Optimizing demographic experience of the RCYC members

Steven Yeretsian, Kiana Ravanasa, Aabid Anas, Elliot Min

April 1, 2021

Introduction

Using statistical analysis we will be providing the Royal Canadian Yacht Club insight on how to best manage their resources to maximize value. The data we are referencing includes 1000 people, represented twice in the year 2017 and 2020, so a sample of 2000 total. Using this data, we will try to learn how certain demographics in the club use various facilities, and we will investigate the following questions:

1. Is the median amount of other spending at RCYC facilities for females the same as the median amount of spending for males?
2. Is the average amount of island spending for people who have docks equal to that of those who do not have a dock?
3. Is the proportion of males in the Royal Canadian Yacht Club equal to that of the proportion of females in the Royal Canadian Yacht Club?

Objective 1. Is the median amount of other spending at RCYC facilities for females the same as the median amount of spending for males?

We are trying to find out whether the median amount of other spending at RCYC facilities is the same for males and females.

We will be referencing the sex and the amount of other spending for each member from the sample.

- Is the median amount of other spending at RCYC facilities for females the **same** as the median amount of other spending for males?

OR

- Is the median amount of other spending at RCYC facilities for females **different** than the median amount of other spending for males?

Data Summary - 1

- To clean the data, we filtered out all the rows with unavailable data on the sex or the amount of other spending of a person.
- While going through the dataset, we realized that many of the other_spending cells were not available for 2020, and so we chose to filter out 2020 data.
- We also chose to filter out spending that is above $2500 because this indicated outliers and made the data harder to visualize.
- With this data, we created a summary table that shows each gender's median amount of other spending at RCYC facilities.

```
## # A tibble: 2 x 2
##   Sex   median
##   <chr>  <dbl>
## 1 F        400
## 2 M        450
```

Statistical Methods - 1

- In our sample of 1000 RCYC members, we found that the difference in the median other spending between men and women is $50.
- To test whether the median is the same, we randomly redistributed the spendings from the data to see if this $50 discrepancy is likely when redistributed.
- Then, we repeat the simulation 1000 times, with the ultimate goal of comparing our median difference value of $50 to the median difference calculated by these simulations.
- The proportion of simulations as extreme as our difference in the median other spending between men and women will indicate whether there is a true difference in median other spending.

