

**Statistical Analysis of Marino Center**

**Supervised by: Professor Rajesh Jugulum**

**Prepared by:**

**• Deep Vaghani (NUID: 001677703)**

**• Sahil Shah (NUID: 001703724)**

**• Preet Parmar (NUID: 001679848)**

* **Sonam Malhotra (NUID: 001770583)**

**Table of Contents**

[1. ABSTRACT 3](#_Toc437261570)

[2. METHODOLOGY 4](#_Toc437261571)

[ P - Diagram: 4](#_Toc437261572)

[3. DATA SETS 5](#_Toc437261573)

[ Box Plot Diagram 6](#_Toc437261574)

[ Time Series Plot 7](#_Toc437261575)

[4. DESCRIPTIVE ANALYSIS OF PEOPLE USING MARINO CENTER 8](#_Toc437261576)

[5. HYPOTHESIS ANALYSIS: 11](#_Toc437261577)

[ Outcome 11](#_Toc437261578)

[ Result 11](#_Toc437261579)

[6. SCOPE OF IMPROVEMENTS: 12](#_Toc437261580)

[7. References 15](#_Toc437261581)

**Table of Figures**

[**Figure 1: P- Diagram 5**](#_Toc437255442)

[**Figure 2: Box Plot Diagram 6**](#_Toc437255443)

[**Figure 3: Time Series Plot 7**](#_Toc437255444)

[**Figure 4: Histogram of number of people 8**](#_Toc437255445)

[**Figure 5: Individual plot of number of people 9**](#_Toc437255446)

[**Figure 6: Normal probability plot of number of people 10**](#_Toc437255447)

[**Figure 7: Frequency of people entering VS Time 13**](#_Toc437255448)

[**Figure 8: Facility distribution 13**](#_Toc437255449)

[**Figure 9: Equipment distribution 14**](#_Toc437255450)

[**Figure 10: Quantity VS Population 14**](#_Toc437255451)

# **ABSTRACT**

Marino Center at Northeastern University is one of the most attractive locations which see thousands of people entering the premises. All the activities ranging from sports like basketball, hockey, Badminton to a magnificent three storey gymnasium are available in the Marino Center.

Marino Center, being one of the hotspots on campus, it seemed a good idea to do analysis on the usage of the facilities there. The purpose of our project is to focus on analyzing the capacity and demand requirements of Marino Center and accordingly optimizing the quantification of facilities provided. For this, we gathered all the information from the staff of Marino Center, number of people entering the center, details of facilities available and how frequently these are occupied by members of Marino Center.

Statistical analysis allows us to understand ways to optimize the facilities. In the project, we will do that by performing calculations and tracking time based data provided by officials at Marino Center. It highlights use of statistical tools like box plot, histograms, hypothesis testing and Minitab results. From these workings, we came to a conclusion of ways to reduce waiting time and improving the capacity-demand difference.

# **METHODOLOGY**

First, with Marino Center staff members, we gathered the data set regarding what facilities are available, number of equipment / playing fields, daily count of people entering the center, number of people entering the center at different time intervals and total capacity of Marino center.

**Our objective is:**

* To perform hypothesis testing on the people count entering Marino Center with a fixed mean μ = 4500. This testing will show us whether Marino Center is working at its full capacity or not.
* Based on the results from hypothesis testing, to give suggestions in order to better the efficiency at which Marino Center is working.

# **P - Diagram:**

* Number of Equipment
* Non University members
* Holidays
* Timings

Controllable Factors

Provide effective training opportunity

Marino Center Optimization

Input

* Number of people
* Failure of Equipment
* External Factors

Uncontrollable Factors

Figure 1: P- Diagram

# **DATA SETS**

MARINO CENTER LAST THREE MONTHS DATA – shows number of people entering premises on daily basis

