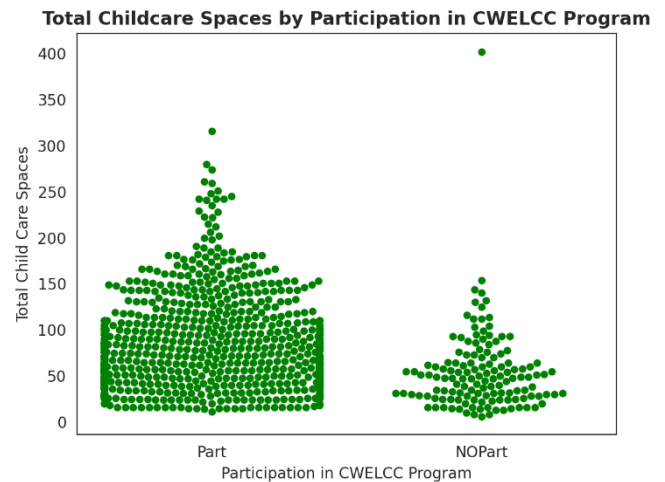


Figure 5

## 4. Discussion

### 4.1. Childcare spaces among operating agencies in Toronto

**Research question 1:** Is there a statistically significant difference in the total childcare spaces among different operating agencies in Toronto, considering the challenges associated with high fees and low availability of childcare spaces?

To respond to the research question, one-way ANOVA test has been employed as it allows for the examination of variations in means across different operating agencies. Additionally, log transformation was applied to the continuous variable `TOTSPACE` to enhance the symmetry of the data distribution and to ensure the robustness of the results.

There is a statistically significant difference between in the total childcare spaces among commercial, non-profit, public agencies in Ontario, with  $F=18.430433$  and  $p=0.000000013$  (Table 1).

	df	sum_sq	mean_sq	F	PR(>F)
C(AUSPICE)	2.0	14.143756	7.071878	18.430433	1.354538e-08
Residual	1060.0	406.728946	0.383707	NaN	NaN

Table 1

We reject the null hypothesis and conclude that at least one operating auspice has a different mean total childcare space compared to the others.

The post-hoc test has been employed to conduct pairwise comparisons between different pairs of operating agencies and identify specific group differences in mean total childcare spaces. In the presented findings, the post-hoc analysis focused on comparing non-profit agencies with both commercial agencies and public agencies. First, the post-hoc test indicates statistically significant differences in mean total childcare spaces compared between non-profit agencies and commercial agencies, with  $p$  value = 0.001. The result indicates a significant mean difference of 0.205 childcare spaces, with a 95% confidence interval ranging from 0.107 to 0.303. Second, statistically significant differences have been also identified in mean total childcare spaces compared between non-profit agencies and public agencies, with  $p$  value = 0.001. The result indicates a significant mean difference of 0.424 childcare spaces, with a 95% confidence interval ranging from 0.185 to 0.663. However, the post-hoc shows that

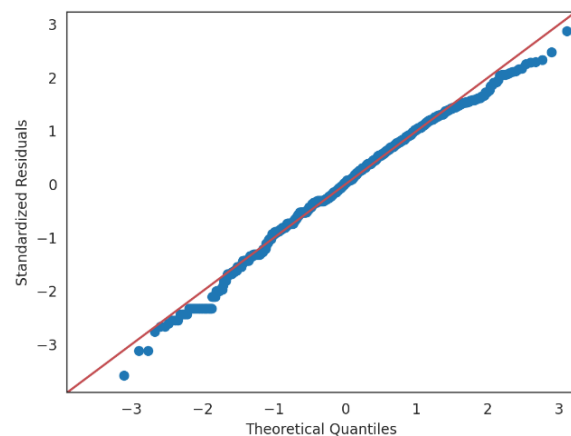
there are no statistically significant differences in mean total childcare spaces compared between commercial and public agencies, with p values 0.094383 (Table 2).