Results - 1a

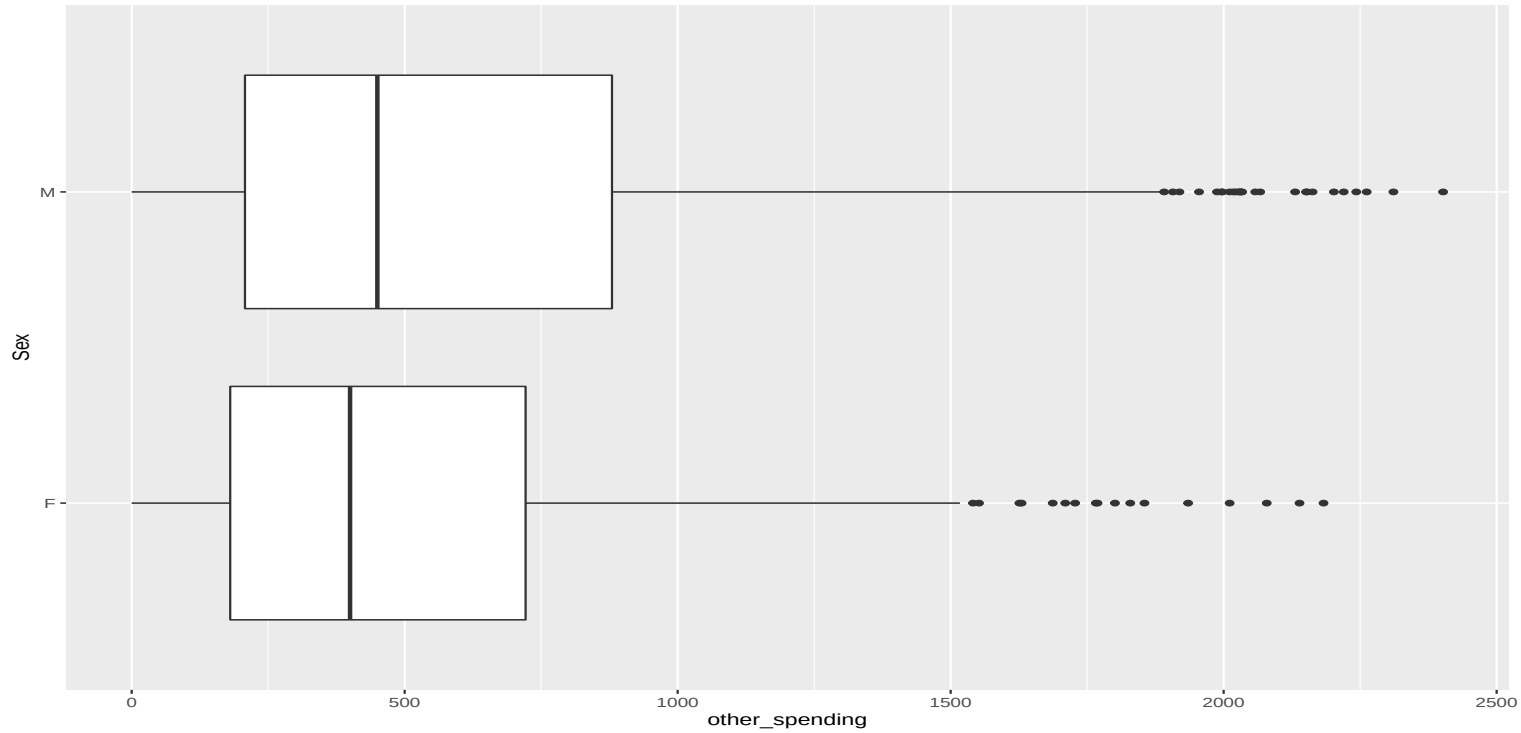

Figure 1: Distribution of Other Spending Between Males and Females

- We can see the median value in the middle of the box, and median spending for both sexes appears relatively close.
- One limitation of this data is the outliers. We have already excluded any spending above 2500 because so few people spend this much, however, there is still a large amount of outliers.

Results - 1b

```
## # A tibble: 1 x 2
##   test_stat p_value
##       <dbl>   <dbl>
## 1        50   0.169
```

- Based on the simulations discussed earlier, we find that there is a 16.9% probability of obtaining our sample result of a $50 difference if there is really no difference in spending between males and females. While 16.9% may seem unlikely, in statistical terms, this is sufficient evidence to show that our result is possible.
- We conclude that there is no evidence suggesting the median spending between males and females is different.

Objective 2. Is the average amount of island spending for people who have docks equal to that of those who do not have a dock?

We have decided to look at the correlation between island dining, age, and having a dock in RCYC members.

This question will allow the RCYC to understand whether members that rent a dock spend more on island dining than those who don't rent a dock, depending on their age. Based on the amount spent by category, the club can determine whether their prices for food and dock prices are just right or too high.

Data Summary - 2

- This test requires observations from the variable “dock” and “island_dining.” “island_dining” is only available for 2017, so we were limited to using data for both variables from 2017.
- We also filtered out any data that did not have a value for dock usage, island dining, or age, since we need these variables to find a relationship between the three.
- We also chose to filter out spending that is above 2500 because this indicated outliers and made the data harder to visualize.

Statistical Methods - 2

- Using the data as described earlier, we will first look at a boxplot to view the relationship between dock and island dining. This will help us get an initial idea for what the relationship may look like.
- We will then look at a linear regression model, which will observe the relationship between age, dock, and island spending by plotting a line of best fit.

