Statistics as Applied to Economics and Business

Econ 249

Professor Schiro Withanachchi

Group 1 Fall Project (Team Report)

Group 1 Members:

Burgos, Jay

Clarke, Shanice

Louis, Loubens

Noguera, Eliana

Tulsiram, Rajin

Zehra, Aasiya

**Part 1. Explanation of selected variables with a source of data**

Two variables (X and Y) need to be selected from the data source with at least 30 observations for each variable to analyze the possible association

**Team Topic**

Sustainable Goal number 3 – Good Health and Well Being

**Topic: Beer Consumption**

How are income and consumption correlated?

The Cost Per Capita depends on the Price Per Drink.

X variables:- Cost Per Drink

Y Variable:- Cost Per Capita

\*Consumption by States in the U.S.

**Part 2. Using data from the Y variable, create the following:**

• Frequency Distribution Table

**span.cut Freq Rel\_FreqCum\_Freq**

**1 [590,690) 5 0.10 5**

**2 [690,790) 23 0.46 28**

**3 [790,890) 14 0.28 42**

**4 [890,990) 4 0.08 46**

**5 [990,1.09e+03) 3 0.06 49**

**6 [1.09e+03,1.19e+03) 1 0.02 50**

**7 [1.19e+03,1.29e+03) 0 0.00 50**

**8 [1.29e+03,1.39e+03) 0 0.00 50**

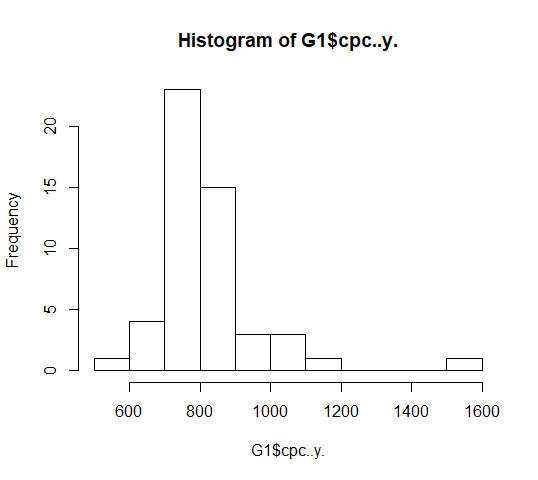
**9 [1.39e+03,1.49e+03) 0 0.00 50**

Span cut 100

Lowest Y value 592

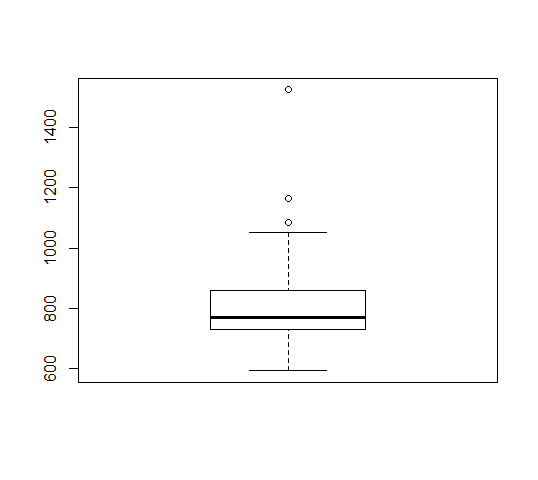
Highest Y value 1526

• Histogram with explanation (shape, the category with highest/lowest frequency, etc.)



The overall shape appears somewhat right-skewed and unimodal at first glance. However, a closer look tells a different story, note that the overall range is 5−10=5 drinks. That is a very small range, only about 5% of the mean. The shape is deceiving in large part because the vertical axis does not start at 0, which exaggerates the differences between the classes. The highest Frequency is approximately 25 and the lowest is approximately 1.

• Boxplot with explanation (shape, outliers, variability, etc.)



In the above box plot, the Cost Per Capita are skewed right. The bottom part of the box represents the median. We currently have one extreme outlier. From the boxplot above the figure shows the descriptive statistics for Cost Per Capita, the variability in cost peak, are measured by the IQR, is Q3-Q1= 49.6 - (-75.28) = 124.88. From the data, cost per capita were closest to the median, half of them were spending $125 per bill.

**Part 3. Using data from the X variable, indicate the summary statistics:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Mean**  **1.983** | **Std. Dev.**  **0.347** | **Min**  **0.920** | **Max**  **2.770** | **Q1**  **1.715** | **Q3**  **2.225** | **Median**  **2.050** |

• What falls in the 75th percentile? Why/why not?

Cases with a cost-per-drink higher than 2.25 are above 75% of the cases in the study. When arranged in an ascending order, the bottom 11 cases fall in the 75th percentile: AL, AK, NY, WY, KY, CA, OK, UT, NM.

