One/Two-Way ANOVA for Toronto Children's Day Care

Xiaoxin Zhou 1003549742

1. Introduction

In Toronto, the high fees and low availability of spaces are critical issues for children's daycare. Most childcare workers can expect to pay up to $1,774 every month [1] to send a child to a licensed childcare center. Ontario does not have enough licensed childcare centers to meet the current demand. As a result, thousands of new parents are forced to send their children to unlicensed child care centers to reduce costs, despite the significantly lower quality of childcare services, because they need to work. However, according to a survey by Toronto Children's Services, 75% of families still cannot afford childcare. Without parents to take care of their babies, it becomes an essential problem for child education and safety. As a result, the provincial government pledged 100,000 new child care spaces for 2016 to 2026 [2]. Because of the need to control budgets, the government needs to know exactly how many families currently need this benefit. We mainly focusing on those questions:

One-Way ANOVA:
- Does the average daily capacity of childcare centers differ by age group (infants, toddlers, preschoolers)?

Two-Way ANOVA:
- Effects of the City of Toronto's Child Care Licensing System and subsidy availability on the average daily capacity of childcare centers

2. Method

The dataset comprises 1063 rows and 17 columns. Upon review, we identified some features that could be removed as they are not necessary for the scope of our analysis.

2.1 Dataset features

LOC_ID: Unique identifier for the location
LOC_NAME: Name of the child care centre
AUSPICE: Operating auspice (Commercial, Non-Profit, or Public)
ADDRESS: Address of the child care centre
PCODE: Postal code
ward: Ward number
bldg_type: Building type
BLDGNAME: Building name
IGSPACE: Infant spaces
TGSPACE: Toddler spaces
PGSPACE: Preschool spaces
KGSPACE: Kindergarten spaces
SGSPACE: School-age spaces
TOTSPACE: Total spaces
subsidy: Indicates if subsidy is available (Y for yes)

cwelcc_flag: Indicates if the center is part of the City of Toronto's Child Care Licensing System (Y for yes)

After reviewing all the data and features, I identified columns such as '_id', 'LOC_ID', 'ADDRESS', 'PCODE', and 'BLDGNAME', which consist of identifiers and addresses, as not useful for ANOVA analysis. Additionally, I will remove columns with missing values.

2.2 The features we keep:
AUSPICE: This categorical variable indicates the operating auspice and has three levels, making it suitable for a one-way ANOVA or as one of the factors in a two-way ANOVA.
ward: As a categorical variable representing different wards, it could be used to compare groups based on geographic location.
bldg_type: This categorical variable represents the type of building in which the childcare centre operates and could influence the number of spaces available.
IGSPACE, TGSPACE, PGSPACE, KGSPACE, SGSPACE, TOTSPACE: These are all quantitative variables representing the number of spaces for different age groups and the total number of spaces available, which are suitable for comparison in ANOVA tests.
subsidy: Indicates whether a subsidy is available or not and is a binary categorical variable.
cwelcc_flag: Indicates if the centre is part of the City's Childcare Growth Strategy and is also a binary categorical variable.

2.3 Table of Describe and Visualization Data

	ward	IGSPACE	TGSPACE	PGSPACE	KGSPACE	SGSPACE	TOTSPACE
count	1063.00	1063.00	1063.00	1063.00	1063.00	1063.00	1063.00
mean	12.511	3.896	11.600	24.258	14.257	21.661	75.674
std	7.03	6.09	12.08	18.577	20.493	30.423	47.816
min	1	0	0	0	0	0	6
25%	6	0	0	0	0	0	43
50%	12	0	10	24	0	0	62
75%	19	10	15	32	26	30	97
max	25	30	90	144	130	285	402

Table I The summary of ward IGSPACE, TGSPACE, PGSPACE, KGSPACE, SGSPACE, TOTSPACE, includes number of counts, mean, std, min, max, and 25%, 50% 75%. The

interesting information from this table, IGSPACE, KGSPACE, and SGSPACE at 50% are ZERO, which means that more than half of the centres do not have these types of spaces available.

