Date:01/07/2022

Name: Karthik Murthy

Data Science with R Programming

Project No 7:Healthcare Cost Analysis

DESCRIPTION

Background and Objective:

A nationwide survey of hospital costs conducted by the US Agency for Healthcare consists of hospital records of inpatient samples. The given data is restricted to the city of Wisconsin and relates to patients in the age group 0-17 years. The agency wants to analyze the data to research on healthcare costs and their utilization.

**Domain:** Healthcare

**Dataset Description:**

Here is a detailed description of the given dataset:

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| Age | Age of the patient discharged |
| Female | A binary variable that indicates if the patient is female |
| Los | Length of stay in days |
| Race | Race of the patient (specified numerically) |
| Totchg | Hospital discharge costs |
| Aprdrg | All Patient Refined Diagnosis Related Groups |

**Analysis to be done:**

**1. To record the patient statistics, the agency wants to find the age category of people who frequently visit the hospital and has the maximum expenditure.**

**2. In order of severity of the diagnosis and treatments and to find out the expensive treatments, the agency wants to find the diagnosis-related group that has maximum hospitalization and expenditure.**

**3. To make sure that there is no malpractice, the agency needs to analyze if the race of the patient is related to the hospitalization costs.**

**4. To properly utilize the costs, the agency has to analyze the severity of the hospital costs by age and gender for the proper allocation of resources.**

**5. Since the length of stay is the crucial factor for inpatients, the agency wants to find if the length of stay can be predicted from age, gender, and race.**

**6. To perform a complete analysis, the agency wants to find the variable that mainly affects hospital costs.**

Project Code & Output

Karthik

2022-06-28

library(readxl)  
 Hospital <- read\_excel("C:/Users/HP/Downloads/Hospital.xlsx")  
View(Hospital)  
library(dplyr) ## For Data Manipulation

##   
## 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(ggplot2) # For Data visualization  
library(data.table) ## For Data manipulation

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

library(psych)# For addditional summary Statistics

##   
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

glimpse(Hospital)

## Rows: 500  
## Columns: 6  
## $ AGE <dbl> 17, 17, 17, 17, 17, 17, 17, 16, 16, 17, 17, 15, 15, 15, 15, 15,…  
## $ FEMALE <dbl> 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, …  
## $ LOS <dbl> 2, 2, 7, 1, 1, 0, 4, 2, 1, 2, 2, 2, 2, 4, 7, 4, 1, 4, 3, 3, 1, …  
## $ RACE <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …  
## $ TOTCHG <dbl> 2660, 1689, 20060, 736, 1194, 3305, 2205, 1167, 532, 1363, 1245…  
## $ APRDRG <dbl> 560, 753, 930, 758, 754, 347, 754, 754, 753, 758, 758, 753, 751…

colnames(Hospital)

## [1] "AGE" "FEMALE" "LOS" "RACE" "TOTCHG" "APRDRG"

rownames(Hospital)

