Data-570-Lab-2

Dhun Sheth

2023-10-26

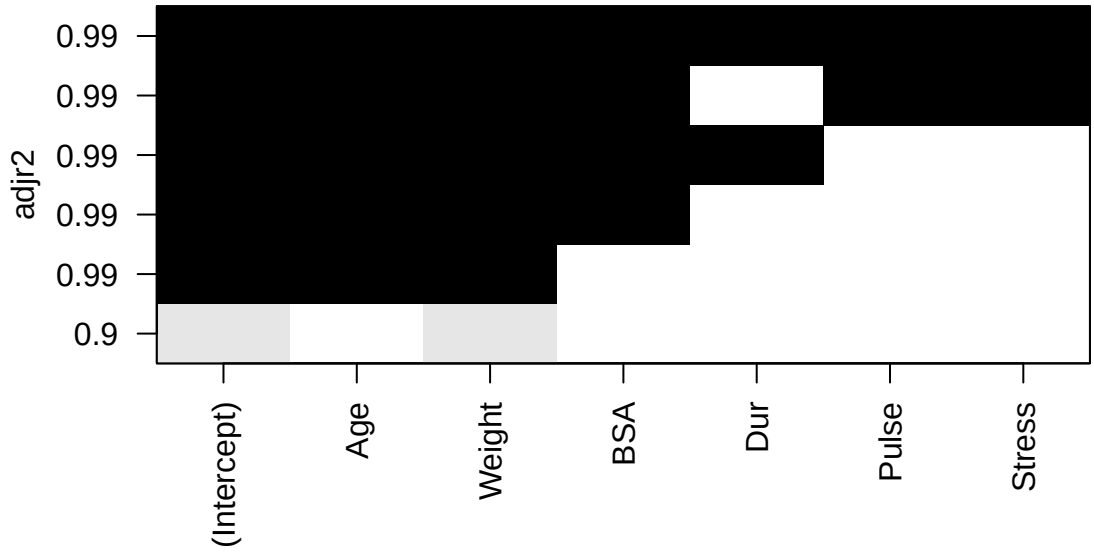
Question 7

Part 7.1

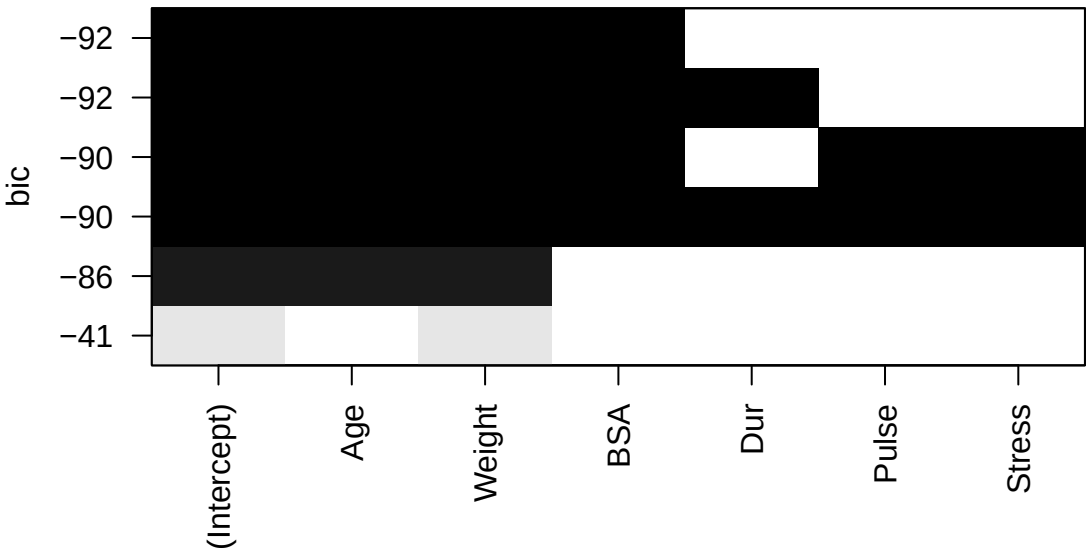
```
library(leaps)

bp=read.table("bloodpress.txt", header = TRUE, sep = "" )

bp_fit <- lm(BP~., data = bp)

leaps=regsubsets(BP~.,data=bp)
bp_reg_summary = summary(leaps)
bp_reg_summary$adjr2
```

```
## [1] 0.8972190 0.9903724 0.9935193 0.9938453 0.9939744 0.9943734
```

```
plot(leaps, scale = "adjr2")
```




```
plot(leaps, scale = "bic")
```



