

Exploring Spaces of Child Care Center in Ontario

1. Introduction

In Ontario, it could be difficult to find a proper licensed or unlicensed child care center due to high fees and low space availability. Moreover, the government has found that most families in Ontario could not afford the price of child care. Hence, the government has pledged 100,000 new child care spaces from 2016 to 2026. Based on the dataset of *INF_2178_A1.xlsx*, this research has conducted a quantitative analysis using both one-way and two-way ANOVA, and then conducted the corresponding post-hoc analysis and some interaction plots. Based on these analysis, we hope to find which variables are affecting the total space in the child care of Toronto. By doing this, the government may have a better arrangement on assigning those new child care spaces in the future.

Since the method of this research relied on the use of ANOVA, it is necessary to locate those categorical variables in the dataset first. This dataset has included following categorical variables:

1. *LOC_ID* & *LOC_NAME*: Represent the unique identification and name for care center
2. *AUSPICE*: Operating AUSPICE, including commercial, nonprofit and public
3. *Ward*: City ward number
4. *Bldg_type* & *BLDGNAME*: Building type and their name

Moreover, there are some binary variables like *subsidy* in the dataset representing if the center has a free subsidy contract or not. The numeric value, which will be the dependent variable in our ANAVO analysis, is all about the space availability of different ages while this research will focus on the total space availability. In this research, the one-way anova will focus on investigating if there is a statistically significant difference among means of different AUSPICE. And then the two way ANOVA will involve *subsidy* based on the one way ANOVA to see the effects of *subsidy* and *AUSPICE* on the total space availability and if there is an interaction effect between these two variables.

One-way ANOVA

Null Hypothesis: The total space mean of different groups in AUSPICE is same

After conducting the one-way anova, we have got the following table:

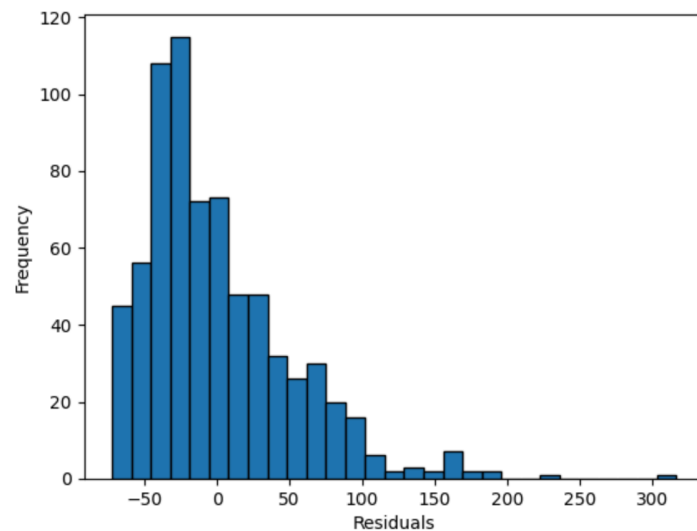
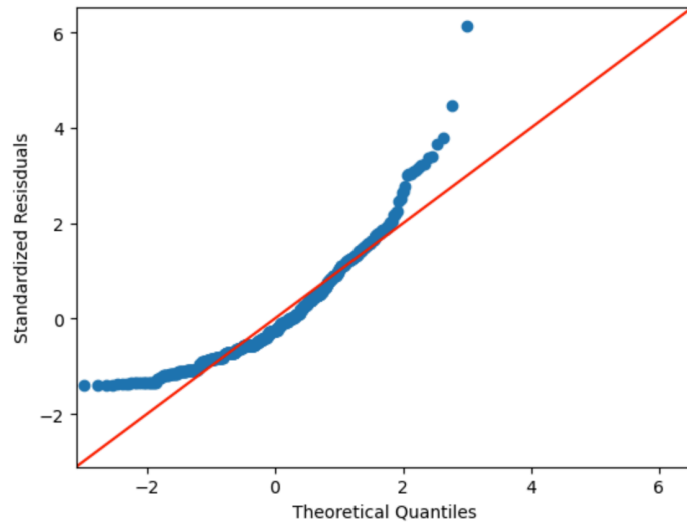
index	sum_sq	df	F	PR(>F)
AUSPICE	96112.11429024815	2.0	21.843051025265375	5.057716322707518e-10
Residual	2332065.2648254596	1060.0	NaN	NaN

