Length of Stay in Hospitals

Code

Anurag Dinesh Karmarkar

Introduction:

The problem focuses on the statistics of large and medium hospitals available with Australian Institute of Health and Welfare (AIHW). The dataset is about the average length of stay of people in the hospitals. The data can be used to derive a focussed analysis based on what relation the hospitals' peer group has on the basis average length of stay in the hospital in terms of number of days.

Hide

```
View(alos_data)
```

Hide

```
library(dplyr)
```

```
Registered S3 method overwritten by 'dplyr':
  method           from
  print.rowwise_df

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
```

Hide

```
colnames(alos_data)
```

```
 [1] "Reporting unit"                   "Reporting unit type"              "State"
 [4] "Local Hospital Network (LHN)"     "Peer group"                      "Time period"
 [7] "Category"                         "Total number of stays"           "Number of overnight
stays"
[10] "Percentage of overnight stays"    "Average length of stay (days)"   "Peer group average
(days)"
[13] "Total overnight patient bed days"
```

Hide

```
#Renaming Columns
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Reporting unit"] <- "reporting_unit"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Reporting unit type"] <- "reporting_unit_type"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"state"] <- "state"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Local Hospital Network (LHN)"] <- "local_hospital_network"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Peer group"] <- "peer_group"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Time period"] <- "time_period"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Category"] <- "category"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Total number of stays"] <- "total_number_of_stays"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Number of overnight stays"] <- "number_of_overnight_stays"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Percentage of overnight stays"] <- "percentage_of_overnight_stays"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Average length of stay (days)"] <- "alos_days"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Peer group average (days)"] <- "peer_group_average_days"
names(average_length_of_stay_multilevel_data)[names(average_length_of_stay_multilevel_data) ==
"Total overnight patient bed days"] <- "Total_overnight_patient_bed_days"
```

Hide

```
View(average_length_of_stay_multilevel_data)
```

Hide

```
alos_data_filtered <- select(average_length_of_stay_multilevel_data,State,local_hospital_networ
k,peer_group,time_period,category,alos_days)
```

Hide

```
View(alos_data_filtered)
```

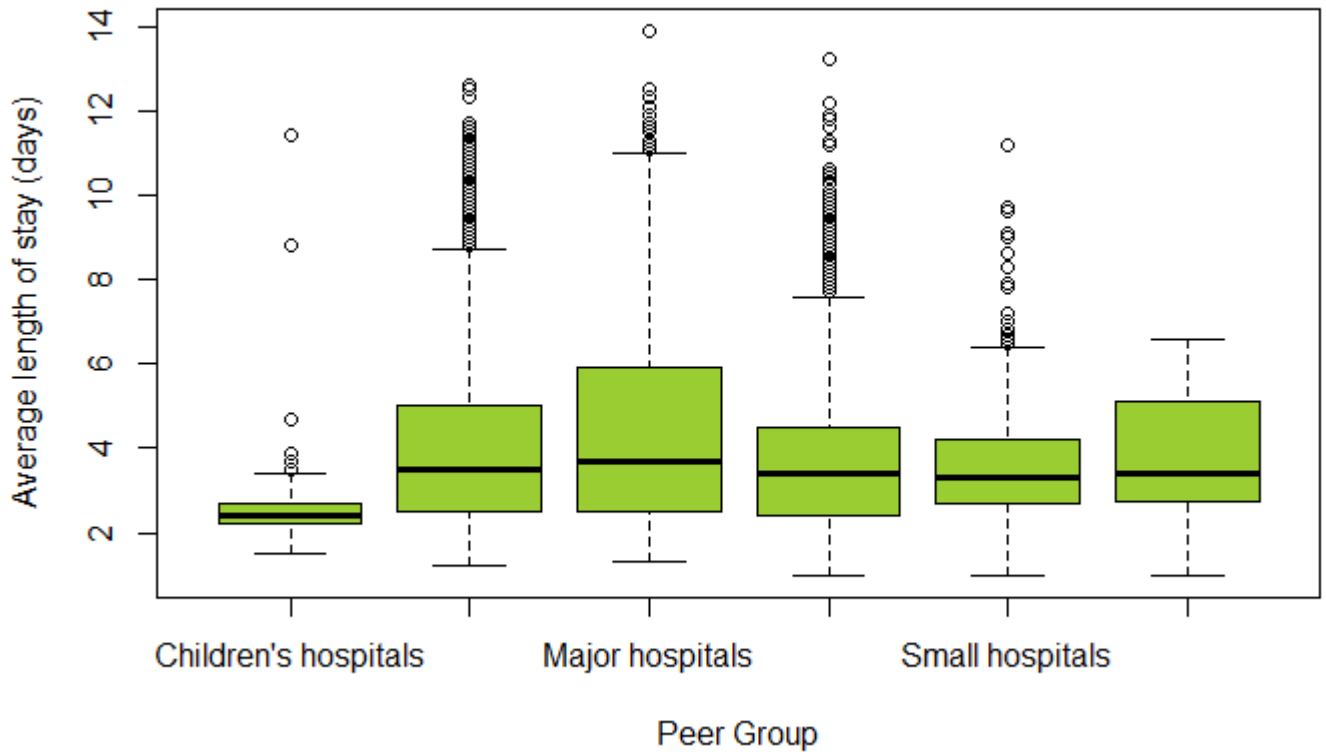
Hide

```
#Deleting alos values with np and -

alos_data_filtered<-alos_data_filtered[!(alos_data_filtered$alos_days=="NP"),]
alos_data_filtered<-alos_data_filtered[!(alos_data_filtered$alos_days=="-"),]
View(alos_data_filtered)
```