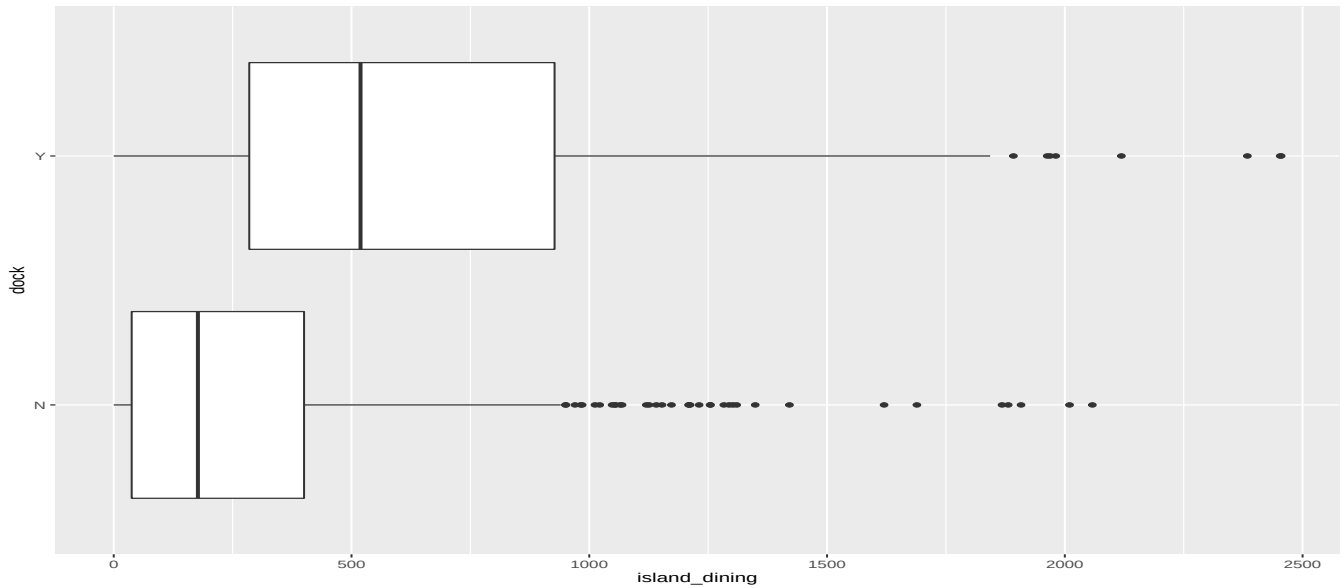

Figure 2: Distribution of docks and island dining

Results - 2a

- We can look at the boxplot to get an initial sense of the correlation between having a dock and island spending.
- We can see that the median amount spent on island dining for members who rent a dock is greater than the amount spent by the majority of members who do not rent a dock.
- Those who can afford to rent a dock for their yacht must be very wealthy, which may explain why they're able to spend so much on island dining compared to those who don't own yachts or rent docks.

