Linear Relationship Between Boday Fat Percentage and Body Part Circumference

Ruiying Liu

2023-03-17

Table of Contents

[SUMMARY 1](#_Toc129976833)

[INTRODUCTION 2](#_Toc129976834)

[Data description 2](#_Toc129976835)

[Summary Statistics 3](#_Toc129976836)

[Data anaylysis 4](#_Toc129976837)

[Scatter plots 5](#_Toc129976838)

[R-square 6](#_Toc129976839)

[Factor analysis 7](#_Toc129976840)

[Determine the number of factors 7](#_Toc129976841)

[Extract the factors and Interprest the factors 7](#_Toc129976842)

[Linear model and visualization 8](#_Toc129976843)

[Model Diagnostics 9](#_Toc129976844)

[Conclusion 9](#_Toc129976845)

[Citation 9](#_Toc129976846)

[APPENDIX: The Code 10](#_Toc129976847)

# SUMMARY

Body fat percentage(BFP) is a measure of the proportion of fat tissue in the human body. It is a very important physiological indicator for maintaining health, improving body composition and optimizing athletic performance, so BFP measurement is essential for our daily life. While in the laboratory, BFP can be accurately measured by the techniques such as densitometry, hydrometry, dual energy X-ray absorptiometry, and computed comography (CT) or magnetic rosonance imaging(MRI) imaging (Deurenberg,1999). However, these methods can be costly and inconvenient, and are not practical for field use. To estimate BFP for field purposes, simpler methods that rely on statistical relationships between easily measurable parameters are commonly used. These methods includ skin-fold thickness measurement, bio-electrical impedance, weight-height index(also known as body mass index, BMI) (Deurenberg,1999) and underwater weighing method(Katch and McArdle, 1977). However, these methods require specialized measurement tools and sophistical calculations that may not be easily conducted at home. Moreover also BMI method is race- and sex-specific(Deurenberg,1999), limiting its applicability in diverse populations. Body part circumference, particularly waist circumference (also refered to as abdomen circumference) is a critical indicator of fat mass in obesity studies. Therefore, it may be a useful and practical way for estimating BFP in field settings or at home, as it only requires as measuring tape. However, the conversion from circumference measurement into BFP can be challenging. Therefore, the objective of this paper is to develop a straightforward linear model that explains the relationship between circumference measurements and BFP, making it easy for daily use as a health assessment tool.

# INTRODUCTION

## Data description

Data (source: <http://lib.stat.cmu.edu/datasets/bodyfat>) in this paper was collected from 252 male adults aged from 22 to 81. BFP listed in this data was accurately estimated using underwater weighing method. 15 variables are included, and from left to right represented: Density determined from underwater weighing, Percent body fat from Siri’s (1956) equation, Age (years), Weight (lbs), Height (inches), Neck circumference (cm), Chest circumference (cm), Abdomen 2 circumference (cm), Hip circumference (cm), Thigh circumference (cm), Knee circumference (cm), Ankle circumference (cm), Biceps (extended) circumference (cm), Forearm circumference (cm) and Wrist circumference (cm). Density determined from underwater weighing (DDUW) is the body density estimated by equation (Katch and McArdle, 1977):

where WA = Weight in air(kg), WW = Weight in water(kg), c.f. = water correction factor, LV = Residual Lung Volume(liters).

Percent body fat (BFP) is calculated by equation (Siri, 1956) dependent on DDUW:

Other measurements were based on the standards list in Benhke and Wilmore (1974).

Here in this study, only 13 variables were included while the two variables DDUW and Age were not taken into consideration.

## Summary Statistics

The summary statistics are shown below:

