Jingxian Chen

1005354572

INF2178: Assignment2

Background and Introduction

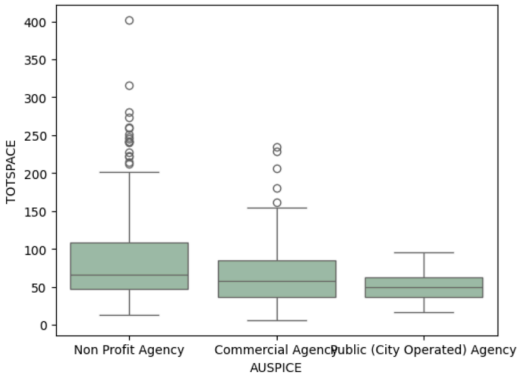
The accessibility of childcare in Ontario, particularly in Toronto, presents significant challenges for families. Responding to this need, the provincial government has committed to creating 100,000 new childcare spaces by the end of 2026—a move towards addressing this critical shortfall.

Research Questions

1 - Are there significant differences in the total space/capacity **(TOTSPACE)** of these institutions or centers under the influence of different management types **(AUSPICE)**?
2 - How does the type of operating agency **(AUSPICE)** and the receipt of government subsidies **(subsidy)** interact to influence the total available spaces **(TOTSPACE)** in childcare centers across Toronto?

Box Plots Observations

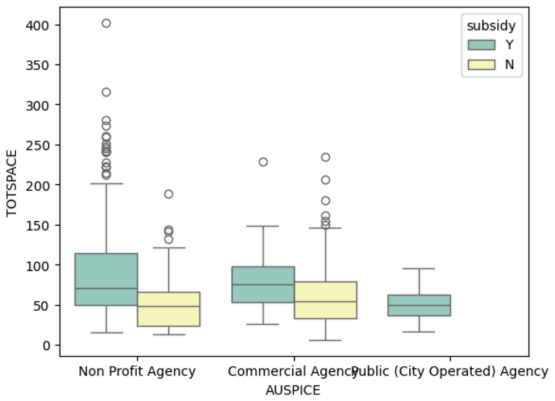
The boxplot visualization presents a comparative view of the total available spaces (**TOTSPACE**) in childcare centers across different types of operating agencies (**AUSPICE**): Non-Profit, Commercial, and Public (City Operated). Non Profit Agencies display wide variability in childcare spaces with many outliers, Commercial Agencies have moderate and more consistent spaces, and Public Agencies show the least variability, indicating uniformity in space allocation.



(Figure 1: Boxplot of Different Types of Operating Agencies by Total Available Childcare Spaces)

From the boxplot below, subsidized agencies tend to offer more TOTSPACE, with Non-Profit ones showing the greatest variability, especially among subsidized groups. Commercial agencies present less TOTSPACE variability, and Public (City

Operated) agencies show little difference between subsidized and unsubsidized groups, suggesting uniform space allocation. Outliers across categories indicate that while capacities typically conform to a range, some agencies significantly exceed this norm.



(Figure 2: Boxplot of Comparison of Childcare Space Capacity by Agency Type and Subsidy Status)

Results of testing the assumptions for running One-way ANOVA

Statistical Analysis

A one-way ANOVA was conducted to examine variances in **TOTSPACE** across different **AUSPICE** categories, followed by Tukey's HSD for post-hoc analysis.

One-Way ANOVA Results Interpretation

Given the very small p-value (far less than any standard significance level like 0.05 or 0.01), we can reject the null hypothesis that there is no effect of AUSPICE on TOTSPACE. This means that there is a statistically significant difference in TOTSPACE between the different AUSPICE groups.

Post-hoc Results

Since we find a significant effect of different types of operating agencies on the total space/capacity, we want to know which specific group means are different.

The post-hoc analysis reveals significant differences in the mean values between Non Profit Agencies and both Commercial Agencies and Public (City Operated) Agencies, but not between Commercial and Public (City Operated) Agencies. This suggests that the type of agency (Non Profit vs. Commercial vs. Public) has an impact on the variable of interest, with specific differences detailed by the mean differences.

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	16.806538	3.993722	29.619353	4.356853	0.006071
1	Non Profit Agency	Public (City Operated) Agency	36.177966	8.673910	63.682022	4.369046	0.005901
2	Commercial Agency	Public (City Operated) Agency	19.371429	−10.141900	48.884757	2.180132	0.272554

(Table 1: Post-hoc test among AUSPICE and TOTSPACE)

Assumption Analysis in One-way ANOVA

In the analysis of variance conducted to investigate the influence of different AUSPICE categories on TOTSPACE, the examination of residuals played a crucial role in validating the assumptions underlying ANOVA.