Hide

```
alos_data_filtered %>% boxplot(as.numeric(alos_days) ~ peer_group, data = .,col = "yellowgreen",
ylab="Average length of stay (days)",xlab="Peer Group") #Side-by-side boxplot
```



Hide

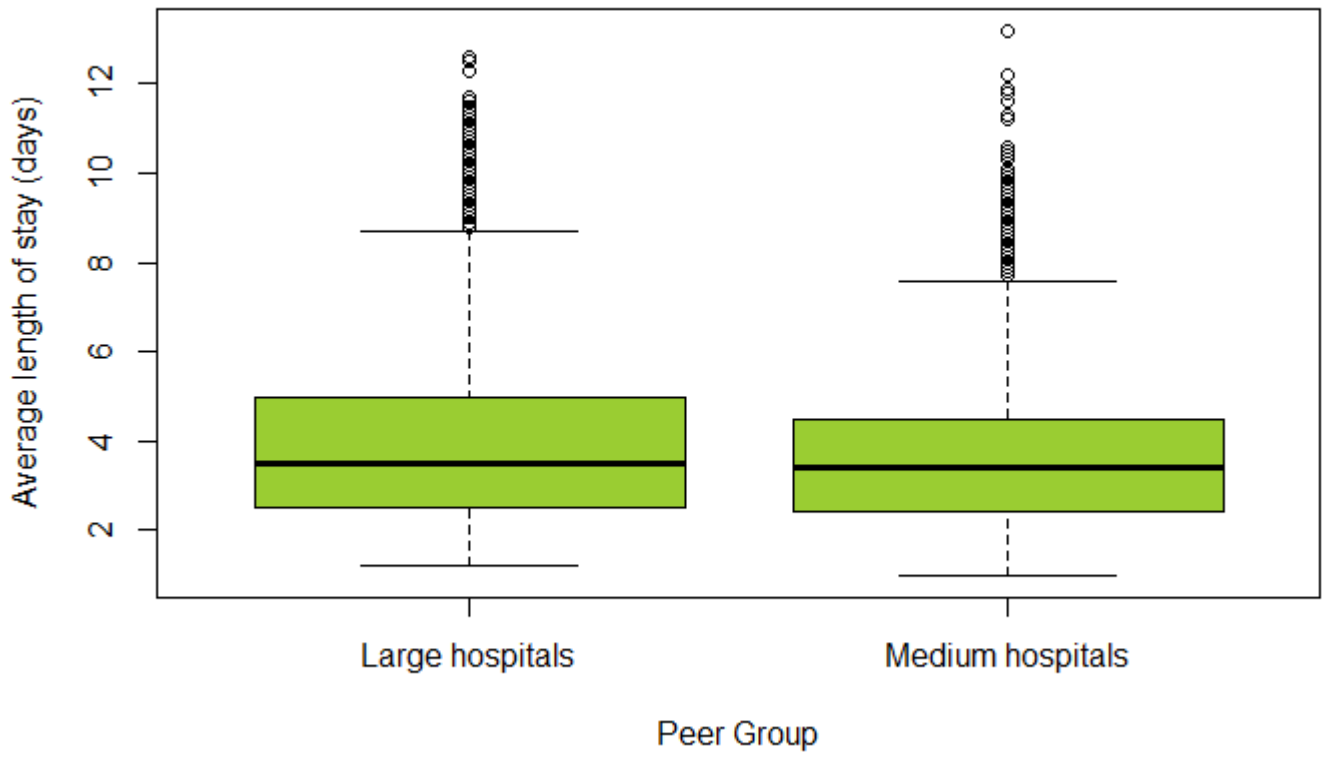
```
NA
```

Hide

```
alos_data_filtered2 <- alos_data_filtered[(alos_data_filtered$peer_group=="Large hospitals") |
 (alos_data_filtered$peer_group=="Medium hospitals"),]
```

Hide

```
alos_data_filtered2 %>% boxplot(as.numeric(alos_days) ~ peer_group, data = .,col = "yellowgreen"
, ylab="Average length of stay (days)",xlab="Peer Group") #Side-by-side boxplot
```



Hide

```
#alos_data_filtered2 <- transform(alos_data_filtered2, alos_days = as.numeric(alos_days))
View(alos_data_filtered2)
```

Hide

```
alos_data_filtered2 %>% group_by(peer_group) %>% summarise(Min = min(alos_days,na.rm = TRUE),
Q1 = quantile(alos_days,probs = .25,na.rm = TRUE),
Median = median(alos_days, na.rm = TRUE),
Q3 = quantile(alos_days,probs = .75,na.rm = TRUE),
Max = max(alos_days,na.rm = TRUE),
Mean = mean(alos_days, na.rm = TRUE),
SD = sd(alos_days, na.rm = TRUE),
n = n(),
Missing = sum(is.na(alos_days)))
```

peer_group	…	Q1	Median	Q3	Max	Mean	SD	n	Missing
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>	<int>
Large hospitals	1.2	2.5	3.5	5.0	12.6	3.986874	1.978679	4411	0
Medium hospitals	1.0	2.4	3.4	4.5	13.2	3.706049	1.852530	2182	0

