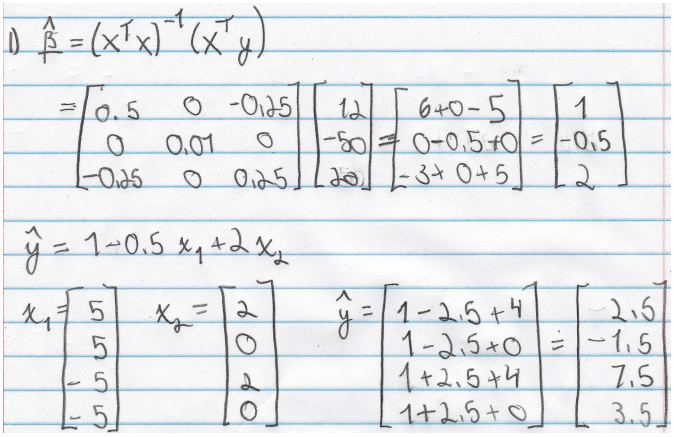
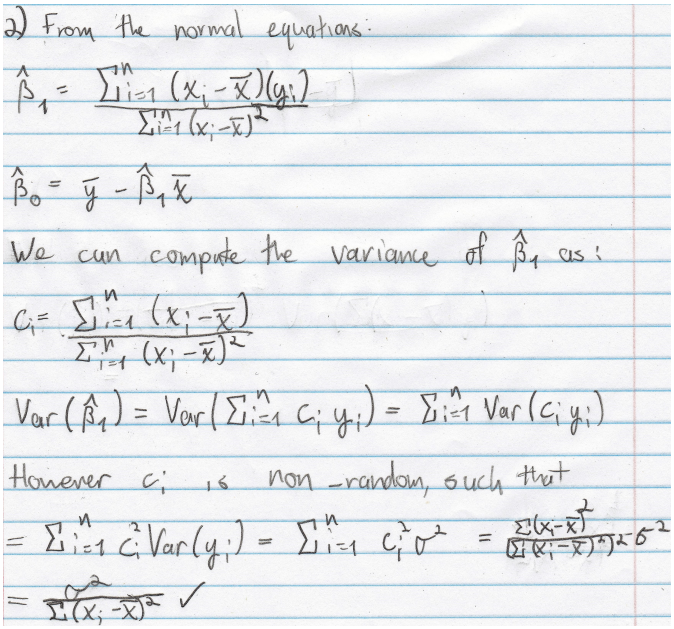
Kyle Salitrik (kps168)
STAT 462
Homework 2 Report

Theoretical Solutions

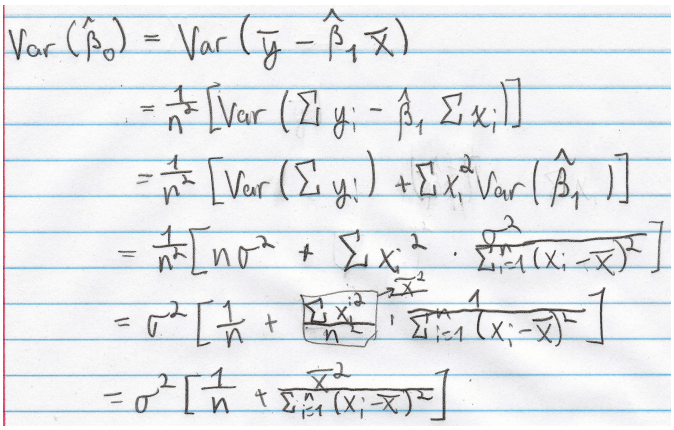
Numerical Computations



Proof for $\hat{\beta}_1$



Proof for $\hat{\beta}_0$



Applied Analysis

Part A

The following reduction on the data was performed:

- Samples below 2% and above 40% body fat were excluded. The minimum was obtained through the chart on the (ACE) website and maximum was set to limit outliers.
- The minimum height accepted was cut to be over 50 inches to remove outliers.
- The maximum weight was also clipped to be below 300 pounds. Although the outlier may have been legitimate, it could have also been an error.

For each individual linear model created between the response of body fat percentage and predictors weight, height and abdomen circumference, the following hypotheses were assumed.

$$H_0 : \beta_0 = \beta_1 = 0$$
$$H_1 : \exists \beta_i \neq 0$$

To test these hypotheses, a T-distribution test was performed on each model with the resultant p-values:

- Weight Predictor: p $<$ 2e-16
- Height Predictor: p = 0.805
- Abdomen Circumference Predictor: p $<$2e-16

After performing the tests, we cannot reject the null hypothesis for the height as a predictor of body fat percentage. We can however reject the null hypotheses for the weight and abdomen circumference, implying that there is some non-zero slope relationship between the body fat percentage and these predictors.

