**Data Visualization Approach and One-Way ANOVA:**

**A Case Study of Li Ka Shing Library Entry Data**

Ren Mengxi; Wang Sijia; Wang Tianjing

Singapore Management University

# Abstract

# The purpose of this study is to provide an insightful understanding of usage levels for SMU Li Ka Shing Library. By analyzing individual’s preferred time and visiting frequency, this study seeks to draw relationship between students’ demographic such as nationality and their visiting behavior.

The dataset consists of 481,648 entries generated by 8241 unique users. This study uses Tableau to get exploratory insights and JMP to test for statistical validity. R is used to automate data preparation process.

The one-way analysis of variance (ANOVA) tests the hypothesis of equal means of a number of data.

While this is not common, there are a number of advantages of such an approach, including greater statistical power due to increased precision and more informative interpretation of the results. The objective of our project is to apply statistical analysis to prove the insights drawn by the data visualization.

1. Introduction

Li Ka Shing Library (LKS library) is the first library of Singapore Management University, officially opened on 24 February 2006. The Library is named after Hong Kong businessman Dr. Li Ka-shing, and the Li Ka Shing Foundation donated and endowment to the library for collections. The main objective of the library is to offer an interactive study and research space for SMU community.

The library includes four floors that comprise about 8,800 square meters with 1,800 seats. Inside the library, there are a variety of spaces including open spaces for individual and collaborative use, learning commons which opens 24/7, quiet areas that for individuals to focus on their work, project rooms with LCD panels, investment studio, postgraduate lounges etc. As a modern library, it is also well equipped with high-speed wireless network, color printers, scanners, public computers with professional financial software available, up-to-date newspapers and magazines, collections of lifestyle videos and games, and this is also the reason why LKS library is so attractive for SMU community.

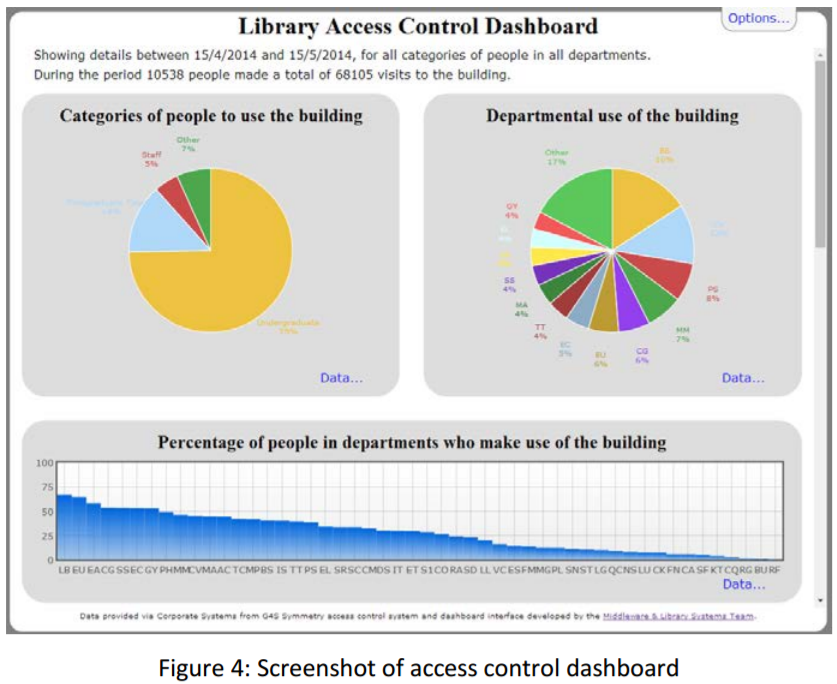
In our project, our focus is on analyzing the library entry information from the card reader logs. The card readers are located at the entrance of the library gantries, both located at the main entrance of LKS library and at the linkbridge side entrance. Students need to tap their card whenever they enter the library. This provides us with the entry information, which includes timestamp and basic information about the student. To better understand the library usage, the library management team is interested to know whether we could find any usage pattern for library of a particular user group (e.g. Dean’s List student), and see if any other business insights could be drawn from the data. We will also work on statistical analysis in order to confirm on our insights.

We use R to build a web application to clean the raw data and use Tableau to do data visualization to compare the usage level for dean’s list and non-dean’s list students, Singaporean students and international students. Then, we will do one-way ANOVA confirmative analysis using SAS JMP.

1.1 Literature Review

Georgia State University(GSU) Library has required students to swipe their campus ID card to enter since 2002 for security control, and the data generated through card swiping could be collected. Since January 2009, a new analytics system by Advanced Campus Services (ACS) has been implemented to make use of data generated by the system in order to help the library to provide better service. This system can generate reports on the total number of swipes and the number of swipes for unique visitors. Therefore, the user will be able to see how many unique students of certain admission year have visited the library during certain time period. Our group think we can apply the same logic into our project as well by analyzing the percentage of each type of students and how many times does each student enter the library with various breakdowns. Another point that GSU team analyzed is the address of the student (stay in the dorm or not). If we could get related information and further look into this aspect, we can draw insights about the relationship between the living area and the frequency the student comes to the library, and figure out whether students living nearer to school are tend to visit library more often.

There is a group of students from Loughborough University publishing a visualizing model to understand the usage of Pilkington Library. Their main focus is on categories of people who enter the library and departmental use of the building. The final presentation inspired us to include a dynamic dashboard in our final product which allows the management team to look at the breakdown by defining a certain time period.



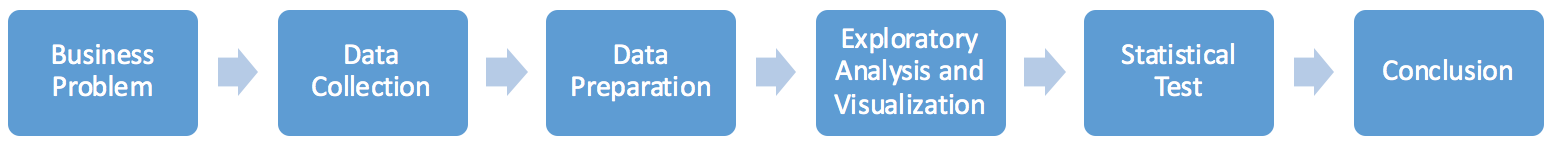
(source:<https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/15522/3/visualising-access-final.pdf>)

We also found a space utilization study of Princeton Theological Seminary Library Besides the gate counts, the research team also collected data like room reservation information, observational data and other qualitative data through surveys. This gives them advantages to have a broader view about the space utilization analysis. Although the paper does not give detailed analysis except key findings, it will be a good move forward direction of our study if library could cooperate with other departments to get related data all together in the future.

2. Methodology

With the business problem identified, we collected related data from our sponsor. Not all datasets provided are used in this project after careful examining of our scope. Details about the data chosen and data preparation process are described in the following section.

After preparation, we proceed to exploratory data analysis (EDA) to get insights from our data. However, conclusions gathered from visualization may not be so accurate, and that is why we proceed to next step to use statistical test for confirmation purpose. ANOVA is used as the main method for such confirmative analysis because it provides more precise and informative interpretation of the results. Final conclusion will only be drawn when it passed the confirmative test.

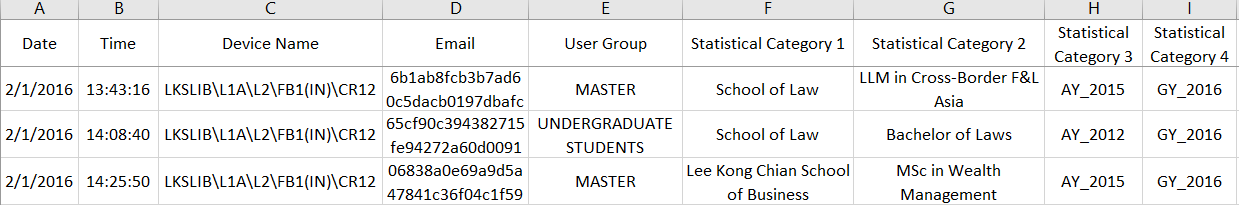


*Figure 1: Data exploration process*

3. Data

3.1 Data Selected

|  |  |
| --- | --- |
| Dataset Name | Description |
| Main Entry Data | The dataset carries out entry data through the main gate of SMU LKS library for the duration of six months from January to June 2017. |
| Linkbridge Data | The dataset carries out entry data through the linkbridge entrance of SMU LKS library for the duration of six months from January to June 2017. |
| Dean’s lister Data | The dataset consists of a list of dean’s listers in academic year 2015-2016. |
| Citizenship Data | The data consists of a list of SMU currently enrolled students and graduated students indicated whether they are Singaporean or international student. |



*Figure 2: Snapshot of main entry data*

The original datasets include:

* Date: the date of the entry in format of d/m/y
* Time: the time of the entry in format of hh:mm:ss
* Device Name: gantry number
* Email: the hashed email address of the entered student
* User Group: which group the student is in (undergraduate/master/phd)
* Statistical Category 1: the school of the student (LKCSB/SOA/SOE/SIS/SOL/SOSS)
* Statistical Category 2: the major of the student
* Statistical Category 3: the admission year of the student
* Statistical Category 4: the graduation year of the student. If the student has not graduated yet, this field will be shown as 0.
* Dean’s List and Non Dean’s List: weather the student is on Dean’s List or not. 1 for dean’s list students and 0 for non dean’s last student.
* Singapore Citizens: whether the student is Singaporean Citizen or not based on the initial letter of student’s ID. 1 for Singaporean students and 0 for international students.

