Module 4 Team Assignment

Team Assignment Expectations

This is a team assignment. Team assignments should be completed by each team as a group effort. There are expectations that all students will contribute their complete participation in the team assignments. Members of different teams are not to discuss these assignments with each other. If your team has questions or needs clarification, please contact me.

Data Details

What indicators are significant when modeling which ICU patients have the highest risk of death?

Data Details icu data from Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013) Applied Logistic Regression, 3rd ed., New York: Wiley. This data set available in the aplore3 R package.

An older version of this data set from 2000 is used in Exercises 17.4 and 17.5 in the textbook.
The 2013 updated dataset has 200 observations and 21 variables.
The code below will access the updated data set. You can confirm you are using the updated data set by referencing the first variable, this should be id.

```
knitr::opts_chunk$set(echo = TRUE)
#install.packages("HH")
#install.packages("aplore3")

library(HH)
```

```
## Loading required package: lattice

## Loading required package: grid

## Loading required package: latticeExtra

## Loading required package: multcomp

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##
## Attaching package: 'TH.data'
```

```
## The following object is masked from 'package:MASS':
##
##     geyser

## Loading required package: gridExtra
```

```
library(aplore3)
```

```
## Warning: package 'aplore3' was built under R version 4.0.5
```

```
summary(icu)
```

```
##        id            sta            age            gender        race
##  Min.   :  4.0   Lived:160   Min.   :16.00   Male  :124   White:175
##  1st Qu.:210.2   Died : 40   1st Qu.:46.75   Female: 76   Black: 15
##  Median :412.5               Median :63.00                Other: 10
##  Mean   :444.8               Mean   :57.55
##  3rd Qu.:671.8               3rd Qu.:72.00
##  Max.   :929.0               Max.   :92.00
##        ser        can        crn        inf        cpr           sys
##  Medical : 93   No :180   No :181   No :116   No :187   Min.   : 36.0
##  Surgical:107   Yes: 20   Yes: 19   Yes: 84   Yes: 13   1st Qu.:110.0
##                                                         Median :130.0
##                                                         Mean   :132.3
##                                                         3rd Qu.:150.0
##                                                         Max.   :256.0
##       hra           pre             type        fra          po2           ph
##  Min.   : 39.00   No :170   Elective : 53   No :185   > 60 :184   >= 7.25:187
##  1st Qu.: 80.00   Yes: 30   Emergency:147   Yes: 15   <= 60: 16   < 7.25 : 13
##  Median : 96.00
##  Mean   : 98.92
##  3rd Qu.:118.25
##  Max.   :192.00
##     pco         bic          cre           loc
##  <= 45:180   >= 18:185   <= 2.0:190   Nothing:185
##  > 45 : 20   < 18 : 15   > 2.0 : 10   Stupor :  5
##                                       Coma   : 10
##
##
##
```

```
head(icu)
```

```
##   id   sta age gender  race      ser can crn inf cpr sys hra pre      type fra
## 1  4  Died  87 Female White Surgical  No  No Yes  No  80  96  No Emergency Yes
## 2  8 Lived  27 Female White  Medical  No  No Yes  No 142  88  No Emergency  No
## 3 12 Lived  59   Male White  Medical  No  No  No  No 112  80 Yes Emergency  No
## 4 14 Lived  77   Male White Surgical  No  No  No  No 100  70  No  Elective  No
## 5 27  Died  76 Female White Surgical  No  No Yes  No 128  90 Yes Emergency  No
## 6 28 Lived  54   Male White  Medical  No  No Yes  No 142 103  No Emergency Yes
##     po2      ph   pco   bic    cre     loc
```

```
## 1 <= 60  < 7.25  > 45 >= 18 <= 2.0 Nothing
## 2  > 60 >= 7.25 <= 45 >= 18 <= 2.0 Nothing
## 3  > 60 >= 7.25 <= 45 >= 18 <= 2.0 Nothing
## 4  > 60 >= 7.25 <= 45 >= 18 <= 2.0 Nothing
## 5  > 60 >= 7.25 <= 45 >= 18 <= 2.0 Nothing
## 6  > 60 >= 7.25 <= 45 >= 18 <= 2.0 Nothing
```