## [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12"   
## [13] "13" "14" "15" "16" "17" "18" "19" "20" "21" "22" "23" "24"   
## [25] "25" "26" "27" "28" "29" "30" "31" "32" "33" "34" "35" "36"   
## [37] "37" "38" "39" "40" "41" "42" "43" "44" "45" "46" "47" "48"   
## [49] "49" "50" "51" "52" "53" "54" "55" "56" "57" "58" "59" "60"   
## [61] "61" "62" "63" "64" "65" "66" "67" "68" "69" "70" "71" "72"   
## [73] "73" "74" "75" "76" "77" "78" "79" "80" "81" "82" "83" "84"   
## [85] "85" "86" "87" "88" "89" "90" "91" "92" "93" "94" "95" "96"   
## [97] "97" "98" "99" "100" "101" "102" "103" "104" "105" "106" "107" "108"  
## [109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"  
## [121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"  
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143" "144"  
## [145] "145" "146" "147" "148" "149" "150" "151" "152" "153" "154" "155" "156"  
## [157] "157" "158" "159" "160" "161" "162" "163" "164" "165" "166" "167" "168"  
## [169] "169" "170" "171" "172" "173" "174" "175" "176" "177" "178" "179" "180"  
## [181] "181" "182" "183" "184" "185" "186" "187" "188" "189" "190" "191" "192"  
## [193] "193" "194" "195" "196" "197" "198" "199" "200" "201" "202" "203" "204"  
## [205] "205" "206" "207" "208" "209" "210" "211" "212" "213" "214" "215" "216"  
## [217] "217" "218" "219" "220" "221" "222" "223" "224" "225" "226" "227" "228"  
## [229] "229" "230" "231" "232" "233" "234" "235" "236" "237" "238" "239" "240"  
## [241] "241" "242" "243" "244" "245" "246" "247" "248" "249" "250" "251" "252"  
## [253] "253" "254" "255" "256" "257" "258" "259" "260" "261" "262" "263" "264"  
## [265] "265" "266" "267" "268" "269" "270" "271" "272" "273" "274" "275" "276"  
## [277] "277" "278" "279" "280" "281" "282" "283" "284" "285" "286" "287" "288"  
## [289] "289" "290" "291" "292" "293" "294" "295" "296" "297" "298" "299" "300"  
## [301] "301" "302" "303" "304" "305" "306" "307" "308" "309" "310" "311" "312"  
## [313] "313" "314" "315" "316" "317" "318" "319" "320" "321" "322" "323" "324"  
## [325] "325" "326" "327" "328" "329" "330" "331" "332" "333" "334" "335" "336"  
## [337] "337" "338" "339" "340" "341" "342" "343" "344" "345" "346" "347" "348"  
## [349] "349" "350" "351" "352" "353" "354" "355" "356" "357" "358" "359" "360"  
## [361] "361" "362" "363" "364" "365" "366" "367" "368" "369" "370" "371" "372"  
## [373] "373" "374" "375" "376" "377" "378" "379" "380" "381" "382" "383" "384"  
## [385] "385" "386" "387" "388" "389" "390" "391" "392" "393" "394" "395" "396"  
## [397] "397" "398" "399" "400" "401" "402" "403" "404" "405" "406" "407" "408"  
## [409] "409" "410" "411" "412" "413" "414" "415" "416" "417" "418" "419" "420"  
## [421] "421" "422" "423" "424" "425" "426" "427" "428" "429" "430" "431" "432"  
## [433] "433" "434" "435" "436" "437" "438" "439" "440" "441" "442" "443" "444"  
## [445] "445" "446" "447" "448" "449" "450" "451" "452" "453" "454" "455" "456"  
## [457] "457" "458" "459" "460" "461" "462" "463" "464" "465" "466" "467" "468"  
## [469] "469" "470" "471" "472" "473" "474" "475" "476" "477" "478" "479" "480"  
## [481] "481" "482" "483" "484" "485" "486" "487" "488" "489" "490" "491" "492"  
## [493] "493" "494" "495" "496" "497" "498" "499" "500"

#Descriptive Statistics  
dim(Hospital)

## [1] 500 6

summary(Hospital)

## AGE FEMALE LOS RACE   
## Min. : 0.000 Min. :0.000 Min. : 0.000 Min. :1.000   
## 1st Qu.: 0.000 1st Qu.:0.000 1st Qu.: 2.000 1st Qu.:1.000   
## Median : 0.000 Median :1.000 Median : 2.000 Median :1.000   
## Mean : 5.086 Mean :0.512 Mean : 2.828 Mean :1.078   
## 3rd Qu.:13.000 3rd Qu.:1.000 3rd Qu.: 3.000 3rd Qu.:1.000   
## Max. :17.000 Max. :1.000 Max. :41.000 Max. :6.000   
## NA's :1   
## TOTCHG APRDRG   
## Min. : 532 Min. : 21.0   
## 1st Qu.: 1216 1st Qu.:640.0   
## Median : 1536 Median :640.0   
## Mean : 2774 Mean :616.4   
## 3rd Qu.: 2530 3rd Qu.:751.0   
## Max. :48388 Max. :952.0   
##

str(Hospital)

## tibble [500 × 6] (S3: tbl\_df/tbl/data.frame)  
## $ AGE : num [1:500] 17 17 17 17 17 17 17 16 16 17 ...  
## $ FEMALE: num [1:500] 1 0 1 1 1 0 1 1 1 1 ...  
## $ LOS : num [1:500] 2 2 7 1 1 0 4 2 1 2 ...  
## $ RACE : num [1:500] 1 1 1 1 1 1 1 1 1 1 ...  
## $ TOTCHG: num [1:500] 2660 1689 20060 736 1194 ...  
## $ APRDRG: num [1:500] 560 753 930 758 754 347 754 754 753 758 ...