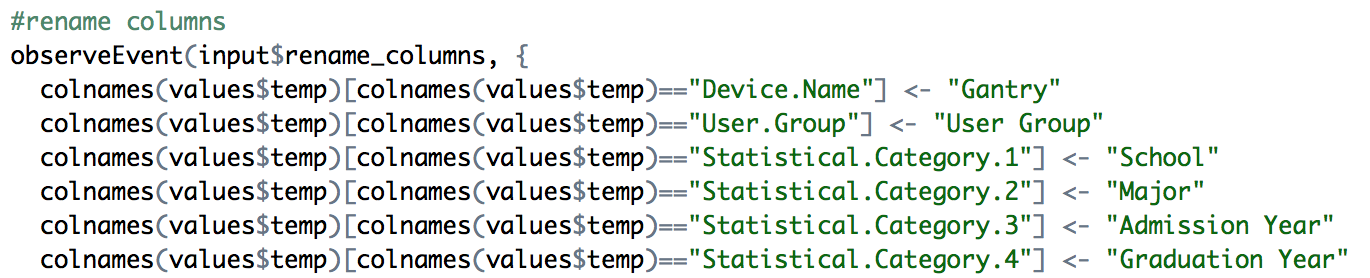
3.2 Data Preparation

The original datasets were presented to us in csv format and thus essential data preparation work is required to allow us to focus on the key objectives better and provide more meaningful analysis.

We concatenated and joined all tables with the hashed email address as the primary key, giving us a dataset consisting of 481,648 entries generated by 8241 unique users. R is used to automate data preparation process as demonstrated as following.

**Rename columns**

To better facilitate our visualization and analysis later, we gave more meaningful names to some columns.

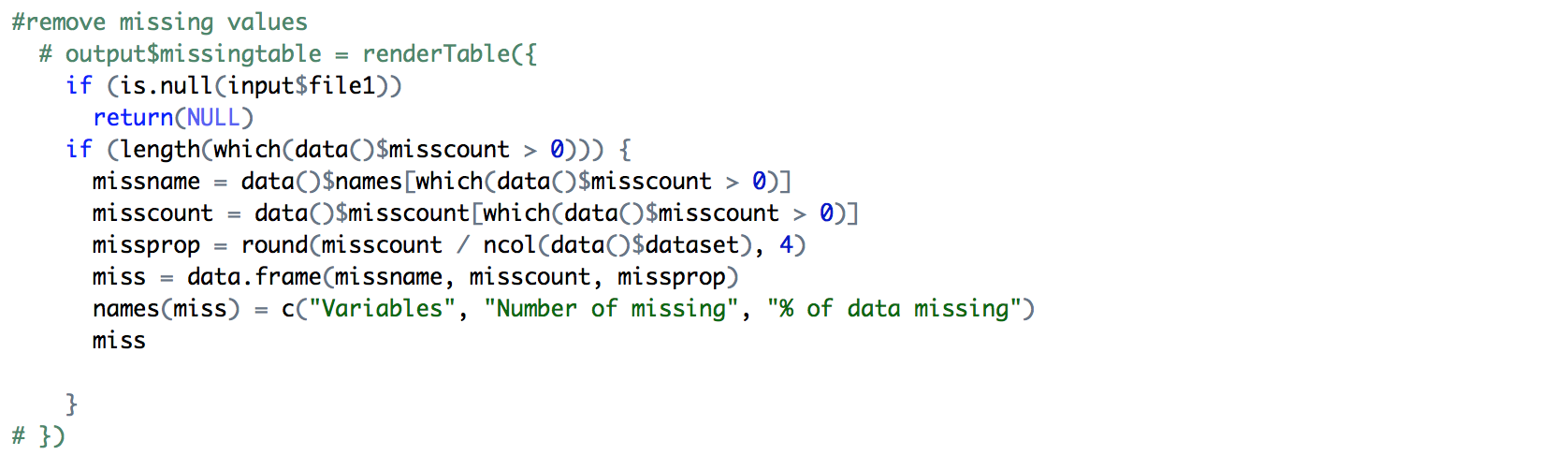


*Figure 3: R code for renaming columns*

**Detect missing value**

Missing data as one of the major dirty data forms will affect distribution badly. In our case, as most of the data is about the demographic information of students, there is no way for us to manipulate the missing data. Therefore, all records with missing value will be excluded from the dataset.

We provided a table for user to view what are the missing values before they are removed. There are 10 records removed.



*Figure 4: R code for removing missing value data*

**Detect outliers**

There are no extreme values in the dataset, but there are values violating our business logic. For example, there are 36 entries through main gate at 12AM and 1AM, which is outside the opening hours. As our dataset covers only the first half year of 2016, any records falling outside the period were treated as outliers as well. These data were removed from our dataset.

**Recode column value**

**Gantry:** Give a short form for each gantry.

**User Group:** Replace “UNDERGRADUATE STUDENTS” with “UNDERGRADS” to make it consistent with the rest of values.

**School:** Use the short form name for each school.

**Major:** Replace “0” value as “Unknown”.

**Admission Year:** Remove “AY\_” in front of year and convert string to number.

**Graduation Year**: Remove “GY\_”in front of year and convert string to number.

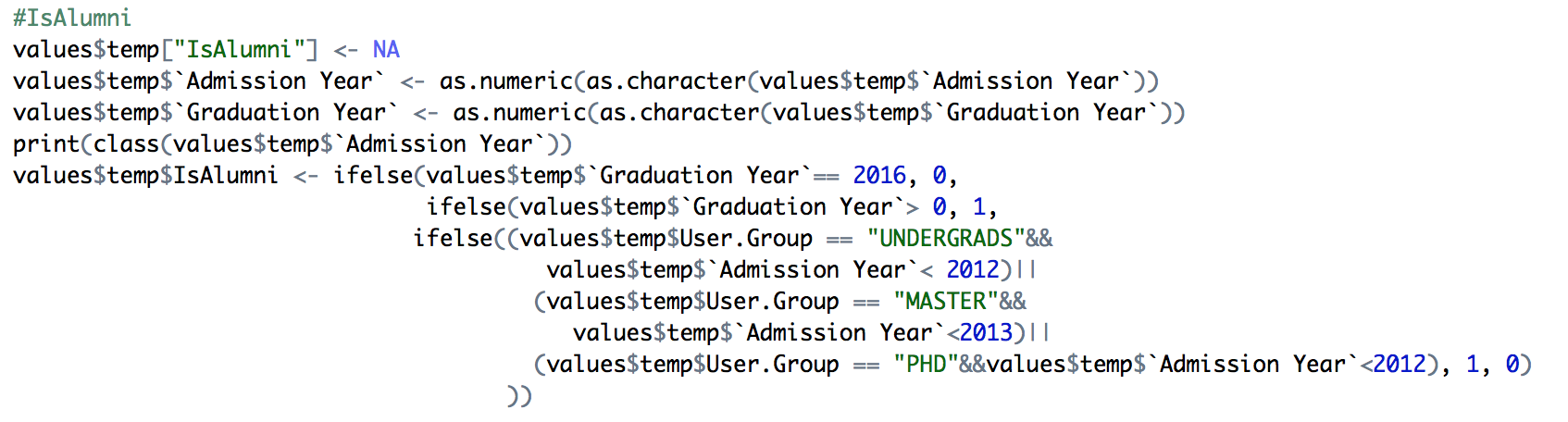
*Figure 5: R code for recoding column value*

**Derive new columns**

We derived new variables ‘Month’ from ‘Date’, ‘Day of Week’ and ‘Hour’ from ‘Time’ to analyze monthly, weekly and hourly usage pattern respectively.

Based on admission and graduation year, we segregated students into different batches and alumni in order to analyze usage pattern across different year of study. We have done a separate study for alumni as a special user group.





