Analysis of Annual Average Spending on Dining at the RCYC's restaurants in the city of Toronto (mainland) for 2017

Haojun Qiu, Jasmine Zhuang, Junru Lin, Xiaoqi Liu – Group 44

March 19, 2021

Introduction

- Main Goal – We want to help RCYC's managers better understand the **spending behavior of different demographic groups** of RCYC members at the RCYC's restaurants in the city of Toronto (city dining) for 2017 (we will approach it by investigating three research questions).
- Target Population – We will target the members who spent money dining at the RCYC's restaurants in the city of Toronto (mainland) for 2017.
- The dataset we will use is provided by RCYC, which is a random sample of 1,000 RCYC members.

Data Summary and Wrangling

Data Summary

- The variable we will use are sex, year (for which the age, and spending variables are measured), age, and city dining (yearly amount spent on dining at the RCYC's restaurants in the city of Toronto (mainland), which is only available for 2017).

Data Wrangling

- Question 1 – We take the original data frame given, filter, and store the observations spending less than $2000 in 2017.
- Question 2 – We take the data frame from the first question, filter and store the observations without missing value for the variable sex.
- Question3 – We first filter and store the observations without missing value for the variable Age. Then, we create a new categorical variable - age group, where different values correspond to different age groups (e.g. '20-40', '41-60', etc.), and store this dataframe for later use.

Research Question 1

Question: What is a **range of plausible values** of the mean yearly spending on dining for the RCYC members at the RCYC's restaurants in the city of Toronto (mainland) for 2017?

- The answer to this question can help us gain insight into the average spending level of the members at the RCYC's restaurants in the city of Toronto (mainland) for 2017.
- Then, we can predict future sales, plan for future pricing as well as prepare for the plan to control future expenditure cost.

