**STAT 40001/STAT 50001 Statistical Computing Fall 2024**

**Lab -17**

1. The percentage of a person’s weight that is made up of body fat is often used as an indicator of health and fitness. However, accurate methods of measuring percent body fat are difficult to implement. One method involves immersing the body in water to estimate its density and then applying a formula to estimate percent body fat. An alternative is to develop a model for percent body fat that is based on body characteristics such as height and weight that are easy to measure. The dataset *BodyFat* in *Lock5Data* contains such measurements for a sample of 100 men.
2. Import the data and Access the variable names included in the dataset.
3. Fit a model to predict Bodyfat using Height and Weight. Comment on whether either of the predictors appears to be important in the model.
4. Add Abdomen as a third predictor to the model (b) and repeat the assessment of the effectiveness of each predictor.
5. Interpret the coefficient of Abdomen you get in the model part (c).

> #Q1

> install.packages("Lock5Data")

> library(Lock5Data)

> data(BodyFat)

> head(BodyFat)

Bodyfat Age Weight Height Neck Chest Abdomen Ankle Biceps Wrist

1 32.3 41 247.25 73.50 42.1 117.0 115.6 26.3 37.3 19.7

2 22.5 31 177.25 71.50 36.2 101.1 92.4 24.6 30.1 18.2

3 22.0 42 156.25 69.00 35.5 97.8 86.0 24.0 31.2 17.4

4 12.3 23 154.25 67.75 36.2 93.1 85.2 21.9 32.0 17.1

5 20.5 46 177.00 70.00 37.2 99.7 95.6 22.5 29.1 17.7

6 22.6 54 198.00 72.00 39.9 107.6 100.0 22.0 35.9 18.9

> dim(BodyFat)

[1] 100 10

> names(BodyFat)

[1] "Bodyfat" "Age" "Weight" "Height" "Neck" "Chest" "Abdomen" "Ankle"

[9] "Biceps" "Wrist"

> attach(BodyFat)

> model1 <- lm(Bodyfat~Weight+Height)

> model1

Call:

lm(formula = Bodyfat ~ Weight + Height)

Coefficients:

(Intercept) Weight Height

71.4825 0.2316 -1.3357

> cat("Fitted Model: Bodyfat = 71.4825 + 0.2316\*Weight-1.3357\*Height")

Fitted Model: Bodyfat = 71.4825 + 0.2316\*Weight-1.3357\*Height

> summary(model1)

Call:

lm(formula = Bodyfat ~ Weight + Height)

Residuals:

Min 1Q Median 3Q Max

-12.7697 -3.9527 -0.5364 4.0473 13.2829

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 71.48247 16.20086 4.412 2.65e-05 \*\*\*

Weight 0.23156 0.02382 9.721 5.36e-16 \*\*\*

Height -1.33568 0.25891 -5.159 1.32e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.754 on 97 degrees of freedom

Multiple R-squared: 0.494, Adjusted R-squared: 0.4836

F-statistic: 47.35 on 2 and 97 DF, p-value: 4.48e-15

> cat("Both the Height as well as the Weight are significant")

Both the Height as well as the Weight are significant

> model2 <- lm(Bodyfat~Weight+Height+Abdomen)

> model2

Call:

lm(formula = Bodyfat ~ Weight + Height + Abdomen)

Coefficients:

(Intercept) Weight Height Abdomen

-56.1329 -0.1756 0.1018 1.0747

> cat("Fitted Model: Bodyfat = -56.1329 -0.1756\*Weight+ 0.1018\*Height + 1.0747\*Abdomen")

Fitted Model: Bodyfat = -56.1329 -0.1756\*Weight+ 0.1018\*Height + 1.0747\*Abdomen

> summary(model2)

Call:

lm(formula = Bodyfat ~ Weight + Height + Abdomen)

Residuals:

Min 1Q Median 3Q Max

-9.5219 -2.9969 0.0378 2.8933 9.2859

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -56.1329 18.1372 -3.095 0.002580 \*\*