describe(Hospital)

## vars n mean sd median trimmed mad min max range skew  
## AGE 1 500 5.09 6.95 0.0 4.26 0.00 0 17 17 0.73  
## FEMALE 2 500 0.51 0.50 1.0 0.52 0.00 0 1 1 -0.05  
## LOS 3 500 2.83 3.36 2.0 2.31 1.48 0 41 41 7.09  
## RACE 4 499 1.08 0.51 1.0 1.00 0.00 1 6 5 7.43  
## TOTCHG 5 500 2774.39 3888.41 1536.5 1894.24 646.41 532 48388 47856 5.53  
## APRDRG 6 500 616.36 178.32 640.0 656.37 0.00 21 952 931 -2.02  
## kurtosis se  
## AGE -1.34 0.31  
## FEMALE -2.00 0.02  
## LOS 65.68 0.15  
## RACE 57.52 0.02  
## TOTCHG 45.97 173.89  
## APRDRG 3.55 7.97

table(Hospital$AGE)

##   
## 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17   
## 307 10 1 3 2 2 2 3 2 2 4 8 15 18 25 29 29 38

cor(Hospital$TOTCHG,Hospital$APRDRG)

## [1] -0.3300012

cor(Hospital$RACE,Hospital$TOTCHG)

## [1] NA

sum(is.na(Hospital))

## [1] 1

head(Hospital)

## # A tibble: 6 × 6  
## AGE FEMALE LOS RACE TOTCHG APRDRG  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 17 1 2 1 2660 560  
## 2 17 0 2 1 1689 753  
## 3 17 1 7 1 20060 930  
## 4 17 1 1 1 736 758  
## 5 17 1 1 1 1194 754  
## 6 17 0 0 1 3305 347

tail(Hospital)

## # A tibble: 6 × 6  
## AGE FEMALE LOS RACE TOTCHG APRDRG  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 0 1 3 1 1886 640  
## 2 0 1 6 1 5881 636  
## 3 0 1 2 1 1171 640  
## 4 0 1 2 1 1171 640  
## 5 0 1 2 1 1086 640  
## 6 0 0 4 1 4931 640

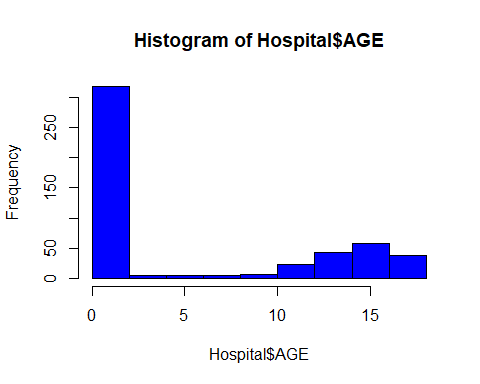
#Q1  
hist(Hospital$AGE,col="blue",title="age category of people who frequently visit the hospital ")

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical  
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...): "title"  
## is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter



# This shows that the frequency of thr age of patient visiting hospital is lying between 0-2 years   
Expense = aggregate(TOTCHG ~ AGE, FUN=sum, data=Hospital)  
which.max(tapply(Expense$TOTCHG, Expense$TOTCHG, FUN=sum))

## 678118   
## 18

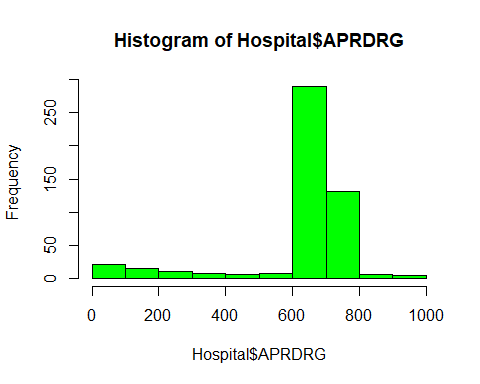
#Q2  
hist(Hospital$APRDRG,col="green",title="frequency of diagnosis related group that has maximum hospitalization and expenditure")

## Warning in plot.window(xlim, ylim, "", ...): "title" is not a graphical  
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...): "title"  
## is not a graphical parameter