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non-Profit Agency	Commercial Agency	0.205307	0.107371	0.303243	6.958243	0.001000
1	Non-Profit Agency	Public (City Operated) Agency	0.423834	0.184660	0.663008	5.881919	0.001000
2	Commercial Agency	Public (City Operated) Agency	0.218527	-0.028014	0.465067	2.942074	0.094383

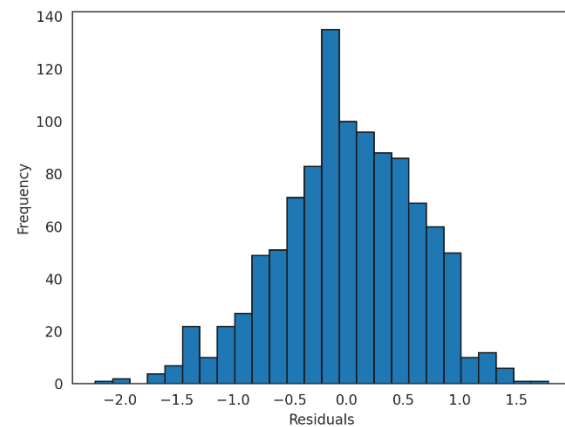
**Table 2**

The assumption tests have been utilized to evaluate two critical assumptions: normality and homogeneity of variances. The assumption check suggests the following:

- The skewness in the histogram and deviations from the diagonal line in the QQ plot indicate that the standardized residuals are not normally distributed, but close to be normally distributed (Figure 6, 7).



**Figure 6**



**Figure 7**

- The null hypothesis in Levene's test is that the variances across different groups are equal. The small p-value (0.0045) suggests that there is strong evidence to reject the null hypothesis. The low p-value indicates that variances are significantly different among the operating auspices. The variances are not homogeneous across the different operating auspices.

Based on the assumptions above, the robustness of the results should be reconsidered. Specifically, the violation of normality assumption affects the reliability of p-values and confidence intervals. The violation of the homogeneity of variances assumption impacts the validity of the one-way ANOVA results.

#### 4.2. Effect of Fee Subsidy contract and CWELCC program on childcare spaces

**Research question 2.1:** Is there a statistically significant joint effect of having a fee subsidy contract and being part of a CWELCC program on the total childcare spaces available for all age groups in childcare centers?

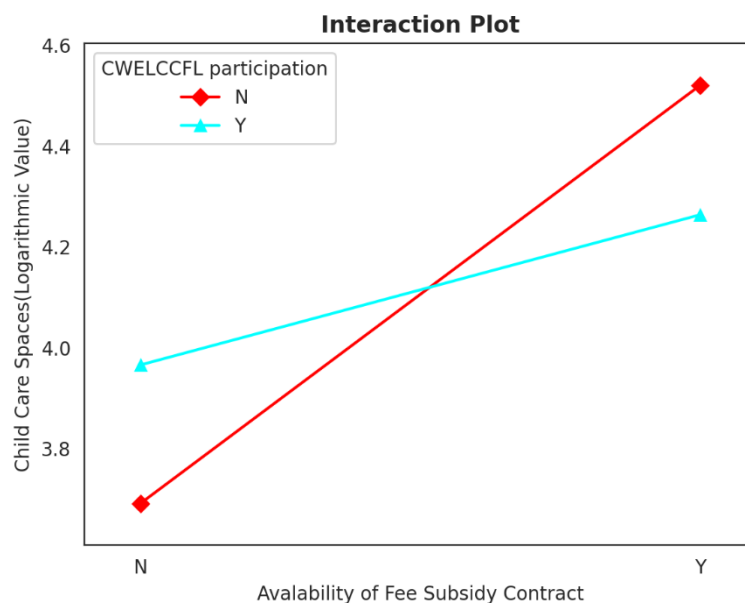
**Research question 2.2:** Do fee subsidy contracts and CWELCC program interact in a way that significantly influences the variation in childcare spaces?