*Figure 6: R code for deriving new columns*

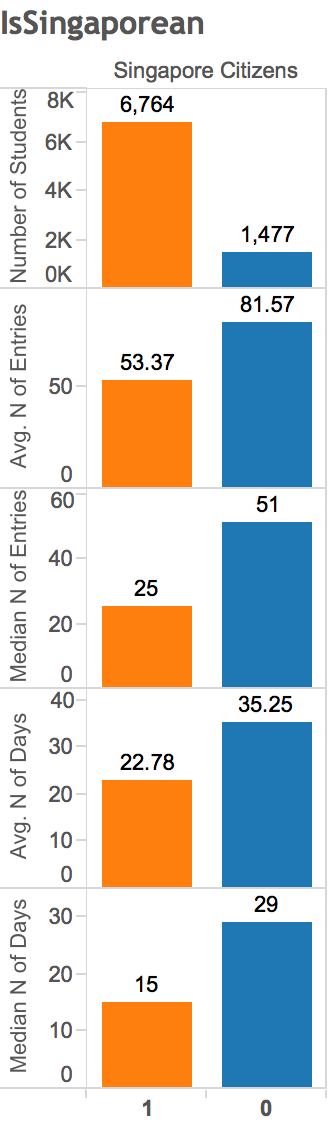
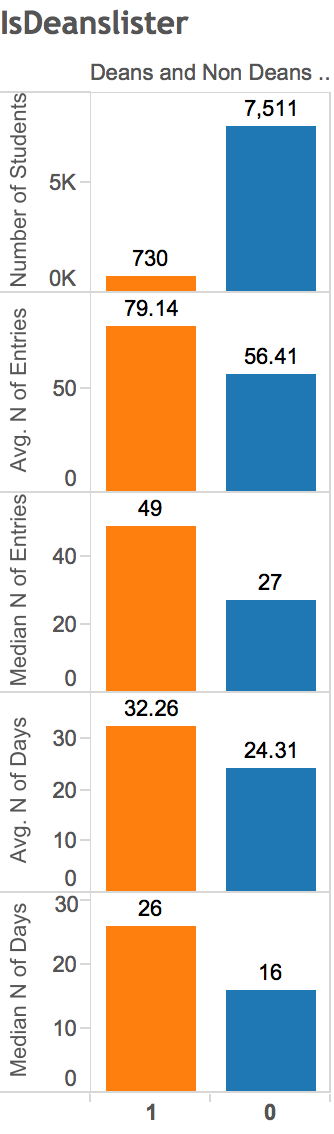
After cleaning, there are 468,891 entries used for exploratory analysis. Below is a summary table for the variables we have modified.

|  |  |
| --- | --- |
| Variable Name | Details of Measures |
| Gantry | Which gantry does the student pass through?  “Main”: Library main entrance  “Linkbridge”: access to the Learning Commons which is available after hours |
| School | Which school does this student from?  "SOE": School of Economics  "SOB":Lee Kong Chian School of Business  "SOL": School of Law  "SIS": School of Information Systems  "SOA": School of Accountancy  "SOSS": School of Social Science |
| IsDeanslister | Is this student a dean’s list student?  0 = No  1 = Yes |
| IsSingaporean | Is this student a Singaporean Citizen?  0 = No  1 = Yes |
| IsAlumni | Is this student an alumni?  0 = No  1 = Yes |

4. Exploratory Data Analysis

Exploratory data analysis (EDA) is an approach to analysis data sets to summarize their main characteristics. Bar charts are usually used to present the grouped data with rectangular bars with lengths proportional to the values that they represent in order to show the comparison among categories. Primarily EDA is often used for seeing what the data can tell us and explore the data to see the possible hypotheses that could lead to model fitting and hypothesis testing. In our study, we use EDA to have an overview of the analysis and draw a primary conclusion from that.

4.1 EDA for All Day Entries Data (main gate + linkbridge)



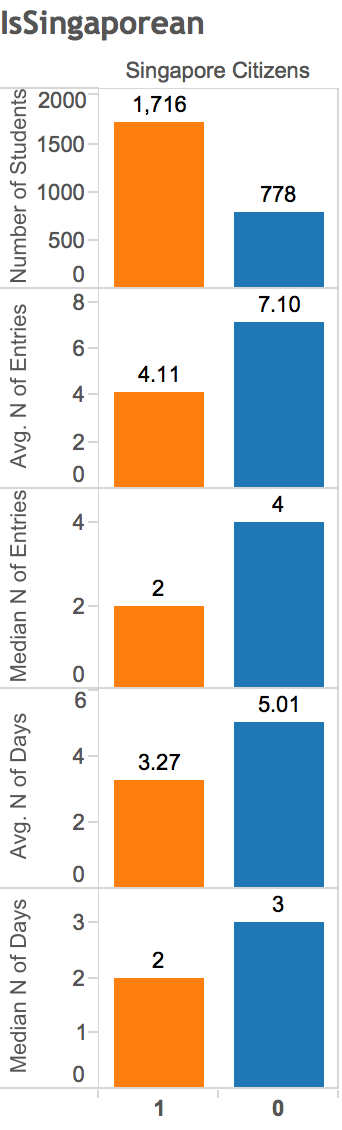
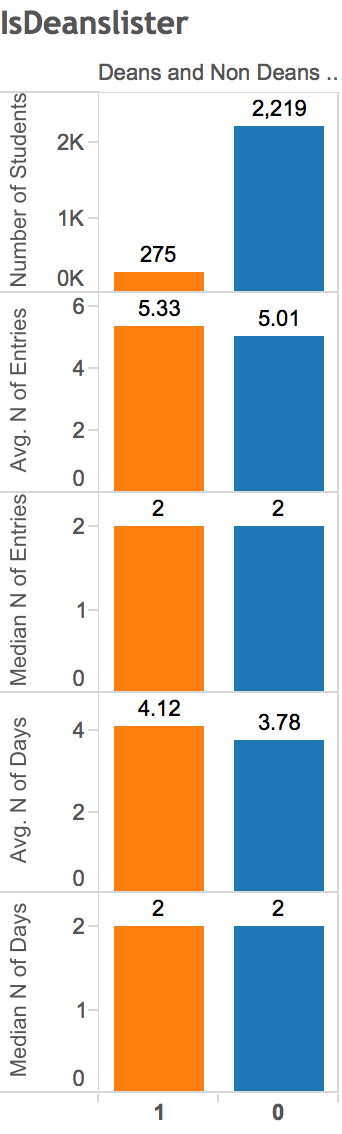
*Figure 7: Comparison between dean’s list and non-dean’s list students, Singaporean and international students via main gate and linkbridge*

There are 730 dean's list students and 7511 non-dean's list students have been to the library during the past six months. Although the absolute number of dean’s list students is low, they have been to library more than non-dean’s list students in terms of both number of entries and number of days.

There are 6764 Singaporean students and 1477 international students have been to the library during the past six months. Although the absolute number of international students is low, they have been to library more than Singaporean students in terms of both number of entries and number of days.

4.2 EDA for Night Entries Data through Linkbridge

Linkbridge entrance is located at the side of LKS Library. It only opens from 12AM to 8AM when main gate is closed. Unlike the main entrance where security guards sit, no one monitors the wave through linkbridge at night. That is why we have put special focus on this set of data.



*Figure 8: Comparison between dean’s list and non-dean’s list students, Singaporean and international students via link bridge only*

Among the 730 dean’s list students who have been to the library during the past six months, 275 of them have used the linkbridge entrance. Among the 7511 dean’s list students who have been to the library during the past six months, 2219 of them have used the linkbridge entrance. Proportion wise is around the same. However, when we compare number of entries and number of days, there is no much difference between dean’s list and non-dean’s list students. Therefore, we conclude that their usage level at night is quite similar.

Among the 6764 Singaporean students who have been to the library during the past six months, 1716 of them have used the linkbridge entrance. Among the 1477 international students who have been to the library during the past six months, 778 of them have used the linkbridge entrance. An interesting pattern can be found here that almost half of the international student group have visited library at night, the proportion is obviously higher than Singaporean student group. In terms of number of entries and number of days, the conclusion will be the same as all day entries that international students have been to library more than Singaporean students.

