MATH1324 Assignment 2

Comparison in the average length of stay between large hospitals and medium hospitals

Akshat Vijayvargia s3826627

5/10/2020

INTRODUCTION

Australian Institute of Health and Welfare was founded on 5 June'1987 and is a national agency, which collects and provides facts, information and statistics for Australia's health wealth and welfare. In this project, we will look on the average stay of patients, and the average means the single term for the presentation of the data and one of the most used method for central tendency. Average or mean or arithmetic mean tells about the that main value around which other numbers are oscillating. It is conventionally calculated as sum of elements in a set divided by the number of elements.So, the Average length of stay means quotient number of days, around which patients stay in hospitals. If we consider Large hospitals, then we can understand back in the day, large hospitals, like Mass General had several thousands of beds but, just as airliner size is "downsizing" due to the changing economy, I think that hospital size is also changing. Generally hospitals are sorted into a set of dependent variables indicative of organizational formalism and a number of independent variables separately measuring hospital size (number of beds) and various types of complexity commonly associated with size. A series of regressions are performed on the data to determine whether size alone has significant influence. Throughout the investigation, by evaluating two hypothesis, we will try to determine people's preference between large hospitals and medium hospitals.

PROBLEM STATEMENT

This investigation looks to compare the average length of stay of patients in large hospitals and medium hospitals and will try to seek what patients tendency are in terms of hospitals. We can draw conclusions by comparing the average of stay of both the hospitals by firstly using summary () function, sd() function, and IQR() function. Secondly, we can use Box Plot for getting interesting features of the data set. And in the end, we will use two sample t test() function to determine the average stay of patients in both large hospitals and medium hospitals.

DATA

I picked the open data set from Australian Institute of Health and Welfare(https://www.aihw.gov.au/reports-data/myhospitals/sectors/admitted-patients) and then looked for the data named "Average Length Of Stay Multiple Data". Firstly, I will import the data into R, through read_excel() function and then will convert every NP value to NA for the usage of na.rm. Secondly, will add filters on the 'Peer Group' by %>% filter() function and will assign a new name for Large hospitals and Medium hospitals, ALS_LARGE AND ALS_MEDIUM respectively for better understanding and usability. While importing the dataset into R, I noticed that these columns have wrong class, so we will turn the character class column to numeric class column by using as.numeric() function because if we don't they will disrupt our working.

DESCRIPTIVE STATISTICS AND VISUALISATION

While reprocessing the data for hypothesis testing, I noticed that some columns are not in the correct datatype. To change their datatype from character to numeric, I used as.numeric()function. And for the NP's, I converted them to NA's. After making the data ready for the summary statistics, we will now firstly filter out the Large hospitals and Medium hospitals by naming them ALS_LARGE and ALS_MEDIUM respectively. Secondly, we will use the summary() function, sd() function and IQR() function for Large hospitals with na.rm=TRUE. Thirdly, we will use all the functions for the Medium hospitals with na.rm=TRUE. Fourthly, we will use all this data for visualization by using box-plot under ggplot2.