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Date From** | **Date To** | **Sun** | **Mon** | **Tue** | **Wed** | **Thu** | **Fri** | **Sat** |
| 9/6/2015 | 9/12/2015 | 1811 | 2602 | 3172 | 3732 | 3209 | 2962 | 1894 |
| 9/13/2015 | 9/19/2015 | 1879 | 3834 | 3623 | 3282 | 3127 | 2837 | 1676 |
| 9/20/2015 | 9/26/2015 | 1761 | 3504 | 3311 | 2981 | 2938 | 2447 | 1558 |
| 9/27/2015 | 10/3/2015 | 1473 | 3046 | 3072 | 2693 | 2664 | 2479 | 1597 |
| 10/4/2015 | 10/10/2015 | 1549 | 3013 | 2997 | 2815 | 2571 | 2033 | 1298 |
| 10/11/2015 | 10/17/2015 | 1293 | 2191 | 2800 | 2767 | 2727 | 2290 | 1438 |
| 10/18/2015 | 10/24/2015 | 1465 | 2796 | 2923 | 2628 | 2538 | 2264 | 1534 |
| 10/25/2015 | 10/31/2015 | 1460 | 2838 | 2758 | 2399 | 2350 | 2048 | 1256 |
| 11/1/2015 | 11/7/2015 | 1406 | 2833 | 2760 | 2412 | 2541 | 2362 | 1655 |
| 11/8/2015 | 11/14/2015 | 1481 | 2981 | 2941 | 2593 | 2559 | 2311 | 1702 |
| 11/15/2015 | 11/21/2015 | 1486 | 2870 | 2829 | 2617 | 2479 | 2268 | 1637 |
| 11/22/2015 | 11/28/2015 | 1477 | 2294 | 1675 | 702 | closed | 405 | 536 |
| 11/29/2015 | 12/5/2015 | 1226 | 2771 | 2843 | 2594 | 2552 |  |  |

## **Box Plot Diagram**

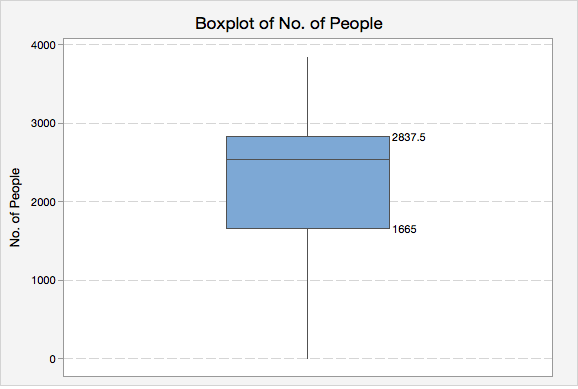
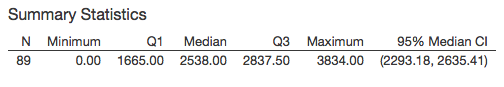


Figure 2: Box Plot Diagram

**

Box and whisker plots are uniform in their use of the box: the bottom and top of the box are always the first and third [quartiles](https://en.wikipedia.org/wiki/Quartile), and the band inside the box is always the second [quartile](https://en.wikipedia.org/wiki/Quartile) (the [median](https://en.wikipedia.org/wiki/Median))

Box Plots are the important graphs used for identifying outliners and for comparing distributions. A box plot is a graphical display that simultaneously describes several important features of data set, such as center, spread, departure from symmetry, and identification of unusual observations or outliners. There are three quartiles, the minimum and the maximum of the data on a rectangular box aligned either horizontally or vertically. Here we can see the mean of the people using Marino center and the variation between them.

## **Time Series Plot**

The graphical representation of Histogram and Box plot does not include an important factor which is time. The Time-series or Time-sequence plot is the graphical representation of observation with respect to when they occur. A Time-Series plot is a graph in which the horizontal axis denotes the time and the vertical axis denotes the observation, the biggest advantage of this graph is that we can see the trends, cycles, or other broad features which we cannot observe otherwise.

Figure 3: Time Series Plot

In this graph we can see the time plot of the people using Marino Center. We see a lot of variations in the amount of people using Marino center for last 3 months. It is true for every facility to be used as per their capacity to make the most use of it, but here due to lots of noise factors like human factors, waiting time and other external uncontrollable factors Marino center is not used fully as per its designed capacity. This is seen clearly from the graph.