From the above table, it can easily be seen that a F-statistics of 21.843051 for AUSPICE indicates that there is a significant amount of variance between the different AUSPICE groups when compared to the variance within the groups. On the other hand, the p-value of 5.057716e-10 for AUSPICE is extremely small, indicating that there is a statistically significant difference in the mean of total space across different AUSPICE categories.

In conclusion, the ANOVA analysis has rejected the null hypothesis, suggesting that not all groups means are equal. Given the F-statistic and p-values, we can conclude that there are meaningful differences in child care center capacities among different management types. Since the ANOVA has found significant differences, some follow-up analysis like post-hoc test by using Tukey's HSD should be performed to determine which specific pairs of AUSPICE categories differ from each other in terms of total space availability. The following is the result of doing Tukey's HSD test:

group1	group2	Diff	Lower	Upper	q-value	p-value
Non Profit Agency	Commercial Agency	16.80653753026634	3.993721765112518	29.61935329542016	4.35685314786073	0.006071389357186763
Non Profit Agency	Public (City Operated) Agency	36.177966101694906	8.673909797025185	63.682022406364624	4.369046357484494	0.005901208365050925
Commercial Agency	Public (City Operated) Agency	19.371428571428567	-10.141900057062916	48.884757199920045	2.180131598320647	0.27255408638555445

From the above table, we can see that groups of nonprofit agencies and commercial agencies and groups of nonprofit agencies and public agencies all have a low value, which suggests a statistically significant difference between each of them. The p-value of 0.27 suggests that there is no such a difference in the number of spaces available between commercial and public agency, which may imply that these two types of management might operate with similar constraints or capabilities regarding available spaces. However, when we compare the first two sets of groups, we may find that nonprofit agencies tend to have significantly more spaces available than commercial and public agencies, which may imply that nonprofit agencies might have different operational capacities or priorities that allow for more child care spaces. In conclusion, the overall ANOVA analysis did give us an insight that nonprofit agency's operational management may lead to more available spaces for child care. However, to examine the effectiveness of this insight, the last step of conducting a diagnostic check to validate the reliability of the ANOVA results.



The first graph created here is the Quantile-Quantile plot to see if empirical distribution of the residuals closely matches the theoretical distribution while the second graph is a histogram of the distribution of residuals. Theoretically, we want to see if the dot in the QQ plot can fit that 45-degree angle line and the histogram of residuals display a normal distribution. However, we can see the distribution of residuals can not fit the theoretical line very well and it doesn't have a normal distribution. Hence, it gave the conclusion that normality assumption of the residuals isn't satisfied. To prove it in further, I kept doing the Shapiro-Wilk test for normality on the residuals of a model, and get the results of **0.902** as test statistic and an extremely small p-value, which strongly suggests that the residuals of the model do not follow a normal distribution. Lastly, I have conducted a Levene's test to check if the assumption of homogeneity of variances is satisfied. As a result, I have got **9.1994** as the test result and still a very small p-value.