## Warning in axis(1, ...): "title" is not a graphical parameter

## Warning in axis(2, at = yt, ...): "title" is not a graphical parameter



#frequency of All Patient Refined Diagnosis Related Groups   
table(Hospital$APRDRG)

##   
## 21 23 49 50 51 53 54 57 58 92 97 114 115 137 138 139 141 143 204 206   
## 1 1 1 1 1 10 1 2 1 1 1 1 2 1 4 5 1 1 1 1   
## 225 249 254 308 313 317 344 347 420 421 422 560 561 566 580 581 602 614 626 633   
## 2 6 1 1 1 1 2 3 2 1 3 2 1 1 1 3 1 3 6 4   
## 634 636 639 640 710 720 723 740 750 751 753 754 755 756 758 760 776 811 812 863   
## 2 3 4 267 1 1 2 1 1 14 36 37 13 2 20 2 1 2 3 1   
## 911 930 952   
## 1 2 1

max(table(Hospital$APRDRG))

## [1] 267

diagnosiscost<-aggregate(TOTCHG~APRDRG,FUN = sum,data=Hospital)  
diagnosiscost

## APRDRG TOTCHG  
## 1 21 10002  
## 2 23 14174  
## 3 49 20195  
## 4 50 3908  
## 5 51 3023  
## 6 53 82271  
## 7 54 851  
## 8 57 14509  
## 9 58 2117  
## 10 92 12024  
## 11 97 9530  
## 12 114 10562  
## 13 115 25832  
## 14 137 15129  
## 15 138 13622  
## 16 139 17766  
## 17 141 2860  
## 18 143 1393  
## 19 204 8439  
## 20 206 9230  
## 21 225 25649  
## 22 249 16642  
## 23 254 615  
## 24 308 10585  
## 25 313 8159  
## 26 317 17524  
## 27 344 14802  
## 28 347 12597  
## 29 420 6357  
## 30 421 26356  
## 31 422 5177  
## 32 560 4877  
## 33 561 2296  
## 34 566 2129  
## 35 580 2825  
## 36 581 7453  
## 37 602 29188  
## 38 614 27531  
## 39 626 23289  
## 40 633 17591  
## 41 634 9952  
## 42 636 23224  
## 43 639 12612  
## 44 640 437978  
## 45 710 8223  
## 46 720 14243  
## 47 723 5289  
## 48 740 11125  
## 49 750 1753  
## 50 751 21666  
## 51 753 79542  
## 52 754 59150  
## 53 755 11168  
## 54 756 1494  
## 55 758 34953  
## 56 760 8273  
## 57 776 1193  
## 58 811 3838  
## 59 812 9524  
## 60 863 13040  
## 61 911 48388  
## 62 930 26654  
## 63 952 4833

diagnosiscost[which.max(diagnosiscost$TOTCHG),]

## APRDRG TOTCHG  
## 44 640 437978

#Q3  
#To check if any missing values are present or not   
Hospital<-na.omit(Hospital)  
anyNA(Hospital)

## [1] FALSE

summary(as.factor(Hospital$RACE))

## 1 2 3 4 5 6   
## 484 6 1 3 3 2

#To observe if the Race of Patient is related to the hospital costs   
model<-lm(TOTCHG~RACE,data=Hospital)  
model

##   
## Call:  
## lm(formula = TOTCHG ~ RACE, data = Hospital)  
##   
## Coefficients:  
## (Intercept) RACE   
## 2925.7 -137.3

summary(model)

##   
## Call:  
## lm(formula = TOTCHG ~ RACE, data = Hospital)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2256 -1560 -1227 -258 45600   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2925.7 405.0 7.224 1.92e-12 \*\*\*  
## RACE -137.3 339.1 -0.405 0.686   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3895 on 497 degrees of freedom  
## Multiple R-squared: 0.0003299, Adjusted R-squared: -0.001681   
## F-statistic: 0.164 on 1 and 497 DF, p-value: 0.6856

riaov <- aov(TOTCHG ~ RACE, data=Hospital)  
riaov

## Call:  
## aov(formula = TOTCHG ~ RACE, data = Hospital)  
##   
## Terms:  
## RACE Residuals  
## Sum of Squares 2488459 7539623326  
## Deg. of Freedom 1 497  
##   
## Residual standard error: 3894.903  
## Estimated effects may be unbalanced

summary.aov(riaov)

