Heart Pulse

Tarek El-Hajjaoui

2023-02-09

Loading the dataset

```
file_path = '/Users/Tarek/Documents/UCI_MDS_Coding/Stats210P/R_Statistical_Modeling/HeartPulse/pulse.tx
df = read.table(file_path, header=TRUE, sep="", dec=".")
```

Summary of data set

```
str(df)
```

```
## 'data.frame':    232 obs. of  7 variables:
##  $ Active  : int  97 82 88 106 78 109 66 68 100 70 ...
##  $ Rest    : int  78 68 62 74 63 65 43 65 63 59 ...
##  $ Smoke   : int  0 1 0 0 0 0 0 0 0 0 ...
##  $ Gender  : int  1 0 0 0 1 0 1 0 0 1 ...
##  $ Exercise: int  1 3 3 3 3 3 3 3 1 2 ...
##  $ Hgt     : int  63 70 72 72 67 74 67 70 70 65 ...
##  $ Wgt     : int  119 225 175 170 125 188 140 200 165 115 ...
```

Correcting categorical column data types.

```
categorical_cols <- c('Smoke','Gender','Exercise')
df[categorical_cols] <- lapply(df[categorical_cols], as.factor)
```

Verifying column transformations

```
str(df)
```

```
## 'data.frame':    232 obs. of  7 variables:
##  $ Active  : int  97 82 88 106 78 109 66 68 100 70 ...
##  $ Rest    : int  78 68 62 74 63 65 43 65 63 59 ...
##  $ Smoke   : Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 1 1 1 ...
##  $ Gender  : Factor w/ 2 levels "0","1": 2 1 1 1 2 1 2 1 1 2 ...
##  $ Exercise: Factor w/ 3 levels "1","2","3": 1 3 3 3 3 3 3 3 1 2 ...
##  $ Hgt     : int  63 70 72 72 67 74 67 70 70 65 ...
##  $ Wgt     : int  119 225 175 170 125 188 140 200 165 115 ...
```

Summary of dataset

```
summary(df)
```

```
##      Active           Rest        Smoke   Gender  Exercise      Hgt
##  Min.   : 51.0   Min.   : 43.00   0:206   0:122   1: 41    Min.   :60.00
##  1st Qu.: 79.0   1st Qu.: 62.00   1: 26   1:110   2: 91    1st Qu.:65.00
##  Median : 88.5   Median : 68.00                   3:100    Median :68.00
##  Mean   : 91.3   Mean   : 68.35                            Mean   :68.25
##  3rd Qu.:102.0   3rd Qu.: 74.00                            3rd Qu.:71.00
##  Max.   :154.0   Max.   :106.00                            Max.   :78.00
##       Wgt
##  Min.   :102.0
##  1st Qu.:135.0
##  Median :150.0
##  Mean   :157.9
##  3rd Qu.:175.0
##  Max.   :260.0
```

*Model Creation below

Creating a model in which the response, Y, is Rest (resting heart rate), and the covariates, X, are Hgt (Height in inches), Wgt (Weight in pounds) and Smoke (smoking status - 1 for smokers and 0 for non-smokers), and an interaction between height and weight.

$$\hat{Y} = \hat{Rest}$$
$$X_1 = Hgt$$
$$X_2 = Wgt$$
$$X_3 = Smoke$$
$$X_4 = Hgt * Wgt$$

Multicorviate Linear Regression Full Model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 (X_1 * X_2)$$

```
full_model <- lm(Rest ~ Hgt + Wgt + Smoke + Hgt * Wgt, data=df)
```

Summary of full model

```
summary(full_model)
```

```
##
## Call:
## lm(formula = Rest ~ Hgt + Wgt + Smoke + Hgt * Wgt, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.405  -6.300  -0.815   5.667  34.342
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 181.484803  55.771278   3.254  0.00131 **
## Hgt          -1.611175   0.811581  -1.985  0.04832 *
## Wgt          -0.496353   0.371181  -1.337  0.18249
## Smoke1        5.751786   2.011254   2.860  0.00463 **
## Hgt:Wgt       0.006861   0.005251   1.307  0.19264
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.528 on 227 degrees of freedom
## Multiple R-squared:  0.09871,    Adjusted R-squared:  0.08283
## F-statistic: 6.215 on 4 and 227 DF,  p-value: 9.194e-05
```