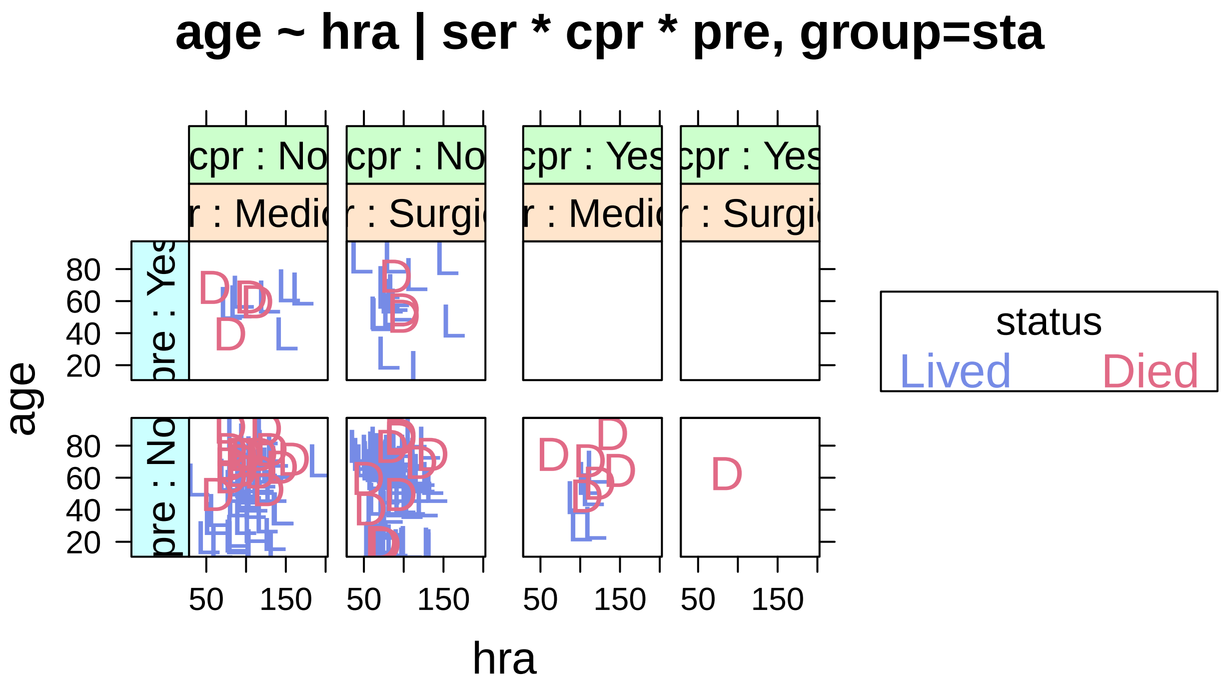
```
# Note that the factors in this data set are not binary but rather coded as 1 & 2
as.numeric(icu$sta)
```

```
##   [1] 2 1 1 1 2 1 1 1 1 1 1 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 2 1 2 2
##  [38] 1 1 1 1 1 1 1 1 2 2 2 1 2 1 1 1 2 1 1 2 2 1 1 1 1 1 2 1 1 1 1 2 1 1 1 1 2
##  [75] 1 1 1 1 2 1 1 1 2 1 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1 2 1 1 2 1 1 1 2 1 2 1 2
## [112] 2 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 2 1
## [149] 1 2 1 1 1 2 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1
## [186] 2 1 1 1 1 1 1 1 1 1 2 1 1 1 1
```

Part 1: Data Visual

The goal of this study is to better understand how a patients age, service at ICU, CPR, heart rate and previous admission relate to a patient's survival status (lived or died).
To start, create a useful visual. Plot the response variable STA (survival status) against the 5 selected predictors in R or Python. Discuss your insights.

```
col2 <- likertColor(2)[2:1]
useOuterStripsT2L1(
  xyplot(age ~ hra | ser * cpr * pre,
         data=icu,
         group=sta, pch=levels(icu$sta),
         col=col2, cex=2.2,
         layout=c(4, 2),
         main=list("age ~ hra | ser * cpr * pre, group=sta",
                   cex=1.6),
         aspect=1,
         between=list(x=c(.5, 1, .5), y=1),
         scales=list(cex=1, alternating=FALSE),
         xlab=list(cex=1.4), ylab=list(cex=1.4),
         par.strip.text=list(cex=1.25),
         key=list(space="right",
                  text=list(levels(icu$sta),
                            cex=1.5, adj=1,
                            col=col2),
                  columns=2, border=1,
                  title="status", cex.title=1.25,
                  cex=1)))
```



Insights

We gained the following insights from our visual:

General Data: The majority of the observations are for patients with no previous admission within 6 months who did not receive CPR. There are no observations where there was both a previous admission within 6 months and CPR performed prior to admission. There was only one observation where CPR was performed prior to admission for a surgical procedure with no previous admission within 6 months.