2 rows

Hide

```
3.986874 - 3.706049
```

```
[1] 0.280825
```

Hide

```
install.packages("car")
```

```
WARNING: Rtools is required to build R packages but is not currently installed. Please download
and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 愀牶C:/Users/smart/OneDrive/Documents/R/win-library/3.6愀牷
(as 愀牶lib愀牷 is unspecified)
also installing the dependencies 愀牶forcats愀牷, 愀牶zip愀牷, 愀牶SparseM愀牷, 愀牶MatrixModels愀牷, 愀牶sp愀牷, 愀牶haven愀牷, 愀牶data.table愀牷, 愀牶openxlsx愀牷, 愀牶minqa愀牷, 愀牶nloptr愀牷, 愀牶statmod愀牷, 愀牶RcppEigen愀牷, 愀牶carData愀牷, 愀牶abind愀牷, 愀牶pbkrtest愀牷, 愀牶quantreg愀牷, 愀牶maptools愀牷, 愀牶rio愀牷, 愀牶lme4愀牷

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/forcats_0.5.0.zip'
Content type 'application/zip' length 356671 bytes (348 KB)
downloaded 348 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/zip_2.0.4.zip'
Content type 'application/zip' length 443372 bytes (432 KB)
downloaded 432 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/SparseM_1.78.zip'
Content type 'application/zip' length 1070197 bytes (1.0 MB)
downloaded 1.0 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/MatrixModels_0.4-1.zip'
Content type 'application/zip' length 356874 bytes (348 KB)
downloaded 348 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/sp_1.4-1.zip'
Content type 'application/zip' length 1878758 bytes (1.8 MB)
downloaded 1.8 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/haven_2.2.0.zip'
Content type 'application/zip' length 1044795 bytes (1020 KB)
downloaded 1020 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/data.table_1.12.8.zip'
Content type 'application/zip' length 2276913 bytes (2.2 MB)
downloaded 2.2 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/openxlsx_4.1.4.zip'
Content type 'application/zip' length 2574746 bytes (2.5 MB)
downloaded 2.5 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/minqa_1.2.4.zip'
Content type 'application/zip' length 674539 bytes (658 KB)
downloaded 658 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/nloptr_1.2.2.1.zip'
Content type 'application/zip' length 1079056 bytes (1.0 MB)
downloaded 1.0 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/statmod_1.4.34.zip'
Content type 'application/zip' length 285605 bytes (278 KB)
```

```
downloaded 278 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/RcppEigen_0.3.3.7.0.zip'
Content type 'application/zip' length 2678965 bytes (2.6 MB)
downloaded 2.6 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/carData_3.0-3.zip'
Content type 'application/zip' length 1817853 bytes (1.7 MB)
downloaded 1.7 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/abind_1.4-5.zip'
Content type 'application/zip' length 63845 bytes (62 KB)
downloaded 62 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/pbkrtest_0.4-8.6.zip'
Content type 'application/zip' length 275841 bytes (269 KB)
downloaded 269 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/quantreg_5.55.zip'
Content type 'application/zip' length 1782151 bytes (1.7 MB)
downloaded 1.7 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/maptools_0.9-9.zip'
Content type 'application/zip' length 2171410 bytes (2.1 MB)
downloaded 2.1 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/rio_0.5.16.zip'
Content type 'application/zip' length 505276 bytes (493 KB)
downloaded 493 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/lme4_1.1-23.zip'
Content type 'application/zip' length 5708612 bytes (5.4 MB)
downloaded 5.4 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/car_3.0-7.zip'
Content type 'application/zip' length 1556446 bytes (1.5 MB)
downloaded 1.5 MB
```

```
package ‘forcats’ successfully unpacked and MD5 sums checked
package ‘zip’ successfully unpacked and MD5 sums checked
package ‘SparseM’ successfully unpacked and MD5 sums checked
package ‘MatrixModels’ successfully unpacked and MD5 sums checked
package ‘sp’ successfully unpacked and MD5 sums checked
package ‘haven’ successfully unpacked and MD5 sums checked
package ‘data.table’ successfully unpacked and MD5 sums checked
package ‘openxlsx’ successfully unpacked and MD5 sums checked
package ‘minqa’ successfully unpacked and MD5 sums checked
package ‘nloptr’ successfully unpacked and MD5 sums checked
package ‘statmod’ successfully unpacked and MD5 sums checked
package ‘RcppEigen’ successfully unpacked and MD5 sums checked
package ‘carData’ successfully unpacked and MD5 sums checked
package ‘abind’ successfully unpacked and MD5 sums checked
package ‘pbkrtest’ successfully unpacked and MD5 sums checked
package ‘quantreg’ successfully unpacked and MD5 sums checked
package ‘maptools’ successfully unpacked and MD5 sums checked
package ‘rio’ successfully unpacked and MD5 sums checked
package ‘lme4’ successfully unpacked and MD5 sums checked
package ‘car’ successfully unpacked and MD5 sums checked

The downloaded binary packages are in
    C:\Users\smart\AppData\Local\Temp\Rtmpye98MS\downloaded_packages
```