Data Visualization for Q1

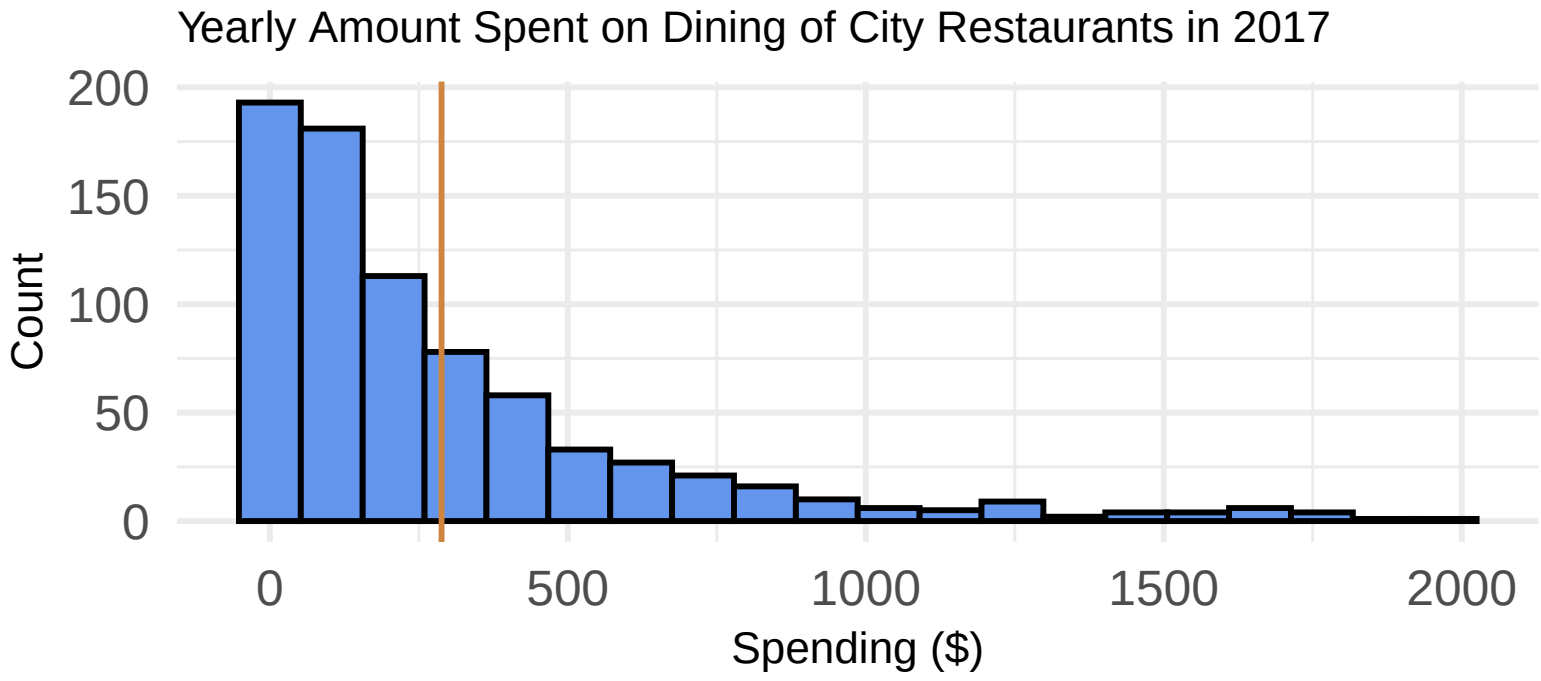

Figure 1

- The distribution of the mean value of the yearly amount spent shows that a majority of members' yearly spendings are less than $500 (right-skewed).
- The vertical line drawn represents the mean value of spending at RCYC's dining restaurant **from the sample**, which is about $288.

Statistical Method for Q1

- The Bootstrap method allows us to estimate the range of plausible values of average yearly dining spending by resampling from the original sample:
- Firstly, we randomly draw a sample of the size of the original sample with replacement (i.e. put back after drawing out) from the original sample (the data we have in hand).
- Next, we calculate the mean value of city dining spending from the sample drawn.
- Then, we repeat this process 2000 times to get 2000 samples and the corresponding statistics (sample means), so we have the distribution of these statistics.
- Finally, we are 95% confident that the true average yearly amount spent of all RCYC members at RCYC's city dining restaurant is within the middle 95% range (which is also called the 95% confidence interval) of our sample means calculated from these 2000 samplings.

Result for Q1

- The 95% confidence interval is \$263~\$312. In other words, we are 95% confident that the interval \$263 to \$312 captures the true mean city dining spending.
- The manager can use this result, combining with the change in price level and the number of members, to predict the future total revenue gained by their city dining restaurants and plan for future possible input to maximize profit.

Research Question 2

Question: Is the mean value of the yearly amount spent on dining at the RCYC's restaurants **different between Female and Male members** in the city of Toronto (mainland) for 2017?

- This can help determine if different marketing strategies are required to be designed for specific gender groups (e.g. ladies special menu).

Data Visualization for Q2

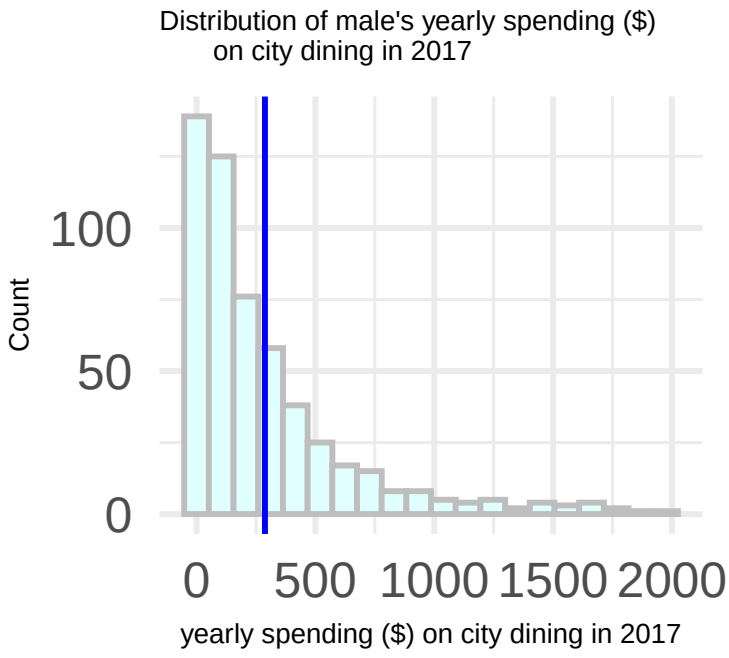

Figure 2

Figure 3

- The blue and red vertical lines in the graphs above represent the mean values of males' and females' yearly city dining spending in 2017 from the sample given (test statistics), which are $288 and $289, respectively.
- The distributions of these two groups' spendings are similar to each other, and they both lined with the overall distribution (see Figure 1).

Statistical Method for Q2

- The two-group hypothesis test helps test if there is a difference in mean yearly spending on city dining between females and males in 2017.