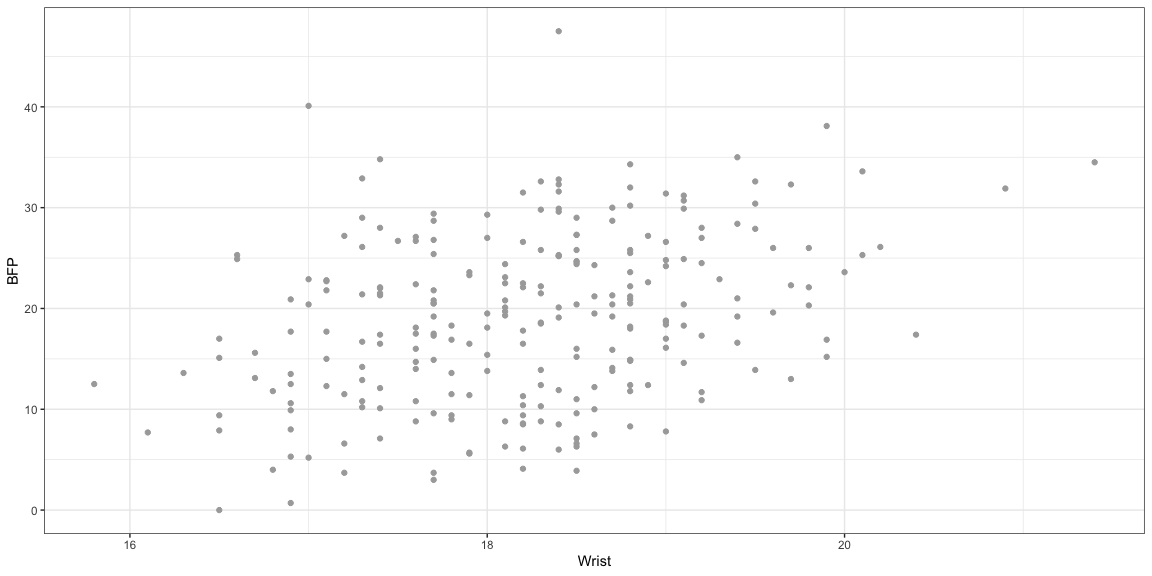
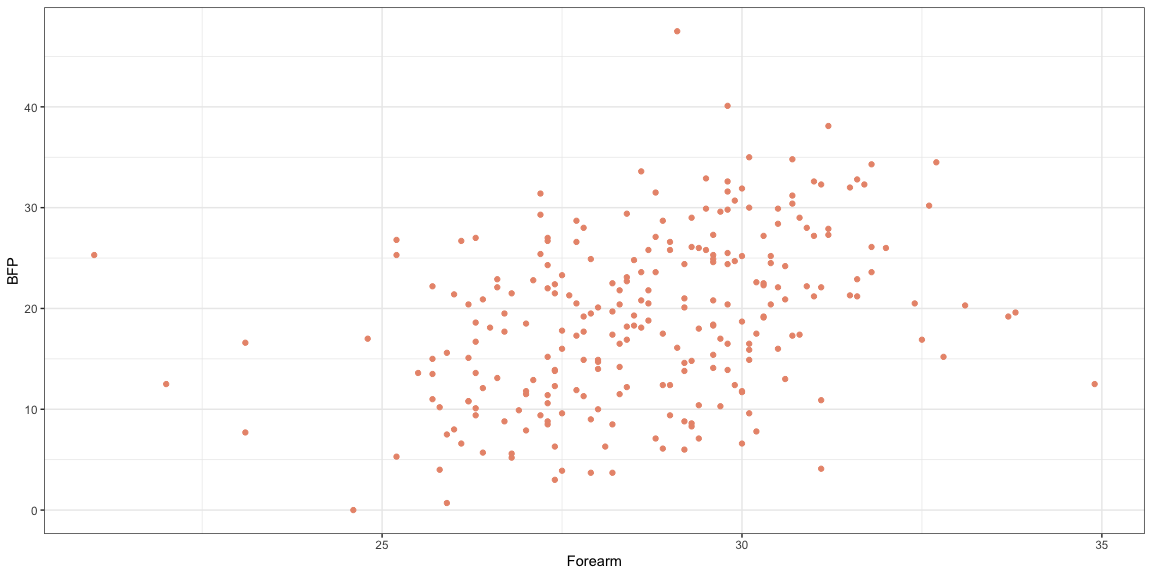
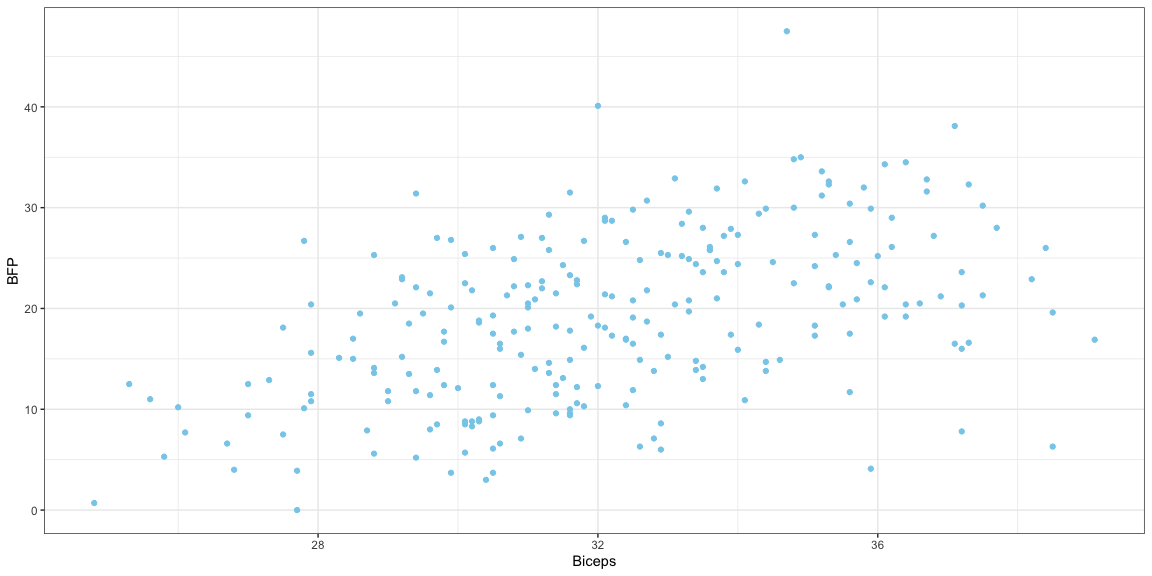
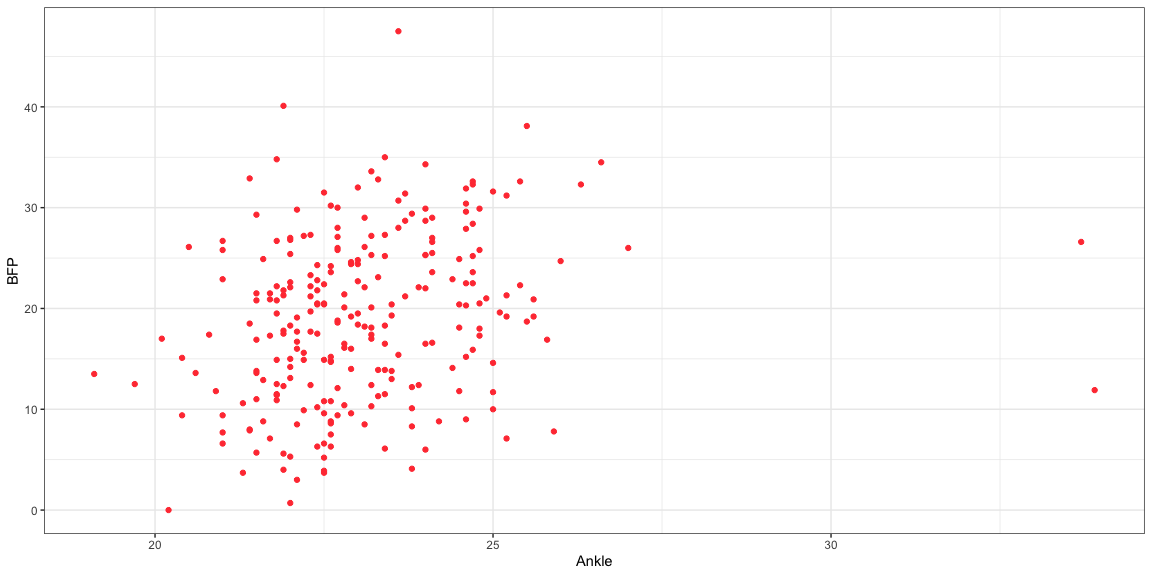
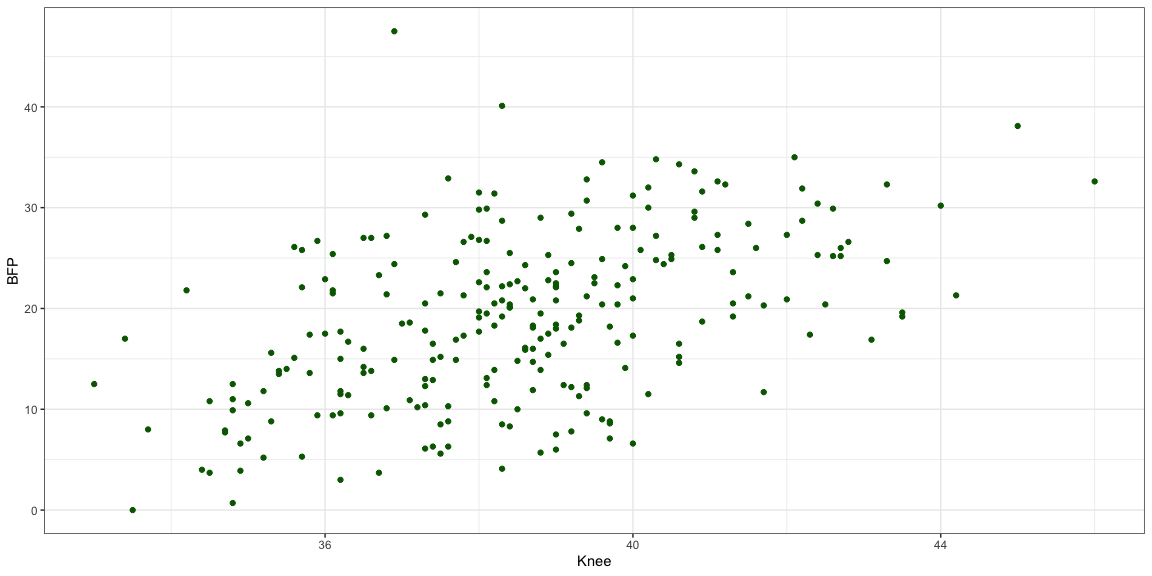
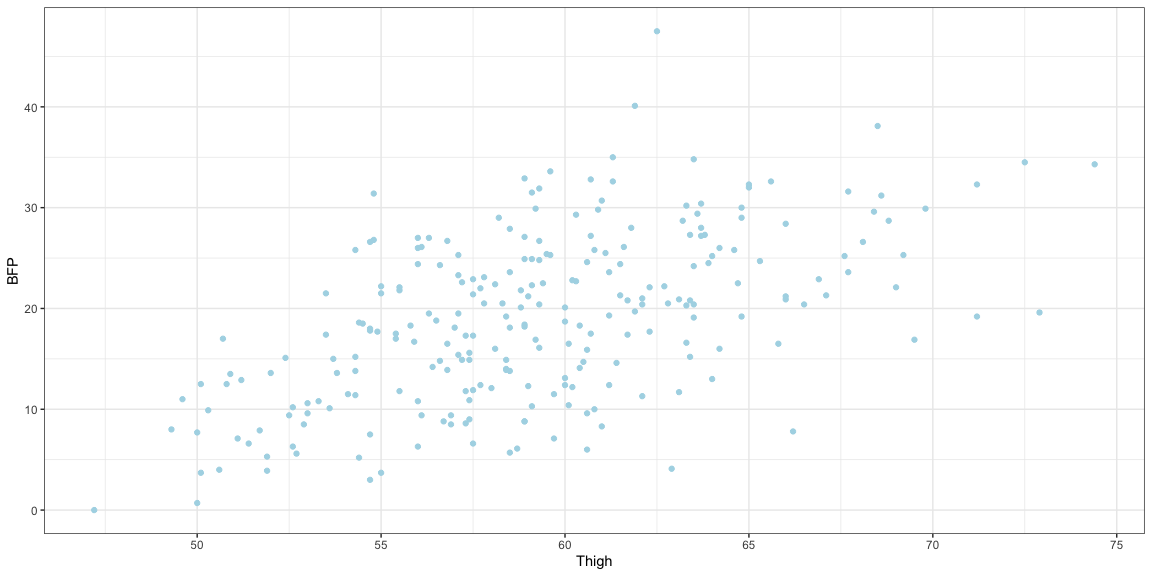
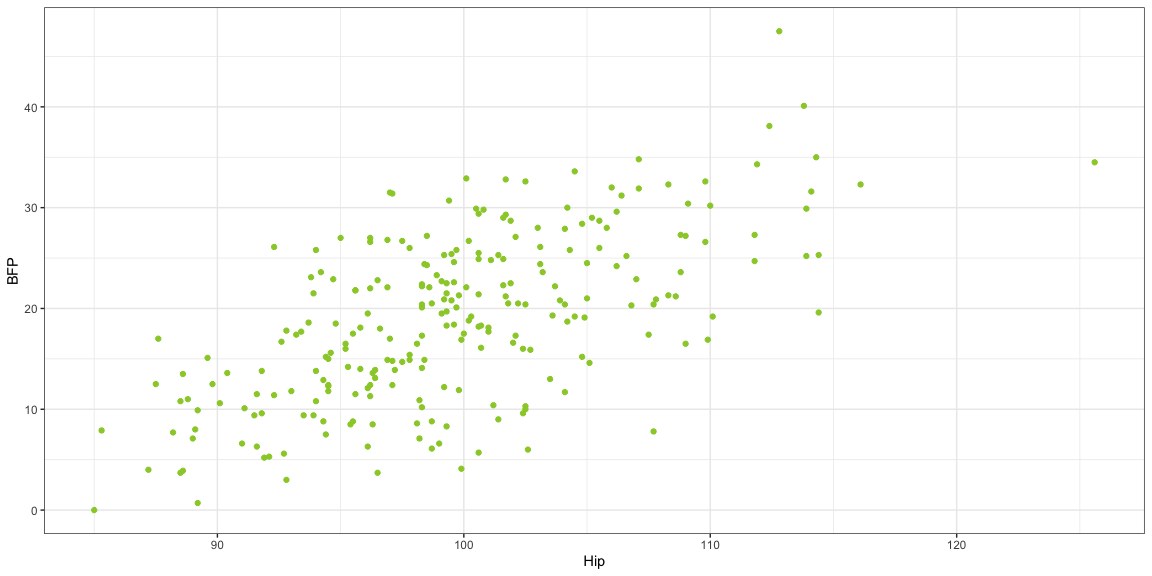
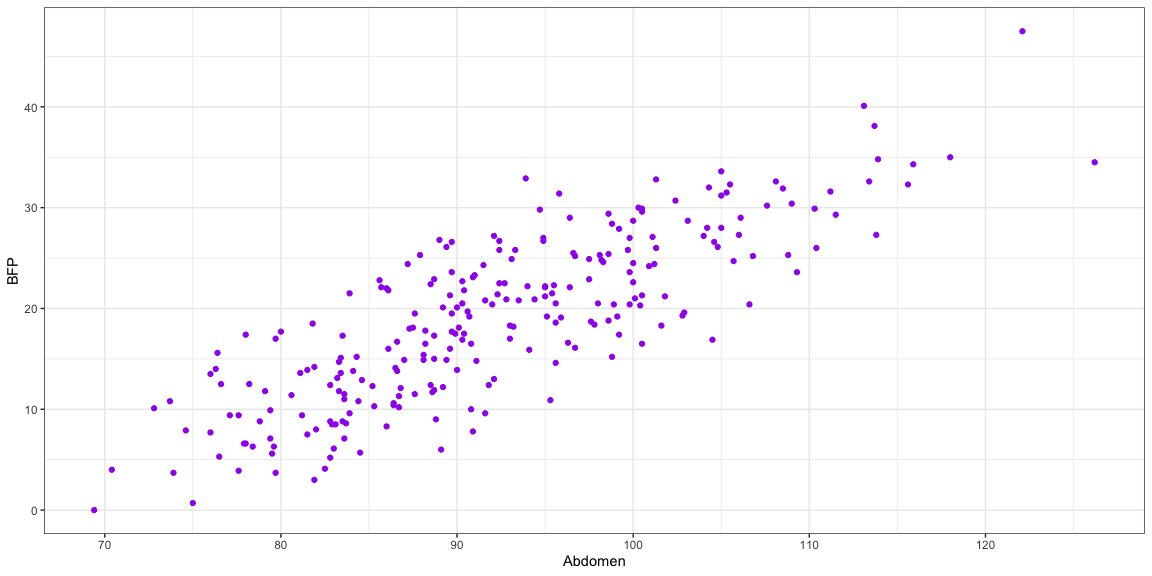