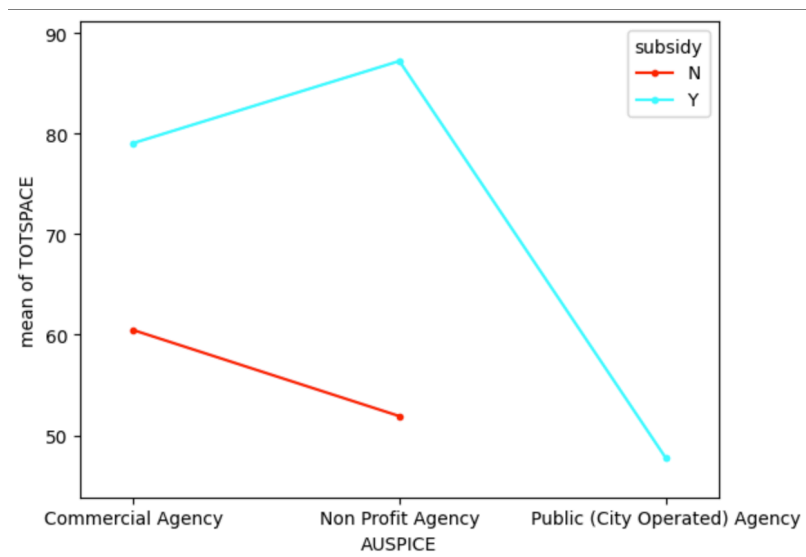
Combining two tests together, the conclusion is that both assumptions of normality and homogeneity of variances are not satisfied, and the ANOVA results can indeed be affected. For future implementation of this ANOVA, some potential solutions may include doing Kruskal-Wallis test which does not assume normal distribution of residuals and Welch's ANOVA which is an adaptation of ANOVA that does not require the assumption of equal variances.

Two-way ANOVA

Similar to the step of one-way ANOVA, I have still created an two-way ANOVA table to examine how selected two factors (*AUSPICE* and *subsidy*) independently and jointly affect the total space availability.

index	sum_sq	df	F	PR(>F)
C(AUSPICE)	8567.996108745856	2.0	2.0575860838979505	0.1282729503898439
C(subsidy)	83527.4417869589	1.0	40.117875793386546	3.529094165783121e-10
C(AUSPICE):C(subsidy)	56034.454073261026	2.0	13.456555238426358	1.6942822007381886e-06
Residual	2202809.3876588224	1058.0	NaN	NaN

From the above table, we can get the conclusion that the independent variable of *AUSPICE* alone did not have an important effect on the total space availability while the *subsidy* and the interaction between these two factors both have a statistically important effect on the total space availability. Based on this result, I stepped forwards to create an interaction plot using the statsmodel library to visualize the interaction between two factors, which are useful for identifying whether the effect of one factor on the response of total space differs across the levels of another.



First of all, we can find the lines in the graph are not parallel, which indicates an interaction effect between *AUSPICE* and *subsidy* on the total spaces available. If the lines were parallel, it would suggest no interaction effect which means that the effect of receiving subsidies on total spaces does not depend on the type of agency. For commercial agencies and nonprofit agencies, those that receive subsidies tend to have more spaces available than those that do not. However, for public agencies, the trend reverses that those without subsidies tend to have more spaces than those with subsidies. In conclusion, subsidies appear to have a smaller effect on total spaces available in commercial agencies compared to nonprofit and public agencies while nonprofit agencies show the highest mean total spaces when subsidized and a significant drop in total spaces when not subsidies. Lastly, public agencies show an opposite trend that more total spaces are available when not receiving subsidies.

Similar to the one-way ANOVA, the same post-hoc analysis is still done through the Tukey's Honestly Significant (HSD) tests on the results of the two-way ANOVA analysis.

index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency	Commercial Agency	16.80653753026634	4.406270182009921	29.20680487852276	4.5018295685724405	0.004332721615780799
1	Non Profit Agency	Public (City Operated) Agency	36.177966101694906	9.559488353271004	62.796443850118806	4.514428513213716	0.004204720483586888
2	Commercial Agency	Public (City Operated) Agency	19.371428571428567	-9.191626734927869	47.934483877785	2.2526765441975174	0.24965039587209503

index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Y	N	34.121178057553955	24.87300623779995	43.36934987730796	10.244061678532105	0.001