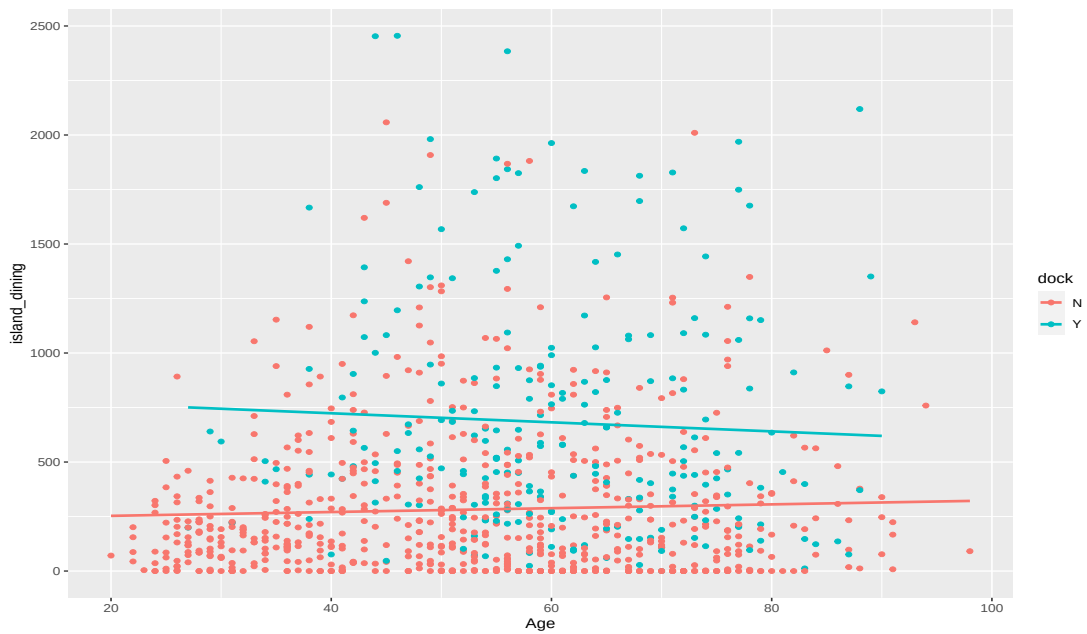

Figure 3: Correlation between age, island dining, and dock

Results - 2b

- The model shows a stronger relationship between age and island dining for those that have a dock compared to those that do not have a dock.
- Those with a dock have a negative relationship, meaning as they get older, they tend to spend less.
- Those without a dock seem to have a positive relationship, but it is much weaker.

One big limitation of this model is the fact that there are so many low values. We can see from the plot that many people spent close to or under $250, which makes it harder to distinguish a relationship.

Another limitation is that we could only use 2017 data, as the data was unavailable for 2020, which could indicate some issues if there have been changes made between 2017 and now.

Objective 3. Is the proportion of males in the Royal Canadian Yacht Club equal to that of the proportion of females in the Royal Canadian Yacht Club?

We are trying to find out whether the proportion of males in the RCYC is equal to the proportion of females in the RCYC.

We will be referencing the sex of the members in the sample.

- Is the proportion of males **equal** to the proportion of females in the RCYC?

OR

- Is the proportion of males **not equal** to the proportion of females in the RCYC?

Data Summary - 3

- To clean the data, we filtered out the rows with data that is unavailable on the sex of the person.
- We also made sure to only include the year 2017 as most of the 2020 data is just overlap of the same member just in a different year.
- With this data, we created a summary table that shows the proportion of males in the RCYC.

```
## # A tibble: 1 x 1
##   prop_male
##       <dbl>
## 1     0.698
```

Statistical Methods - 3

- In our sample of 1000 RCYC members, we found that about 69.77% of the members in that sample are male. At a first glance, we can already see that this is a very high number.
- One limitation of this data could be that it is not entirely representative. Since we only know that the data is random, there's a possibility that it isn't representative.
- We want to test if this proportion of 69.77% from the sample is possible if there is really an even split between males and females.
- To test this, we ran 2000 simulations under the assumptions that there is an equal amount of males as there are females.
- From these simulations, we ended up with 2000 new proportions that we can compare to our data to see the likelihood of getting a 69.77%.