Hide

```
library(car)
```

```
Loading required package: carData
Registered S3 method overwritten by 'data.table':
  method           from
  print.data.table

Attaching package: 勘笣car勘笮

The following object is masked from 勘笣package:dplyr勘笮:

    recode
```

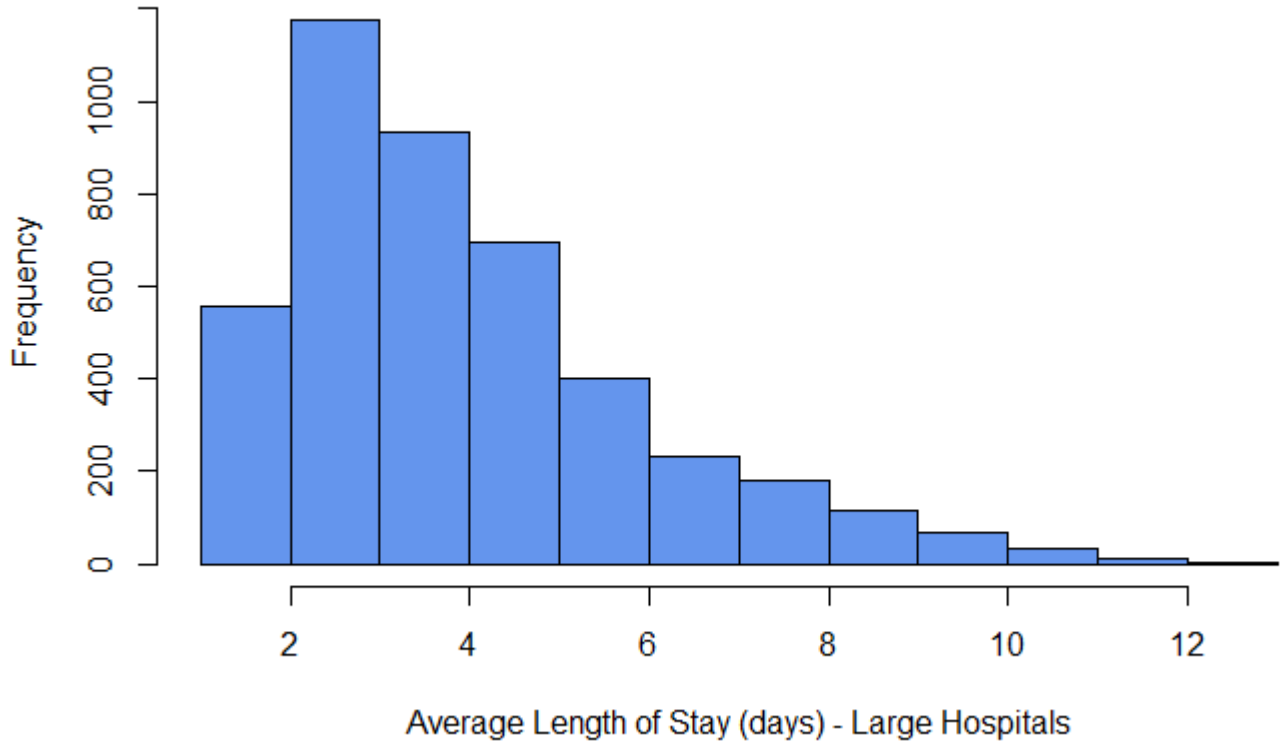
Hide

```
alos_data_lh <- alos_data_filtered2[alos_data_filtered2$peer_group %in% "Large hospitals",]
View(alos_data_lh)

alos_data_mh <- alos_data_filtered2[alos_data_filtered2$peer_group %in% "Medium hospitals",]
View(alos_data_mh)
```