Age: Across all categories, the majority of deaths were for patients over ~40 and the deaths were more prevalent in patients over ~60. There are only two observations for deaths among young patients (~20), which were both admitted for surgical services with no CPR performed and no previous admissions; however, most young patients undergoing surgery lived.

Heart Rate: There is no noticeable trend between patients who lived or died and their heart rates. Both statuses generally seem to span the whole range. For patients with previous admissions who did not receive CPR, the data set does not include any deaths with a heart rate over ~110 for medical services and ~100 for surgical services, but there are limited death observations in each of these categories in the data set. When CPR was performed, the three highest heart rate observations resulted in deaths; however, the lowest heart rate observation also did.

Service: The ratio of patients who lived appears to be higher for surgical procedures compared to medical procedures.

CPR: A larger proportion of patients receiving CPR died compared to those not receiving CPR. 50% of the patients getting a medical procedure died, and the only observation of a patient getting CPR with a surgical procedure died.

Previous Admission: We don't see any clear relationship between status and previous admissions.

Logistic Regression Model

To answer the questions given, you and your team will need to create a logistic regression model that can be used to explain factors of patient survival status. For your model, you are not required to create (partial) residual or diagnostics plots; however, the final model should be evaluated and determined to be an acceptable model before used to answer any question.

```
# full model
icu1.glm <- glm(sta ~ ser + cpr + pre + hra + age,
                data = icu, family = binomial)
summary(icu1.glm)
```

```
##
## Call:
## glm(formula = sta ~ ser + cpr + pre + hra + age, family = binomial,
##     data = icu)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3797  -0.6762  -0.5376  -0.2604   2.6018
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.473174   1.077048  -2.296  0.02166 *
## serSurgical -0.989969   0.427675  -2.315  0.02063 *
## cprYes       1.554969   0.636663   2.442  0.01459 *
## preYes       0.485816   0.500901   0.970  0.33210
## hra         -0.006645   0.007588  -0.876  0.38117
## age          0.032530   0.011809   2.755  0.00587 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 177.84  on 194  degrees of freedom
## AIC: 189.84
##
## Number of Fisher Scoring iterations: 5
```

```
# heart rate is deleted - high p value (0.33210)
icu2.glm <- glm(sta ~ ser + cpr + pre + age,
                data = icu, family = binomial)
summary(icu2.glm)
```

```
##
## Call:
## glm(formula = sta ~ ser + cpr + pre + age, family = binomial,
##     data = icu)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3971  -0.6662  -0.5382  -0.2622   2.6148
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.13167    0.78713  -3.979 6.93e-05 ***
## serSurgical -0.85444    0.39639  -2.156  0.03112 *
## cprYes       1.51560    0.63426   2.390  0.01687 *
## preYes       0.45833    0.49798   0.920  0.35737
## age          0.03163    0.01169   2.705  0.00683 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 178.62  on 195  degrees of freedom
## AIC: 188.62
##
## Number of Fisher Scoring iterations: 5
```

```r
#previous admission is deleted - high p value (0.35737)
#lowest aic value (187.43)
icu3.glm <- glm(sta ~ ser + cpr + age,
                data = icu, family = binomial)
summary(icu3.glm)
```

```
##
## Call:
## glm(formula = sta ~ ser + cpr + age, family = binomial, data = icu)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3960  -0.6770  -0.5410  -0.2754   2.5775
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.05728    0.77311  -3.955 7.67e-05 ***
## serSurgical -0.82774    0.39392  -2.101  0.03561 *
## cprYes       1.44173    0.62759   2.297  0.02160 *
## age          0.03158    0.01157   2.730  0.00633 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 179.43  on 196  degrees of freedom
## AIC: 187.43
##
## Number of Fisher Scoring iterations: 5
```