## Df Sum Sq Mean Sq F value Pr(>F)  
## RACE 1 2.488e+06 2488459 0.164 0.686  
## Residuals 497 7.540e+09 15170268

#Q4 Creating a linear regression model  
  
model1<-lm(TOTCHG~AGE+FEMALE,data = Hospital)  
model1

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + FEMALE, data = Hospital)  
##   
## Coefficients:  
## (Intercept) AGE FEMALE   
## 2719.45 86.04 -744.21

summary(model1)

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + FEMALE, data = Hospital)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3403 -1444 -873 -156 44950   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2719.45 261.42 10.403 < 2e-16 \*\*\*  
## AGE 86.04 25.53 3.371 0.000808 \*\*\*  
## FEMALE -744.21 354.67 -2.098 0.036382 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3849 on 496 degrees of freedom  
## Multiple R-squared: 0.02585, Adjusted R-squared: 0.02192   
## F-statistic: 6.581 on 2 and 496 DF, p-value: 0.001511

#Conclusion which can be drawn from the model can be written as follows  
# As Age is an important factor in Hospital Costs as we can see from the   
#p values and level of signifiicance Gender seems to have an impact on the  
#given model and the negative coefficient values shown that females inccur  
# less cost than males .  
#Q5  
  
riaov1=aov(LOS~AGE+FEMALE+RACE,data=Hospital)  
riaov1

## Call:  
## aov(formula = LOS ~ AGE + FEMALE + RACE, data = Hospital)  
##   
## Terms:  
## AGE FEMALE RACE Residuals  
## Sum of Squares 26.907 16.510 1.165 5599.939  
## Deg. of Freedom 1 1 1 495  
##   
## Residual standard error: 3.363482  
## Estimated effects may be unbalanced

summary.aov(riaov1)

## Df Sum Sq Mean Sq F value Pr(>F)  
## AGE 1 27 26.907 2.378 0.124  
## FEMALE 1 17 16.510 1.459 0.228  
## RACE 1 1 1.165 0.103 0.748  
## Residuals 495 5600 11.313

model2<-lm(LOS~AGE+FEMALE+RACE,data=Hospital)  
model2

##   
## Call:  
## lm(formula = LOS ~ AGE + FEMALE + RACE, data = Hospital)  
##   
## Coefficients:  
## (Intercept) AGE FEMALE RACE   
## 2.94377 -0.03960 0.37011 -0.09408

summary(model2)

##   
## Call:  
## lm(formula = LOS ~ AGE + FEMALE + RACE, data = Hospital)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.22 -1.22 -0.85 0.15 37.78   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.94377 0.39318 7.487 3.25e-13 \*\*\*  
## AGE -0.03960 0.02231 -1.775 0.0766 .   
## FEMALE 0.37011 0.31024 1.193 0.2334   
## RACE -0.09408 0.29312 -0.321 0.7484   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.363 on 495 degrees of freedom  
## Multiple R-squared: 0.007898, Adjusted R-squared: 0.001886   
## F-statistic: 1.314 on 3 and 495 DF, p-value: 0.2692

table(Hospital$FEMALE)

##   
## 0 1   
## 244 255

#The conclusion which is drawn from the model is that the p values of Age Female Race is very high   
#There is no relationship between the variables Hence we cannot predict the length of stay on the basis of   
#the variables present in the model  
#Q6  
model3=lm(TOTCHG~.,data=Hospital)  
model3

##   
## Call:  
## lm(formula = TOTCHG ~ ., data = Hospital)  
##   
## Coefficients:  
## (Intercept) AGE FEMALE LOS RACE APRDRG   
## 5218.677 134.695 -390.692 743.152 -212.429 -7.791

summary(model3)

##   
## Call:  
## lm(formula = TOTCHG ~ ., data = Hospital)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6377 -700 -174 122 43378   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5218.6769 507.6475 10.280 < 2e-16 \*\*\*  
## AGE 134.6949 17.4711 7.710 7.02e-14 \*\*\*  
## FEMALE -390.6924 247.7390 -1.577 0.115   
## LOS 743.1521 34.9225 21.280 < 2e-16 \*\*\*  
## RACE -212.4291 227.9326 -0.932 0.352   
## APRDRG -7.7909 0.6816 -11.430 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2613 on 493 degrees of freedom  
## Multiple R-squared: 0.5536, Adjusted R-squared: 0.5491   
## F-statistic: 122.3 on 5 and 493 DF, p-value: < 2.2e-16