5. Statistical Test and Analysis

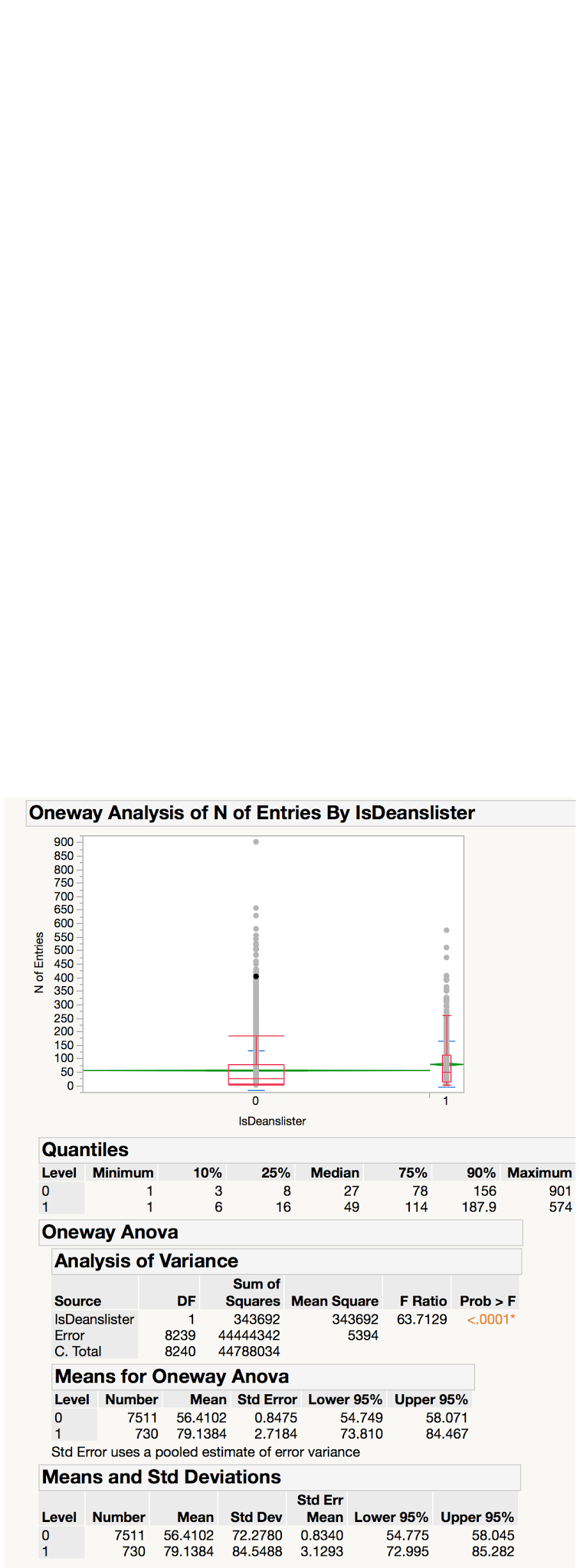
Correlation between the number of entries, number of days, nationality, and whether they are dean’s list students is statistically validated by using parametric ANOVA in this study. The statistical data and graphs are processed with the use of SAS JMP. ANOVA provides a statistical test of whether or not the means of the number of entries and days made by several groups of students are equal. The one-way ANOVA was used to determine whether there are differences at the level of the entries by different groups of student.

The null hypothesis says that means are equal (H0: μ1 = μ2) and the alternative hypothesis says that two means are not equal. The null hypothesis is rejected or accepted on the basis of statistical significance (the significance level α = 0.05). In this case, our null hypothesis is that there is no significant difference between the mean of the two groups.

To further find out business insights from our data, we proceed with statistical test for two levels: number of entries and number of days per student comes to the library.

5.1 ANOVA Test for All Day Entries Data (main gate + linkbridge)

**Comparison between dean’s list / non-dean’s list (number of entries):**



*Figure 10: One way analysis for number of entries for dean’s list and non-dean’s list students via all entries*

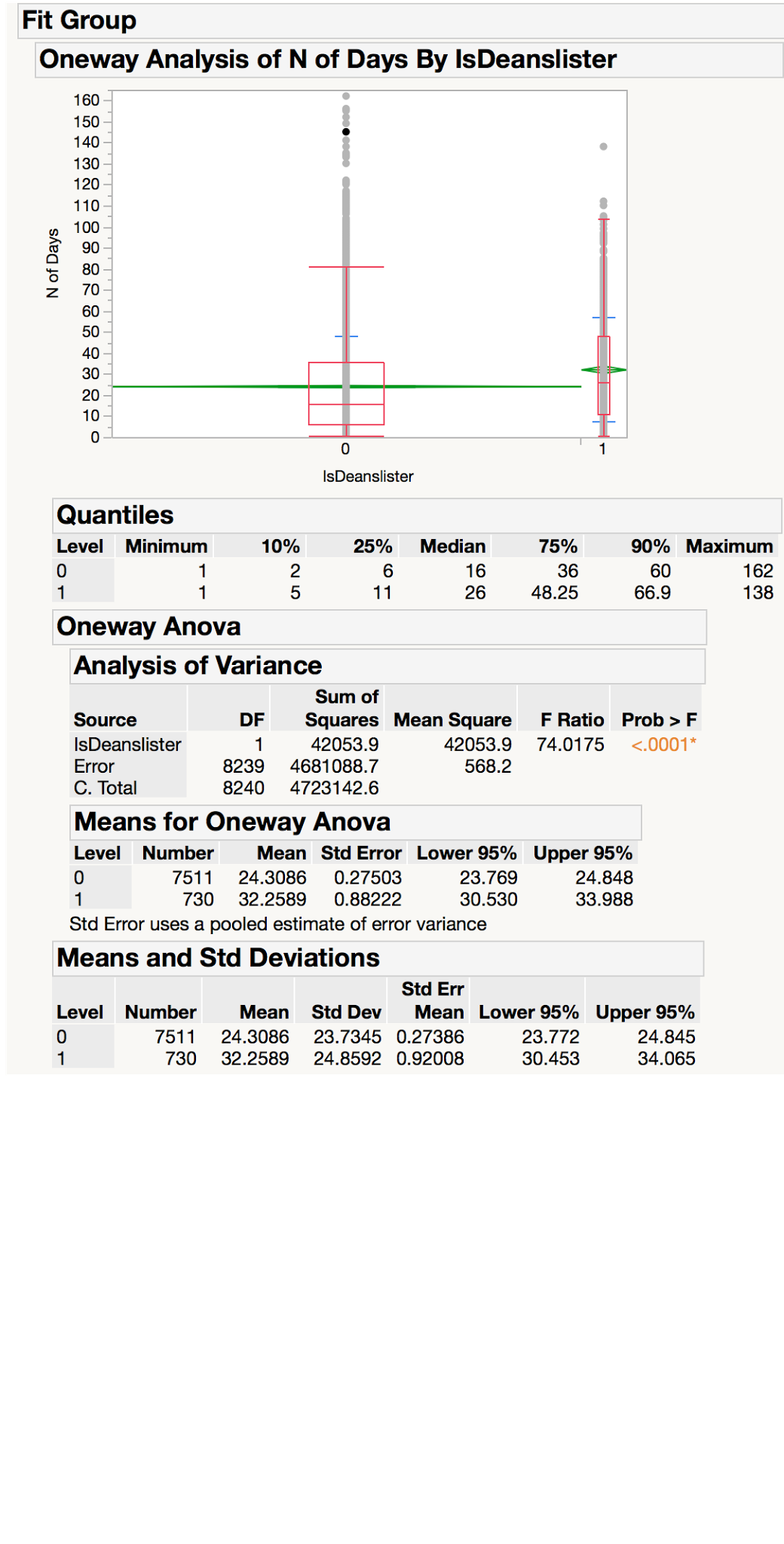
We notice that IsDeanslister = 0 has consistently lower number of entries than IsDeanslister = 1. We also notice that the *x-*axis ticks are unequally spaced. The length between the ticks is proportional to the number of observations for dean's list students and non-dean's list students.

Mean diamonds representing confidence intervals appear: The line near the center of each diamond represents the group mean. At a glance, we can see that the mean for dean's list students and non-dean's list students looks significantly different. The mean of dean's list students is about 79 entries. The mean of non-dean's list students is about 56. The mean of dean's list students is about 13 entries higher than the mean of non-dean's list students. The vertical span of each diamond represents the 95% confidence interval for the mean of each group. We can see dean’s list students has higher interval than non-dean’s list students.

The quantiles provide overall summary information about the analysis: The median of dean's list students is about 49 entries. The median of non-dean's list students is about 27 entries. The median of dean's list students is about 22 entries more than the median of non-dean's list students.

Our null hypothesis is that there is no significant difference in the mean of number of entries between dean’s list and non-dean’s list students. The Analysis of Variance report shows the standard ANOVA information: The Prob > F (the *p*-value) is <0.0001, which is lower than our significance level 0.05. We would reject our null hypothesis and conclude there are significant differences between the number of entries made by dean's list students and by non-dean's list students, which supports our visual conclusion.