Results - 3a

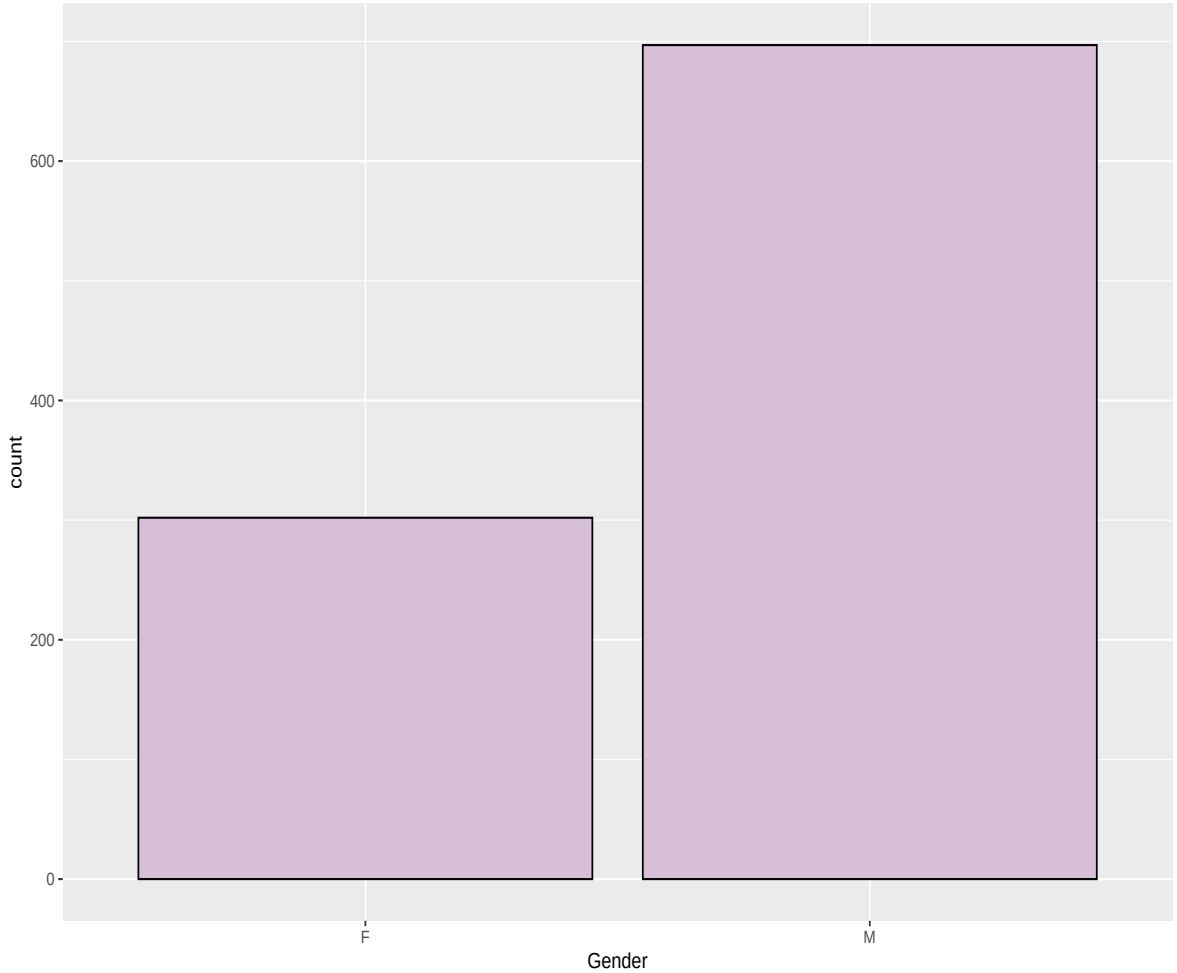

Figure 4: Sex Distribution

As we can see from the bar plot, there are more than twice as many males as females.

Results - 3b

- Based on the simulations discussed earlier, we find that there is a 0% probability of obtaining our sample result of a 69.77% proportion if there is really an equal (50%) proportion.
- While we see a 0% probability, this value is obviously not accurate. If we took an infinite number of samples rather than 2000, we would see that this value is possible. However, a proportion of 69.77% never showed up in our 2000 simulations.
- We conclude that there is very strong evidence suggesting the proportion of males and females is different

Conclusion

Q1: Since we concluded that the median value for men and women is the same, we can conclude that there is little to no discrimination for men and women in other spending, so the RCYC is doing a good job in this area.

Q2: Since we see more spending amongst those with a dock in general, as well as younger dock users, the RCYC can consider either catering specifically to those with docks because they spend more money, or they can try to cater to those without docks to even out spending between groups.

Q3: Because we see that there are many more male members at the RCYC than there are female members, we would suggest that the RCYC make changes to attract more female members. This could include implementing facilities such as a spa, or if their marketing strategy is mostly male oriented, perhaps they should change their marketing technique to appeal more to women, which would incentivize women to join the club.