#Based on the model output it shows that the Age and LOS shows that both of the variables will be   
#affecting the cost And cost is directly proportional to the length of the stay And we can observe that   
#for increase of 1LOS Unit=743.1521 units of it also the increase of 1 Unit of Age is 134.6949 units  
  
  
  
model4=lm(TOTCHG~AGE+LOS+APRDRG,data=Hospital)  
model4

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + LOS + APRDRG, data = Hospital)  
##   
## Coefficients:  
## (Intercept) AGE LOS APRDRG   
## 4960.170 128.552 740.806 -8.006

summary(model4)

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + LOS + APRDRG, data = Hospital)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6603 -719 -169 124 43350   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4960.1705 433.6579 11.44 < 2e-16 \*\*\*  
## AGE 128.5519 17.0946 7.52 2.59e-13 \*\*\*  
## LOS 740.8057 34.9161 21.22 < 2e-16 \*\*\*  
## APRDRG -8.0055 0.6643 -12.05 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2617 on 495 degrees of freedom  
## Multiple R-squared: 0.5506, Adjusted R-squared: 0.5479   
## F-statistic: 202.2 on 3 and 495 DF, p-value: < 2.2e-16

model5=lm(TOTCHG~AGE+FEMALE+RACE+LOS,data=Hospital)  
model5

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + FEMALE + RACE + LOS, data = Hospital)  
##   
## Coefficients:  
## (Intercept) AGE FEMALE RACE LOS   
## 735.9 115.8 -1027.3 -114.1 742.0

summary(model5)

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + FEMALE + RACE + LOS, data = Hospital)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4363 -1114 -644 137 41631   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 735.86 362.16 2.032 0.042699 \*   
## AGE 115.79 19.54 5.925 5.86e-09 \*\*\*  
## FEMALE -1027.30 271.23 -3.788 0.000171 \*\*\*  
## RACE -114.14 255.92 -0.446 0.655802   
## LOS 742.04 39.24 18.911 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2936 on 494 degrees of freedom  
## Multiple R-squared: 0.4353, Adjusted R-squared: 0.4307   
## F-statistic: 95.19 on 4 and 494 DF, p-value: < 2.2e-16

model6=lm(TOTCHG~AGE+RACE,data=Hospital)  
model6

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + RACE, data = Hospital)  
##   
## Coefficients:  
## (Intercept) AGE RACE   
## 2567.63 73.59 -153.08

summary(model6)

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + RACE, data = Hospital)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3060 -1319 -1002 -291 44722   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2567.63 419.79 6.116 1.94e-09 \*\*\*  
## AGE 73.59 24.91 2.954 0.00329 \*\*   
## RACE -153.08 336.51 -0.455 0.64937   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3865 on 496 degrees of freedom  
## Multiple R-squared: 0.01761, Adjusted R-squared: 0.01365   
## F-statistic: 4.446 on 2 and 496 DF, p-value: 0.01219

model8=lm(TOTCHG~AGE+LOS,data=Hospital)  
model8

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + LOS, data = Hospital)  
##   
## Coefficients:  
## (Intercept) AGE LOS   
## 200.66 97.96 734.27

summary(model8)

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + LOS, data = Hospital)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4783 -1103 -458 -133 41382   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 200.66 203.48 0.986 0.325   
## AGE 97.96 19.21 5.101 4.83e-07 \*\*\*  
## LOS 734.27 39.66 18.512 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2973 on 496 degrees of freedom  
## Multiple R-squared: 0.4188, Adjusted R-squared: 0.4164   
## F-statistic: 178.7 on 2 and 496 DF, p-value: < 2.2e-16

#The conclusions which can be drawn from the model are as follows   
#As it is apparent from the coefficient values, Age, Length of stay (LOS) and patient refined diagnosis related groups(APRDRG) have three stars (\*\*\*) next to it. So they are the ones with statistical significance  
#Also, RACE is the least significant. build a model after removing RACE   
#We observe that APRDRG has t values which is negative value and hence we decided to drop it   
#And in the model 6 it shows that the Standard Error Increases to 3865 on 496 degrees of freedom

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.