**Comparison between dean’s list / non-dean’s list (number of days):**



*Figure 11: One way analysis for number of days for dean’s list and non-dean’s list students for all entries*

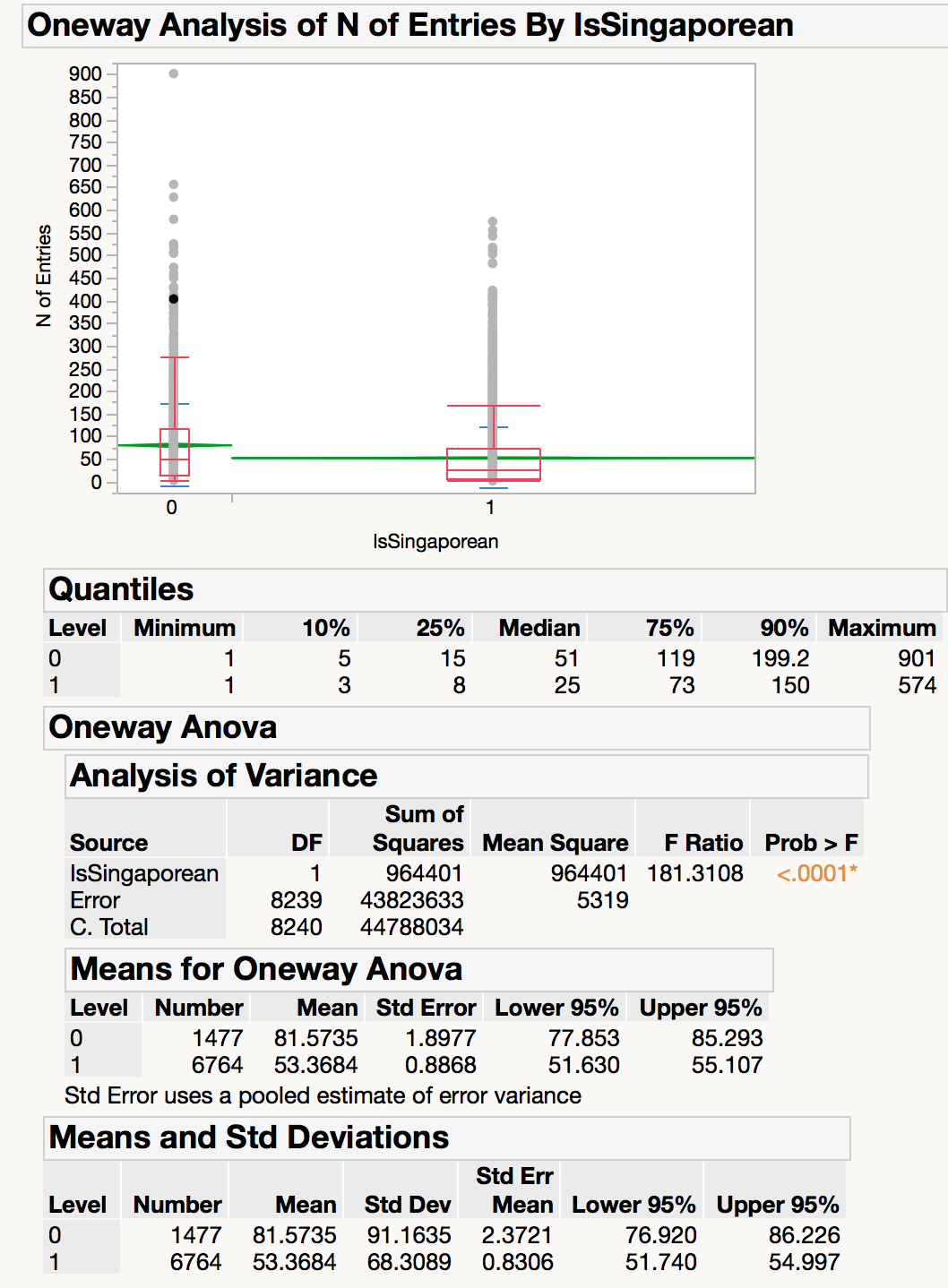
We notice that IsDeanslister = 0 has consistently lower number of days than IsDeanslister = 1. We also notice that the *x-*axis ticks are unequally spaced. The length between the ticks is proportional to the number of observations for dean's list students and non-dean's list students.

Mean diamonds representing confidence intervals appear: The line near the center of each diamond represents the group mean. At a glance, we can see that the mean for dean's list students and non-dean's list students looks significantly different. The mean of dean's list students is about 32 days. The mean of non-dean's list students is about 24. The mean of dean's list students is about 8 days higher than the mean of non-dean's list students.The vertical span of each diamond represents the 95% confidence interval for the mean of each group. We can see dean’s list students has higher interval than non-dean’s list students.

The quantiles provides overall summary information about the analysis:The median of dean's list students is about 26 days. The median of non-dean's list students is about 16 days. The median of dean's list students is about 10 days higher than the median of non-dean's list students.

Our null hypothesis is that there is no significant difference in the mean number of days between dean’s list and non-dean’s list students. The Analysis of Variance report shows the standard ANOVA information: The Prob > F (the *p*-value) is <0.0001, which is lower than our significance level 0.05. We would reject our null hypothesis and conclude that there are significant differences between the number of days by dean's list students and non-dean's list students, which supports our visual conclusion.

**Comparison between Singaporean / international students (number of entries):**



*Figure 12: One way analysis for number of entries for Singaporean and international students for all entries*

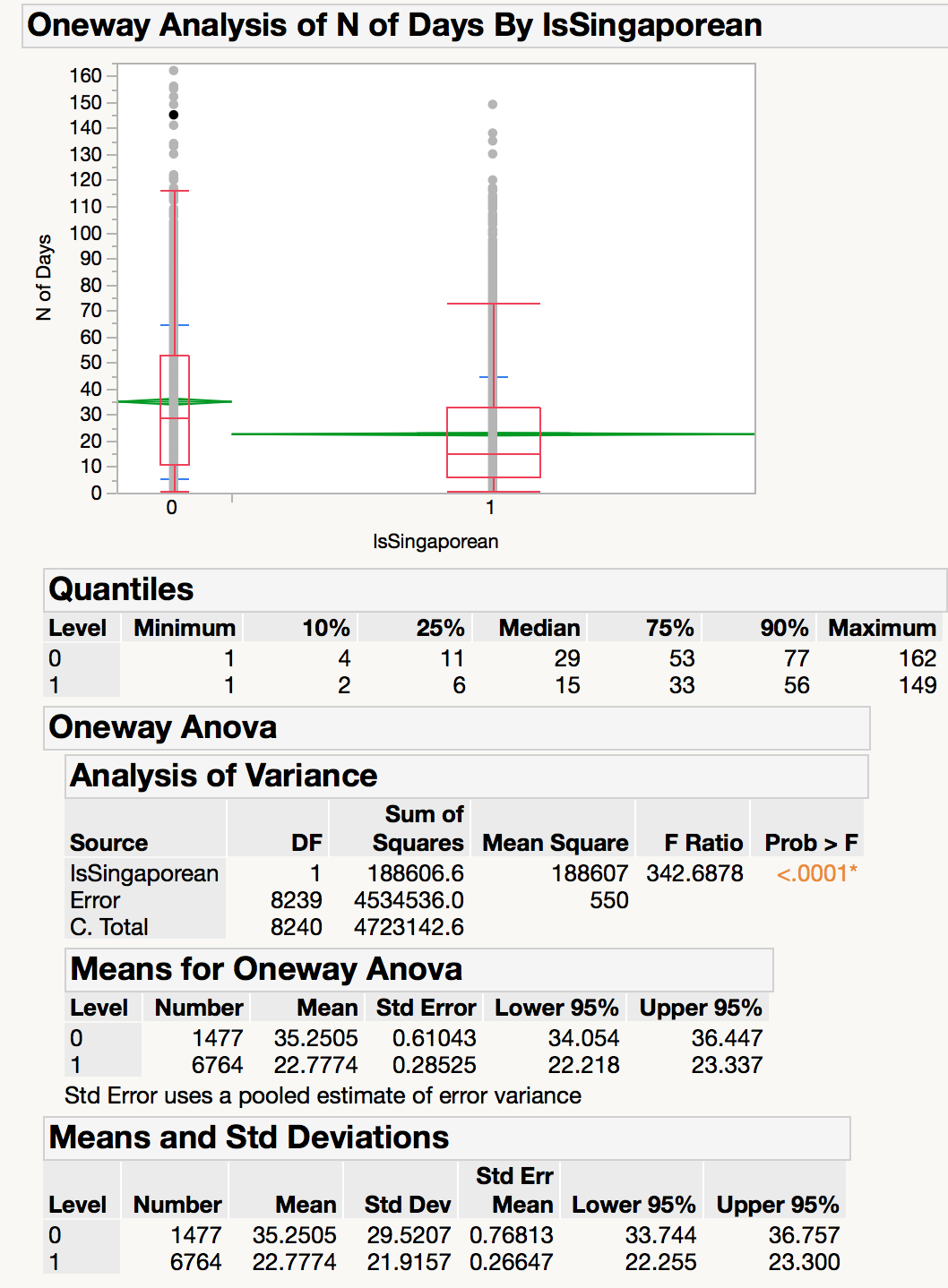
We notice that IsSingaporean = 0 has consistently higher number of entries than IsSingaporean = 1. We also notice that the *x-*axis ticks are unequally spaced. The length between the ticks is proportional to the number of observations for Singaporean and non-Singaporean.

