March 9th, 2024

INF2178

Shion Guha

Benjamin Rabishaw

INF2178 Technical Assignment 2:

ANOVAs

I. Introduction

In 2021, the federal government of Canada budgeted a nation-wide childcare subsidy for children under the age of 6, with the goal for families to pay no more than $10 a day for high-quality community-based childcare (PMO, 2021). In doing so, the government hoped to free women in particular to join or re-join the workforce following the initial years of the COVID-19 pandemic, thereby encouraging Canada's broader social and economic recovery (PMO, 2021). The first province to sign on to the childcare agreement with the federal government was British Columbia (PMO, 2021), while the province of Ontario joined the next year (PMO, 2022).

Nearly two years afterward, demand from parents in Ontario for subsidized childcare space has grown significantly (Jones, 2024). This increased demand resulting from more affordable fees has led to a shortage in childcare spaces and an increase in wait times for community-based childcare, with waitlists in some communities nearly doubling in estimated duration – long enough that some children may age out of these spaces completely, before they are able to enter them at all (Jones, 2024). Thus, supply for community-based childcare spaces has not yet met the increased demand for them, even though Ontario has added more than 41,000 spaces since 2019 (Jones, 2024). To better understand this bottleneck and how to fulfill the promise of national affordable childcare, this technical assignment examines a dataset of childcare centres in Toronto, Ontario. Specifically, this assignment investigates two main research questions:

1. Is there a significant difference in childcare capacity per type of childcare centre?
2. Is there a significant interaction between type of childcare centre and participation within the federal childcare subsidy program on capacity?

II. The Data and Data Cleaning

The data was retrieved from the INF2178 Quercus module as a Microsoft Excel file with the xlsx file format. The data originates from the City of Toronto's Open Data initiative. The name of the file is "INF2178_A2_data.xlsx", and features a length of 1,063, with 17 columns, including relevant variables such as:

- LOC_NAME, meaning the name of the childcare centre

- AUSPICE, meaning the type of operating organization, between Commercial, Non-Profit, or Public
- Ward, meaning the City ward where the centre is located
- IGSPACE, meaning capacity for infants (up to 18 months-old)
- TGSPACE, meaning capacity for toddlers (up to 30 months-old)
- TOTSPACE, meaning capacity for children of all age groups
- Cwelcc_flag, meaning whether the centre is part of the Canada-wide childcare subsidy program (Y/N)

Only columns relevant for the two research questions above were retained within the dataframe – others were removed.

III. Research Question #1: One-Way ANOVA

To explore Research Question 1, this technical assignment performed a one-way ANOVA test using total childcare capacity, or TOTSPACE, as the dependent variable and type of childcare centre organization, or AUSPICE, as the independent variable. A one-way ANOVA features two assumptions which must be checked before continuing with proper caution. These are, first, the assumption of residual normality for the test's dependent variable, and, second, the assumption of variance homogeneity.

To test for the first assumption, we performed a Shapiro-Wilk test, of which the results were a test statistic of 0.90 and a p-value of <0.0001. Because the p-value is so small, we reject the null hypothesis, indicating that the values do not appear normally distributed. So, we applied a Box-Cox transformation of the data, which iterates transformations until a valid one is reached. A rerun of the Shapiro-Wilk test using the Box-Cox transformed data produced results of a 0.997 test statistic and a p-value of 0.044, which is very close to greater than the alpha of 0.05. As ANOVA tests are somewhat robust to imperfectly-distributed data, we accepted the normality assumption under this transformation.

Then, to test for the second assumption of variance homogeneity, we found that we were unable to run traditional Barlett or Levene tests, as the number of childcare centres under each Auspice was not close to equal, as can be seen from the unequal number of red dots plotted in Figure 1, below:

