**Hospital Data Analysis**

**S Somsekhar**

08-06-2022

hosp<-read.csv("E:/HospitalCosts.csv")

> hosp

AGE FEMALE LOS RACE TOTCHG APRDRG

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

7 17 1 4 1 2205 754

8 16 1 2 1 1167 754

9 16 1 1 1 532 753

10 17 1 2 1 1363 758

11 17 1 2 1 1245 758

12 15 0 2 1 1656 753

13 15 1 2 1 1379 751

14 15 1 4 1 2346 758

15 15 1 7 1 4006 753

16 15 1 4 1 2181 758

17 14 1 1 1 628 754

18 14 1 4 1 2463 758

19 15 1 3 1 1956 753

20 14 1 3 1 1802 758

21 13 1 1 1 3188 812

22 17 1 2 1 2129 566

23 12 0 1 1 7421 249

24 15 1 1 1 1122 422

25 13 1 2 4 1173 754

26 12 0 2 1 3625 812

27 11 1 2 1 3908 50

28 15 0 1 1 3994 139

29 11 0 0 1 1033 753

30 10 0 2 1 2860 141

31 11 0 2 1 3814 420

32 7 0 0 1 1132 139

33 16 1 2 6 1163 751

34 17 1 1 1 610 751

35 6 0 3 1 9530 97

36 15 1 1 1 1268 811

37 17 1 4 1 2582 753

38 16 1 2 1 1287 755

39 17 1 3 1 6594 930

40 13 1 0 1 909 755

41 7 0 0 1 2530 347

42 11 1 2 2 1534 753

43 3 0 5 1 14243 720

44 16 1 3 1 1699 754

45 2 0 2 1 7298 53

46 16 1 1 1 636 754

47 15 1 1 1 626 754

48 1 0 2 1 3782 53

49 14 1 2 1 1444 753

50 14 1 2 1 1183 754

51 14 1 5 1 3045 754

52 14 1 5 1 3624 754

53 14 1 12 1 6810 760

54 1 0 1 1 1409 249

55 13 0 2 1 1211 754

56 1 0 4 1 9606 53

57 1 1 1 1 1411 249

58 15 1 0 1 607 754

59 1 0 1 1 2932 249

60 1 0 3 1 5075 139

61 14 1 1 1 762 753

62 16 1 6 1 6329 753

63 17 1 1 1 1226 753

64 3 1 4 1 8223 710

65 17 0 2 1 1193 776

66 13 1 2 1 1076 754

67 12 1 6 1 17434 115

68 12 1 2 1 1647 753

69 14 1 7 1 3865 754

70 13 1 1 1 628 754

71 15 1 1 1 806 755

72 0 1 41 1 29188 602

73 0 0 2 1 4717 138

74 0 0 12 1 15129 137

75 0 1 2 1 1085 640

76 0 0 3 1 1607 640

77 0 1 3 1 1499 640

78 0 1 3 1 7648 53

79 0 1 2 1 1527 640

80 0 0 2 1 1483 640

81 0 1 4 1 2844 640

82 0 1 3 1 3124 640

83 0 0 3 1 1760 640

84 0 1 2 1 1278 640

85 0 1 2 1 1620 640

86 0 1 2 1 1220 640

87 0 1 2 1 1134 640

88 16 1 0 1 1235 754

89 0 0 3 1 1656 640

90 0 0 4 5 4072 639

91 0 0 2 5 1393 143

92 0 0 0 5 615 254

93 16 1 1 1 779 755

94 0 0 2 1 1385 640

95 0 0 2 1 1224 640

96 0 1 3 1 1779 640

97 0 0 2 1 1526 640