Mean diamonds representing confidence intervals appear: The line near the center of each diamond represents the group mean. At a glance, we can see that the mean for Singaporean and non-Singaporean looks significantly different. The mean of international students is about 82 entries. The mean of Singaporeans is about 53. The mean number of entries made by international students is about 29 entries more than Singaporean students. The vertical span of each diamond represents the 95% confidence interval for the mean of each group. We can see international student has higher interval than Singaporean student.

The quantiles provide overall summary information about the analysis: The median of international students is about 51 entries. The mean of Singaporeans is about 25 entries. The median of international student is about 26 entries more than the median of Singaporean student.

Our null hypothesis is that there is no significant difference in the mean of number of entries between Singaporean and international students. The Analysis of Variance report shows the standard ANOVA information: The Prob > F (the *p*-value) is <0.0001, which is lower than our significance level 0.05. We would reject our null hypothesis and conclude that there are significant differences between the number of entries made by Singaporean students and by international students, which supports our visual conclusion.

**Comparison between Singaporean / international students (number of days):**



*Figure 13: One way analysis for number of days for Singaporean and international students for all entries*

We notice that IsSingaporean = 0 has consistently higher number of days than IsSingaporean = 1. We also notice that the *x-*axis ticks are unequally spaced. The length between the ticks is proportional to the number of observations for Singaporean students and international students.

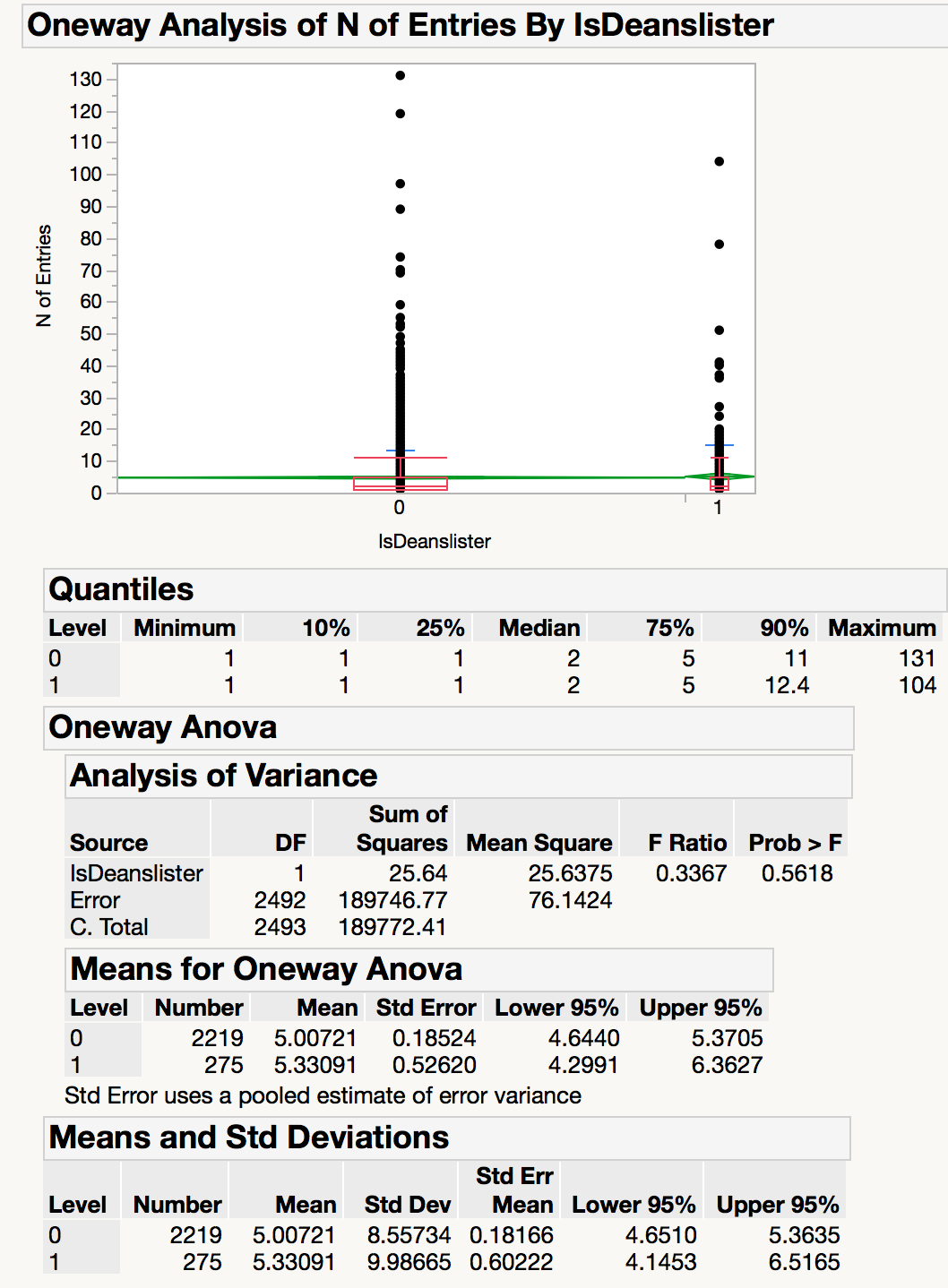
Mean diamonds representing confidence intervals appear: The line near the center of each diamond represents the group mean. At a glance, we can see that the mean for Singaporean students and international students looks significantly different. The mean of international students is about 35 days. The mean of Singaporeans is about 23. The mean of international students is about 12 days higher than the mean of Singaporeans. The vertical span of each diamond represents the 95% confidence interval for the mean of each group. We can see international student has higher interval than Singaporean student.

The quantiles provide overall summary information about the analysis: The mean of international students is about 29 days. The mean of Singaporeans is about 15 days. The median of international student is about 14 days higher than the median of Singaporean student.

Our null hypothesis is that there is no significant difference in the mean of number of days between Singaporean and international students. The Analysis of Variance report shows the standard ANOVA information. The Prob > F (the *p*-value) is <0.0001, which is lower than our significance level 0.05. We would reject our null hypothesis and conclude that there are significant differences between the number of days by Singaporean students and international students, which supports our visual conclusion.

5.2 ANOVA Test for Night Entries Data through Linkbridge

**Comparison between dean’s list / non-dean’s list (number of entries):**



*Figure 14: One way analysis for number of entries for dean’s list and non dean’s list students via linkbridge*

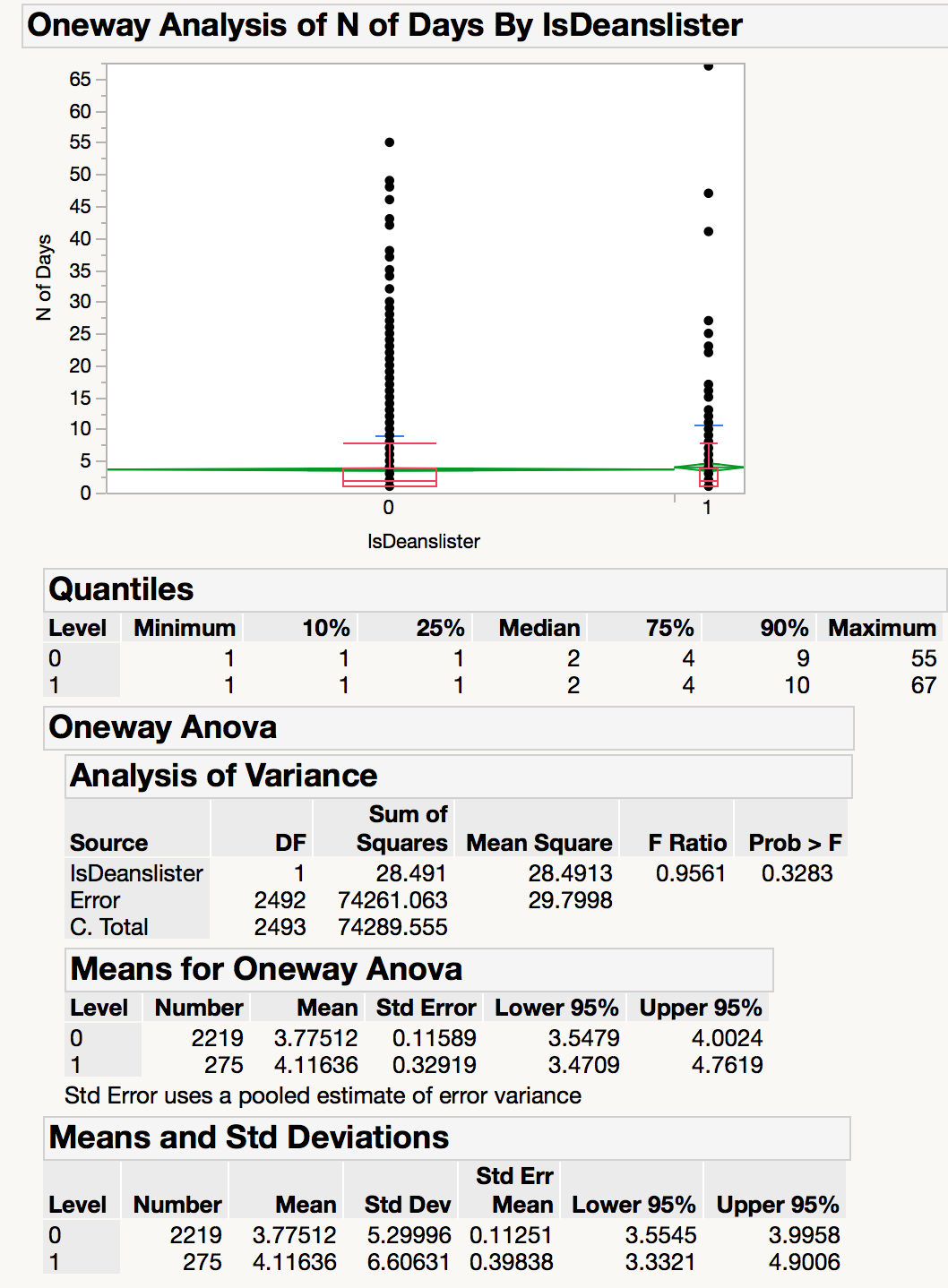
We notice that IsDeanslister = 0 has consistently lower number of entries than IsDeanslister = 1. We also notice that the *x-*axis ticks are unequally spaced. The length between the ticks is proportional to the number of observations for dean's list students and non-dean's list students.

