Chinemerem-R-project-Health.R

cugagu

2020-07-14

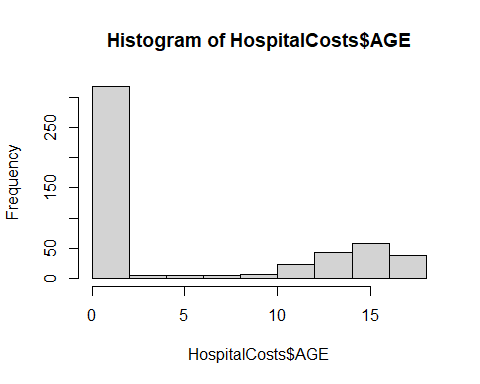
setwd("C:/Users/cugagu/Documents/Chinemerem")  
getwd()

## [1] "C:/Users/cugagu/Documents/Chinemerem"

HospitalCosts<-read.csv("Hospital.csv")  
 summary(HospitalCosts)

## 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   
##

hist(HospitalCosts$AGE)



table(as.factor(HospitalCosts$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

summary(as.factor(HospitalCosts$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

#Finding the age of people who frequent the hospital,and has the max expenditure.  
 aggregate(TOTCHG ~AGE,FUN = sum,data = HospitalCosts)

## AGE TOTCHG  
## 1 0 678118  
## 2 1 37744  
## 3 2 7298  
## 4 3 30550  
## 5 4 15992  
## 6 5 18507  
## 7 6 17928  
## 8 7 10087  
## 9 8 4741  
## 10 9 21147  
## 11 10 24469  
## 12 11 14250  
## 13 12 54912  
## 14 13 31135  
## 15 14 64643  
## 16 15 111747  
## 17 16 69149  
## 18 17 174777

which.max(summary(as.factor(HospitalCosts$APRDRG)))

## 640   
## 44

# determining the diagnosis related groups with max hospitalization & treatment  
 diagnosiscost <- aggregate(TOTCHG ~APRDRG,FUN = sum,data = HospitalCosts)  
 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

# Comparing the race of patients by expenditure  
 sapply(HospitalCosts, function(x) sum(is.na(x)))

## AGE FEMALE LOS RACE TOTCHG APRDRG   
## 0 0 0 1 0 0

HospitalCosts <- na.omit(HospitalCosts)  
 summary(as.factor(HospitalCosts$RACE))

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

HospitalCosts$RACE <- as.factor(HospitalCosts$RACE)  
 modl <- aov(TOTCHG ~ RACE,data = HospitalCosts)  
 modl

## Call:  
## aov(formula = TOTCHG ~ RACE, data = HospitalCosts)  
##   
## Terms:  
## RACE Residuals  
## Sum of Squares 18593279 7523518505  
## Deg. of Freedom 5 493  
##   
## Residual standard error: 3906.493  
## Estimated effects may be unbalanced

# summary(modl)  
 summary(HospitalCosts$RACE)

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

# what is the relationship btw hospitals cost by age and gender?  
 mod2 <- lm(TOTCHG ~ AGE+FEMALE,data= HospitalCosts)  
 summary(mod2)

##   
## Call:  
## lm(formula = TOTCHG ~ AGE + FEMALE, data = HospitalCosts)  
##   
## 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

# predicting if the length of stay on age, gender and race  
 mod3 <- lm(LOS ~ AGE+FEMALE+RACE,data = HospitalCosts)  
 summary(mod3)

##   
## Call:  
## lm(formula = LOS ~ AGE + FEMALE + RACE, data = HospitalCosts)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.211 -1.211 -0.857 0.143 37.789   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.85687 0.23160 12.335 <2e-16 \*\*\*  
## AGE -0.03938 0.02258 -1.744 0.0818 .   
## FEMALE 0.35391 0.31292 1.131 0.2586   
## RACE2 -0.37501 1.39568 -0.269 0.7883   
## RACE3 0.78922 3.38581 0.233 0.8158   
## RACE4 0.59493 1.95716 0.304 0.7613   
## RACE5 -0.85687 1.96273 -0.437 0.6626   
## RACE6 -0.71879 2.39295 -0.300 0.7640   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.376 on 491 degrees of freedom  
## Multiple R-squared: 0.008699, Adjusted R-squared: -0.005433   
## F-statistic: 0.6156 on 7 and 491 DF, p-value: 0.7432

# What are the main variables that affect the hospital cost  
 mod4 <- lm(TOTCHG ~.,data = HospitalCosts)  
 summary(mod4)

##   
## Call:  
## lm(formula = TOTCHG ~ ., data = HospitalCosts)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6367 -691 -186 121 43412   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5024.9610 440.1366 11.417 < 2e-16 \*\*\*  
## AGE 133.2207 17.6662 7.541 2.29e-13 \*\*\*  
## FEMALE -392.5778 249.2981 -1.575 0.116   
## LOS 742.9637 35.0464 21.199 < 2e-16 \*\*\*  
## RACE2 458.2427 1085.2320 0.422 0.673   
## RACE3 330.5184 2629.5121 0.126 0.900   
## RACE4 -499.3818 1520.9293 -0.328 0.743   
## RACE5 -1784.5776 1532.0048 -1.165 0.245   
## RACE6 -594.2921 1859.1271 -0.320 0.749   
## APRDRG -7.8175 0.6881 -11.361 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2622 on 489 degrees of freedom  
## Multiple R-squared: 0.5544, Adjusted R-squared: 0.5462   
## F-statistic: 67.6 on 9 and 489 DF, p-value: < 2.2e-16