Visualization of the distribution of the quantitative variables using histograms
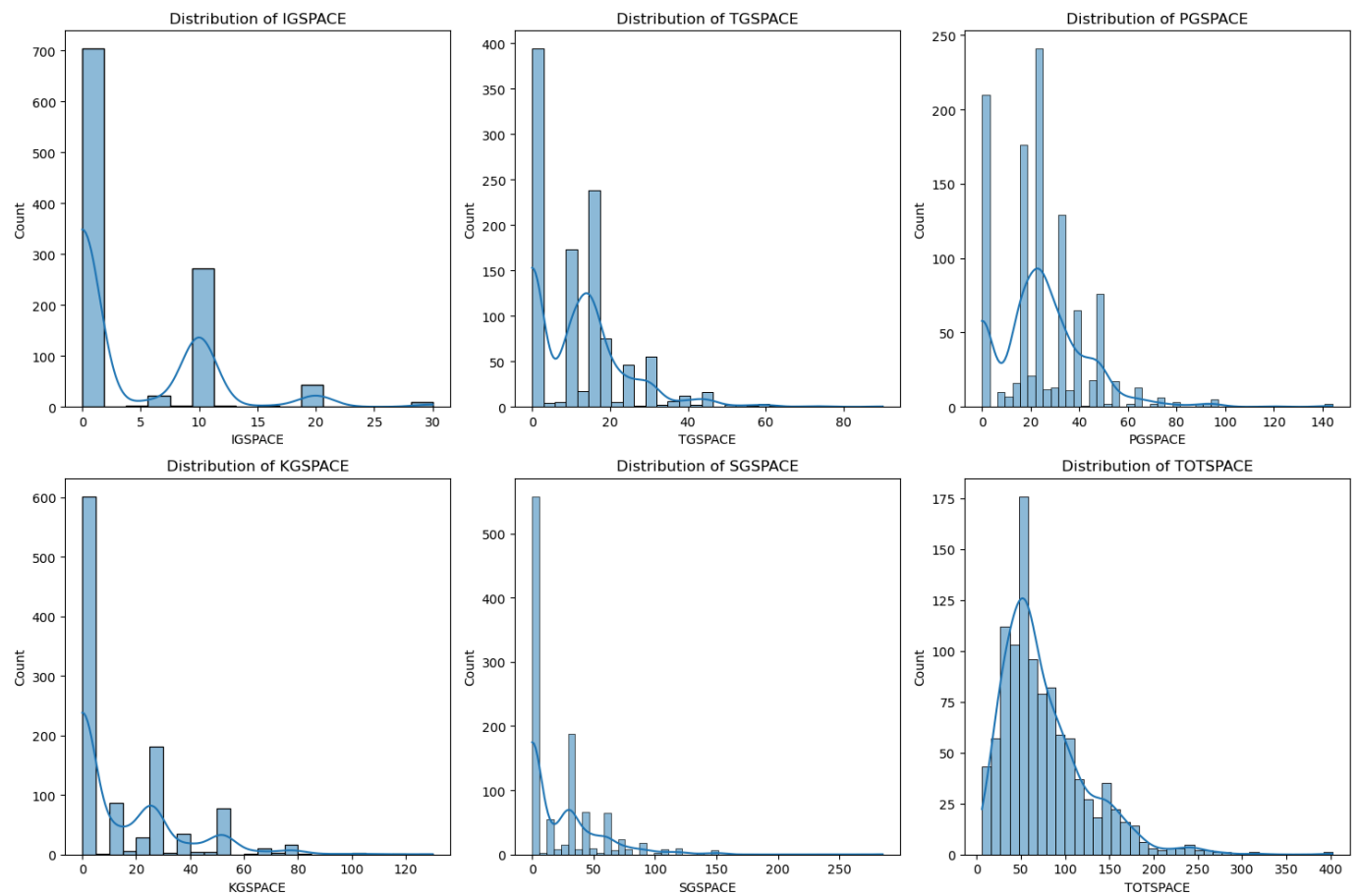

Fig 1 Visualization of the SPACE, TGSPACE, PGSPACE, KGSPACE, SGSPACE, TOTSPACE, and all the distributions are right-skewed and not normality assumption of ANOVA.

