

## Technical Assignment 2

Ying Du (Amelia)

Faculty of Information, University of Toronto

INF2178 LEC0101

Shion Guha

February 26, 2024

In this data analysis, my primary focus will be examining key indicators of Toronto's child care centre space. The columns I will concentrate on include: 'AUSPICE', 'TOTSPACE', and 'subsidy'.

## Data Examination

My initial step in this analytical process will be to meticulously filter and check any missing data from all categories, especially my focused key indicators.

Column Name	Null Count
AUSPICE	0
TOTSPACE	0
subsidy	0

Figure 1. Missing Data Summary for Key Indicators

Figure 1 shows that there is non-missing data for the three focused categories. Since the overall centre capacity numbers are the same as 'TOTSPACE', I will use 'TOTSPACE' as my continuous variable for this analysis.

## Research Questions

1. Is there a difference in total space availability across different operation auspices?
2. Does the availability of subsidies and the auspice agencies have a joint effect on the total number of spaces in child care centres?

## One-way ANOVA Result - Research Question 1

Source	Sum of Squares	df	F	PR(>F)
C(AUSPICE)	96,100	2	21.84	5.06e-10
Residual	2,332,000	1060	NaN	NaN

Figure 2. One-way ANOVA Result with Auspice

The results of a one-way ANOVA, as shown in Figure 2, revealed a significant difference in the total number of spaces available in child care centers based on the different agencies of auspice, as the p-value < 0.001 and the F-statistic = 21.84. The analysis indicated that different agencies operating the child care center impact the total spaces available.

## Two-way ANOVA Result - Research Question 2

Source	Sum of Squares	df	F	PR(>F)
<b>C(AUSPICE)</b>	8,568	2	2.06	0.13
<b>C(subsidy)</b>	83,500	1	40.12	3.53e-10
<b>C(AUSPICE):C(subsidy)</b>	56,000	2	13.46	1.69e-06
<b>Residual</b>	2,203,000	1058	NaN	NaN

Figure 3. Two-way ANOVA Result

Figure 3 above considers both the auspice type and subsidy availability. Firstly, the main effect of auspice type on total spaces is not shown (p-value > 0.05), which is against previous findings from the one-way ANOVA. I also noticed that the main effect of subsidy availability on total spaces is proven to be significant with p-value < 0.001 and F-statistic = 40.12. Thus, I conducted another one-way ANOVA test with the subsidy.

Source	Sum of Squares	df	F	PR(>F)
<b>C(subsidy)</b>	160,000	1	75.23	1.6e-17
<b>Residual</b>	2,300,000	1061	NaN	NaN

Figure 4. One-way ANOVA Result with Subsidy

Based on Figure 4, it shows that the availability of subsidies has a significant impact on the total number of spaces available in child care centers (p-value < 0.001, F-statistic = 75.23), and the F-statistic is much larger than the one for different types of auspice. Therefore, we can explain that the lack of effect of auspice type in the

two-way ANOVA test result might be due to the overpowering effect of subsidy availability, which makes it difficult to distinguish the effects of different types of auspice.

Moreover, the joint effect between auspice and subsidy availability on total spaces is significant with  $p < 0.001$  and F-statistic = 13.46. This suggests that the impact of different auspice agencies on available spaces varies depending on subsidy availability of the child care centre.

### Interaction Plot

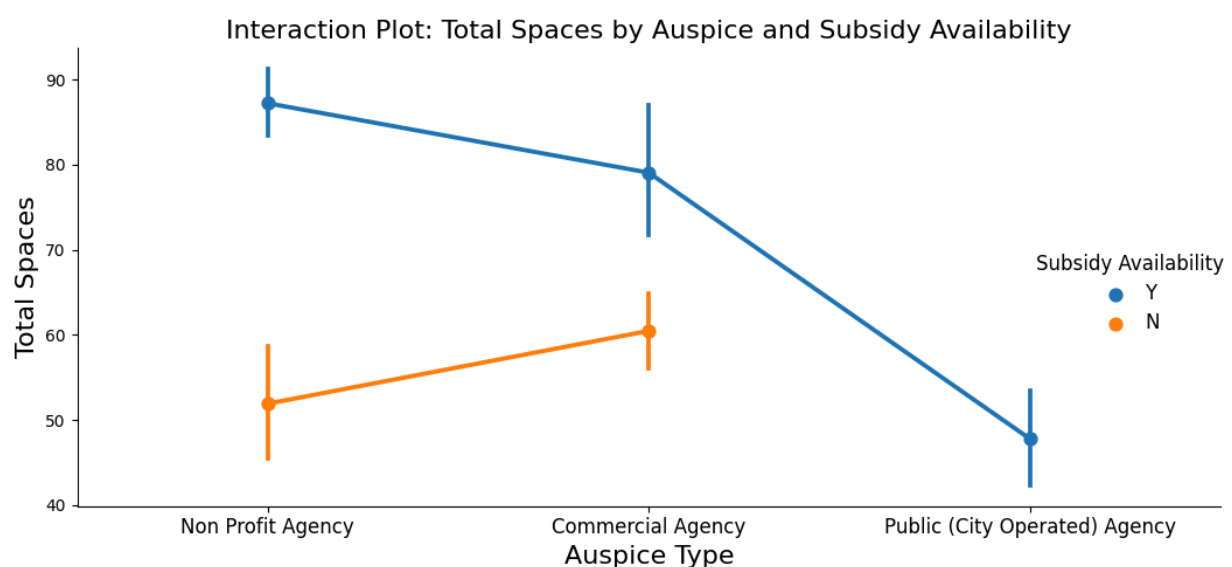


Figure 5. Interactive Plot of Total Spaces by Auspice and Subsidy Availability

Based on the interaction plot above, we can observe that the total spaces available differ across various auspice types. In general, child care centers with subsidies offer more space than those that do not for both non-profit and commercial agencies. The difference in space between non-profit agencies with and without subsidies is more significant than the difference in space between commercial agencies with and without subsidies. Thus, with subsidies, non-profit agencies have more space than commercial agencies, even though when without subsidies, commercial agencies have more. This indicates that the availability of subsidies has a varying and crucial impact on the total spaces in each type of center compared to others.