The two-group hypothesis test:

- **Null hypothesis:** There is no difference in mean yearly spending on city dining between females and males in 2017.
- **Alternative hypothesis:** There is a difference in the mean yearly spending on city dining between females and males in 2017.
- Using the test statistic ($289 - $288 = $1), we stimulate the samples under the assumption that the null hypothesis is true and calculate the statistic for each sample. Then, we obtain a p-value to see the strength of the evidence against our null hypothesis.

Result for Q2

- The p-value calculated is 0.965.
- There is no evidence against the null hypothesis, which means that there is no significant difference between the mean values of female and male members' yearly spending at the RCYC's restaurants in the city of Toronto in 2017.
- Therefore, advertisers don't have to target differently between men and women. Marketing for both groups is equally important, and marketing funds should be evenly distributed.

Research Question 3

Question: What's the relationship between **the value of the mean** yearly spending on dining at the RCYC restaurants in the city of Toronto (mainland) for 2017 and **age groups**?

- Noticing that members of different age groups have different levels of consumption (see Figure 4 on the next page), we want to explore the relation between the age groups and the mean yearly spending on dining in the city restaurants.
- The answer to this question helps understand the contribution of different age groups to restaurant turnover and can be used as a reference to develop some marketing plans targeting different age groups.

Data Visualization for Q3

Figure 4: Yearly Amount Spent on Dining of City Restaurants in 2017

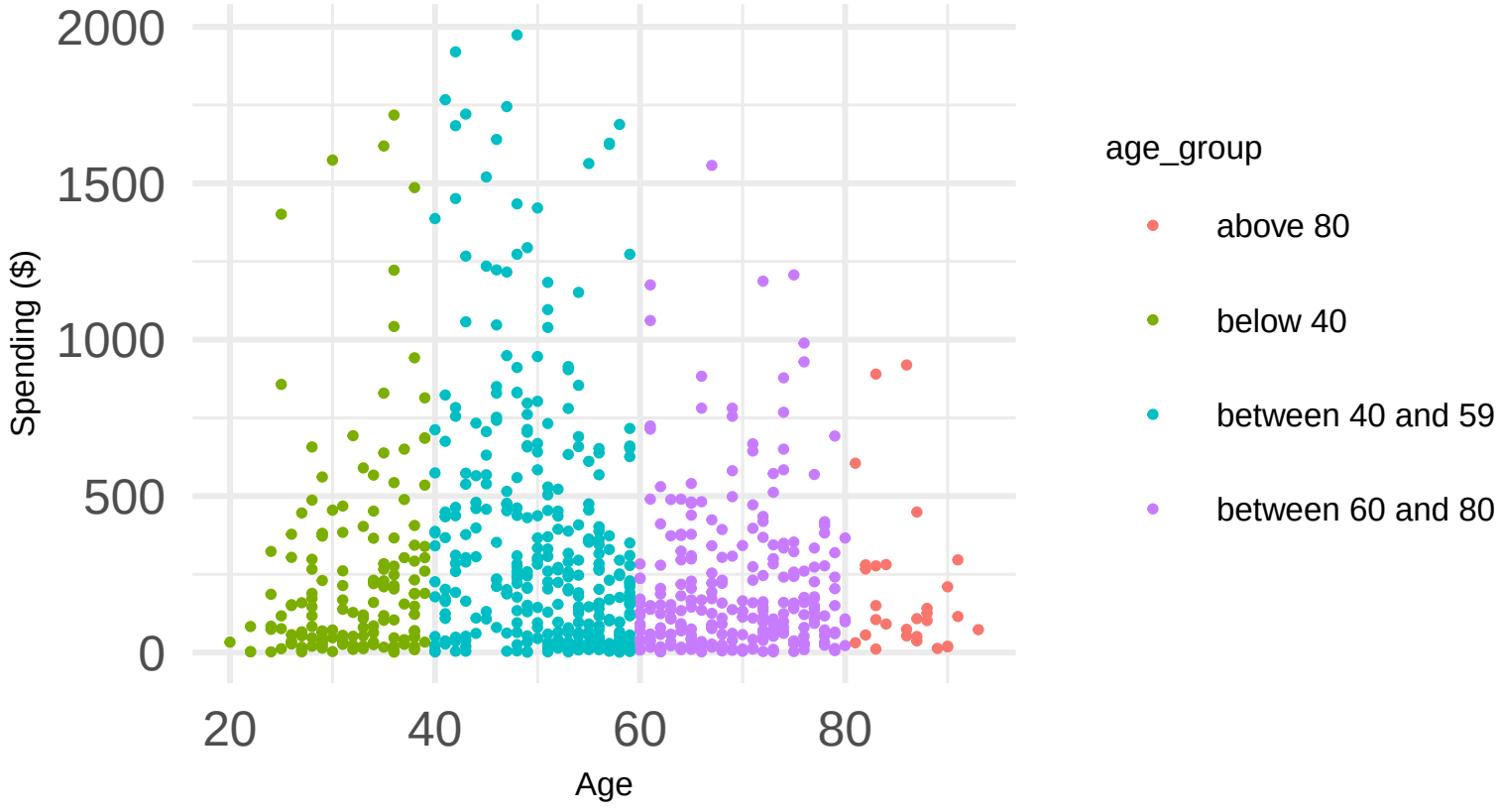

Figure 4

- Each point corresponds to a member's age (horizontal axis) and yearly spending on city dining (vertical axis) from the sample.

Statistical Method for Q3

- We will use the linear regression model shown below to calculate the mean value of yearly spending of each age group, **in the sample**. Also, we can use it to test whether there are differences in **true** mean annual spending among all age groups.
- From the sample, we get the mean of one group and the three differences in means between this group and the other three groups (i.e., β_0, β_1, β_2, β_3, in the model below will be estimated).
- The regression model is:

$$\begin{aligned} \textit{spending}_i = \beta_0 &+ \beta_1 \times I(\text{age group of individual i is below } 40) \\ &+ \beta_2 \times I(\text{age group of individual i is } 40 \text{ to } 59) \\ &+ \beta_3 \times I(\text{age group of individual i is } 60 \text{ to } 80) \\ &+ \textit{error}_i \end{aligned}$$

  where $I(x)$ evaluetes to 1 when x is true, 0 otherwise.
- The baseline value is the age group above 80.

Result for Q3

The Mean Spending ($) of Each Age Group from the Sample