Assumption 1: Normality Assumption

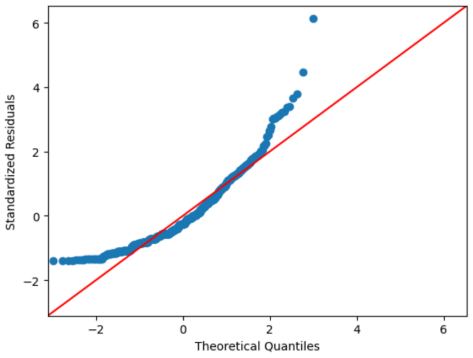
1- Normality of Residuals

A Q-Q plot (quantile-quantile plot) was generated to visually assess the normality of the standardized residuals obtained from the ANOVA. The plot revealed that while the residuals align with the theoretical quantiles in the middle of the distribution, there is a clear deviation in the tails. This suggests the presence of outliers or non-normality in the data.

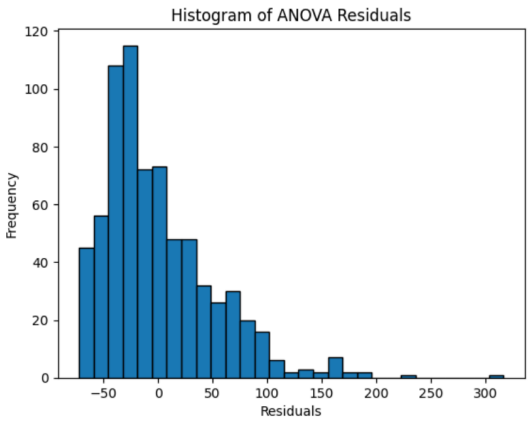
Additionally, a histogram of the ANOVA residuals was created. This further implies a departure from the normal distribution.

2 - Statistical Test for Normality

The Shapiro-Wilk test, a more formal method for testing normality, was applied to the residuals. The test yielded a W statistic of approximately 0.9018 and a p-value of less than 0.001. This highly significant p-value indicates a strong rejection of the null hypothesis of normality.



(Figure 3: QQ Plot of One-way ANOVA Assumption Test)



(Figure 4: Histogram of One-way ANOVA Residuals)

Assumption 2: Homogeneity Test Interpretation

A p-value of 0.0001 is well below the conventional alpha level of 0.05, indicating that there is a statistically significant difference in variances across the groups. This suggests that the assumption of equal variances is violated for the TOTSPACE data across the AUSPICE categories.

Implications and Insights

The residual analysis and the violation of homogeneity of variances suggest caution when interpreting the results of the one-way ANOVA, as this assumption underpins the validity of the F-test used in ANOVA. With this assumption not being met, the results of the ANOVA may not be reliable.

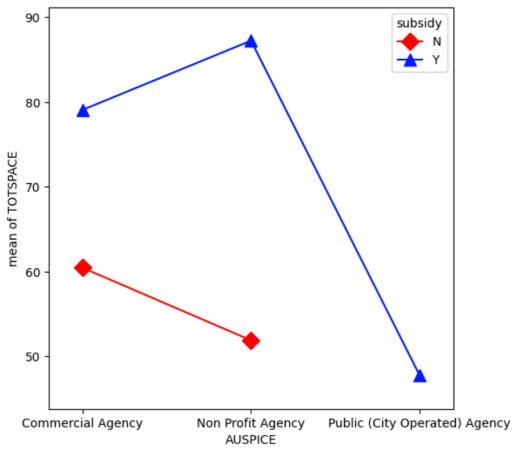
Results of testing the assumptions for running two-way ANOVA

Two-Way ANOVA Results Interpretation

While the type of operating agency alone did not significantly influence the capacity of childcare spaces, subsidy reception showed a strong independent effect, suggesting centers that receive subsidies have notably different capacities compared to those that do not. This implies that the strategy for allocating subsidies might be optimized considering the type of agency to maximize the increase in childcare spaces.

Interaction Plot

The interaction plot shows that subsidized Non Profit Agencies offer the most childcare spaces, while Public (City Operated) Agencies show little change in space availability regardless of subsidy status. Subsidies seem to significantly increase spaces in Non Profit Agencies but have a varied impact on Commercial and Public Agencies.