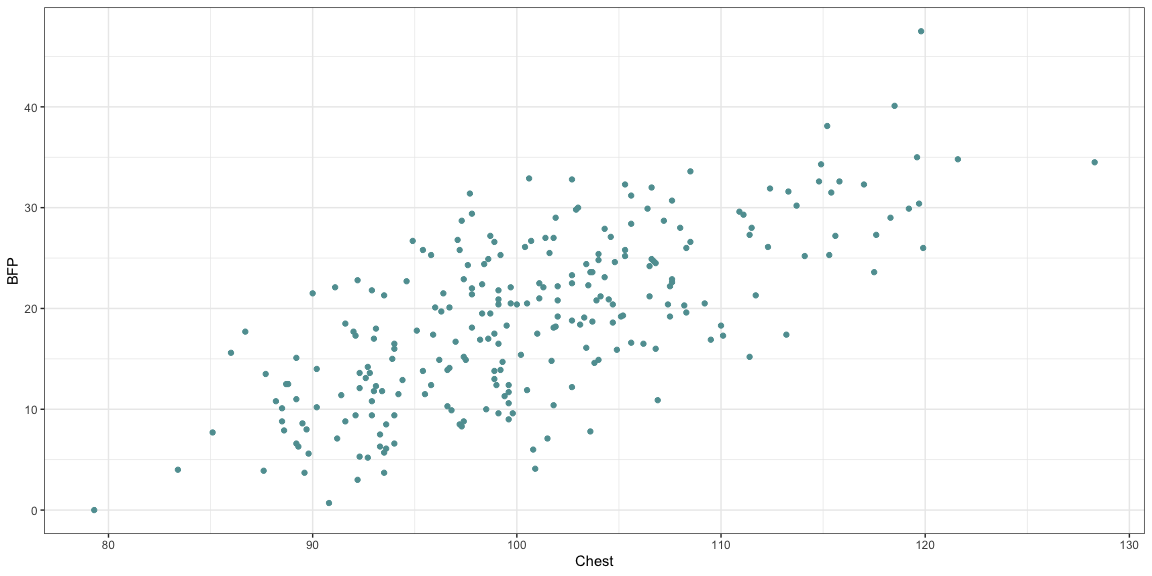
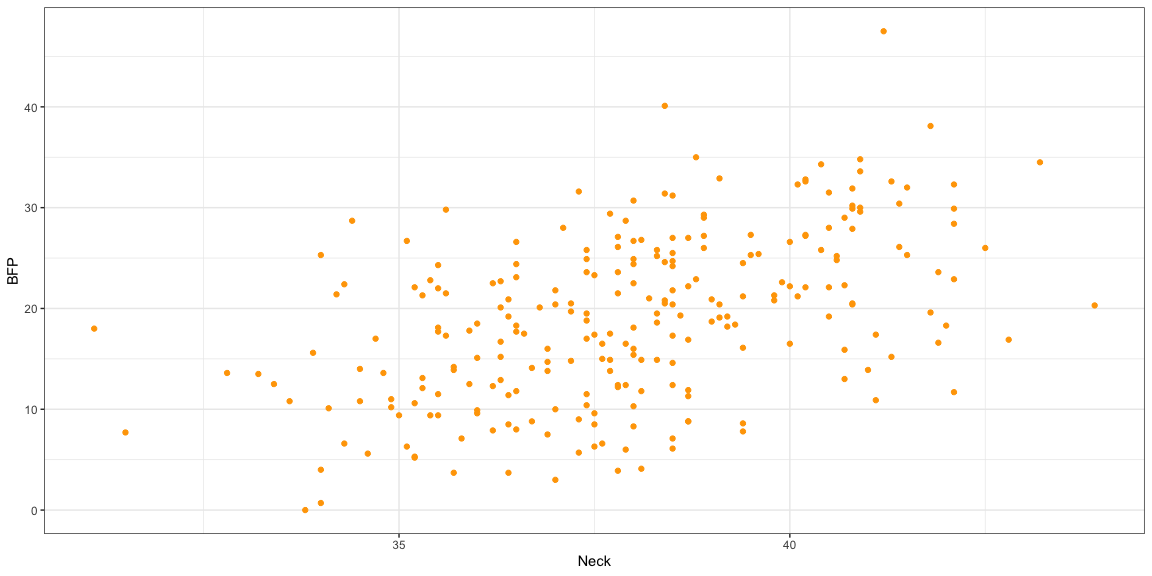
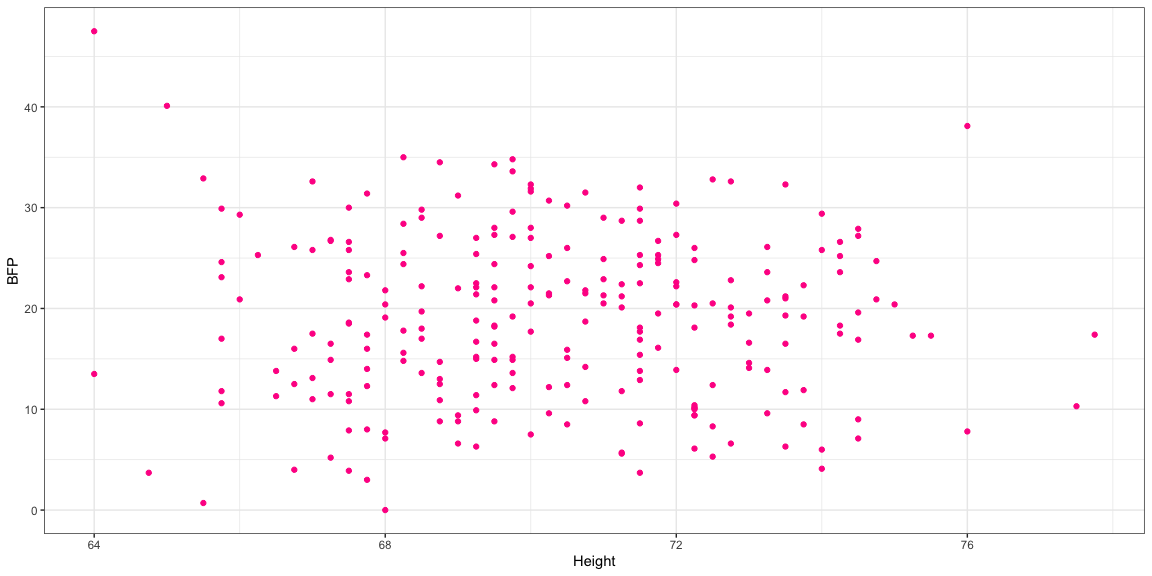
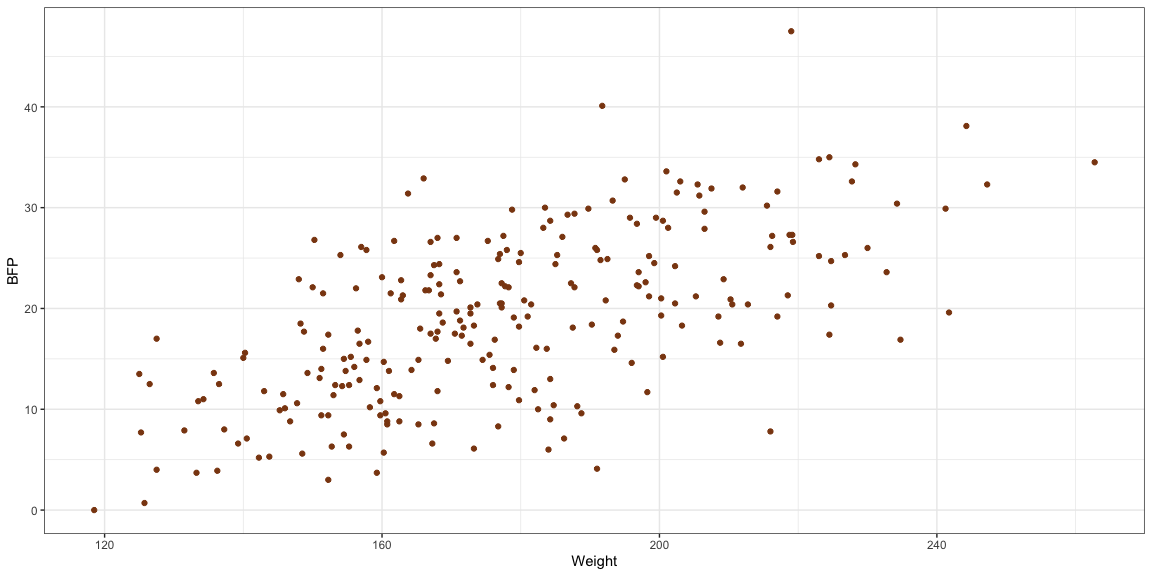
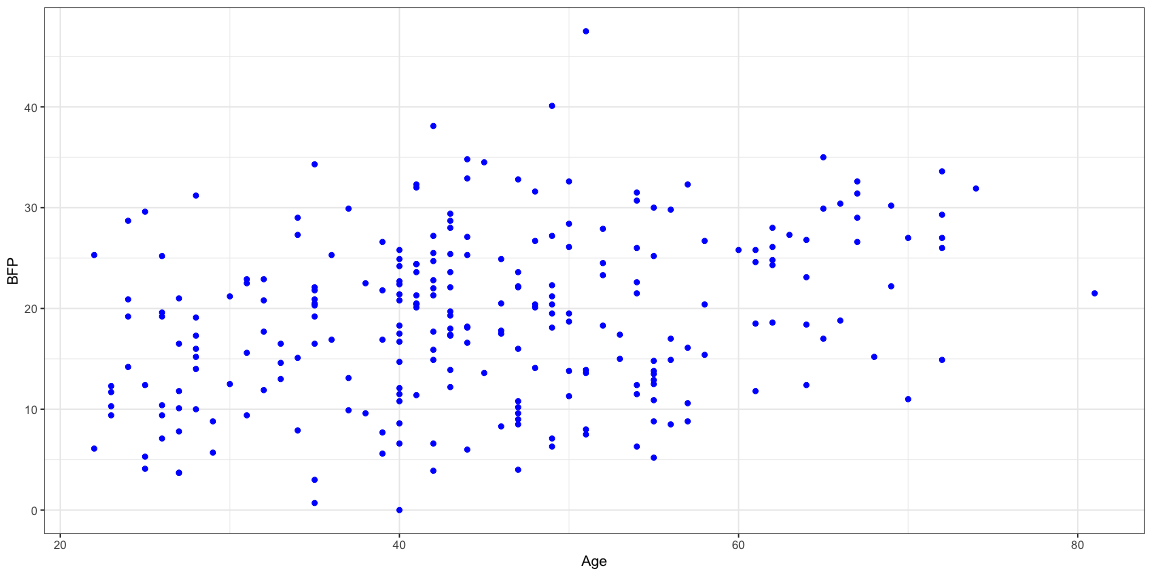
DDUW BFP Age Weight   
 Min. :0.995 Min. : 0.00 Min. :22.00 Min. :118.5   