Based on our regression models, we can obtain the following 99% confidence intervals for β_1 of each model:

- For each pound increase in weight, we are 99% confident that, on average, the body fat percentage will increase between 0.142 and 0.221 percent.
- For each inch increase in height, we are 99% confident that, on average, the body fat percentage will increase between 0.571 and 0.471 percent.
- For each cm increase in abdomen circumference, we are 99% confident that, on average, the body fat percentage will increase between 0.574 and 0.729 percent.

As we can see from the confidence intervals above, not rejecting the null hypothesis for the height as a predictor is further reinforced due to the inclusion of 0 within the interval.

Looking at the coefficients of determination for each model, the values for the weight($R^2 = 0.3658$) and abdomen circumference ($R^2 = 0.6607$) models are significantly higher than that of the height model ($R^2 = 0.0611$). Again, looking at the confidence intervals, we can see that the abdomen circumference has a lower bound that is much higher (more than twice) than the upper bound for the weight, indicating that abdomen circumference is the strongest predictor out of the three.

Part B

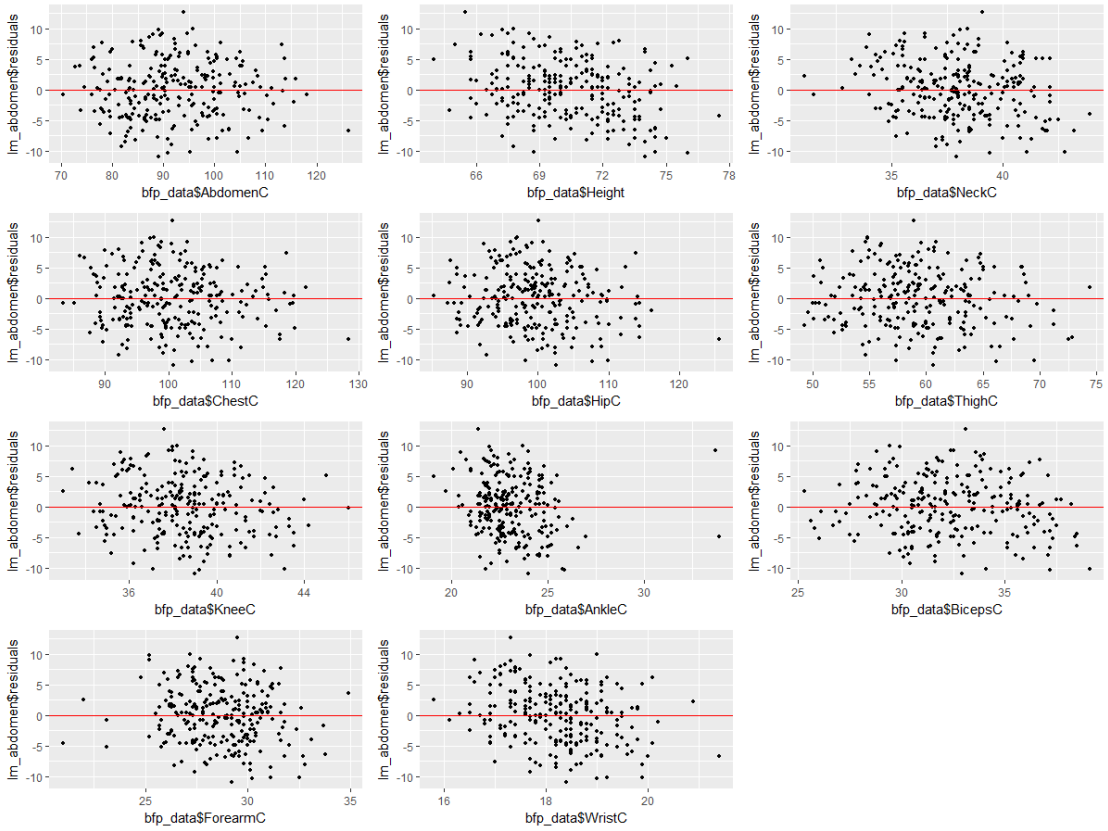
Next, we are interested whether or not there is evidence to indicate that the average increase in body fat percentage for each cm increase of abdomen circumference is actually greater than 0.5%. A t-test was performed on the following hypotheses.

$$H_0 : \beta_1 \leq 0.5$$
$$H_1 : \beta_1 > 0.5$$

The t-value obtained was 5.067856 and an underlying normal distribution was assumed for β_1. As such, the p-value was calculated using $1 - pnorm(t)$ in R to obtain a p-value of $2.011611 * 10^-7$. With the p-value confirming that our confidence interval for the slope is correct, and that the actual value for β_1 is over 0.5.

Part C

The following figure includes the residuals from the abdomen circumference predictor model plotted against weight, height, neck circumference, chest circumference, hip circumference, thigh circumference, knee circumference, ankle circumference, biceps circumference, forearm circumference, and wrist circumference.



From these plots, we can see that there may potentially be a relationship between the body fat percentage and the ankle circumference. It may make sense that wrist circumference could indicate an increase in fat percentage, based on where the wrist is defined to lie. If the measurements were taken further up the forearm where one would wear a watch, as opposed to the actual wrist joint, then the measurement may indicate something about the body fat percentage.