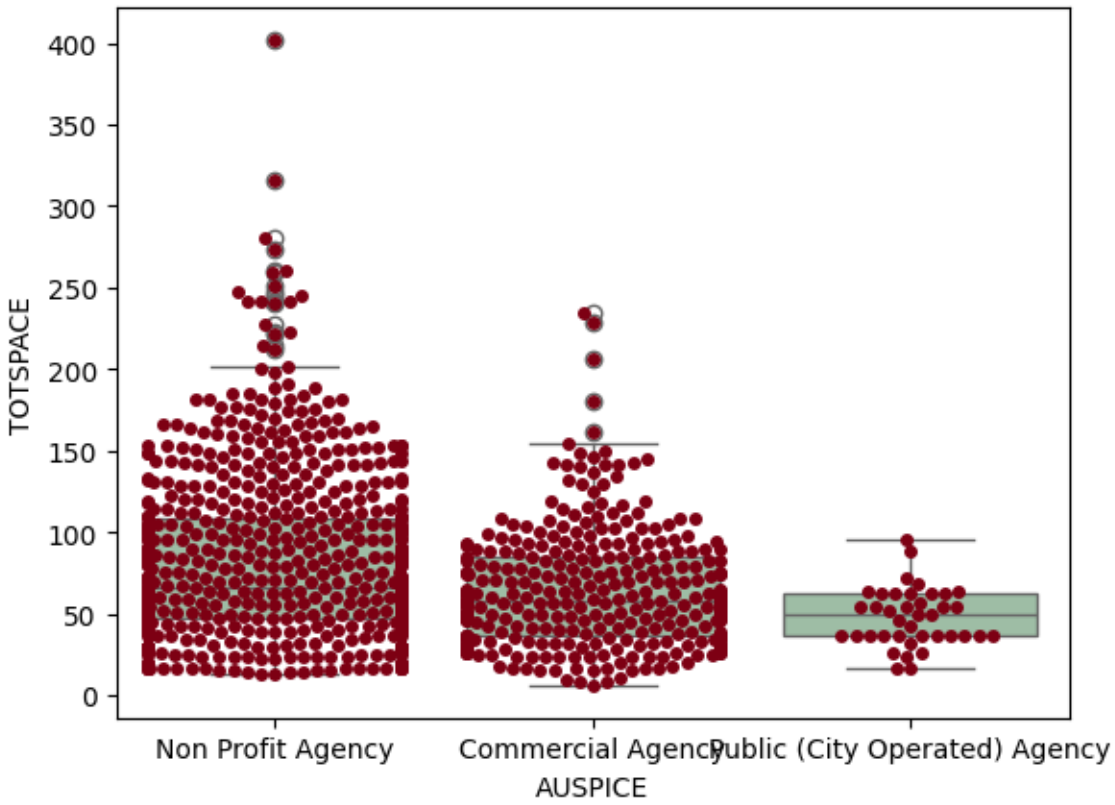

Figure 1: A Boxplot of total number of childcare spaces per Auspice in Toronto

So, rather than a traditional Barlett or Levene test, we ran a variation of the Levene test called the Brown-Forsythe test, which produced a test statistic of 7.65 and a p-value of 0.00049. As the p-value is quite small, we reject the Null Hypothesis and find that the variances do not appear homogenous. Therefore, we further transformed the data by calculating the square roots of the Box-Cox-transformed capacity values, and reran the Brown-Forsythe test. With the further transformed data, our second test produced a test statistic of 5.00 and a p-value of 0.0068. From there, we proceeded with the one-way ANOVA with the understanding that this p-value is closer to the alpha of 0.05, that ANOVAs are somewhat robust to imperfect variance homogeneity, and that the assumption of normality is a more consequential one for one-way ANOVAs than the homogeneity assumption.

Proceeding to the one-way ANOVA itself, we produced the following table (Figure 2):

	Sum_sq	df	F	PR(>F)
C(AUSPICE)	1.98	2.0	18.25	1.615e-08
Residual	57.518	1060.0	NaN	NaN

Figure 2: ANOVA table for Research Question 1

In this table, we see that the F-statistic is 18.25 and the p-value is <0.0001. This indicates that the group means do not seem the same, and that at least one is significantly different.

To determine which of the Auspices has a significant effect on the number spaces available, we conducted a post-hoc Tukey test, which produced the results below (Figure 3):

Group 1	Group 2	MeanDiff	P-Adj	Lower	Upper	Reject
Commercial	Non-Profit	0.077	0.0	0.0402	0.1139	True
Commercial	City-Operated	0.081	0.1011	-0.1737	0.0118	False
Non-Profit	City-Operated	-0.158	0.0001	-0.2479	-0.068	True

Figure 3: Tukey HSD test results for Research Question 1

As we see from the Reject column of the results table from Figure 3, the Tukey test indicates that the pairwise relationship between Commercial and Non-Profit Auspices do differ significantly with regard to number of childcare spaces available, and the same is true between Non-Profit and City-Operated Auspices. However, there is no significant difference in mean between Commercial and City-Operated Auspices.

Therefore, we can answer Research Question 1: there are indeed significant differences between amount of childcare space capacity between different types of childcare centres.

IV. Research Question #2 Two-Way ANOVA

For the second research question, this technical assignment considered the total number of spaces as the dependent variable, similarly to the first research question. Therefore, the Shapiro-Wilk and Brown-Forsythe tests run for Research Question 1 would be application for Research Question 2 as well, fulfilling both the assumptions of normally-distributed residuals and variance homogeneity with transformed values for the TOTSPACES variable.

For this research question, we are interested in determining whether there is a significant interaction between type of childcare centre (Auspice) and participation within the federal childcare subsidy program on capacity. The ANOVA table for this two-way ANOVA test is the following (Figure 4):

	Sum_sq	df	F	PR(>F)
C(AUSPICE)	1.63	2.0	15.76	7.67e-05
C(cwellcc_flag)	2.12	1.0	41.16	2.11e-10
C(AUSPICE):C(cwelcc_flag)	1.03	2.0	9.98	5.06e-05
Residual	54.59	1058.0	NaN	NaN

Figure 4: ANOVA table for Research Question 2

In this table, we see that the p-values associated with each variable are quite small – all well below 0.0001. Therefore, we reject the Null Hypotheses that the variables have no significant effect on each other, and conclude that the effects of Auspices and participation in the Canada-wide subsidy program are both significant with regard to total number of spaces available. On top of this, there is also a significant interaction between Auspice and participation in the subsidy program.

Therefore, we can also answer Research Question 2: There are indeed significant effects between type of childcare centre, participation in the subsidy program, and childcare capacity.

This technical assignment represents just the start of a multi-faceted and complex investigation into the largest opportunities for increasing childcare capacity across Toronto and the country more broadly. As time and the investigation continue, we endeavor to further approach the promise of affordable community-based childcare for every Canadian.

V. References

Jones, Allison. (2024, March 1). *Child-care wait lists balloon in many Ontario regions amid $10-a-day program*. CBC. https://www.cbc.ca/news/canada/toronto/child-care-wait-lists-ontario-1.7130656

Office of the Prime Minister of Canada. (2021, July 8). *Canada announced historic first early learning and child care agreement*. Office of the Prime Minister of Canada. https://www.pm.gc.ca/en/news/news-releases/2021/07/08/canada-announces-historic-first-early-learning-and-child-care#:~:text=British%20Columbia%20and%20Canada%20agree%20on%20the%20goal,6%20by%20the%20end%20of%20the%20five-year%20agreement

Office of the Prime Minister of Canada. (2022, March 28). *Announcing $10-a-day child care for families in Ontario*. Office of the Prime Minister of Canada. https://www.pm.gc.ca/en/videos/2022/03/28/announcing-10-day-child-care-families-ontario