Mean diamonds representing confidence intervals appear: The line near the center of each diamond represents the group mean. At a glance, we can see that the mean for dean's list students and non-dean's list students does not look significantly different. The mean of dean's list students is about 5 entries. The mean of non-dean's list students is about 5.33. The vertical span of each diamond represents the 95% confidence interval for the mean of each group. We can see dean’s list students has slightly higher interval than non-dean’s list students.

The quantiles provide overall summary information about the analysis: The median of dean's list students is about 2 entries. The median of non-dean's list students is about 2 entries.

Our null hypothesis is that there is no significant difference in the mean number of entries via link bridge between dean’s list and non-dean’s list students. The Analysis of Variance report shows the standard ANOVA information: The Prob > F (the *p*-value) is <0.56, which is higher than our significance level 0.05. Therefore, we accept our null hypothesis and conclude there is no significance difference between the mean number of entries of dean’s list and non-dean’s list students.

**Comparison between dean’s list / non-dean’s list (number of days):**



*Figure 15: One way analysis for number of days for dean’s list and non-dean’s list students via linkbridge*

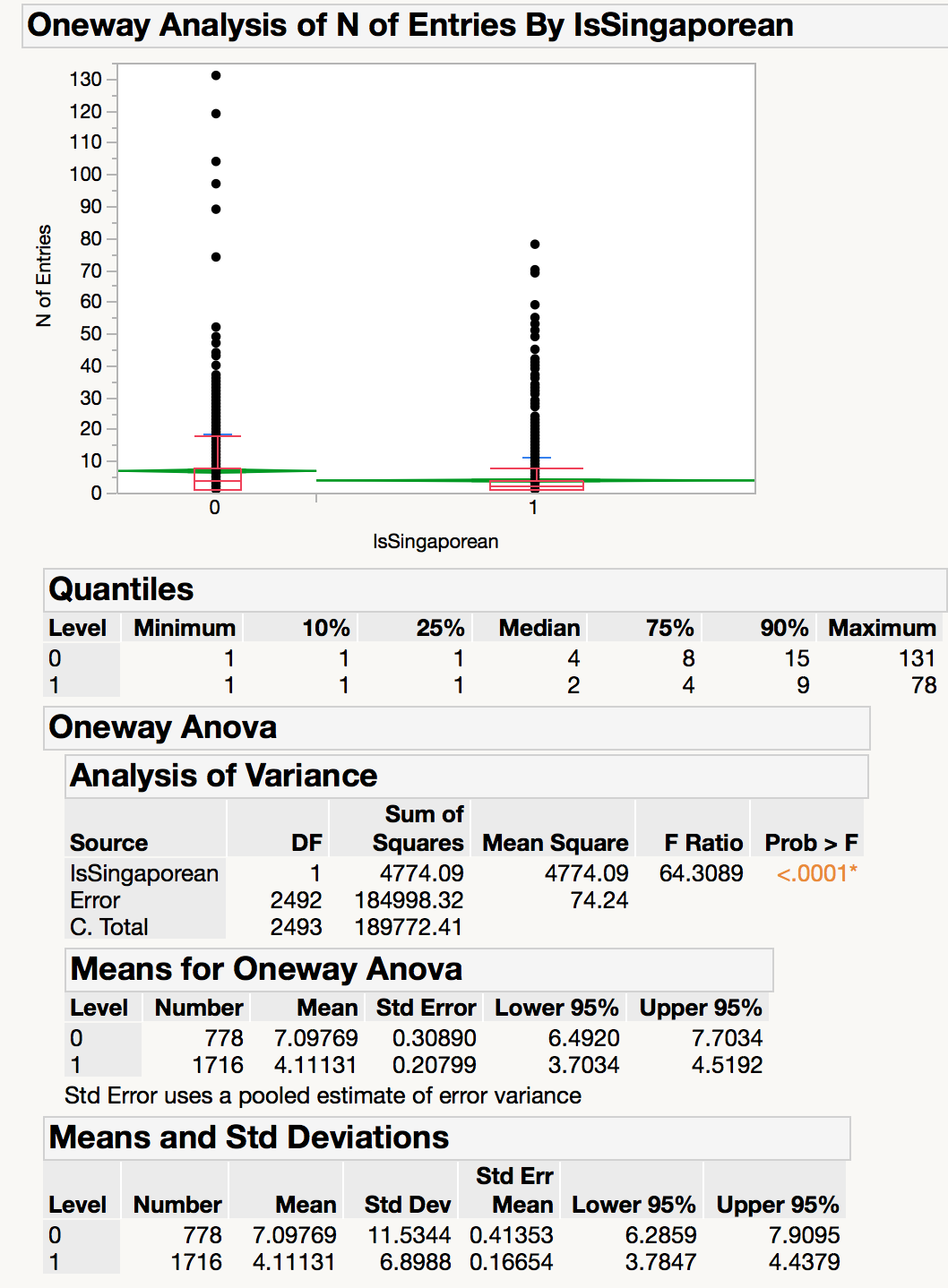
We notice that IsDeanslister = 0 has consistently lower number of days than IsDeanslister = 1. We also notice that the *x-*axis ticks are unequally spaced. The length between the ticks is proportional to the number of observations for dean's list students and non-dean's list students.

Mean diamonds representing confidence intervals appear: The line near the center of each diamond represents the group mean. At a glance, we can see that the mean for dean's list students and non-dean's list students does not looks significantly different. The mean of dean's list students is about 4.1 days. The mean of non-dean's list students is about 3.7. The vertical span of each diamond represents the 95% confidence interval for the mean of each group. We can see dean’s list students has slightly higher interval than non-dean’s list students.

The quantiles provide overall summary information about the analysis: The median of dean's list students is about 2 days. The median of non-dean's list students is about 2 days.

Our null hypothesis is that there is no significant difference in the mean number of days via link bridge between dean’s list and non-dean’s list students. The Analysis of Variance report shows the standard ANOVA information: The Prob > F (the *p*-value) is <0.33, which is higher than our significance level 0.05. Therefore, we accept our null hypothesis and conclude there is no significance difference between the number of days of dean’s list and non-dean’s list students.