2.4 Remove outlier
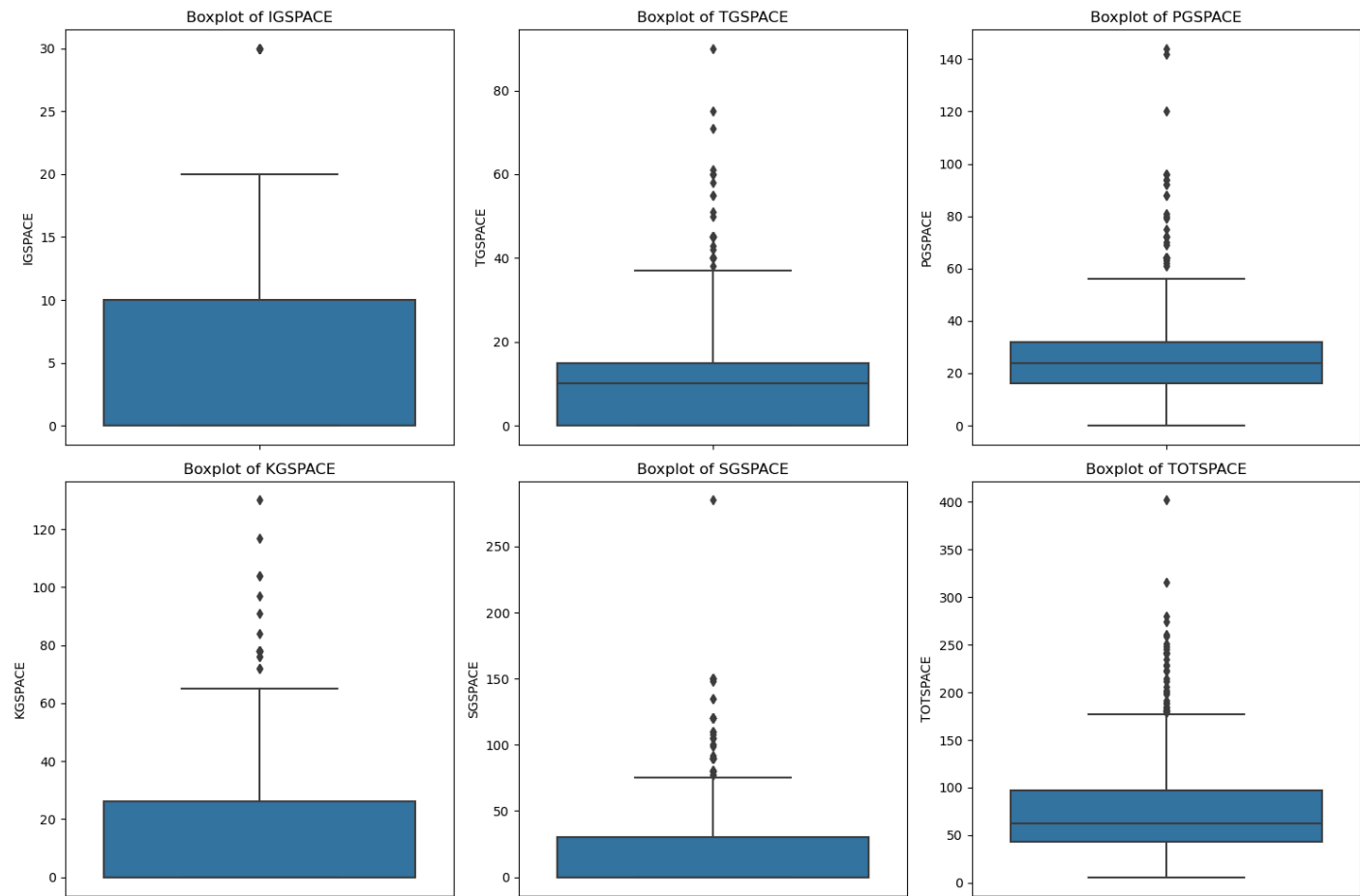Visualization of the outlier of quantitative variables using boxplots

Fig 2 To visualize outliers in quantitative variables, we can use a boxplot. This method easily identifies outliers in all quantitative variables, which we will remove before conducting further analysis. This step is crucial because outliers can significantly affect the mean and variance of the groups.

2.5 One-Way ANOVA

2.5.1 Does the average daily capacity of childcare centers differ by age group (infants, toddlers, preschoolers)?
Null Hypothesis (H0): There is no difference in the total number of child care spaces among different wards.
Alternative Hypothesis (H1): There is a difference in the total number of child care spaces among different wards.