(Figure 5: Interaction Plot of Two-way ANOVA test)

Post-hoc test results

The analysis indicates a substantial influence of subsidy reception on the capacity of Non-Profit Agencies, highlighting the significant role of subsidies in these settings. Interestingly, the same effect was not observed within the Commercial Agency category, regardless of subsidy status. Moreover, a significant difference was noted between subsidized NonProfit and Public (City Operated) Agencies, emphasizing the variability in how subsidies may affect different agency types.

	group1	group2	Diff	Lower	Upper	q-value	p-value
0	(Non Profit Agency, Y)	(Non Profit Agency, N)	44.986164	25.584897	64.387430	9.370204	0.001000
1	(Non Profit Agency, Y)	(Commercial Agency, Y)	0.145138	−28.465689	28.755965	0.020500	0.900000
2	(Non Profit Agency, Y)	(Commercial Agency, N)	28.370552	11.192465	45.548638	6.674107	0.001000
3	(Non Profit Agency, Y)	(Public (City Operated) Agency, Y)	40.752830	8.307960	73.197700	5.075887	0.004757
4	(Non Profit Agency, Y)	(Public (City Operated) Agency, N)	0.000000	−inf	inf	0.000000	0.900000

(Table2: Post-hoc test results of Two-way ANOVA)

Assumption Analysis in Two-way ANOVA

Assumption 1: Normality Assumption

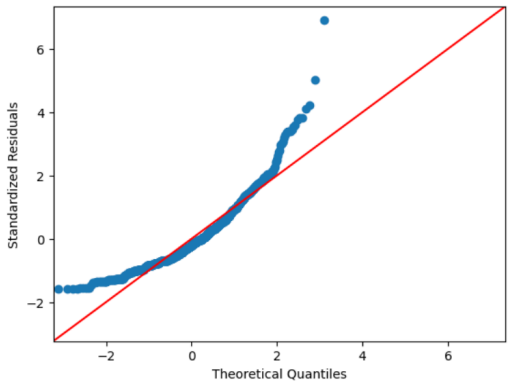
1 - Normality of Residuals

In the Q-Q plot, while the majority of points seem to follow the line, there is a noticeable deviation at both ends. This pattern indicates some departure from normality, particularly with some potential outliers or heavy tails, as evidenced by the points that stray from the line in the upper quantile range.

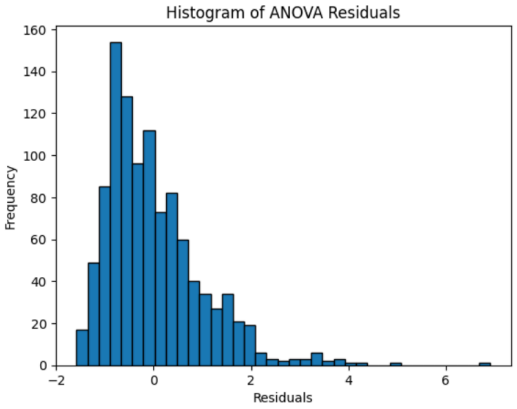
Also, the shape of the histogram seems to approximate a normal distribution, but there appears to be a slight right skew, as indicated by a longer tail to the right.

2 - Statistical Test for Normality

In The Shapiro-Wilk test's results, the p-value is extremely low (less than 0.05), which leads to the rejection of the null hypothesis, suggesting that the residuals do not follow a normal distribution.



(Figure 6: QQ Plot of Two-way ANOVA)



(Figure 7: Histogram of Two-way ANOVA Residuals)

Assumption 2: Homogeneity Test Interpretation

In this case, since the p-value is much lower than 0.05, we would reject the null hypothesis of Levene's test, indicating that there is statistically significant evidence that the variance is not the same across the groups.

Conclusion

This report acknowledges that the use of ANOVA on the dataset pertaining to childcare centers in Toronto, particularly the count variable TOTSPACE, may not have been fully appropriate. TOTSPACE, not being a continuous variable, led to a violation of the normal distribution assumption critical to ANOVA.

Despite a comprehensive application of both one-way and two-way ANOVAs, and Tukey's HSD post-hoc tests, the large sample size could invoke the Central Limit Theorem to justify the use of ANOVA by approximating a normal distribution of sample means. However, the observed non-normality in the residuals and the Levene's test results call for a cautious interpretation of the findings.

Future analyses should consider alternative non-parametric methods that do not assume normality to provide more reliable insights into the effects of AUSPICE and subsidy on the availability of childcare spaces.