• How many observations are more than 3 standard deviations from the mean? Explain

There is no observation more than 3 standard deviations away from the mean because a cost-per-drink with 3 standard deviations away from the mean would be 3.024. The nearest observations to that are Wisconsin and Wyoming (2.74 and 2.77 consecutively) with more than 2 standard deviations away but not 3.

**Part 4. Using both X and Y variables, indicate the following:**

• explanatory and response variables

* Explanatory Variable: Cost per Drink (X)
* Response Variable:Cost per Capita (Y)

• scatterplot with regression line and equation commenting on direction, form, strength, and outliers

Equation:

***Y = a + bX***

* The scatterplot shows a moderate, positive, linear association between the cost per capita and cost per drink. There are a few potential outliers in the data, for:
* The District of Columbia - Utah

cost per drink = 2.14, cost per drink = 2.74

cost per capita = 1526 cost per capita = 592

• strength of the coefficient of correlation between the explanatory and response variables

* Coefficient:

(Intercept) 556.81 119.41 4.663 2.43e-05 \*\*\*

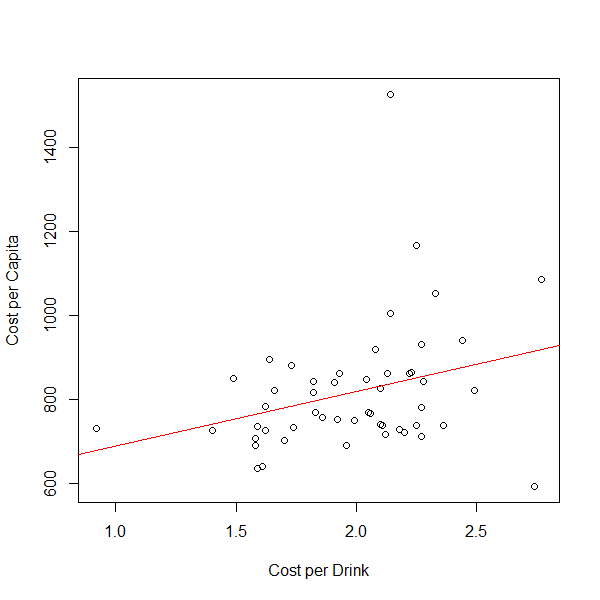
Cost.per.price.x 130.46 59.33 2.199 0.0326 \*

* Based off the results from Rstudio, Correlation is 0.2996744. It would be stronger if we had less outliers in the data. But, the majority of the observations show a strong correlation between Cost-per-drink and Cost-per-capita.

Since, Cost per Capita (*y*) is dependent of Cost per Drink (*x*): The two variables being compared have a positive relationship; when the *x* variable moves higher, the *y* variable moves in the same direction with the same magnitude.

• predict a value for Y using a best guess value for X

* Y= 790
* X= 2.09



**Part 5. Now, estimate the average for the X variable in the future:**

• Using data for the X variable, calculate the 95% confidence interval and interpret the interval



 = **2.082**

**1.884 per drink < U < 2.082 per drink**

**I am 95% confident that the average cost per drink in the United States is between 1.884 per drink and 2.082 per drink**

• Is your best guess of X (from Part 4) within the interval? Explain consistency

Our guess of 2.09 did not fall within the interval. The guess came very close to the upper interval of 2.082. Had we done a 90% confidence, it would have surely fallen within the bounds. This is understandable because it was simply a prediction.

**Part 6. Discuss the following:**

• The economic implications of the relationship between these 2 variables

• What perspectives/experiences did your team have in relation to your topic?

• Specific recommendations to improve this study or alleviate this economic problem

• Each team member needs to reflect on the process of working with your team at:

<https://bit.ly/2nF0sJJ>

There are various economic implications involved regarding the relationship between income and consumption of alcohol. We discovered that there was a direct relationship, as you’d expect, between income and the consumption of alcohol. In essence, the more money people made annually, the more they spent on alcohol. As a group, we discussed how having more money can result in an increase of it being spent on alcohol and juxtaposed the matter into our daily lives. Simply, with the freedom and luxury of having more money, we would be able to spend it whenever we are going out for a drink or for different occasions which may warrant the purchase of alcohol.

Our team was split regarding the topic. There were some members that did not agree with the data or believed that the data was not fairly representative because they do not consume alcohol themselves. With that said, the Cost per Capita does not represent the population as a whole. Others believed that the reason why Alaska’s Cost Per Capita was so high was because there are not many social activities to partake in, leading to an inducement of alcohol as a means of doing so. Ultimately, the variability of responses had reasonable explanations all stemming from social, cultural, and environmental factors or contributions.

​Collectively, we agree that the direct relationship between income and alcohol consumption is not a serious problem, and other studies can be explored regarding the two variables being discussed. Although enjoying a drink or two after work is not a major problem, the average American does have a drinking problem and to fix this, we would suggest either having Americans limit the number of drinks that they drink or substitute high alcoholic beverages with ones with lower alcohol content.