Slide with R Output

```
#Load Packages

library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readxl)
library(readr)
library(magrittr)
library(ggplot2)


#Read the dataset
library(readxl)
ALS <- read_excel("AA/average-length-of-stay-multilevel-data.xlsx",
    skip = 12)
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting logical in O30023 / R30023C15: got 'â€¡'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting logical in O30024 / R30024C15: got 'â€¡'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting logical in O30025 / R30025C15: got 'â€¡'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting logical in O30026 / R30026C15: got 'â€¡'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting logical in O30027 / R30027C15: got 'â€¡'
```

```
## Warning in read_fun(path = enc2native(normalizePath(path)), sheet_i = sheet, :
## Expecting logical in O30028 / R30028C15: got 'â€¡'
```

```
## New names:
## * `` -> ...9
## * `` -> ...11
## * `` -> ...13
## * `` -> ...15
## * `` -> ...17
## * ...
```

```
# Converting NPs to NAs

ALS[ALS=="NP"]<-NA

# Separating Large and Medium hospitals

ALS_LARGE<-ALS %>% filter(`Peer group`=="Large hospitals")
ALS_MEDIUM<-ALS %>% filter(`Peer group`=="Medium hospitals")

# Summary Statistics of Large hospitals
```

```
ALS_LARGE$`Average length of stay (days)`<-as.numeric(ALS_LARGE$`Average length of stay (days)`)
```

```
## Warning: NAs introduced by coercion
```

```
summary(ALS_LARGE$`Average length of stay (days)`, na.rm=TRUE)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.200   2.500   3.500   3.987   5.000  12.600    1281
```

```
'Standard Deviation of Large hospitals'
```

```
## [1] "Standard Deviation of Large hospitals"
```

```
sd(ALS_LARGE$`Average length of stay (days)`, na.rm=TRUE)
```

```
## [1] 1.978679
```

```
'Inter-Quartile of Large hospitals'
```

```
## [1] "Inter-Quartile of Large hospitals"
```

```
IQR(ALS_LARGE$`Average length of stay (days)`, na.rm=TRUE)
```

```
## [1] 2.5
```

```
# Summary Statistics of  Medium hospitals

ALS_MEDIUM$`Average length of stay (days)`<-as.numeric(ALS_MEDIUM$`Average length of stay (days)`)
```

```
## Warning: NAs introduced by coercion
```

```
summary(ALS_MEDIUM$`Average length of stay (days)`, na.rm=TRUE)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   2.400   3.400   3.706   4.500  13.200    1695
```

```
'Standard Deviation of Medium hospitals'
```

```
## [1] "Standard Deviation of Medium hospitals"
```

```
sd(ALS_MEDIUM$`Average length of stay (days)`, na.rm=TRUE)
```

```
## [1] 1.85253
```

```
'Inter-Quartile of Medium hospitals'
```

```
## [1] "Inter-Quartile of Medium hospitals"
```

```
IQR(ALS_MEDIUM$`Average length of stay (days)`, na.rm=TRUE)
```

```
## [1] 2.1
```

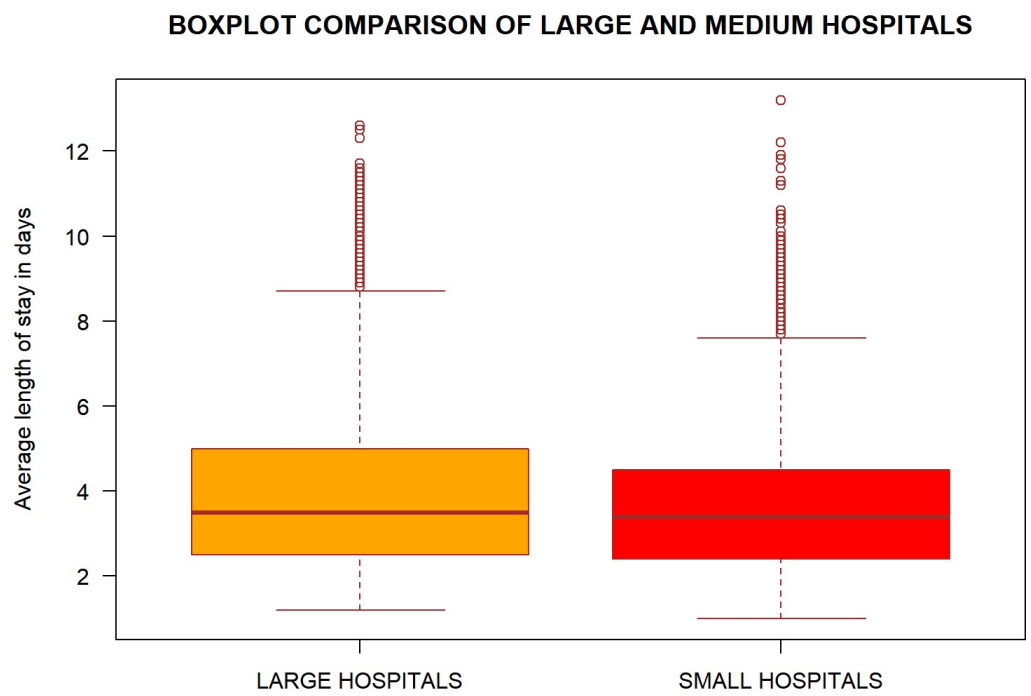
```r
#After calculating Summary(), Sd() and IQR() of both the Large hospitals and medium hospitals. Now, our data is set for
  visualization. We will have to convert the Average length of stay column to numerics and that can be done by
  as.numeric() function.


ALS$`Average length of stay (days)`<-as.numeric(ALS$`Average length of stay (days)`)
```

```
## Warning: NAs introduced by coercion
```

```r
#I have created box plot from the above imported and converted data to actually visually see, whether Large hospital or
  Medium hospital has more patients and their length of stay.

boxplot(ALS_LARGE$`Average length of stay (days)`,ALS_MEDIUM$`Average length of stay (days)`,
main = "BOXPLOT COMPARISON OF LARGE AND MEDIUM HOSPITALS",
ylab="Average length of stay in days",
at = c(1,2),
names = c("LARGE HOSPITALS","SMALL HOSPITALS"),
las = 1,
by=1,
col = c("orange","red"),
border = "brown",
horizontal = FALSE,
notch = FALSE
)
```




```r
# According to the boxplot and summary statistics, the Average length of stay in days of Large hospitals is slightly
  greater than the Average length of stay in days of Medium hospitals.
```