index	group1	group2	Diff	Lower	Upper	q-value	p-value
0	Non Profit Agency,Y	Non Profit Agency,N	44.986163522012575	25.58489740798289	64.38742963604226	9.370204265045107	0.001
1	Non Profit Agency,Y	Commercial Agency,Y	0.1451378809869368	-28.465689282209816	28.75596504418369	0.020499843129401694	0.9
2	Non Profit Agency,Y	Commercial Agency,N	28.370551707666586	11.192464915807506	45.548638499525666	6.674106955701865	0.001
3	Non Profit Agency,Y	Public (City Operated) Agency,Y	40.75283018867924	8.307960147717104	73.19770022964138	5.075886581186364	0.004757438918105983
4	Non Profit Agency,Y	Public (City Operated) Agency,N	0.0	-Infinity	Infinity	0.0	0.9
5	Non Profit Agency,N	Commercial Agency,Y	44.84102564102564	11.398077665668467	78.28397361638281	5.418401187414243	0.0019205648842343104
6	Non Profit Agency,N	Commercial Agency,N	16.61561181434599	-7.775739196796728	41.006962825488706	2.752838000448803	0.3757287266038358
7	Non Profit Agency,N	Public (City Operated) Agency,Y	4.233333333333334	-32.54330232431011	41.00996899097678	0.4651688691416603	0.9
8	Non Profit Agency,N	Public (City Operated) Agency,N	0.0	-Infinity	Infinity	0.0	0.9
9	Commercial Agency,Y	Commercial Agency,N	28.22541382667965	-3.9787521815201856	60.42957983487948	3.541835893080001	0.12440332903345075
10	Commercial Agency,Y	Public (City Operated) Agency,Y	40.607692307692304	-1.7561039625898545	82.97148857797447	3.87358965507854	0.06919877437104338
11	Commercial Agency,Y	Public (City Operated) Agency,N	0.0	-Infinity	Infinity	0.0	0.9
12	Commercial Agency,N	Public (City Operated) Agency,Y	12.382278481012655	-23.27159184601213	48.03614880803744	1.4034406774663803	0.9
13	Commercial Agency,N	Public (City Operated) Agency,N	0.0	-Infinity	Infinity	0.0	0.9
14	Public (City Operated) Agency,Y	Public (City Operated) Agency,N	0.0	-Infinity	Infinity	0.0	0.9

From the result of the above 3 tables, we can find that nonprofit agencies significantly differ in their mean scores compared to both commercial and public agencies, with nonprofit agencies having higher scores. Also there is no statistically significant difference between commercial agencies and public agencies. For the post-hoc analysis on the aspect of *subsidy*, the result indicates a statistically significant difference between centers with subsidies and without subsidies on the total space available. Centers with subsidy will have a notably higher mean value compared to those centers without subsidy. The confidence interval provides an estimate of

the range within which the true mean difference likely falls, and the very low p-value suggests that this finding is unlikely to be due to random chance. Lastly, significant differences were found within nonprofit agencies based on the *subsidy* factor, and between nonprofit agencies with subsidy and both commercial agencies without subsidy and public agencies with subsidy. There are also significant differences between all nonprofit agencies and commercial agencies with subsidy, suggesting that the type of agency and subsidy received interact in complex ways to affect the total space available. There are many comparisons involving public agencies without subsidies that show no significant difference, which could indicate that this group does not significantly differ from others based on the total space available.

Conclusion

This research focused on investigating the relationship between different child care center management types and the total space available in Ontario by using the quantitative analysis of ANOVA. Moreover, it kept investigating more interactive relationships between management types and subsidies received by using two-way ANOVA. As a result, it did find some statistical significant relationships among those factors. However, some ANOVA results may not be reliable because the dataset distribution may not satisfy some assumptions of ANOVA. In the future stage, this research should consider some new quantitative analysis such as Welch's ANOVA which doesn't require those assumptions. Besides that, this research should at least give the government some insights on how to make arrangements for child care centers in the future.