In order to process the One-Way ANOVA I group TOTSPACE by ward and apply the One-Way ANOVA from the stats library.

ANOVA result for TOTSPACE among wards: ANOVA result for different age IGSPACE TGSPACE PGSPACE KGSPACE SGSPAC: **F_onewayResult(statistic=218.9536487483347, pvalue=1.8354486402613428e-172)**

Through the examination of childcare center data, we have calculated the average capacities for different age groups, revealing variability in the distribution of spaces allocated to infants, toddlers, preschoolers, kindergarteners, and school-age children. The subsequent statistical analysis, a one-way ANOVA, produced a significant result $p < 0.05$, leading to the rejection of the null hypothesis, which posited no difference in the total number of child care spaces among different age groups.

This analysis confirms that the average number of childcare spaces varies significantly by age group, supporting the alternative hypothesis. The differences in average capacities underscore the distinct space allocation preferences or needs across age groups within childcare centers. When planning to open new daycare facilities, we need to pay attention to the demand across different age groups.

2.6 Two-Way ANOVA

2.5.1 Effects of the City of Toronto's Child Care Licensing System and subsidy availability on the average daily capacity of childcare centers.

	sum_sq	df	F	PR(>F)
C(licensed)	9.312739e+03	1.0	8.234504	4.203142e-03
C(subsidy_available)	7.198723e+04	1.0	63.652500	4.342758e-15
C(licensed)&C(subsidy_available)	8.161992e+02	1.0	0.721699	3.958055e-01
Residual	1.052906e+06	931.0	NaN	NaN

Table II Two-Way ANOVA for how licensing and subsidy availability effects the total daily capacity of childcare centers

Main Effects
Licensing
Null Hypothesis (H0): The average daily capacity of childcare centers does not differ based on whether they are licensed by the City of Toronto's Child Care Licensing System.
Alternative Hypothesis (H1): The average daily capacity of childcare centers differs based on whether they are licensed by the City of Toronto's Child Care Licensing System.

The p-value is approximately 4.203142e-03, which is less than the alpha level of 0.05. This indicates a statistically significant main effect of licensing status on the average daily capacity of childcare centers. Thus, we reject the null hypothesis for licensing and accept the alternative

hypothesis that the average daily capacity differs based on whether childcare centers are licensed by the City of Toronto's Child Care Licensing System.

Subsidy Availability
Null Hypothesis (H0): The average daily capacity of childcare centers does not differ based on the availability of subsidies.
Alternative Hypothesis (H1): The average daily capacity of childcare centers differs based on the availability of subsidies.

With a p-value of about 4.342758e-15, this result is highly significant, indicating a significant main effect of subsidy availability on the average daily capacity. Therefore, we reject the null hypothesis for subsidy availability and accept the alternative hypothesis, suggesting a significant difference in average daily capacities based on the availability of subsidies.

Interaction Effect
Licensing and Subsidy Availability Interaction
Null Hypothesis (H0): There is no interaction effect between the childcare center's licensing status and the availability of subsidies on the average daily capacity.
Alternative Hypothesis (H1): There is an interaction effect between the childcare center's licensing status and the availability of subsidies on the average daily capacity.

The p-value is approximately 0.396, which is above the commonly used significance level of 0.05. This indicates that there is no statistically significant interaction effect between the childcare center's licensing status and the availability of subsidies on the average daily capacity. Hence, we fail to reject the null hypothesis for the interaction effect, suggesting that the effect of licensing status on daily capacity does not vary depending on the availability of subsidies (and vice versa).

There is a statistically significant difference in the average daily capacity of childcare centers based on whether they are licensed and whether subsidies are available, with both factors independently affecting capacities. However, there is no statistically significant interaction between licensing status and subsidy availability. This indicates that the impact of one factor on average daily capacity is not dependent on the level of the other factor.

3. Reference:
[1] Ywca Niagara Region(February 10 2022) ,Affordable childcare: what happens when mom or dad can't afford childcare? ,
https://www.ywcaniagararegion.ca/blog/affordable-childcare

[2] CD19.2 Appendix A (2016),TORONTO'S LICENSED CHILD CARE GROWTH STRATEGY For children under 4 2017-2026,
9791-Torontos-Licensed-Child-Care-Growth-Strategy.pdf