Part 7.2

```
alpha_enter = 0.1
alpha_remove = 0.1

# Step 1
summary(lm(BP~Age , data=bp))$coefficients[,4]
```

```
## (Intercept)         Age
## 0.028943443 0.001573655
```

```
summary(lm(BP~Weight , data=bp))$coefficients[,4]
```

```
##  (Intercept)       Weight
## 8.019513e-01 1.527885e-10
```

```
summary(lm(BP~BSA , data=bp))$coefficients[,4]
```

```
##  (Intercept)          BSA
## 1.400279e-04 8.114254e-07
```

```r
summary(lm(BP~Dur , data=bp))$coefficients[,4]
```

```
##  (Intercept)          Dur
## 2.207353e-16 2.102216e-01
```

```r
summary(lm(BP~Pulse , data=bp))$coefficients[,4]
```

```
##  (Intercept)        Pulse
## 0.0178708540 0.0003307037
```

```r
summary(lm(BP~Stress, data=bp))$coefficients[,4]
```

```
##  (Intercept)       Stress
## 5.556118e-21 4.898895e-01
```

```r
# Weight has lowest p-value, so on first step enter Weight
g1 <- lm(BP~ Weight, data=bp)
summary(update(g1,. ~ . +Age))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age
## 3.803805e-05 6.859831e-17 2.217640e-10
```

```r
summary(update(g1,. ~ . +BSA))$coefficients[,4]
```

```
##  (Intercept)       Weight          BSA
## 5.551796e-01 4.870718e-05 3.496199e-01
```

```r
summary(update(g1,. ~ . +Dur))$coefficients[,4]
```

```
##  (Intercept)       Weight          Dur
## 7.271130e-01 3.932098e-10 1.618116e-01
```

```r
summary(update(g1,. ~ . +Pulse))$coefficients[,4]
```

```
##  (Intercept)       Weight        Pulse
## 8.645697e-01 5.883065e-08 8.548800e-02
```

```r
summary(update(g1,. ~ . +Stress))$coefficients[,4]
```

```
##  (Intercept)       Weight       Stress
## 8.350993e-01 1.200453e-10 7.272597e-02
```

```r
# Age has the smallest p-value after regressing BP over Weight and Age
# Enter Age into step wise model
# Go back and check t-test P-value for Weight - it is 6.86e-17 which is smaller
# than alpha_remove so continue on to 3rd step
g2 <- update(g1,. ~ . +Age)
summary(update(g2,. ~ . +BSA))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age          BSA
## 9.415636e-05 3.198252e-12 3.001622e-11 7.764304e-03
```

```
summary(update(g2,. ~ . +Dur))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age          Dur
## 9.036004e-05 4.544005e-16 1.471097e-09 4.446388e-01
```

```
summary(update(g2,. ~ . +Pulse))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age        Pulse
## 3.404702e-05 3.400595e-15 1.355336e-09 1.947189e-01
```

```
summary(update(g2,. ~ . +Stress))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age       Stress
## 8.067721e-05 4.816343e-16 2.796059e-09 4.654863e-01
```

```
# BSA has smallest p-value, and t-test for both Weight and Age is still below
# alpha_remove so continue.
g3 <- update(g2,. ~ . +BSA)
summary(update(g3,. ~ . +Dur))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age          BSA          Dur
## 2.106578e-04 8.883014e-12 1.580481e-10 5.304674e-03 1.941837e-01
```

```
summary(update(g3,. ~ . +Pulse))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age          BSA        Pulse
## 1.482528e-04 2.616257e-10 1.146635e-09 2.192909e-02 6.575790e-01
```

```
summary(update(g3,. ~ . +Stress))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age          BSA       Stress
## 1.822496e-04 8.994316e-12 3.628000e-10 7.264139e-03 3.003656e-01
```

```
# All p-values for the remaining predictors are above our alpha_enter
# value, so we stop our step wise regression procedure.
```

The final regression model based on the step wise procedure contains only Weight, Age, and BSA predictors.