98 15 1 1 1 882 754

99 0 0 1 1 2075 581

100 0 0 17 1 12042 633

101 0 0 2 1 1309 640

102 0 0 2 1 1290 640

103 0 0 2 1 1280 640

104 0 0 3 1 1719 640

105 0 1 2 1 1102 640

106 0 1 3 1 1543 640

107 0 1 2 1 1174 640

108 0 1 2 1 1105 640

109 0 0 2 1 1335 640

110 0 0 2 1 1550 640

111 0 0 4 1 2473 640

112 0 0 2 1 1322 640

113 0 0 4 1 2553 640

114 15 0 5 1 2835 753

115 0 1 2 1 1191 640

116 0 0 2 1 1439 640

117 0 1 2 1 1237 640

118 0 0 2 1 1265 640

119 0 1 4 1 2280 640

120 0 0 2 1 1096 640

121 0 1 2 1 1156 640

122 0 0 2 1 1199 640

123 13 1 10 1 5615 754

124 0 1 4 1 2518 640

125 15 0 0 1 625 754

126 0 1 2 1 1246 640

127 0 1 3 1 1821 640

128 0 0 5 1 3101 626

129 12 1 2 1 1293 754

130 0 1 2 1 1176 640

131 0 0 3 1 1891 640

132 5 1 2 1 10584 53

133 13 1 3 1 2373 754

134 0 0 1 1 935 640

135 0 0 2 1 1395 640

136 0 0 2 1 1561 640

137 0 1 7 1 6912 636

138 12 1 2 1 1157 754

139 0 0 3 1 2197 640

140 0 0 4 1 2288 640

141 16 1 4 1 2348 754

142 0 0 2 1 1320 640

143 0 1 2 1 1139 640

144 0 1 4 1 2134 639

145 0 0 2 1 1407 640

146 0 0 2 1 1982 640

147 0 0 4 1 2539 640

148 0 0 2 1 1528 640

149 0 1 2 1 1513 640

150 0 1 2 1 1191 640

151 0 0 2 1 1280 640

152 0 0 2 1 3977 139

153 0 1 2 1 1269 640

154 0 0 2 1 1501 640

155 0 1 2 1 1396 640

156 0 0 3 1 1777 640

157 0 1 1 1 833 640

158 0 1 1 1 715 640

159 17 1 5 1 2936 751

160 0 0 2 1 1375 640

161 0 0 2 1 1330 640

162 0 0 2 1 1628 640

163 0 0 2 1 1368 640

164 12 1 1 1 622 755

165 17 0 2 1 14174 23

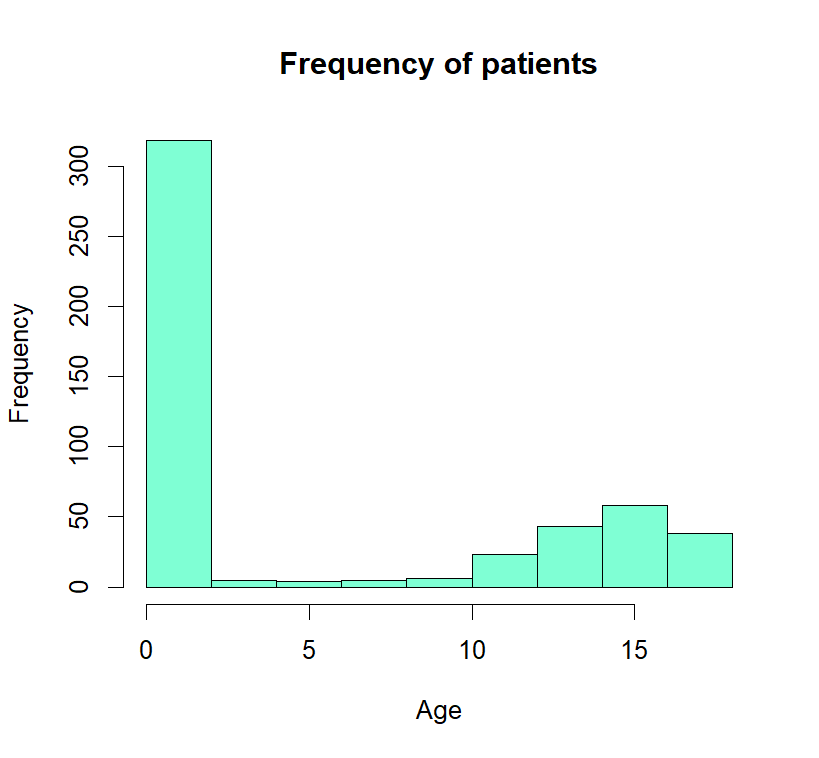
166 7 0 1 1 6425 57

[ reached 'max' / getOption("max.print") -- omitted 334 rows ]

>

> ##1Record patient stats-frequency of patients and max expenditure

> hist(hosp$AGE,main = "Frequency of patients",col = "aquamarine",xlab = "Age")



> attach(hosp)

The following object is masked \_by\_ .GlobalEnv:

AGE

> AGE<-as.factor(AGE)

> summary(AGE)#Thus infants have max hospital visits

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

> aggregate(TOTCHG~AGE,FUN=sum,data = hosp)

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(aggregate(TOTCHG~AGE,FUN=sum,data=hosp))

