bmi_final

Qianshu Wang

2017/12/9

Clean Data

```r
GHP = read_excel("./GHProject_Dataset.xlsx") %>%
  filter(ICU_Flag != 1)

GHP = GHP %>%
  separate(AdmitDtm, into = c("weekday", "date", "year"), sep = ",") %>%
  select(-year) %>%
  mutate(date = paste(date, " 2016"), date = as.Date(date, format = " %B %d %Y")) %>%
  mutate(Cindex = if_else(Cindex == 1 & Cindex ==2, 1, Cindex),
         Cindex = if_else(Cindex == 3 & Cindex == 4, 2, Cindex),
         Cindex = if_else(Cindex >= 5, 3, Cindex)) %>% # 0 = normal, 1 = mild, 2 = moderate, 3 = severe
  mutate(MEWS = if_else(MEWS == 1, 0, MEWS),
         MEWS = if_else((MEWS == 2) + (MEWS == 3) == 1, 1, MEWS),
         MEWS = if_else((MEWS == 4) + (MEWS == 5) == 1, 2, MEWS),
         MEWS = if_else(MEWS > 5, 3, MEWS)) # 0 = normal, 1 = increase caution, 2 = further deteriorati
GHP = GHP %>%
  group_by(PatientID) %>%
  filter(date == min(date)) %>%
  clean_names()

#write.csv(GHP, "GHP.csv")
```

From the plot we know that LOSDays2 is the same as LOSHour

```r
#plot(GHP$LOSHours, GHP$LOSDays2)
```

```r
mult.reg = lm(losdays2 ~ mews + is30dayreadmit + cindex + evisit + ageyear + gender + race + religion +
  summary(mult.reg)
```

```
##
## Call:
## lm(formula = losdays2 ~ mews + is30dayreadmit + cindex + evisit +
##     ageyear + gender + race + religion + maritalstatus + insurancetype,
##     data = GHP)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -7.936 -2.979 -1.409  1.168 82.832
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          6.282982   8.097113   0.776 0.437833
## mews                 0.507390   0.139168   3.646 0.000271 ***
## is30dayreadmit       1.167261   0.299746   3.894 0.000101 ***
## cindex               0.466437   0.086911   5.367 8.57e-08 ***
## evisit               0.241223   0.068008   3.547 0.000395 ***
```

```
## ageyear                        0.020539    0.007935    2.588 0.009684 **
## genderMale                      0.812582    0.210047    3.869 0.000112 ***
## raceAsian                      -0.170178    0.528289   -0.322 0.747374
## raceNative Amer/Alaskan        -0.269339    1.337643   -0.201 0.840435
## raceNatv Hawaii/Pacf Isl       -2.593695    2.874443   -0.902 0.366949
## raceOther/Multiracial          -0.041782    0.372491   -0.112 0.910696
## raceWhite                       0.110991    0.313225    0.354 0.723101
## religionCatholic               -1.271840    5.716027   -0.223 0.823935
## religionChristian              -1.026726    5.720640   -0.179 0.857574
## religionHebrew                 -1.882514    8.089345   -0.233 0.815997
## religionHindu                  -0.718155    5.751737   -0.125 0.900643
## religionIslam                  -2.398080    5.751369   -0.417 0.676737
## religionJewish                 -1.344090    5.720694   -0.235 0.814261
## religionMormon                 -3.535869    6.999076   -0.505 0.613459
## religionNo Affiliation         -1.628862    5.732999   -0.284 0.776336
## religionNon Denominational     -4.969233    8.089441   -0.614 0.539069
## religionOther                  -0.683502    5.738924   -0.119 0.905204
## maritalstatusDivorced          -2.071398    5.738613   -0.361 0.718154
## maritalstatusMarried           -2.013650    5.726312   -0.352 0.725125
## maritalstatusSeparated         -2.227036    5.784195   -0.385 0.700247
## maritalstatusSingle            -1.475744    5.724909   -0.258 0.796596
## maritalstatusWidowed           -1.822665    5.733093   -0.318 0.750566
## insurancetypeMedicare          -1.481530    0.527732   -2.807 0.005025 **
## insurancetypePrivate           -1.591008    0.497448   -3.198 0.001395 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.706 on 3253 degrees of freedom
##   (267 observations deleted due to missingness)
## Multiple R-squared:  0.05626,    Adjusted R-squared:  0.04814
## F-statistic: 6.926 on 28 and 3253 DF,  p-value: < 2.2e-16
```

```r
# stepwise <- step(mult.reg, direction = 'both')
# summary(stepwise)

best <- function(model, ...)
{
  subsets <- regsubsets(formula(model), model.frame(model), ...)
  subsets <- with(summary(subsets),
              cbind(p = as.numeric(rownames(which)), which, rss, rsq, adjr2, cp, bic))

  return(subsets)
}

best_subsets = best(mult.reg, nbest = 1)
pander(best_subsets)
```

Table 1: Table continues below

p	(Intercept)	mews	is30dayreadmit	cindex	evisit	ageyear
1	1	0	0	1	0	0
2	1	1	0	1	0	0
3	1	1	1	1	0	0
4	1	1	1	1	1	0

p	(Intercept)	mews	is30dayreadmit	cindex	evisit	ageyear
5	1	1	1	1	1	0
6	1	1	1	1	1	1
7	1	1	1	1	1	1
8	1	1	1	1	1	1

Table 2: Table continues below

genderMale	raceAsian	raceNative Amer/Alaskan	raceNatv Hawaii/Pacf Isl
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0

Table 3: Table continues below

raceOther/Multiracial	raceWhite	religionCatholic	religionChristian
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Table 4: Table continues below

religionHebrew	religionHindu	religionIslam	religionJewish
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Table 5: Table continues below

religionMormon	religionNo Affiliation	religionNon Denominational
0	0	0
0	0	0

religionMormon	religionNo Affiliation	religionNon Denominational
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0

Table 6: Table continues below

religionOther	maritalstatusDivorced	maritalstatusMarried
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0

Table 7: Table continues below

maritalstatusSeparated	maritalstatusSingle	maritalstatusWidowed
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	0	0
0	1	0
0	0	0

Table 8: Table continues below

insurancetypeMedicare	insurancetypePrivate	rss	rsq	adjr2
0	0	109849	0.02119	0.02089
0	0	108737	0.0311	0.03051
0	0	107807	0.03938	0.0385
0	0	107348	0.04347	0.04231
0	0	106941	0.04711	0.04565
0	0	106765	0.04867	0.04693
0	0	106560	0.0505	0.04847
1	1	106416	0.05178	0.04946

cp	bic
95.9	-54.09
63.73	-79.41

cp	bic
37.19	-99.48
25.08	-105.4
14.57	-109.8
11.17	-107.1
6.874	-105.3
4.452	-101.6