Part 7.3

```
alpha_enter = 0.1

# Step 1
summary(lm(BP~Age , data=bp))$coefficients[,4]
```

```
## (Intercept)         Age
## 0.028943443 0.001573655
```

```r
summary(lm(BP~Weight , data=bp))$coefficients[,4]
```

```
##  (Intercept)       Weight
## 8.019513e-01 1.527885e-10
```

```r
summary(lm(BP~BSA , data=bp))$coefficients[,4]
```

```
##  (Intercept)          BSA
## 1.400279e-04 8.114254e-07
```

```r
summary(lm(BP~Dur , data=bp))$coefficients[,4]
```

```
##  (Intercept)          Dur
## 2.207353e-16 2.102216e-01
```

```r
summary(lm(BP~Pulse , data=bp))$coefficients[,4]
```

```
##  (Intercept)        Pulse
## 0.0178708540 0.0003307037
```

```r
summary(lm(BP~Stress, data=bp))$coefficients[,4]
```

```
##  (Intercept)       Stress
## 5.556118e-21 4.898895e-01
```

```r
# Weight has lowest p-value, so on first step enter Weight
g1 <- lm(BP~ Weight, data=bp)
summary(update(g1,. ~ . +Age))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age
## 3.803805e-05 6.859831e-17 2.217640e-10
```

```r
summary(update(g1,. ~ . +BSA))$coefficients[,4]
```

```
##  (Intercept)       Weight          BSA
## 5.551796e-01 4.870718e-05 3.496199e-01
```

```r
summary(update(g1,. ~ . +Dur))$coefficients[,4]
```

```
##  (Intercept)       Weight          Dur
## 7.271130e-01 3.932098e-10 1.618116e-01
```

```r
summary(update(g1,. ~ . +Pulse))$coefficients[,4]
```

```
##  (Intercept)       Weight        Pulse
## 8.645697e-01 5.883065e-08 8.548800e-02
```

```r
summary(update(g1,. ~ . +Stress))$coefficients[,4]
```

```
##  (Intercept)       Weight       Stress
## 8.350993e-01 1.200453e-10 7.272597e-02
```

```r
# Age has the smallest p-value after regressing BP over Weight and Age
# Enter Age into step wise model
# For forward regression procedure do not go back and remove any predictors,
# so continue to next step.
g2 <- update(g1,. ~ . +Age)
summary(update(g2,. ~ . +BSA))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age          BSA
## 9.415636e-05 3.198252e-12 3.001622e-11 7.764304e-03
```

```r
summary(update(g2,. ~ . +Dur))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age          Dur
## 9.036004e-05 4.544005e-16 1.471097e-09 4.446388e-01
```

```r
summary(update(g2,. ~ . +Pulse))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age        Pulse
## 3.404702e-05 3.400595e-15 1.355336e-09 1.947189e-01
```

```r
summary(update(g2,. ~ . +Stress))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age       Stress
## 8.067721e-05 4.816343e-16 2.796059e-09 4.654863e-01
```

```r
# BSA has smallest p-value.
g3 <- update(g2,. ~ . +BSA)
summary(update(g3,. ~ . +Dur))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age          BSA          Dur
## 2.106578e-04 8.883014e-12 1.580481e-10 5.304674e-03 1.941837e-01
```

```r
summary(update(g3,. ~ . +Pulse))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age          BSA        Pulse
## 1.482528e-04 2.616257e-10 1.146635e-09 2.192909e-02 6.575790e-01
```

```r
summary(update(g3,. ~ . +Stress))$coefficients[,4]
```

```
##  (Intercept)       Weight          Age          BSA       Stress
## 1.822496e-04 8.994316e-12 3.628000e-10 7.264139e-03 3.003656e-01
```

```r
# All p-values for the remaining predictors are above our alpha_enter value,
# so we stop our forward regression procedure.
```

For forward regression, it works the same way as the step wise regression procedure except you never remove a predictor once it has been added. In part 7.2, the step wise regression procedure was performed and at no point were any predictors removed once the procedure began, so the result of the forward regression procedure is the same as the step wise procedure. That is, the final regression model from the forward regression procedure consists of Weight, Age, and BSA.

Part 7.4

```r
alpha_remove = 0.1

fit <- lm(BP~., data = bp)


# Start backward regression procedure
summary(fit)$coefficients
```

```
##                 Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) -12.87047602 2.55664988 -5.034118 2.287593e-04
## Age           0.70325939 0.04960581 14.176958 2.763784e-09
## Weight        0.96991978 0.06310846 15.369094 1.022456e-09
## BSA           3.77649100 1.58015087  2.389956 3.269369e-02
## Dur           0.06838304 0.04844149  1.411663 1.815344e-01
## Pulse        -0.08448469 0.05160898 -1.637015 1.255937e-01
## Stress        0.00557150 0.00341230  1.632770 1.264913e-01
```

```r
# Dur predictor has the largest p-value, so remove this predictor and continue.

fit <- update(fit, .~. - Dur)
summary(fit)$coefficients
```

```
##                  Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) -13.523474760 2.602082666 -5.197173 1.352831e-04
## Age           0.712341206 0.050901000 13.994641 1.269791e-09
## Weight        0.970929100 0.065303569 14.867933 5.720862e-10
## BSA           3.694072454 1.634101902  2.260613 4.023999e-02
## Pulse        -0.074485916 0.052902157 -1.407994 1.809498e-01
## Stress        0.006061205 0.003512922  1.725402 1.064490e-01
```

```r
# Pulse predictor has the largest p-value, so remove this predictor and continue.
fit <- update(fit, .~. - Pulse)
summary(fit)$coefficients
```

```
##                  Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) -13.175722304 2.673818103 -4.927681 1.822496e-04
## Age           0.681748392 0.047514026 14.348361 3.628000e-10
## Weight        0.907477215 0.048784813 18.601634 8.994316e-12
## BSA           4.703863677 1.515632469  3.103565 7.264139e-03
## Stress        0.003137336 0.002924738  1.072689 3.003656e-01
```

```r
# Stress predictor has the largest p-value, so remove this predictor and continue.
fit <- update(fit, .~. - Stress)
summary(fit)$coefficients
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) -13.6672473 2.64663630 -5.164007 9.415636e-05
## Age           0.7016198 0.04395951 15.960593 3.001622e-11
## Weight        0.9058223 0.04898945 18.490151 3.198252e-12
## BSA           4.6273883 1.52106609  3.042201 7.764304e-03
```

```r
# The remaining predictors all have t-value below the alpha_remove threshold,
# so none of them can be removed. Stop backward regression procedure here.
```

The final model after the backward regression procedure has ended includes the Age, Weight, and BSA predictors.