Homework 5

STA 138 | Camden Possinger | Winter 2022

```
largework <- read.csv("largework.csv")

#largework$gender %<>% as.factor
#largework$marriage %<>% as.factor
#largework$y %<>% as.factor
```

1. Randomly subset about half of your data as a training set, and using this, report the best model according to forward stepwise selection with AIC that you obtain.

```
set.seed(502)

largework_len <- largework %>% nrow

train_set_ind <- sample(largework_len,largework_len / 2)

train_set <- largework[train_set_ind,]
test_set <- largework[-train_set_ind,]


best_model <- step(glm(y~1,family = "binomial",data = train_set),
                   scope = ~gender*age*marriage*min*chol*sysbp*height,
                   direction = "forward",
                   trace = 0)

print(best_model$formula)
```

```
## y ~ height + chol + sysbp + gender + height:chol + height:sysbp +
##     sysbp:gender
```

2. Using the model obtained above with the testing set, test for evidence of nonzero coefficients, using = 0.1; Interpret your results.

```
best_model_test_sum <- glm(y ~ chol + sysbp + gender + height:chol + height:sysbp + sysbp:gender, "bino

best_model_train_sum <- best_model %>% summary

print("Test Set: ")
```

```
## [1] "Test Set: "
```

```
best_model_test_sum$coefficients[,c(1,4)]
```

```
##                       Estimate  Pr(>|z|)
## (Intercept)       1.907059e+00 0.3710852
```

```
## chol              7.319362e-02 0.7495775
## sysbp            -7.724820e-03 0.9293080
## gendermale       -8.304435e-01 0.7573506
## chol:height      -6.745455e-04 0.6297526
## sysbp:height     -3.121454e-05 0.9547728
## sysbp:gendermale  7.576893e-03 0.7293370
```

```
print("Training Set: ")
```

```
## [1] "Training Set: "
```

```
best_model_train_sum$coefficients[,c(1,4)]
```

```
##                      Estimate    Pr(>|z|)
## (Intercept)      69.290742863 0.013967873
## height           -0.412383454 0.015227097
## chol              0.608727465 0.010045239
## sysbp            -0.715920106 0.003904143
## gendermale       -4.837254976 0.140821432
## height:chol      -0.004113503 0.006307355
## height:sysbp      0.004333275 0.003603149
## sysbp:gendermale  0.046155663 0.080793088
```

In this model using the test set there are no variables that are significant using $\alpha = 0.1$ which using Bonferroni correction would be $\alpha = 0.1/7$. This is interesting because when we use the training set there are variables that are significant.