A two-way ANOVA has been employed to respond to research questions 2.1 and 2.2. To enhance the robustness of the analysis, a logarithmic transformation has been applied to the continuous variable 'TOTSPACE'. The p-values for availability of fee subsidy contract ( $p = 0.00000000000014$ ) and participation in CWELCC program ( $p = 0.0012$ ) are low (both  $p < 0.05$ ), indicating that both variables individually have a significant impact on the log-transformed total childcare spaces. The p-value for the interaction between availability of fee subsidy contract and participation in CWELCC program ( $p = 0.003179021$ ) is also significant. The interaction effect significantly contributes to the variation in childcare spaces (Table 3).

	sum_sq	df	F	PR(>F)
C(SUBCONTR)	19.920678	1.0	56.030970	1.497549e-13
C(CWELCCFL)	3.730290	1.0	10.492201	1.236067e-03
C(SUBCONTR):C(CWELCCFL)	3.107954	1.0	8.741755	3.179021e-03
Residual	376.506028	1059.0	NaN	NaN

**Table 3**

The interaction plot suggests that there is an interaction effect between availability of fee subsidy and participation of the center in CWELCC program, i.e. between categorical variables 'SUBCONTR' and 'CWELCCFL'. The crossing pattern indicates that the combined effect of having a fee subsidy contract and participating in the CWELCC program is not consistently positive or consistently negative across different levels. Instead, the interaction effect changes direction between the levels of the variables, suggesting a relationship between fee subsidies and CWELCC program participation in influencing childcare spaces. This emphasizes the need to consider the joint effects of these variables (Figure 8).



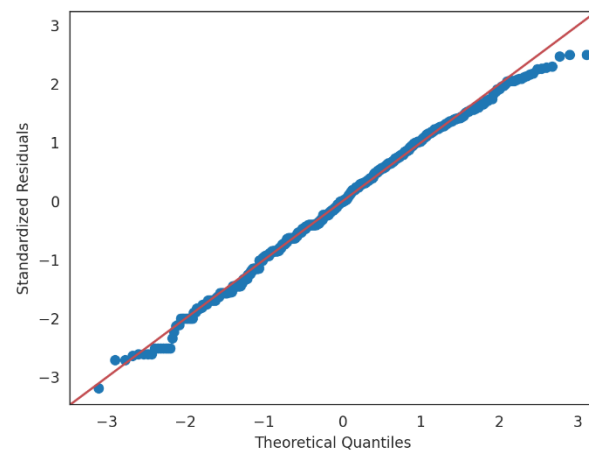
**Figure 8**

The post hoc test results for the two-way ANOVA provide insights into the differences in childcare spaces across different conditions. The post hoc test suggests that there is no significant difference in childcare spaces between centers with fee subsidies that are part of CWELCC program and those that are not the part of the CWELL program. The p value in this instance equals 0.41589. In contrast, in all other pairwise comparisons, the p-values are notably low, suggesting significant differences (Table 4).

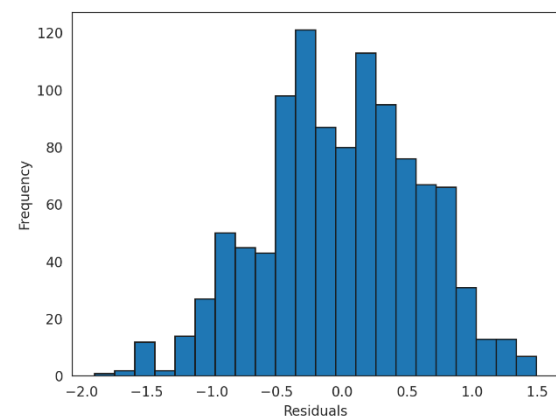
	group1	group2	Diff	Lower	Upper	q-value	p-value
0	(Y, Part)	(Y, NOPART)	0.256861	-0.172573	0.686295	2.176605	0.415859
1	(Y, Part)	(N, Part)	0.297904	0.179623	0.416186	9.165143	0.001000
2	(Y, Part)	(N, NOPART)	0.572653	0.423246	0.722061	13.947515	0.001000
3	(Y, NOPART)	(N, Part)	0.554766	0.116900	0.992631	4.610484	0.006311
4	(Y, NOPART)	(N, NOPART)	0.829514	0.382235	1.276793	6.748748	0.001000
5	(N, Part)	(N, NOPART)	0.274749	0.102600	0.446897	5.807780	0.001000