## Post-Hoc Test Result

Group 1	Group 2	Mean Difference	Adjusted P-Value	Lower Bound	Upper Bound	Reject Null Hypothesis
Commercial Agency	Non-Profit Agency	17.12	0.0	9.70	24.54	True
Commercial Agency	Public (City Operated) Agency	-17.22	0.078	-35.88	1.45	False
Non-Profit Agency	Public (City Operated) Agency	-34.33	0.0	-52.44	-16.22	True

Figure 6. Post-Hoc Test Result for Auspice

From Figure 6, we can tell that, on average, non-profit agencies have 17.12 more space available than commercial agencies, and public agencies have the least space capacity compared to the others.

Group 1	Group 2	Mean Difference	Adjusted P-Value	Lower Bound	Upper Bound	Reject Null Hypothesis
N	Y	26.27	0.0	20.32	32.21	True

Figure 7. Post-Hoc Test Result for Subsidy

From Figure 7, we can tell that, on average, agencies with subsidies have 26.27 more space available than those without.

Group 1	Group 2	Mean Difference	Adjusted P-Value	Lower Bound	Upper Bound	Reject Null Hypothesis
Commercial	Commercial	18.60	0.015	2.3733	34.8249	True

Agency: N	I Agency: Y					
Commercial Agency: N	Non Profit Agency: Y	26.76	<0.001	17.2862	36.243	True
Commercial Agency: Y	Non Profit Agency: N	-27.16	0.0008	-45.9161	-8.4082	True
Commercial Agency: Y	Public (City Operated) Agency: Y	-31.29	0.0045	-55.747	-6.8427	True
Non Profit Agency: N	Non Profit Agency: Y	35.33	<0.001	21.9757	48.6796	True
Non Profit Agency: Y	Public (City Operated) Agency: Y	-39.46	<0.001	-60.0631	-18.8577	True

Figure 8. Post-Hoc Test Result (Reject Null Only)

To understand how the availability of subsidies affects the total number of spaces across different auspices, I also conducted the post-hoc test to determine the interaction effect. From Figure 8 above, I found

- Commercial agencies with subsidies have an average of 18.6 more spaces than their non-subsidized counterparts, and non-profit agencies with subsidies have an average of 35.33 more spaces than their non-subsidized counterparts.
- Public agencies, which are inherently subsidized, generally have fewer spaces than both commercial (-31.29 spaces) and non-profit agencies (-39.46 spaces), suggesting that both commercial and non-profit agencies may have an advantage in capacity.

The differences in the number of spaces between commercial and non-profit agencies are significantly influenced by whether they are subsidized or not, which enhances the fact that subsidy status is a critical factor in determining capacity.

# ANOVA Assumptions Testing

## Normal Distribution Test

To test the normal distribution assumption, I conducted the Shapiro-Wilk test.

- Non Profit Agency: p-value =  $8.34e-22 < 0.05$ ,
- Commercial Agency: p-value =  $1.02e-11 < 0.05$ ,
- Public (City Operated) Agency: p-value =  $0.09 > 0.05$
  
- Y(with subsidies): p-value =  $1.16e-22 < 0.05$
- N(without subsidies): p-value =  $1.09e-13 < 0.05$

The Shapiro-Wilk test results indicate that the distribution of total spaces is not normally distributed for non-profit agencies, commercial agencies, agencies with and without subsidies. However, for public agencies, the p-value is above 0.05, suggesting that the distribution of total space is normal for this group.

## Homogeneity of Variances Test

- Auspice Levene Result: p-value =  $2.20e-08 < 0.05$
- Subsidy Levene Result: p-value =  $1.86e-06 < 0.05$

By utilizing Levene's tests, I found that both p-values are smaller than 0.05. These suggest that the variances of total space are not equal across different auspice agencies and subsidy availabilities, thus violating the ANOVA second assumption.

Thus, implementing ANOVA in this dataset may affect the accuracy of the analysis result.

## Data Reference

*Check normality of distribution using Shapiro-Wilk test in Python.* CodingFleet. (2024, January). <https://codingfleet.com/transformation-details/check-normality-of-distribution-using-shapiro-wilk-test-in-python/>

*Mann-Whitney U test using SPSS statistics.* Mann-Whitney U Test in SPSS Statistics | Setup, Procedure & Interpretation | Laerd Statistics. (n.d.). <https://statistics.laerd.com/spss-tutorials/mann-whitney-u-test-using-spss-statistics.php>

*Post hoc tests: Uncovering hidden patterns in two way anova.* FasterCapital. (2023, December 2). <https://fastercapital.com/content/Post-hoc-tests--Uncovering-Hidden-Patterns-in-Two-Way-ANOVA.html#Introduction-to-Post-hoc-Tests-in-Two-Way-ANOVA>