Part D

Next if we fit a linear model (stated below) using all of these variables to predict the body fat percentage, we will be able to compare it to the reduced (abdomen circumference) only model in order to determine whether the other predictors are

actually necessary. The test statistic used for this was an f-distribution (also noted below).

$$\begin{aligned}
BFP =& \beta_0 + \beta_1 * AbdomenC + \beta_2 * Weight + \beta_3 * Height + \beta_4 * NeckC + \\
& \beta_5 * ChestC + \beta_6 * HipC + \beta_7 * ThighC + \beta_8 * KneeC + \beta_9 * AnkleC + \\
& \beta_10 * BicepsC + \beta_11 * ForearmC + \beta_12 * WristC + \epsilon \\
\widehat{BFP} =& 1.36987591 + 0.98199026 * AbdomenC - 0.04420200 * Weight - \\
& 0.26163775 * Height - 0.38320373 * NeckC - 0.10481085 * ChestC - \\
& 0.20730987 * HipC + 0.06182966 * ThighC + 0.14670469 * KneeC + \\
& 0.12543748 * AnkleC + 0.17961142 * BicepsC + 0.23096461 * ForearmC - \\
& 1.40026669 * WristC \\
\hat{\sigma}^2 =& 17.9776 \\
R^2 =& 0.7304
\end{aligned}$$

The F-value used in the f-test was computed to be the following value, and using the given parameters:

$$\begin{aligned}
H_0 :& \quad \omega : y_i = \beta_0 + \beta_1 * AbdomenC_i \\
H_1 :& \quad \Omega : \text{The full model stated above where } \exists \beta_i \neq 0, i = [2, 12] \\
n =& 246, \quad p = 13, \quad q = 2 \\
F \sim& F_{11,233} \\
F_{value} =& 5.482263 \\
p - value =& 9.011923 * 10^{-8}
\end{aligned}$$

Based on the results of the F-test and looking at the R-squared values, we cannot reject the null hypothesis that the abdomen circumference model is not strong enough to predict the body fat percentage accurately.

R Output

This section contains the output from the R console for important values and summaries.

```
##############################################################################
#### Display Results
##############################################################################

## Part A
summary(lm_weight)
  Residuals:
       Min        1Q    Median        3Q       Max
  -18.1734   -4.5946    0.0303    4.8440   18.5365
  Coefficients:
               Estimate Std. Error t value Pr(>|t|)
  (Intercept) -13.30646    2.76104  -4.819 2.53e-06 ***
  Weight        0.18185    0.01533  11.864  < 2e-16 ***
  ---
  Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
  Residual standard error: 6.355 on 244 degrees of freedom
  Multiple R-squared:  0.3658,  Adjusted R-squared:  0.3632
  F-statistic: 140.8 on 1 and 244 DF,  p-value: < 2.2e-16

summary(lm_height)
  Residuals:
       Min        1Q    Median        3Q       Max
  -16.2242   -6.4851    0.2486    6.0752   20.7394
  Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  22.5848     14.1206    1.599    0.111
Height       -0.0496      0.2006   -0.247    0.805
Residual standard error: 7.979 on 244 degrees of freedom
Multiple R-squared:  0.0002504, Adjusted R-squared:  -0.003847
F-statistic: 0.06111 on 1 and 244 DF,  p-value: 0.805

summary(lm_abdomen)
Residuals:
     Min       1Q   Median       3Q      Max
-10.9111  -3.5650   0.1989   3.1421  12.7618
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -41.03612    2.77484  -14.79   <2e-16 ***
AbdomenC      0.65148    0.02989   21.80   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Residual standard error: 4.648 on 244 degrees of freedom
Multiple R-squared:  0.6607,  Adjusted R-squared:  0.6593
F-statistic:   475 on 1 and 244 DF,  p-value: < 2.2e-16

lm_weight_conf
                    0.5 %     99.5 %
(Intercept) -20.4744889 -6.1384398
Weight        0.1420582  0.2216441

lm_height_conf
                    0.5 %     99.5 %
(Intercept) -14.0740731 59.2436692
Height       -0.5705262  0.4713198

lm_abdomen_conf
                    0.5 %     99.5 %
(Intercept) -48.2399646 -33.8322748
AbdomenC      0.5738825   0.7290846

# Part B
t_abdomen
[1] 5.067856
p_abdomen
[1] 2.011611e-07


# Part D
f
[1] 5.482263
pvalue
[1] 9.011923e-08

# Check results
anova(lm_abdomen,lm_full)
Analysis of Variance Table

Model 1: SiriBFperc ~ AbdomenC
Model 2: SiriBFperc ~ (Density + AbdomenC + Weight + Height + NeckC +
    ChestC + HipC + ThighC + KneeC + AnkleC + BicepsC + ForearmC +
    WristC) - Density
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    244 5272.4
2    233 4188.4 11      1084 5.4823 9.012e-08 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```