RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

Source

Console | Terminal | Jobs

C:/Users/DELL/Desktop/dad/mom photos/ml/new R/

```
> setwd("C:/Users/DELL/Desktop/dad/mom photos/ml/new R")
> library("openxlsx")
> hospital_cost<-read.xlsx("1555054100_hospitalcosts.xlsx")
> summary(hospital_cost)
      AGE             FEMALE            LOS              RACE            TOTCHG
 Min.   : 0.000   Min.   :0.000   Min.   : 0.000   Min.   :1.000   Min.   :  532
 1st Qu.: 0.000   1st Qu.:0.000   1st Qu.: 2.000   1st Qu.:1.000   1st Qu.: 1216
 Median : 0.000   Median :1.000   Median : 2.000   Median :1.000   Median : 1536
 Mean   : 5.086   Mean   :0.512   Mean   : 2.828   Mean   :1.078   Mean   : 2774
 3rd Qu.:13.000   3rd Qu.:1.000   3rd Qu.: 3.000   3rd Qu.:1.000   3rd Qu.: 2530
 Max.   :17.000   Max.   :1.000   Max.   :41.000   Max.   :6.000   Max.   :48388
                                                   NA's   :1
     APRDRG
 Min.   : 21.0
 1st Qu.:640.0
 Median :640.0
 Mean   :616.4
 3rd Qu.:751.0
 Max.   :952.0

> hist(hospital_cost$AGE,main="Histogram of frequency of the patients",
+      xlab="Age",ylab="Frequency")
> age=as.factor(hospital_cost$AGE)
> summary(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
> which.max(summary(age)) #gives age which visits frequent to the hospital
0
1
> expnd=aggregate(TOTCHG~AGE,FUN=sum,data=hospital_cost)
> expnd
   AGE TOTCHG
1    0 678118
2    1  37744
3    2   7298
4    3  30550
5    4  15992
6    5  18507
```

Type here to search

ENG 07:47 PM 17-08-2020

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

Source

Console | Terminal | Jobs

C:/Users/DELL/Desktop/dad/mom photos/ml/new R/

```
> expnd=aggregate(TOTCHG~AGE,FUN=sum,data=hospital_cost)
> expnd
   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
> max(expnd) #gives maximum expenditure
[1] 678118
> hist(hospital_cost$APRDRG,main="Histogram of Diagnosis Related Groups",
+      xlab="Treatment",ylab="Frequency" )
> dgroup=as.factor(hospital_cost$APRDRG)
> summary(dgroup)
 21  23  49  50  51  53  54  57  58  92  97 114 115 137 138 139 141 143 204 206 225
  1   1   1   1   1  10   1   2   1   1   1   1   2   1   4   5   1   1   1   1   2
249 254 308 313 317 344 347 420 421 422 560 561 566 580 581 602 614 626 633 634 636
  6   1   1   1   1   2   3   2   1   3   2   1   1   1   3   1   3   6   4   2   3
639 640 710 720 723 740 750 751 753 754 755 756 758 760 776 811 812 863 911 930 952
  4 267   1   1   2   1   1  14  36  37  13   2  20   2   1   2   3   1   1   2   1
> which.max(summary(dgroup)) #gives the type of APRDRG which has maximum hospitalisation.
640
 44
> expnd2=aggregate(TOTCHG~APRDRG,FUN=sum,data=hospital_cost)
> expnd2
```

Type here to search

ENG 07:47 PM 17-08-2020

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

Source

Console | Terminal | Jobs

C:/Users/DELL/Desktop/dad/mom photos/ml/new R/

```
640
 44
> expnd2=aggregate(TOTCHG~APRDRG,FUN=sum,data=hospital_cost)
> expnd2
   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
```

Type here to search

ENG 07:47 PM 17-08-2020

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

Source

Console | Terminal | Jobs

C:/Users/DELL/Desktop/dad/mom photos/ml/new R/

```
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
```

Type here to search

ENG 07:47 PM 17-08-2020

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

Source

Console | Terminal | Jobs

C:/Users/DELL/Desktop/dad/mom photos/ml/new R/

```
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
> max(expnd2) #gives maximum expenditure
[1] 437978
> expnd2[which.max(expnd2$TOTCHG),] #gives index of maximum expenditure
   APRDRG TOTCHG
44     640 437978
> #To make sure that there is no malpractice, the agency needs to analyze if the race of the patient is related to the hospitalization costs.
> race=as.factor(hospital_cost$RACE)
> summary(race)
   1    2    3    4    5    6 NA's
 484    6    1    3    3    2    1
> hospital_cost=na.omit(hospital_cost)
> Aov=aov(TOTCHG~RACE,data=hospital_cost)
> summary(Aov)
             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
> #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.
> gender=as.factor(hospital_cost$FEMALE)
> reg1=lm(formula=TOTCHG~AGE+FEMALE,data=hospital_cost)
> summary(reg1)
```

Type here to search

ENG 07:48 PM 17-08-2020

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

Source

Console | Terminal | Jobs

C:/Users/DELL/Desktop/dad/mom photos/ml/new R/

```
Call:
lm(formula = TOTCHG ~ AGE + FEMALE, data = hospital_cost)

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

> summary(gender)
  0   1
244 255
> #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.
> race=as.factor(hospital_cost$RACE)
> reg2=lm(formula=LOS~AGE+FEMALE+RACE,data=hospital_cost)
> summary(reg2)

Call:
lm(formula = LOS ~ AGE + FEMALE + RACE, data = hospital_cost)

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 ***
```

Type here to search

ENG 07:48 PM 17-08-2020

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

Source

Console | Terminal | Jobs

C:/Users/DELL/Desktop/dad/mom photos/ml/new R/

```
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

> #To perform a complete analysis, the agency wants to find the variable that mainly affects hospital costs.
> reg3=lm(formula=TOTCHG~.,data=hospital_cost)
> summary(reg3)

Call:
lm(formula = TOTCHG ~ ., data = hospital_cost)

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

> |
```

Type here to search

ENG 07:48 PM 17-08-2020