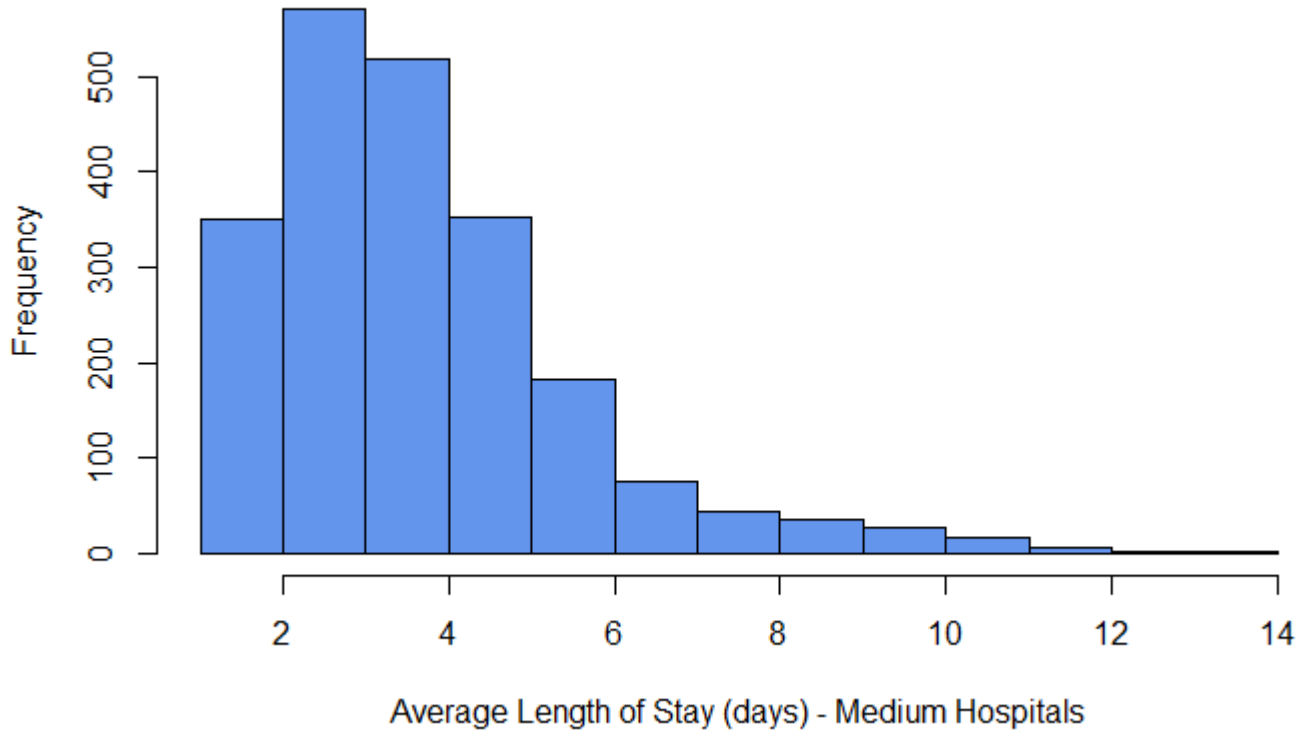
Hide

```
alos_data_lh$alos_days %>% hist(xlab = "Average Length of Stay (days) - Large Hospitals",col =
"cornflowerblue" , main = "")
```



Hide

```
alos_data_mh$alos_days %>% hist(xlab = "Average Length of Stay (days) - Medium Hospitals",col =
"cornflowerblue" , main = "")
```