Calculating the SSE (Sum of Square Errors) for full model

```
SSE <- sum(full_model$resid^2)
SSE
```

```
## [1] 20609.5
```

Comparing to built-in deviance( ) method

```
deviance(full_model)
```

```
## [1] 20609.5
```

Multicorviate Linear Regression Reduced Model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + + \hat{\beta}_3 X_3$$

```
reduced_model = lm(Rest ~ Hgt + Smoke, data=df)
```

Using ANOVA to test difference between the reduced model and full model.

```
anova(reduced_model, full_model)
```

```
## Analysis of Variance Table
##
## Model 1: Rest ~ Hgt + Smoke
## Model 2: Rest ~ Hgt + Wgt + Smoke + Hgt * Wgt
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    229 20781
## 2    227 20610  2    171.57 0.9449 0.3903
```

Sequential Sum of Squares Regression on the full model via ANOVA function.

Full Model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 (X_1 * X_2)$$

The Sequential Sum of Squares Regression on the full model is as follows:

$$anova(fullmodel) \rightarrow SSR(X_1), SSR(X_1|X_2), SSR(X_3|X_1, X_2), SSR(X_4|X_1, X_2, X_3)$$

```
anova_table <- anova(full_model)
anova_table
```

```
## Analysis of Variance Table
##
## Response: Rest
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## Hgt         1  1346.2 1346.18 14.8273 0.0001533 ***
## Wgt         1     0.0    0.03  0.0003 0.9857152
## Smoke       1   756.0  755.99  8.3267 0.0042833 **
## Hgt:Wgt     1   155.0  155.02  1.7075 0.1926369
## Residuals 227 20609.5   90.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Calculating the SSTO of the Sequential Sum of Squares Regression

$$SSTO = SSR(Hgt) + SSR(Wgt|Hgt) + SSR(Smoke|Hgt, Wgt) + SSR(Hgt * Wgt|Hgt, Wgt, Smoke) + SSE$$

```
SSTO <- sum(anova_table[, 2])
SSTO
```

```
## [1] 22866.72
```

Model: Y =Rest and X1=Hgt

```
model <- lm(Rest ~ Hgt, data=df)
```

Model summary

```
summary(model)
```

```
##
## Call:
## lm(formula = Rest ~ Hgt, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.153  -5.982  -0.571   5.565  33.618
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 112.4141    11.6346   9.662  < 2e-16 ***
## Hgt          -0.6457     0.1702  -3.793  0.00019 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.673 on 230 degrees of freedom
## Multiple R-squared:  0.05887,    Adjusted R-squared:  0.05478
## F-statistic: 14.39 on 1 and 230 DF,  p-value: 0.0001902
```

Model 2: Y=Rest , X1=Hgt, and X2=Wgt

```
model_2 <- lm(Rest ~ Hgt + Wgt, data=df)
```

Model 2 summary

```
summary(model_2)
```

```
##
## Call:
## lm(formula = Rest ~ Hgt + Wgt, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -26.159  -5.988  -0.580   5.556  33.630
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.123e+02  1.448e+01   7.755 2.89e-13 ***
## Hgt         -6.422e-01  2.602e-01  -2.468   0.0143 *
## Wgt         -5.383e-04  3.056e-02  -0.018   0.9860
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.694 on 229 degrees of freedom
## Multiple R-squared:  0.05887,    Adjusted R-squared:  0.05065
## F-statistic: 7.163 on 2 and 229 DF,  p-value: 0.0009611
```

Using ANOVA to test difference between the reduced model and full model.

```
anova(model, model_2)
```

```
## Analysis of Variance Table
##
## Model 1: Rest ~ Hgt
## Model 2: Rest ~ Hgt + Wgt
##   Res.Df   RSS Df Sum of Sq     F Pr(>F)
## 1    230 21520
## 2    229 21520  1  0.029168 3e-04  0.986
```

Predict the resting heart rate for someone who is 70 inches tall and weighs 170 pounds:

```
predict(model_2, data.frame(Hgt=70, Wgt=170), se.fit=TRUE)
```

```
## $fit
##        1
## 67.21598
##
## $se.fit
## [1] 0.7037356
##
## $df
## [1] 229
##
## $residual.scale
## [1] 9.694123
```