Weight -0.1756 0.0472 -3.720 0.000335 \*\*\*

Height 0.1018 0.2444 0.417 0.677750

Abdomen 1.0747 0.1158 9.279 5.27e-15 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.199 on 96 degrees of freedom

Multiple R-squared: 0.7332, Adjusted R-squared: 0.7249

F-statistic: 87.96 on 3 and 96 DF, p-value: < 2.2e-16

> cat("Upon introduction of Abdomen, the Height variable becomes not significant")

Upon introduction of Abdomen, the Height variable becomes not significant

> cat("Comparing the co-efficient of the model2, it is found that, for abdomen change in 1 cm, the bodyfat increase by 1.0747%, provided other variable i.e. height and weight variable are kept constant")

Comparing the co-efficient of the model2, it is found that, for abdomen change in 1 cm, the bodyfat increase by 1.0747%, provided other variable i.e. height and weight variable are kept constant

1. A pediatrician wants to determine the relation that may exist between a child‘s head circumference (in centimeters), height (in inches), and weight (in ounces). She randomly selects 14 three-year-old children from her practice and obtains the following data:

|  |  |  |
| --- | --- | --- |
| Height | Weight | Head Circumference |
| 30 | 339 | 47 |
| 26.25 | 267 | 42 |
| 25 | 289 | 43 |
| 27 | 332 | 44.5 |
| 27.5 | 272 | 44 |
| 24.5 | 214 | 40.5 |
| 27.75 | 311 | 44 |
| 25 | 259 | 41.5 |
| 28 | 298 | 46 |
| 27.25 | 288 | 44 |
| 26 | 277 | 44 |
| 27.25 | 292 | 44.5 |
| 27 | 302 | 42.5 |
| 28.25 | 336 | 44.5 |

1. Construct a correlation matrix. Is there any reason to be concerned with multicollinearity?
2. Find the least-squares regression equation with the response variable, head circumference.
3. Construct 95% confidence and prediction intervals for the head circumference of a child whose height is 27.5 inches and weight is 285 ounces.
4. Perform the residual analysis of the model.

> #Q2

> Height <- scan()

1: 30.00 26.25 25.00 27.00 27.50 24.50 27.75 25.00 28.00 27.25 26.00 27.25 27.00 28.25

15:

Read 14 items

> # 30.00 26.25 25.00 27.00 27.50 24.50 27.75 25.00 28.00 27.25 26.00 27.25 27.00 28.25

> Weight <- scan()

1: 339 267 289 332 272 214 311 259 298 288 277 292 302 336

15:

Read 14 items

> # 339 267 289 332 272 214 311 259 298 288 277 292 302 336

> head\_circumference <- scan()

1: 47.0 42.0 43.0 44.5 44.0 40.5 44.0 41.5 46.0 44.0 44.0 44.5 42.5 44.5

15:

Read 14 items

> # 47.0 42.0 43.0 44.5 44.0 40.5 44.0 41.5 46.0 44.0 44.0 44.5 42.5 44.5

> Height; Weight; head\_circumference

[1] 30.00 26.25 25.00 27.00 27.50 24.50 27.75 25.00 28.00 27.25 26.00 27.25 27.00 28.25

[1] 339 267 289 332 272 214 311 259 298 288 277 292 302 336

[1] 47.0 42.0 43.0 44.5 44.0 40.5 44.0 41.5 46.0 44.0 44.0 44.5 42.5 44.5

> data <- data.frame(Height, Weight, head\_circumference)

> data

Height Weight head\_circumference

1 30.00 339 47.0

2 26.25 267 42.0

3 25.00 289 43.0

4 27.00 332 44.5

5 27.50 272 44.0

6 24.50 214 40.5

7 27.75 311 44.0

8 25.00 259 41.5

9 28.00 298 46.0

10 27.25 288 44.0

11 26.00 277 44.0

12 27.25 292 44.5

13 27.00 302 42.5

14 28.25 336 44.5

> dim(data)

[1] 14 3

> cor(data)

Height Weight head\_circumference

Height 1.0000000 0.7847652 0.8708869

Weight 0.7847652 1.0000000 0.7796990

head\_circumference 0.8708869 0.7796990 1.0000000

> cat("Looking at the correleation matrix, there can be a chance of collinearity")