**Table 4**

The assumption check, including an assessment of normality and homogeneity of variances, was conducted to evaluate whether the underlying assumptions of the two-way ANOVA model were met (Figure 9,10). The assumption check suggests that the residuals are not normally distributed. For the second assumption Levene's test for homogeneity of variance has been employed. With the p-value equal to 0.2135, the findings fail to reject null hypothesis. This suggests that there is not enough evidence to conclude that the variances are significantly different among the groups. The violation of the normality assumption affects the reliability of p-values and confidence intervals in the two-way ANOVA model. While Levene's test did not find significant differences in variances, the proximity of the p-value to the significance threshold suggests a potential weakness in the homogeneity of variances assumption.



**Figure 9**



**Figure 10**

## 5. Limitations

The assumption check results indicate potential vulnerabilities in one and two-way ANOVA models. These issues for both models should be considered when interpreting the results, and alternative analyses or robust methods may need to be explored to ensure the reliability of conclusions drawn from the model.

The dataset could benefit from additional data points to facilitate a comprehensive analysis on childcare spaces in Toronto. Firstly, the current dataset is limited to displaying only the supply side, specifically the availability of childcare spaces. However, it lacks data on the demand side, such as the number of children from different age groups. Including this information would provide an opportunity to compare supply and demand and identify any disparities. Additionally, the analysis should explore the geographical distribution of childcare spaces. This would offer a clear understanding of the areas that require more support in creating additional childcare spaces.

## Reference:

Campbell, D (2016). Majority of Toronto families can't afford licensed child care new U of T Scarborough study finds, Retrieved from <https://utsc.utoronto.ca/news-events/archived/majority-toronto-families-cant-afford-licensed-child-care-new-u-t-scarborough-study-finds>

Canada-Wide Early Learning & Child Care (CWELCC) System – Operator Information, Retrieved from <https://www.toronto.ca/community-people/community-partners/early-learning-child-care-partners/canada-wide-early-learning-child-care-agreement-cwelcc/>

Child Care Fee Subsidy – Required Documents, Retrieved from <https://www.toronto.ca/community-people/employment-social-support/child-family-support/child-care-support/child-care-fee-subsidy-eligibility-meeting-documents/>

Childs, D. Z., Hindle, B. J., & Warren, P. H. (Year). APS 240: Data Analysis and Statistics with R, Retrieved from <https://dzchilds.github.io/stats-for-bio/data-transformations.html>

Does your data violate one-way ANOVA assumptions?, Retrieved from [https://quality-control-plan.com/StatGuide/oneway\\_anova\\_ass\\_viol.htm#:~:text=If%20the%20populations%20from%20which,may%20be%20incorrect%20or%20misleading.](https://quality-control-plan.com/StatGuide/oneway_anova_ass_viol.htm#:~:text=If%20the%20populations%20from%20which,may%20be%20incorrect%20or%20misleading.)

Does your data violate one-way ANOVA assumptions?, Retrieved from [https://quality-control-plan.com/StatGuide/oneway\\_anova\\_ass\\_viol.htm#:~:text=If%20the%20populations%20from%20which,may%20be%20incorrect%20or%20misleading.](https://quality-control-plan.com/StatGuide/oneway_anova_ass_viol.htm#:~:text=If%20the%20populations%20from%20which,may%20be%20incorrect%20or%20misleading.)

Friendly, M., Beach, J., & Turiano, M. (2013). Early childhood education and care in Canada. Toronto: Childcare Resource and Research Unit.

Introduction to ANOVA for Statistics and Data Science, Retrieved from <https://www.analyticsvidhya.com/blog/2020/06/introduction-anova-statistics-data-science-covid-python/>

One-way ANOVA, Retrieved from <https://www.pythonfordatascience.org/anova-python/>

Python for Data Science:One way ANOVA, Retrieved from <https://www.pythonfordatascience.org/anova-python/>

Jan, A. U., Jan, D., Yousaf Hayat, A. S., Ali, G., & Fayaz, M. (2014). Logarithmic transformation: a tool for normalizing residuals in Anova models. Sarhad Journal of Agriculture, 30(3), 375-378.