Regression Analysis

Student

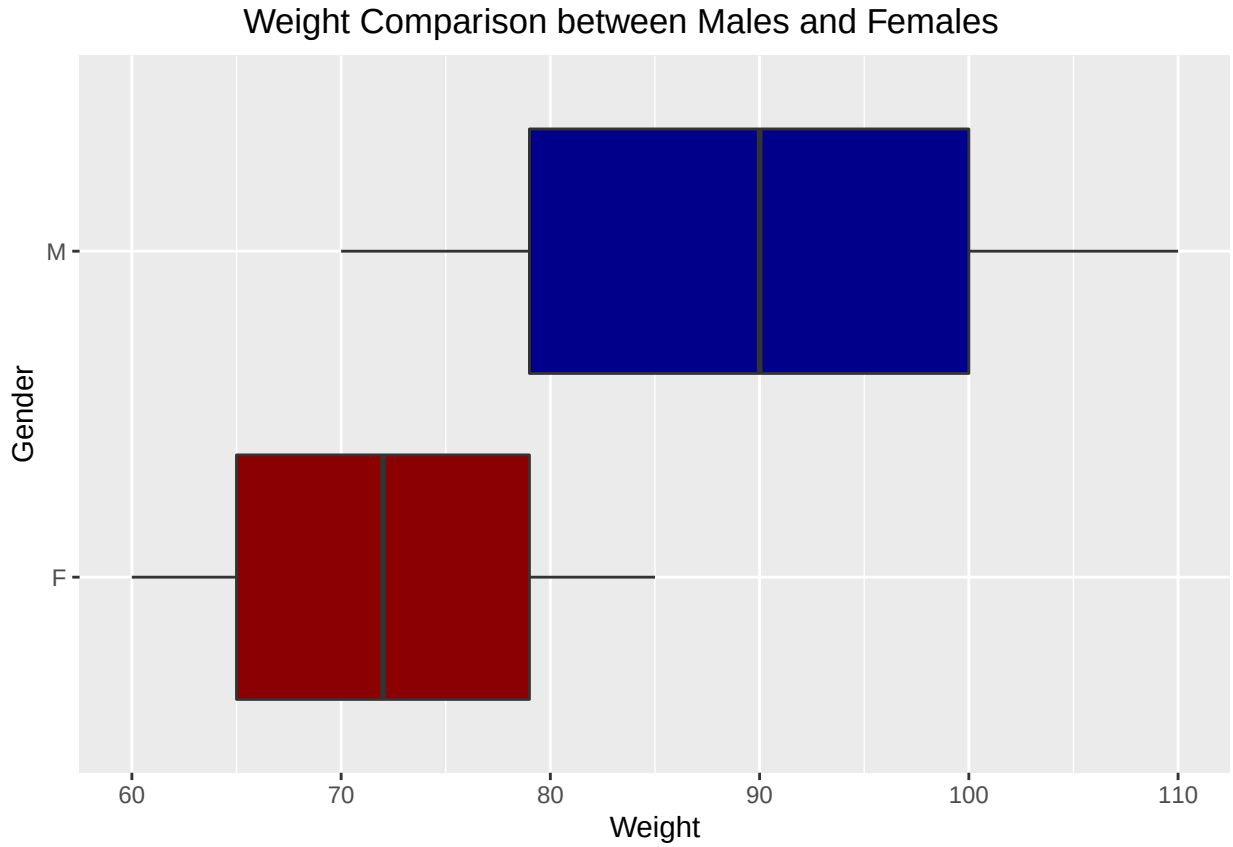
8/25/2022

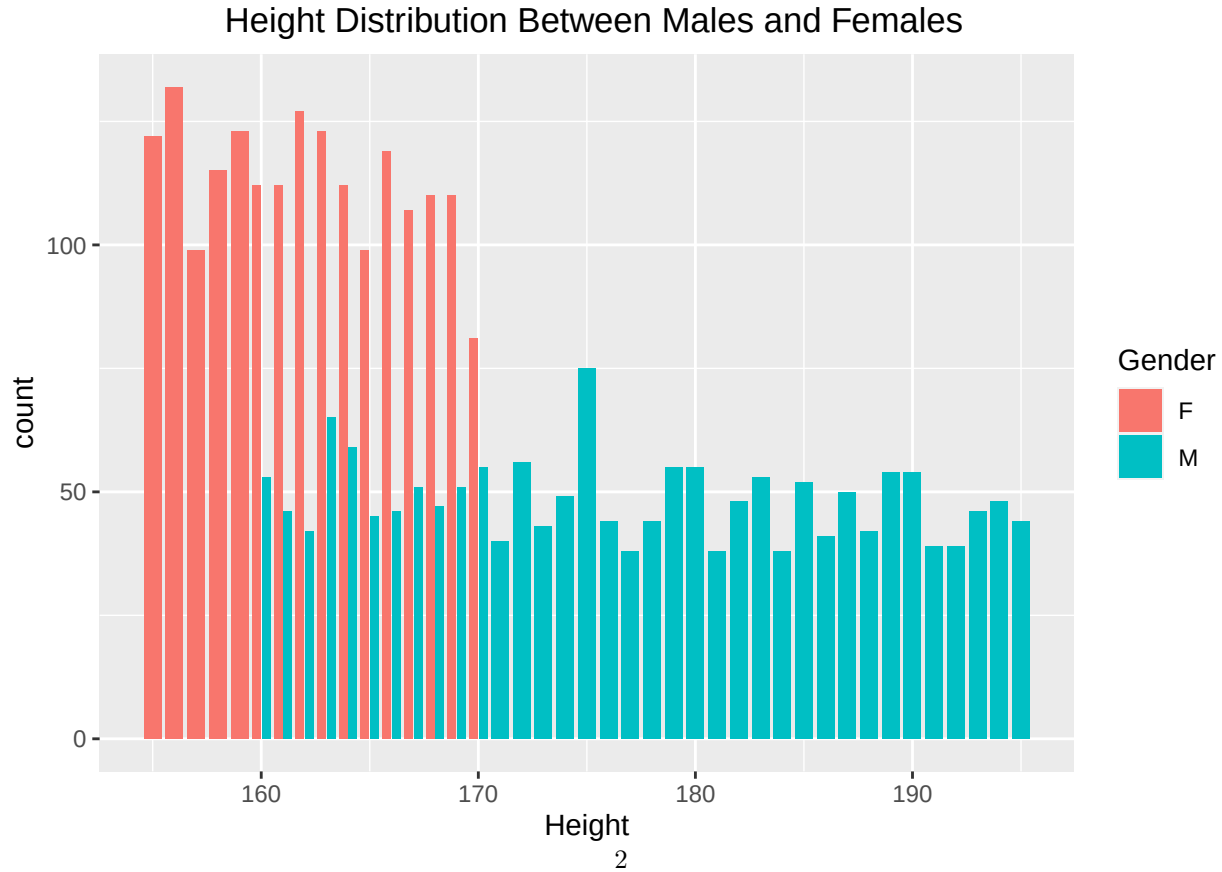
Quick recap of the data:

```
## Rows: 3,548
## Columns: 6
## $ Success  <int> 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0~
## $ Height   <int> 168, 162, 175, 169, 181, 173, 183, 173, 160, 190, 165, 188, 1~
## $ Weight   <int> 69, 99, 81, 75, 99, 74, 73, 83, 63, 96, 61, 82, 71, 83, 80, 8~
## $ Age      <int> 29, 54, 46, 68, 22, 20, 53, 65, 62, 52, 21, 65, 26, 39, 41, 5~
## $ Gender   <chr> "F", "M", "M", "F", "M", "M", "M", "M", "F", "M", "F", "M", "~
## $ Diabetes <int> 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1~
```

```
##   Success Height Weight Age Gender Diabetes
## 1       1    168     69  29      F        0
## 2       0    162     99  54      M        1
## 3       1    175     81  46      M        0
## 4       1    169     75  68      F        0
## 5       1    181     99  22      M        1
## 6       1    173     74  20      M        0
```

Plotting the data:






Correlation between Height & Weight Per Gender
Height
190
180
170
160
60
70
80
90
100
110
Weight
Gender
F
M

Linear Regression:

```
##
## Call:
## lm(formula = Weight ~ Gender + Height + Diabetes, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.1973  -8.3295   0.0426   7.9472  20.5219
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 70.17044    3.42536  20.486   <2e-16 ***
## GenderM     17.43864    0.45760  38.109   <2e-16 ***
## Height       0.01341    0.02105   0.637    0.524
## Diabetes    -0.30351    0.33454  -0.907    0.364
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 9.962 on 3544 degrees of freedom
## Multiple R-squared:  0.4396, Adjusted R-squared:  0.4392
## F-statistic: 926.9 on 3 and 3544 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = Height ~ Gender + Age + Diabetes + Success, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -17.4206  -5.3607  -0.1759   5.5842  18.2771
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 162.546823   0.464771 349.736   <2e-16 ***
## GenderM      14.828938   0.266959  55.548   <2e-16 ***
## Age          -0.008796   0.007584  -1.160    0.246
## Diabetes      0.222717   0.266912   0.834    0.404
## Success       0.042038   0.266969   0.157    0.875
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.948 on 3543 degrees of freedom