Age Group	below 40	40 to 59	60 to 80	above 80
Mean ($)	257	369	214	202

- According to the p-value calculated, there is no significant difference in spending on city dining among groups below 40 (p-value: 0.42), age group 60-80 (p-value: 0.86) and above 80 (baseline group), while there is a significant difference between the age group 40-59 (p-value: 0.01) and the baseline group.
- Furthermore, **in the sample**, the mean spending of members aged 40 to 59 is the highest, while that of the other three groups are similar. This suggests that the **true** mean spending of the group aged 40-60 is probably the highest.
- Therefore, the manager may consider improving the appeal of dishes to other three age groups to boost their consumption.

Limitation

Limitation of the data:

- The 2017 data may not be a good reference to connect with the real situation in 2021.

Limitation of the statistical methods:

- *Limitation of bootstrap method (Q1):* We give a relatively wide range for the true mean spending (but highly confident on it), but if we want to give a narrower range, we will be less confident on it then.
- *Limitation on randomization testing (Q2):* This testing only informs whether the mean yearly amount spent of two groups differs or not, but does not provide any estimation of the true value.
- *Limitations of the linear regression model (Q3):* We only calculate the mean spending of different age groups of the **sample**, but again, no estimations of **true** average spending of different age groups are provided.

Conclusion

Summary of Findings

- Q1: We are 95% confident that the true yearly amount spent on dining for the RCYC members at the RCYC's restaurants in the city of Toronto (mainland) for 2017 lies in the interval $263 to $312.
- Q2: There is no significant difference in the mean spending of males and females at the RCYC's restaurants in the city of Toronto for 2017.
- Q3: The true mean spending of the age group 40-59 is probably the highest, while there is no significant difference among the other three groups.

Connection between the Findings

- The major connection about these findings is under the big topic of spending behavior. Concerning average spending of groups, female and male members behaved similarly, while among all age groups, the 40 to 59 one on average spent the most.

Conclusion

- We encourage the manager to use this as a reference to carry out further marketing plans to boost the consumption of other age groups to maximize the revenue.

Future advice on spending analysis

- We suggest collecting the data of recent years (eg.2019 and 2020) to explore year-on-year changes in consumption.
- We suggest collecting data on the city's dining spending quarterly to explore the difference in consumption at different periods in a year.
- We suggest collecting further data for 2020 and 2021 to investigate how the pandemic has affected the consumption level for gaining more future advice.