Selected Model

```
summary(icu3.glm)
```

```
##
## Call:
## glm(formula = sta ~ ser + cpr + age, family = binomial, data = icu)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3960  -0.6770  -0.5410  -0.2754   2.5775
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.05728    0.77311  -3.955 7.67e-05 ***
## serSurgical -0.82774    0.39392  -2.101  0.03561 *
## cprYes       1.44173    0.62759   2.297  0.02160 *
## age          0.03158    0.01157   2.730  0.00633 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 179.43  on 196  degrees of freedom
## AIC: 187.43
##
## Number of Fisher Scoring iterations: 5
```

Discussion

Our selected logistic regression model includes service, CPR, and age to predict survival status. All selected variables have significant p-values at a 0.05 alpha. The p-values are 0.03561 for service, 0.02160 for CPR, and 0.00633 for age. We removed heart rate and previous ICU admission from our final model due to their high p-values, and our final model resulted in a lower AIC at 187.43 compared to the models that included heart rate and/or previous ICU admission, which further confirmed that the removed variables did not add value to the model. This also aligns with our insights from the visual indicating a relationship between survival status and service, age, and CPR, but no visible relationship between survival status and heart rate or previous admission.

PART 2: Questions

Using the model you and your team selected in part 2, answer the 5 questions given above (also copied below for your convenience). For each question, support your answer with output from your logistic regression. 1. Is there a significant increase in risk for older ICU patients?

Answer: There is a significant increase in risk for older ICU patients. The p-value for age is significant at 0.00633 and age's coefficient in our logistic regression model is positive indicating that risk of death increases with age. For each year of age, the odds of death increase by e^0.03158 = 1.032084 times. For example, at age 20, probability of death is ~8%, but it increases to ~24% at age 60 and ~53% at age 100.

```r
#Age 20, all else 0
antilogit(-3.05728+0.03158*20)
```

```
## [1] 0.08123531
```

```r
#Age 40, all else 0
antilogit(-3.05728+0.03158*40)
```

```
## [1] 0.1425732
```

```r
#Age 60, all else 0
antilogit(-3.05728+0.03158*60)
```

```
## [1] 0.2382169
```

```r
#Age 80, all else 0
antilogit(-3.05728+0.03158*80)
```

```
## [1] 0.3703117
```

```r
#Age 100, all else 0
antilogit(-3.05728+0.03158*100)
```

```
## [1] 0.5251587
```

2. Patient services in the ICU are often classified as medical or surgical. Assume these are mutually exclusive and exhausted groups. Is there a significance difference in risk between patients that receive medical ICU services and patients that receive surgical ICU services?

In the block below include your chi-square analysis.

Answer: There is a significant difference in risk for patients in the ICU for surgical procedures compared to medical procedures. This is confirmed by its significant p-values from both our logistic regression model and our chi-square analysis of 0.03561 and 0.008723 respectively. The negative coefficient in our logistic regression model indicates risk is lower for surgical procedures compared to medical procedures. The patient is e^-0.82774 = 0.437036 times as likely to die when in the ICU for a surgical procedure compared to a medical procedure. Although the residuals squared from the chi-square analysis are below 3.8 for each group, the observations for surgical patients who died were lower than would be expected and the observations for medical patients who died were higher than would expected if there was an independent relationship between the variables. In combination, the relationship between service and survival status is significant, indicating a dependent relationship and a difference in risk.

```r
#p value is significant but none of the residuals squared are in the 95th percentile
icu.ser.chisq <- chisq.test(icu$sta, icu$ser, correct=FALSE)
icu.ser.chisq
```

```
##
##  Pearson's Chi-squared test
##
## data:  icu$sta and icu$ser
## X-squared = 6.8787, df = 1, p-value = 0.008723
```

```r
icu.ser.chisq$observed
```

```
##         icu$ser
## icu$sta Medical Surgical
##   Lived      67       93
##   Died       26       14
```

```r
icu.ser.chisq$expected
```

```
##         icu$ser
## icu$sta Medical Surgical
##   Lived    74.4     85.6
##   Died     18.6     21.4
```

```r
icu.ser.chisq$residuals
```

```
##         icu$ser
## icu$sta    Medical   Surgical
##   Lived -0.8579170  0.7998247
##   Died   1.7158339 -1.5996495
```

```r
icu.ser.chisq$residuals^2
```

```
##         icu$ser
## icu$sta   Medical  Surgical
##   Lived 0.7360215 0.6397196
##   Died  2.9440860 2.5588785
```

3. Is there a significant increase or decrease in risk for patients that received CPR prior to admission to the ICU?

In the block below include your chi-square analysis.