As of now, with the box-plot visualization, we can see that the patients prefer Large hospital than Medium hospital, we can vaguely say that box of Average length of stay for Large hospitals is bit higher than Average length of stay for Medium hospitals. We can be more precise after hypothesis testing.

HYPOTHESIS TESTING

In this step, we will perform hypothesis testing and will try to know which hospital's has more average length of stay of patients. For the our testing, we will use two sample t test() function as we have two separate tables and variable namely 'Large hospitals' and 'Medium hospitals'. Null hypothesis of both Large hospital and Medium hospital= H0 Alternative hypothesis of both Large hospital and Medium hospital= HA

```
##
##  Two Sample t-test
##
## data:  ALS_LARGE$`Average length of stay (days)` and ALS_MEDIUM$`Average length of stay (days)`
## t = 5.5369, df = 6591, p-value = 3.197e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1813999 0.3802485
## sample estimates:
## mean of x mean of y
##  3.986874  3.706049
```

In the hypothesis test, we use two sample t test() function to test the average length of stay of patients in large and medium hospitals. The result of the test, shows our p-value is 3.197e-08 or 0.00000003197 and 95% confidence interval is from 0.1813999 to 0.3802485. According to the result, we know that p-value is less than our significance level which is .05alpha and 95% confidence interval misses our null hypothesis(H0). Therefore, we can reject null hypothesis(H0). According to the test result above, we can say that the average length of stay is more of Large hospitals than Medium hospitals. Moreover, our box-plot of average length of stay of both the hospitals also shows that Average length of stay in Medium hospital is less than Average length of stay in Large hospital.

DISCUSSION

In this investigation, we found that average stay and patient's preference is more inclined towards the Large hospitals than the Medium hospitals.

The strength of my investigation is that I picked the data from AIHW(Australian Institute of Health and Welfare), which is a Government statutory agency and jurisdiction of Commonwealth of Australia, so my dataset is pure and authentic. In addition, this data is a part of huge dataset recorded for admitted patients between 2017 to 2018, so they can provide the brief reflection of length of stay of people.

The limitations of the investigation is that I am restricted to only one factor i.e Average length of stay(days), but to evaluate the number of patients and their stay includes price difference, quality of service, behavior of doctors etc.Therefore my investigation may not fully say which hospital has more average length of stay.

The direction for future investigation is collecting and considering more data and more variables or factors to know the truth or at least get closer to it.

REFERENCES

1)DATASET OF AVERAGE LENGTH OF STAY ->https://www.aihw.gov.au/reports-data/myhospitals/sectors/admitted-patients 2)INFORMATION OF AIHW ->https://en.wikipedia.org/wiki/Australian_Institute_of_Health_and_Welfare 3)MODULE 7 FOR CODING ->https://astral-theory-157510.appspot.com/secured/MATH1324_Module_07.html#paired_samples_(t)-test_visualisation 4)INFORMATION OF AIHW ->https://www.aihw.gov.au/reports/australias-welfare/australias-welfare-2017/contents/table-of-contents 5)HOW TO CLASSIFY LARGE AND MEDIUM HOSPITALS ->https://allnurses.com/what-considered-large-hospital-t309545/ 6)HOW TO CLASSIFY LARGE AND MEDIUM HOSPITALS ->https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1067287/