**Comparison between Singaporean / international students (number of entries):**



*Figure 16: One way analysis for number of entries for Singaporean and international students via linkbridge*

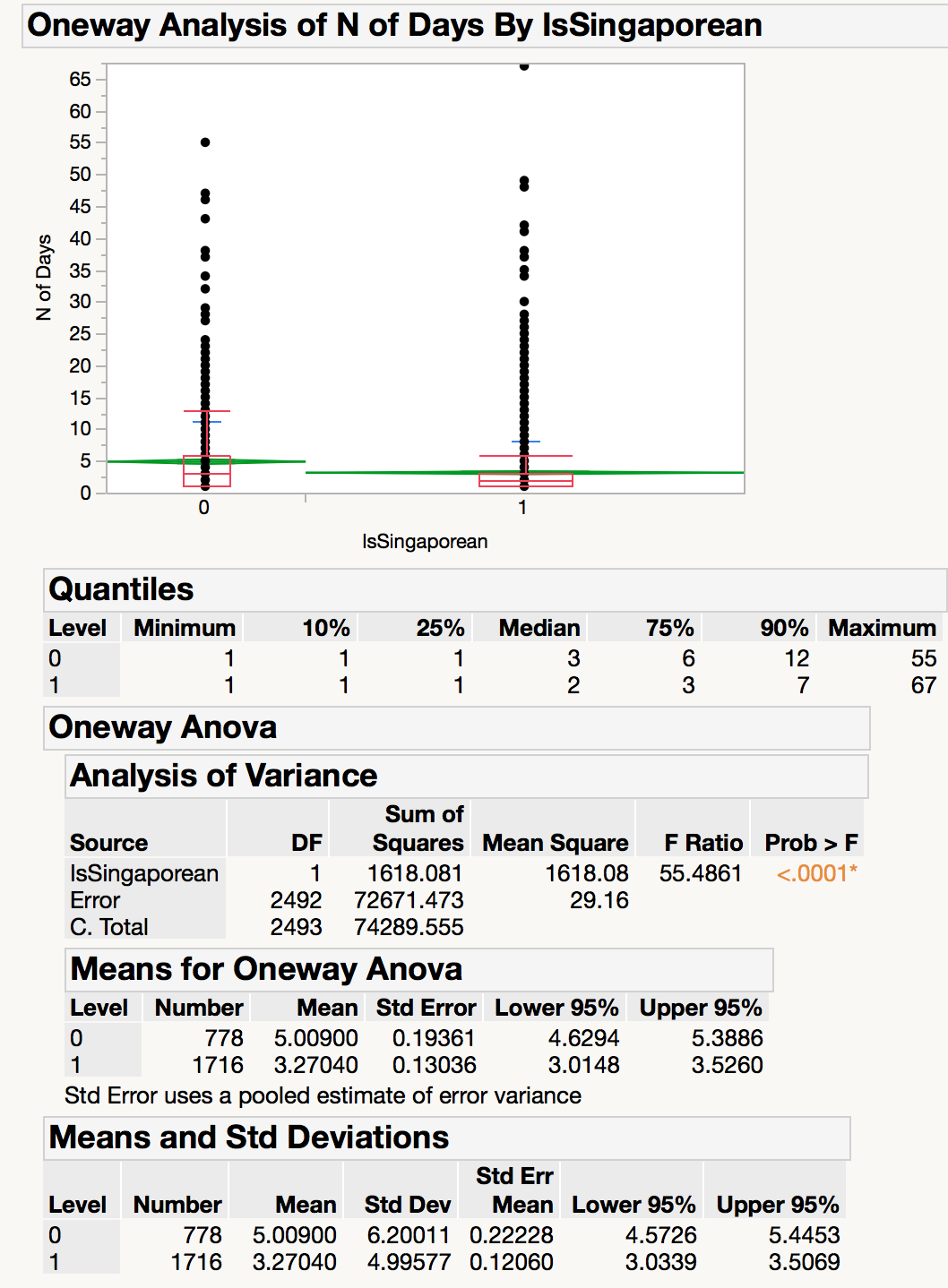
We notice that IsSingaporean = 0 has consistently higher number of entries than IsSingaporean = 1. We also notice that the *x-*axis ticks are unequally spaced. The length between the ticks is proportional to the number of observations for Singaporean and non-Singaporean.

Mean diamonds representing confidence intervals appear: The line near the center of each diamond represents the group mean. At a glance, we can see that the mean for Singaporean and non-Singaporean does not look significantly different. The mean of international students is about 7 entries. The mean of Singaporeans is about 4. The mean number of entries made by international students is about 3 entries more than Singaporean students. The vertical span of each diamond represents the 95% confidence interval for the mean of each group. We can see international student has higher interval than Singaporean student.

The quantiles provides overall summary information about the analysis: The median of international students is about 4 entries. The mean of Singaporeans is about 2 entries. The median of international student is about 2 entries more than the median of Singaporean student.

Our null hypothesis is that there is no significant difference in the mean number of entries via link bridge between Singaporean and international students. The Analysis of Variance report shows the standard ANOVA information: The Prob > F (the *p*-value) is <0.0001, which is lower than our significance level 0.05. We would reject our null hypothesis and conclude that there are significant differences between the number of entries made by Singaporean students and by international students, which supports our visual conclusion.

**Comparison between Singaporean / international students (number of days):**



*Figure 17: One way analysis for number of days for Singaporean and* international *students via linkbridge*

We notice that IsSingaporean= 0 has consistently higher number of days than IsDeanslister = 1. We also notice that the *x-*axis ticks are unequally spaced. The length between the ticks is proportional to the number of observations for Singaporean students and international students.

Mean diamonds representing confidence intervals appear: The line near the center of each diamond represents the group mean. At a glance, we can see that the mean for Singaporean students and international list students looks significantly different. The mean of international students is about 5 entries. The mean of Singaporeans is about 3. The mean of international students is about 2 entries higher than the mean of Singaporeans. The vertical span of each diamond represents the 95% confidence interval for the mean of each group. We can see international student has higher interval than Singaporean student.

The quantiles provides overall summary information about the analysis:The mean of international students is about 3 days. The mean of Singaporeans is about 2 days. The median of international student is about 1 days higher than the median of Singaporean student.

Our null hypothesis is that there is no significant difference in the mean number of days via link bridge between Singaporean and international students. The Analysis of Variance report shows the standard ANOVA information. The Prob > F (the *p*-value) is <0.0001, which is lower than our significance level 0.05. We would reject our null hypothesis and conclude that there are significant differences between the number of days by Singaporean students and international students, which supports our visual conclusion.

# 6. What This Study Adds

Exploratory data analysis (EDA) is used an approach to analysis data to give us an overview of the distributions of the variables and summarize their main characteristics. Here we use bar charts to present the grouped data to show the comparison among dean’s lister students and non-dean’s lister students, and among Singaporean students and international students in terms of the number of entries and the number of days they came to library. Primarily EDA provides us possible hypotheses for following statistical testing.

Compared with EDA data visualization, statistical test in terms of one-way ANOVA brings a number of advantages and great statistical power due to increased precision and more informative interpretation of the results. It provides us with deep insights on the relationships between different variables and statistically prove our hypothesis based on the data. In our case, we use one-way ANOVA test to further validate the conclusion we drawn from EDA process. With our statistical test result, we reject or accept our null hypothesis, and this provides us with statistical support for our final conclusion.

# 7. Limitation and Future Work

During these six months, there are some days that the card reader system underwent upgrading so the entry data are not logged properly. Therefore, we have some data losses inside these six months’ entry data, which compromises the accuracy of our analysis.

The six months includes winter break, Term 2, Term 3A, break and Term 3B. As most of the students only have lessons during Term2, and the operating hours differs from different terms, the user behavior could be very different at different time. More insights could be drawn if we could further look into the usage level for different user groups in different terms/break.

In our current research, we have mapped all students as Singaporean students and international students. However, in SMU, students are from very diverse background and it would be more insightful if we could get the specific nationality for international students and mapped them into different nationalities. This would be useful for our client to see how to better target students from different nations and for the whole SMU community to involve students from all over the world better. Also, we currently mapped student into dean’s list students and non-dean’s list students. However, dean’s list students only make up around 10% of the student population. If our client is interested to see how students with different academic performance differs in their usage level of library, it would be better to split students into more specific groups (e.g. GPA 2.0-2.5).

In this research paper, we focused on two types of user groups (Singaporean and international students, dean’s list and non-dean’s list students). There are more grouping of students that we could look into like school, major, year of study, etc. With further analysis into these groupings, client could have more insights about the usage level of different groups of students and further improve the service.

8. Conclusion

This paper examined the associations between library entry data and students' demographic:

In general, the results of our study support the proposed theoretical framework. Dean’s list students have heavier usage level than non-dean’s list students. International students have heavier usage level than Singaporean students, and international students go through the linkbridge gate at night more than Singaporean students do.

# References

Aluregowda (2013). Organized Retail Strategy – A Study at Loyal World Supermarket. IOSR Journal of Business and Management P75-80

Brewerton, G., & Cooper, J. (2014). Visualising building access data. Ariadne, (73).

Ivan-Damir Aniae & Sonja Radas (2006). The Relationships Between Shopping Trip Type, Purchases

Leighton, P. D., & Weber, D. C. (1999). Planning academic and research library buildings. American library association.

Made on Promotion, and Unplanned Purchases for a High/Low Hypermarket Retailer.

Stanley E Lazic (2008). Why we should use simpler models if the data allow this: relevance for ANOVA designs in experimental biology. BMC Physiology.

Jones, J. L. (2010). Using library swipe-card data to inform decision making. Chicago