# **DESCRIPTIVE ANALYSIS OF PEOPLE USING MARINO CENTER**

We had done hypothesis testing of daily usage of MARINO center for last three months and a descriptive analysis was made

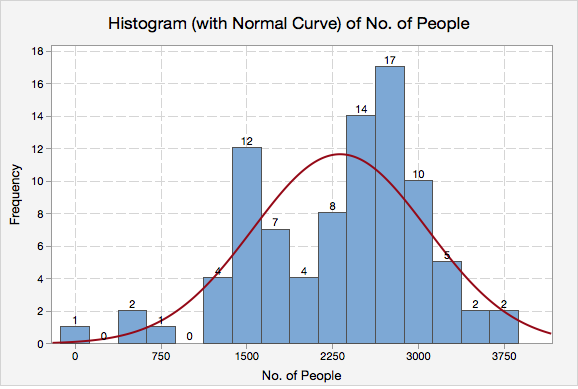
.

Figure 4: Histogram of number of people

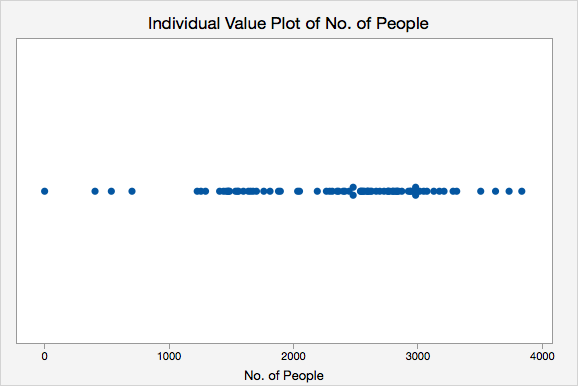
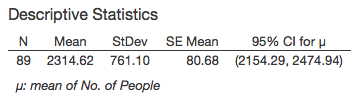
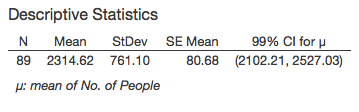
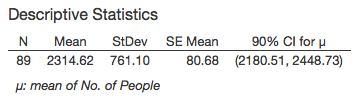


Figure 5: Individual plot of number of people







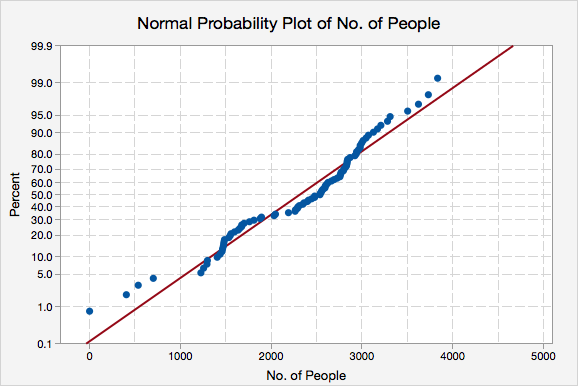
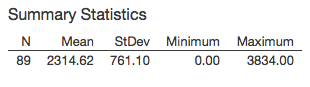
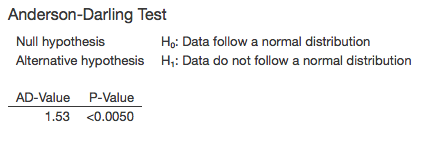


Figure 6: Normal probability plot of number of people





We performed the Anderson-Darling Normality test, from the Histogram and Probability Plot; we conclude that production statistics of PEOPLE USING MARINO CENTER is normally distributed.

# **HYPOTHESIS ANALYSIS:**

We performed the T test for the hypothesis testing on people using MARINO center for keeping their body fit and healthy. We took the mean µ as the daily maximum capacity of MARINO Center which is 4500 people.

H0: µ = 4500

H1: µ ≠ 4500

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| VARIABLE | N | Mean | St. Dev. | SE Mean | 95% CI | T-Value | P Value |
| MARINO | 89 | 2314.62 | 761.10 | 80.68 | (2154.29,2474.94) | -27.09 | <0.0001 |

## **Outcome**

From the Hypothesis we can observe that the p-value is <0.0001 which is less than α=0.05 also after finding the tabulated value for t which equals 2.01, we can see that the calculated is larger than the tabulated, so we reject the null hypothesis. From this we can conclude that MARINO Center is not working as per its capacity. Since the mean and the upper bound of the 95% CI are 2154.29 and 2474.94 people range, we can state that the Marino Center is not fully optimized by the people and improvements are needed in some specific areas so that it can be fully optimized.