Looking at the correleation matrix, there can be a chance of collinearity

> install.packages("car")

Error in install.packages : Updating loaded packages

> install.packages("car")

> library(car)

> model = lm(head\_circumference~Height+Weight)

> model

Call:

lm(formula = head\_circumference ~ Height + Weight)

Coefficients:

(Intercept) Height Weight

18.82425 0.78634 0.01281

> summary(model)

Call:

lm(formula = head\_circumference ~ Height + Weight)

Residuals:

Min 1Q Median 3Q Max

-1.42356 -0.55408 0.06335 0.44183 1.34134

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 18.82425 4.94099 3.810 0.00289 \*\*

Height 0.78634 0.26455 2.972 0.01269 \*

Weight 0.01281 0.01160 1.105 0.29290

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8707 on 11 degrees of freedom

Multiple R-squared: 0.7826, Adjusted R-squared: 0.743

F-statistic: 19.79 on 2 and 11 DF, p-value: 0.0002266

> cat("Found that weight variable is not significant")

Found that weight variable is not significant

> vif(model)

Height Weight

2.603193 2.603193

> # no much collinearity

> # model accepted

> cat("Fitted Model: head\_circumference = 18.82425+ 0.78634\*Height + 0.01281\*Weight ")

Fitted Model: head\_circumference = 18.82425+ 0.78634\*Height + 0.01281\*Weight

> predict(model, data.frame(Height = 27.5, Weight = 285))

1

44.09898

> predict(model, data.frame(Height = 27.5, Weight = 285), interval = "conf", level = 0.95)

fit lwr upr

1 44.09898 43.39962 44.79834

> cat("conf interval range: [43.39962, 44.79834]")

conf interval range: [43.39962, 44.79834]

> predict(model, data.frame(Height = 27.5, Weight = 285), interval = "pred", level = 0.95)

fit lwr upr

1 44.09898 42.05886 46.1391

> cat("Prediction interval range: [42.05886, 46.1391]")

Prediction interval range: [42.05886, 46.1391]

> #residual

> plot(model, 1)

> install.packages("MASS")

Error in install.packages : Updating loaded packages

> install.packages("MASS")

> library(MASS)

> #preform boxcox transformation

> b = boxcox(model)

> y1 <- head\_circumference^(0.7)

> model1 = lm(y1~Height+Weight)

> model1

Call:

lm(formula = y1 ~ Height + Weight)

Coefficients:

(Intercept) Height Weight

8.475787 0.176043 0.002951

> cat("Boxcox Fitted Model: head\_circumference^0.7 = 8.475787+ 0.176043\*Height + 0.002951\*Weight")

Boxcox Fitted Model: head\_circumference^0.7 = 8.475787+ 0.176043\*Height + 0.002951\*Weight

> summary(model1)

Call:

lm(formula = y1 ~ Height + Weight)

Residuals:

Min 1Q Median 3Q Max

-0.32047 -0.12455 0.01760 0.09972 0.30132

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 8.475787 1.113708 7.610 1.05e-05 \*\*\*

Height 0.176043 0.059631 2.952 0.0132 \*

Weight 0.002951 0.002614 1.129 0.2828

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1963 on 11 degrees of freedom

Multiple R-squared: 0.7827, Adjusted R-squared: 0.7432

F-statistic: 19.81 on 2 and 11 DF, p-value: 0.0002257

> cat("Found that weight variable is not significant")