Answer: There is a significant increase in risk for patients who received CPR prior to admission to the ICU. This is confirmed by CPR's positive coefficient of 1.44173 in our logistic regression model and by its significant p-values from both our logistic regression model and our chi-square analysis of 0.02160 and 0.005165 respectively. A patient who received CPR prior to admission to the ICU is e^1.44173 = 4.228 times more likely to die than one who did not receive CPR. The chi-square analysis shows residuals squared of 7.4461538 for patients who received CPR and died, which is above the 99th percentile in the chi-square distribution with 1 degree of freedom, which also confirms there is a significant association between receiving CPR prior to ICU admission and death. There were significantly more observations of patients who died when they had CPR prior to admission than would be expected if the variables were independent.

```r
#Warning message: "Chi-squared approximation may be incorrect"
icu.cpr.chisq <- chisq.test(icu$sta, icu$cpr)
```

```
## Warning in chisq.test(icu$sta, icu$cpr): Chi-squared approximation may be
## incorrect
```

```r
icu.cpr.chisq
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  icu$sta and icu$cpr
## X-squared = 7.8209, df = 1, p-value = 0.005165
```

```r
icu.cpr.chisq$observed
```

```
##        icu$cpr
## icu$sta  No Yes
##   Lived 154   6
##   Died   33   7
```

```r
icu.cpr.chisq$expected
```

```
##        icu$cpr
## icu$sta    No  Yes
##   Lived 149.6 10.4
##   Died   37.4  2.6
```

```r
icu.cpr.chisq$residuals
```

```
##        icu$cpr
## icu$sta         No        Yes
##   Lived  0.3597385 -1.3643821
##   Died  -0.7194769  2.7287642
```

```r
icu.cpr.chisq$residuals^2
```

```
##        icu$cpr
## icu$sta        No       Yes
##   Lived 0.1294118 1.8615385
##   Died  0.5176471 7.4461538
```

4. Is there a significant increase or decrease in risk for patients with a high (or low) heart rate at ICU admission?

Answer: Based on our analysis, there is not a significant change in risk for patients based on their heart rate. The full logistic regression model showed a p-value of 0.38117 for heart rate indicating it is not a statistically significant variable in predicting chance of death; therefore, it was removed from our final model. The visual we created also did not show any meaningful trends for survival status associated with heart rate.

5. Hospital re-admission rates are an increasing concern. Is there a significantly higher risk for patients previous admitted to an ICU within 6 months?

In the block below include your chi-square analysis.

Answer: Based on our analysis, there is not a significant increase in risk for patients who were previously admitted to the ICU within 6 months. After removing heart rate in our stepwise process to determine an appropriate logistic regression model, previous admission had a p-value of 0.35737 indicating it is not a statistically significant variable in predicting chance of death, so it was removed from our final model. Our chi-square analysis also had a high p-value of 0.8045 indicating there is not an association between a previous admission within 6 months and survival status. The observed values are very close to what would be expected under the assumption that the variables are independent.

```
#Not significant p value, no r^2 in 95th percentile
icu.pre.chisq <- chisq.test(icu$sta, icu$pre)
icu.pre.chisq
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  icu$sta and icu$pre
## X-squared = 0.061275, df = 1, p-value = 0.8045
```

```
icu.pre.chisq$observed
```

```
##        icu$pre
## icu$sta  No Yes
##   Lived 137  23
##   Died   33   7
```

```
icu.pre.chisq$expected
```

```
##        icu$pre
## icu$sta  No Yes
##   Lived 136  24
##   Died   34   6
```

```
icu.pre.chisq$residuals
```

```
##        icu$pre
## icu$sta          No         Yes
##   Lived  0.08574929 -0.20412415
##   Died  -0.17149859  0.40824829
```

```
icu.pre.chisq$residuals^2
```

```
##        icu$pre
## icu$sta          No         Yes
##   Lived 0.007352941 0.041666667
##   Died  0.029411765 0.166666667
```