## **Result**

From the hypothesis testing we reject the null hypothesis µ = 4500 people and accept the alternative hypothesis µ ≠ 4500. Also we can observe from the test that sample mean is significantly lower that the tested mean, which states that Marino center is not working at its full capacity and need to focus on some areas to optimize it fully.

# **SCOPE OF IMPROVEMENTS:**

Here we have a graph which shows the number of people at Marino Center at various time slots. The working hour for the center is 5:00 AM – 1:00 AM.

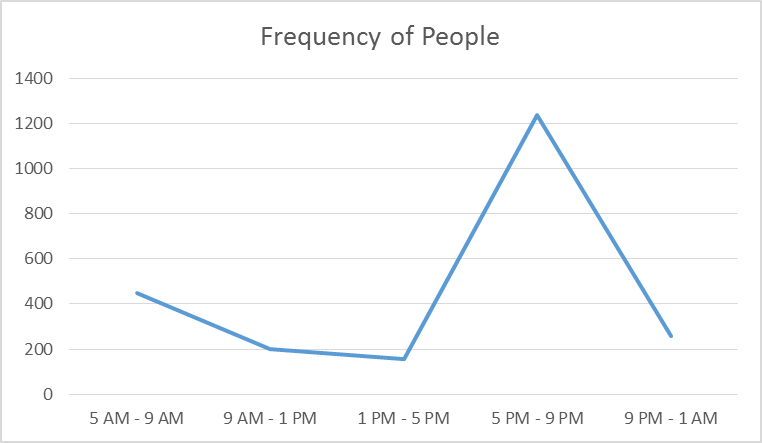


Figure 7: Frequency of people entering VS Time

With the help of the above graph, we can see that 5:00 PM – 9:00 PM time slot is most occupied. Knowing this will help us to make necessary arrangements in the working staff schedule. On contacting the center we came to know that they have an extra team of 5-6 staff members for this time slot.

Here we have a graph which represents the number of facilities and the equipment available at the Marino Centre.

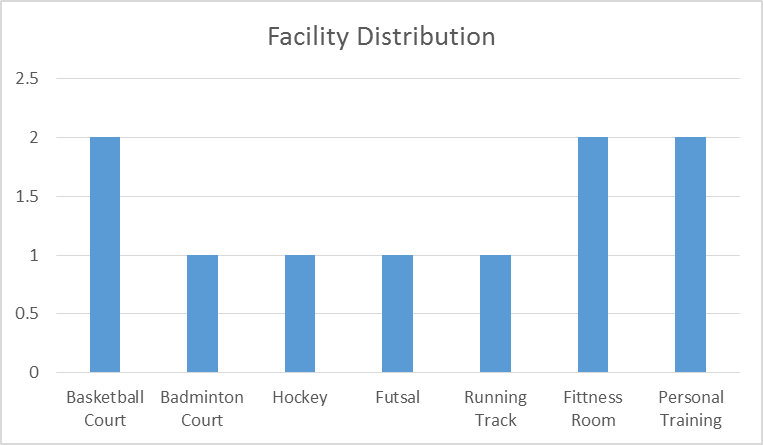


Figure 8: Facility distribution

Figure 9: Equipment distribution

This when compared to the number of people using the equipment is very useful for making various decisions in order to maximize the efficiency of the gym and its services.

Here we can see the comparative analysis of the equipment used with respect the availability of those equipment.

Figure 10: Quantity VS Population

With the help of the above graph we can see that the maximum number of people in the Marino are interested in weight machine, dumbbell section, treadmills and cycles.

On comparing the population using the dumbbell section with the availability for the same, we can see that the availability is very less compared to the demand and when seen for stair master and row boating, it’s the opposite. So a decision can be made to reduce the equipment for stair master and row boating and increase the dumbbell section area. This will increase the efficiency and reduce the waiting time in the dumbbell section.

Various activities such as Group Fitness Workout, Aerobics, Zumba, Cardio Sessions etc. should be marketed more in order to attract people and use the center to its maximum efficiency.

People should be encouraged to use the Marino Center at the times when it is not fully occupied with the help of workshops, sports and various fitness schedule.

# **References**

* + www.wikipedia.com
  + “Applied Statistics and Probability for Engineers” by *Douglas C. Montgomery* & *George C. Runger*
  + Northeastern University, Marino Center