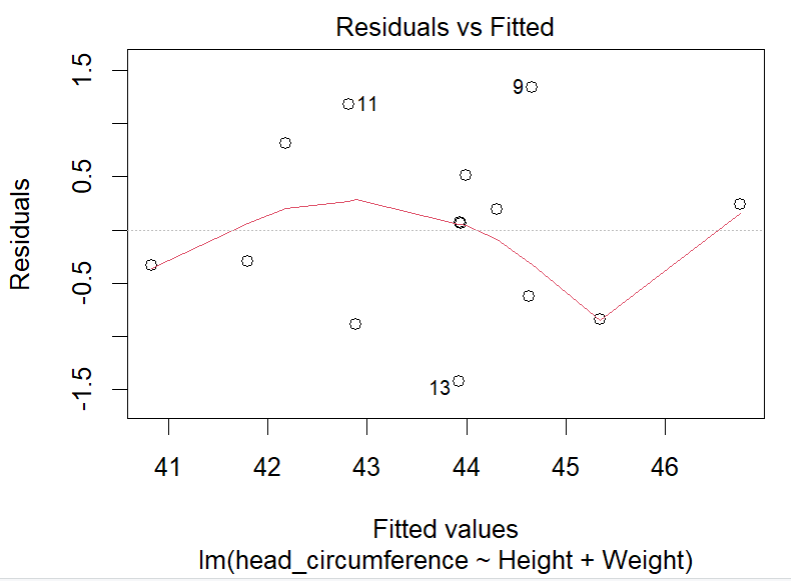
Found that weight variable is not significant

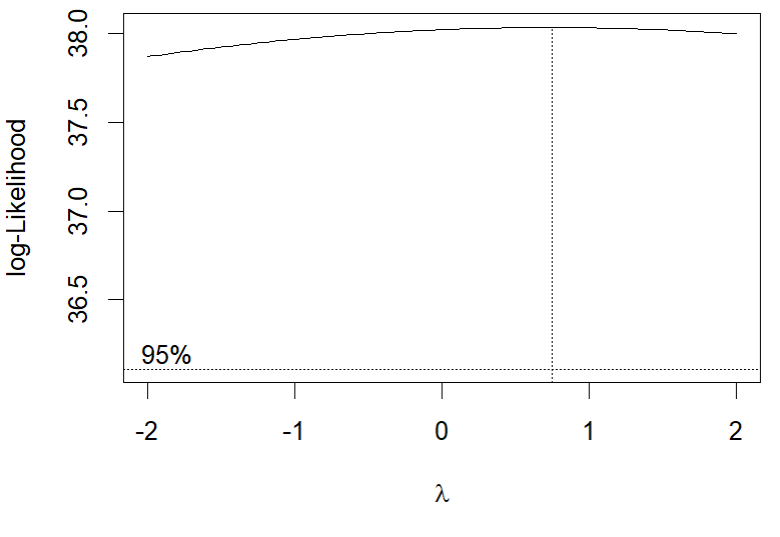
> vif(model1)

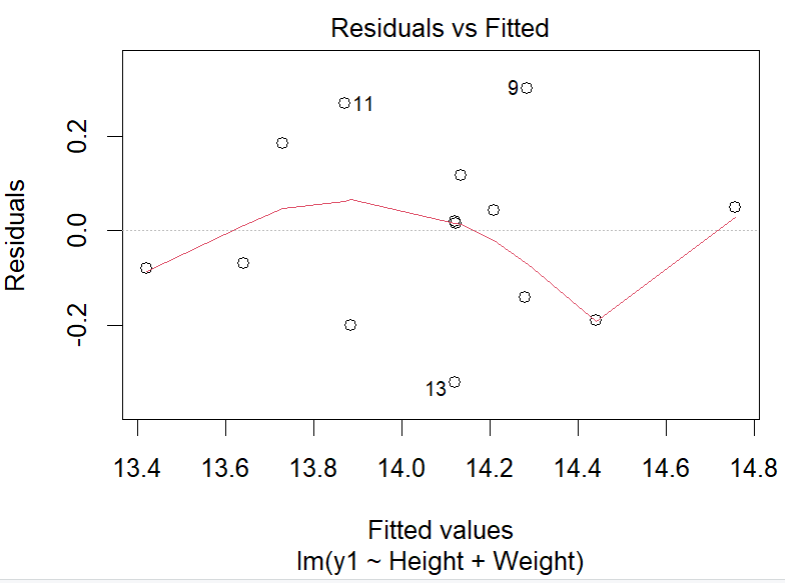
Height Weight

2.603193 2.603193

> plot(model1, 1)





****