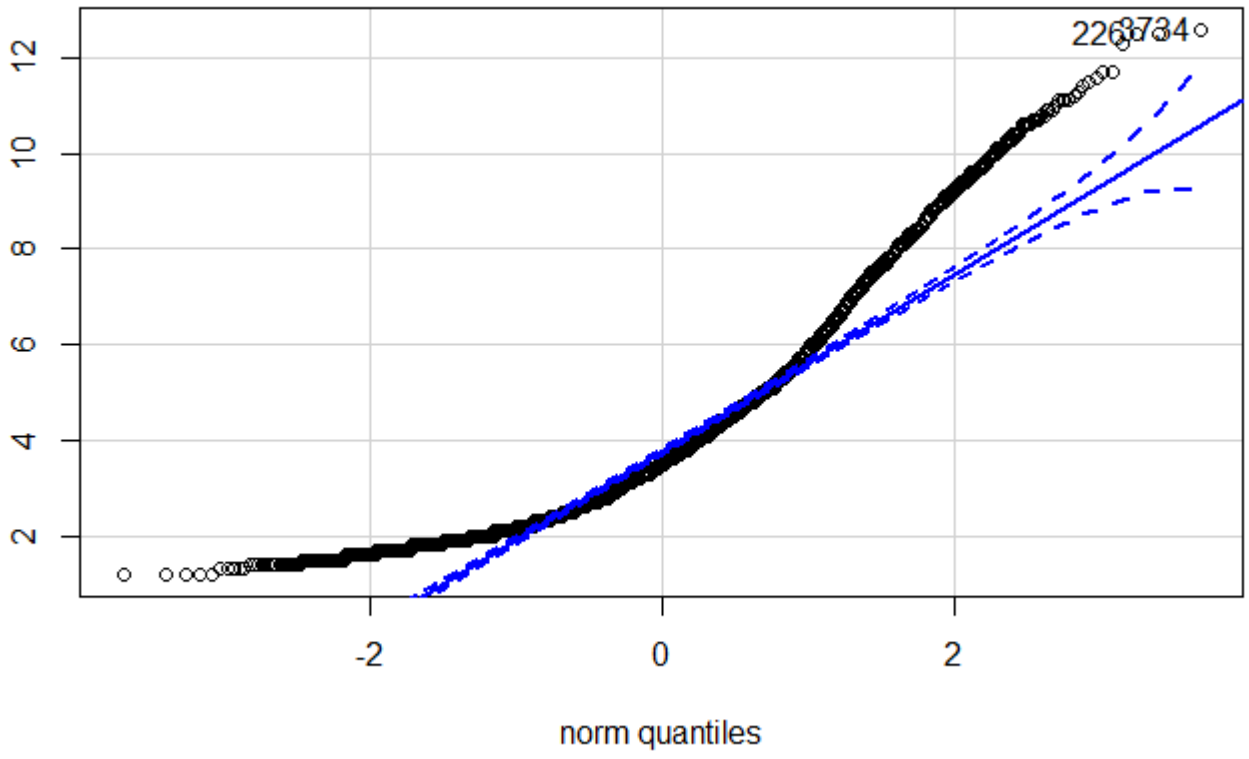


Start hypothesis test

Hide

```
alos_data_lh$alos_days %>% qqPlot(dist="norm")
```

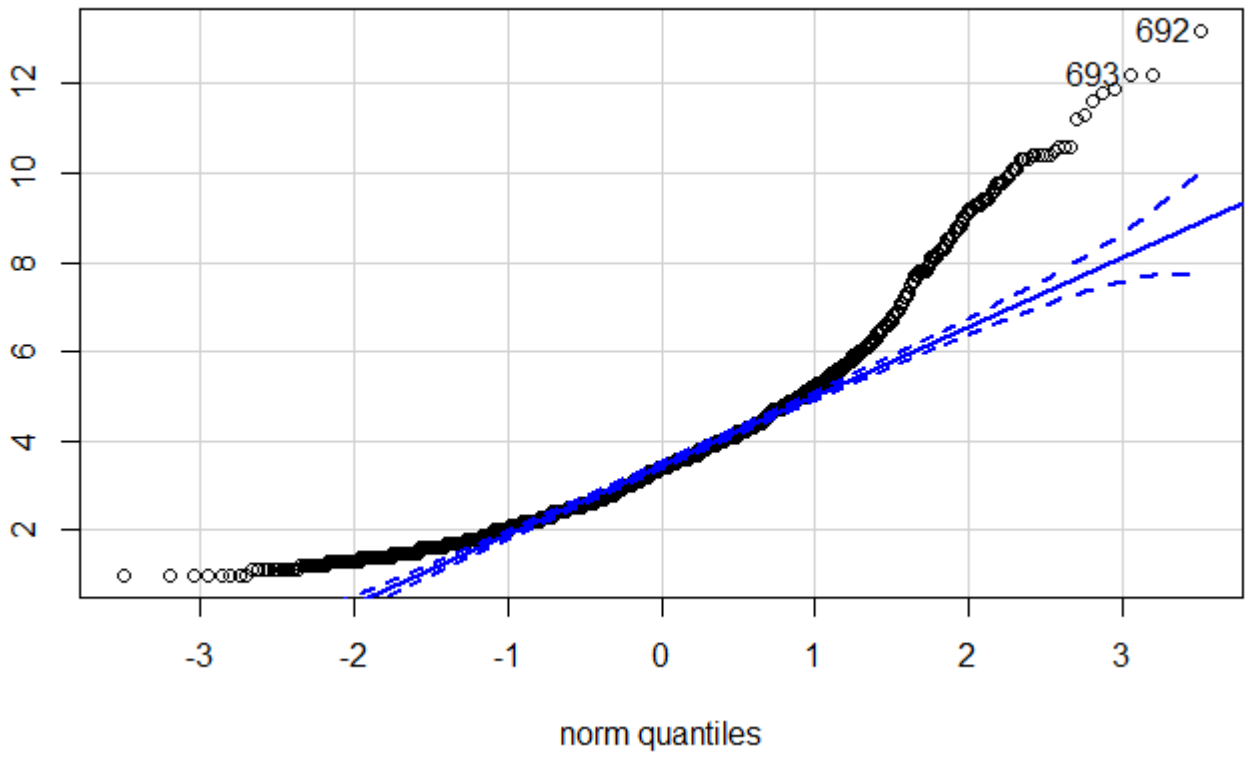
```
[1] 3734  226
```



Hide

```
alos_data_mh$alos_days %>% qqPlot(dist="norm")
```

```
[1] 692 693
```



As many observations lie outside the 95% CI level, we move to Levene's Test

Hide

```
leveneTest(alos_days ~ peer_group, data = alos_data_filtered2)
```

```
group coerced to factor.
```

```
Levene's Test for Homogeneity of Variance (center = median)
        Df F value    Pr(>F)
group    1  16.585 4.707e-05 ***
      6591
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The equal variance cannot be assumed as the probability value is < 0.05

Now we perform the Welch two-sample test

Hide

```
t.test(alos_days ~ peer_group,
       data = alos_data_filtered2,
       var.equal = FALSE,
       alternative = "two.sided")
```

```
    Welch Two Sample t-test

data:  alos_days by peer_group
t = 5.6615, df = 4611, p-value = 1.592e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1835797 0.3780687
sample estimates:
 mean in group Large hospitals mean in group Medium hospitals
                      3.986874                       3.706049
```

Two-sample -test result summary:

We assumed normality as no of samples in both peer groups > 30 Levene Test, p<0.05 variances are not homogeneous Diff bet means 0.280825 95% CI[0.1835797 0.3780687] p value = 1.592e-08, p < alpha

Decision Reject H0 (H0 was that the means are equal)

Conclusion The results of the study found a statistically significant mean difference between large and medium hospitals, t(df=4611) = 5.6615, p=1.592e-08, 95% CI for the difference in means [0.1835797 0.3780687].