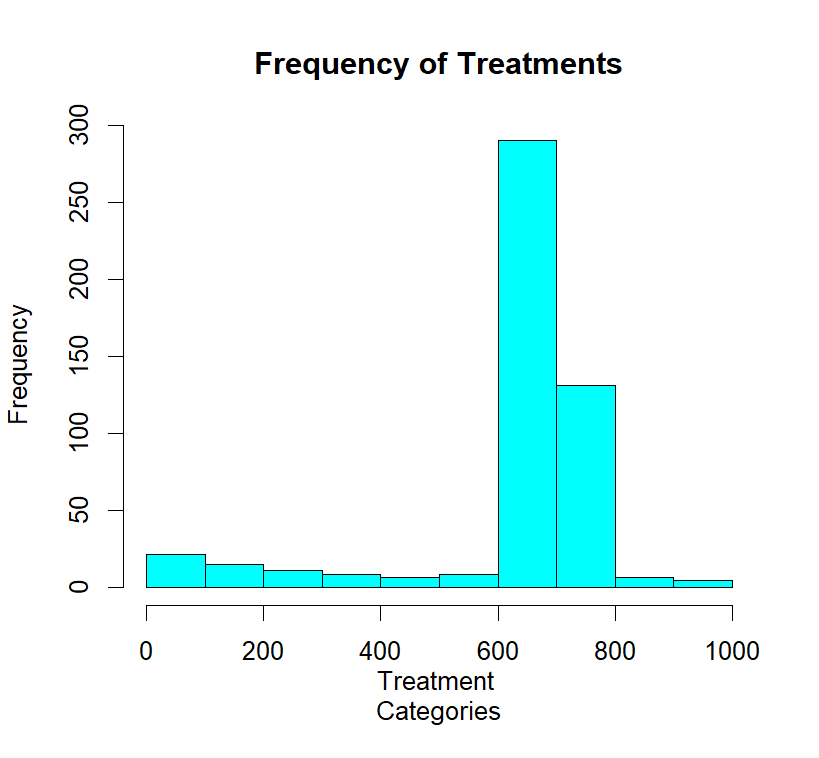
[1] 678118

>

> #2.Most Expensive Treatments

> hist(APRDRG,col = "cyan1",main = "Frequency of Treatments",xlab = "Treatment

+ Categories")



> APRDRG\_fact<-as.factor(hosp$APRDRG)

> summary(APRDRG\_fact)

21 23 49 50 51 53 54 57 58 92 97 114 115 137 138 139 141 143 204 206 225 249 254

1 1 1 1 1 10 1 2 1 1 1 1 2 1 4 5 1 1 1 1 2 6 1

308 313 317 344 347 420 421 422 560 561 566 580 581 602 614 626 633 634 636 639 640 710 720

1 1 1 2 3 2 1 3 2 1 1 1 3 1 3 6 4 2 3 4 267 1 1

723 740 750 751 753 754 755 756 758 760 776 811 812 863 911 930 952

2 1 1 14 36 37 13 2 20 2 1 2 3 1 1 2 1

> which.max(summary(APRDRG\_fact))

640

44

> df<-aggregate(TOTCHG~APRDRG,FUN = sum,data=hosp)

> df

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

> df[which.max(df$TOTCHG),]#category 640 has the highest hospitaliation frequency and cost.

APRDRG TOTCHG

44 640 437978

>

>

> #3.Analyze if race of patients is related to costs

> hosp<-na.omit(hosp)#first we remove "NA"values

> hosp$RACE<-as.factor(hosp$RACE)

> model\_aov<-aov(TOTCHG~RACE,data = hosp)

> model\_aov#ANOVA RESULTS

Call:

aov(formula = TOTCHG ~ RACE, data = hosp)

Terms:

RACE Residuals

Sum of Squares 18593279 7523518505

Deg. of Freedom 5 493

Residual standard error: 3906.493

Estimated effects may be unbalanced

> summary(model\_aov)

Df Sum Sq Mean Sq F value Pr(>F)

RACE 5 1.859e+07 3718656 0.244 0.943

Residuals 493 7.524e+09 15260687

> summary(hosp$RACE)#getting max hospital cost per race

1 2 3 4 5 6

484 6 1 3 3 2

>

>

> #4.Analyze severity of hospital cost by Age and gender

> hosp$FEMALE<-as.factor(hosp$FEMALE)

> model\_lm4<-lm(TOTCHG~AGE+FEMALE,data = hosp)#calling Regression funtion

> summary(model\_lm4)

Call:

lm(formula = TOTCHG ~ AGE + FEMALE, data = hosp)

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 \*\*\*

FEMALE1 -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(hosp$FEMALE)#comapring genders

0 1

244 255

>

>

> #5.Analyze whether length of stay(LOS) is realted to age,gender and race

> hosp$RACE<-as.factor(hosp$RACE)

> model\_lm5<-lm(LOS~AGE+FEMALE+RACE,data = hosp)

> summary(model\_lm5)

Call:

lm(formula = LOS ~ AGE + FEMALE + RACE, data = hosp)

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 .

FEMALE1 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

>

>

> #6.Variable that mainly affect hospitalization cost

> model\_lm6<-lm(TOTCHG~AGE+FEMALE+RACE+LOS+APRDRG,data = hosp)

> summary(model\_lm6)

Call:

lm(formula = TOTCHG ~ AGE + FEMALE + RACE + LOS + APRDRG, data = hosp)

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 \*\*\*

FEMALE1 -392.5778 249.2981 -1.575 0.116

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

LOS 742.9637 35.0464 21.199 < 2e-16 \*\*\*

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

>

>