## Multiple R-squared:  0.4658, Adjusted R-squared:  0.4652
## F-statistic: 772.4 on 4 and 3543 DF,  p-value: < 2.2e-16
```

Comparison Between the 2 models

```
##
## =====================================================================
##                                   Dependent variable:
##                     -------------------------------------------------
##                              Weight                    Height
##                               (1)                        (2)
## ---------------------------------------------------------------------
## GenderM                     17.439***                  14.829***
##                         (16.542, 18.336)          (14.306, 15.352)
##
## Height                       0.013
##                          (-0.028, 0.055)
##
## Age                                                      -0.009
##                                                    (-0.024, 0.006)
##
## Diabetes                    -0.304                       0.223
##                          (-0.959, 0.352)            (-0.300, 0.746)
##
## Success                                                  0.042
##                                                    (-0.481, 0.565)
##
## Constant                    70.170***                 162.547***
##                         (63.457, 76.884)         (161.636, 163.458)
##
## ---------------------------------------------------------------------
## Observations                 3,548                       3,548
## R2                           0.440                       0.466
## Adjusted R2                  0.439                       0.465
## Residual Std. Error     9.962 (df = 3544)          7.948 (df = 3543)
## F Statistic         926.873*** (df = 3; 3544) 772.389*** (df = 4; 3543)
## =====================================================================
## Note:                                     *p<0.1; **p<0.05; ***p<0.01
```

Conclustion:

It seems that model1 and model2 are quite similar, however model2 has a greater level of R^2 and thus, slightly better than model1

Regression Equations:

For Model1:

$y_i = 70.17 + 17.44 * M + 0.01 * Height - 0.303 * Diabetes$

For Model2:

$y_i = 162.55 + 14.83 * M - 0.01 * Age + 0.22 * Diabetes + 0.04 * Success$