 1st Qu.:1.041 1st Qu.:12.47 1st Qu.:35.75 1st Qu.:159.0   
 Median :1.055 Median :19.20 Median :43.00 Median :176.5   
 Mean :1.056 Mean :19.15 Mean :44.88 Mean :178.9   
 3rd Qu.:1.070 3rd Qu.:25.30 3rd Qu.:54.00 3rd Qu.:197.0   
 Max. :1.109 Max. :47.50 Max. :81.00 Max. :363.1   
 Height Neck Chest Abdomen   
 Min. :29.50 Min. :31.10 Min. : 79.30 Min. : 69.40   
 1st Qu.:68.25 1st Qu.:36.40 1st Qu.: 94.35 1st Qu.: 84.58   
 Median :70.00 Median :38.00 Median : 99.65 Median : 90.95   
 Mean :70.15 Mean :37.99 Mean :100.82 Mean : 92.56   
 3rd Qu.:72.25 3rd Qu.:39.42 3rd Qu.:105.38 3rd Qu.: 99.33   
 Max. :77.75 Max. :51.20 Max. :136.20 Max. :148.10   
 Hip Thigh Knee Ankle Biceps   
 Min. : 85.0 Min. :47.20 Min. :33.00 Min. :19.1 Min. :24.80   
 1st Qu.: 95.5 1st Qu.:56.00 1st Qu.:36.98 1st Qu.:22.0 1st Qu.:30.20   
 Median : 99.3 Median :59.00 Median :38.50 Median :22.8 Median :32.05   
 Mean : 99.9 Mean :59.41 Mean :38.59 Mean :23.1 Mean :32.27   
 3rd Qu.:103.5 3rd Qu.:62.35 3rd Qu.:39.92 3rd Qu.:24.0 3rd Qu.:34.33   
 Max. :147.7 Max. :87.30 Max. :49.10 Max. :33.9 Max. :45.00   
 Forearm Wrist   
 Min. :21.00 Min. :15.80   
 1st Qu.:27.30 1st Qu.:17.60   
 Median :28.70 Median :18.30   
 Mean :28.66 Mean :18.23   
 3rd Qu.:30.00 3rd Qu.:18.80   
 Max. :34.90 Max. :21.40

# Data anaylysis

Based on the summary of this data set, two samples with extreme height and weight were removed to tidy up the data.

Scatter plots were used identify variables that show linear association with BFP, and R-square calculation were performed to quantify the strength of this association.

## Scatter plots



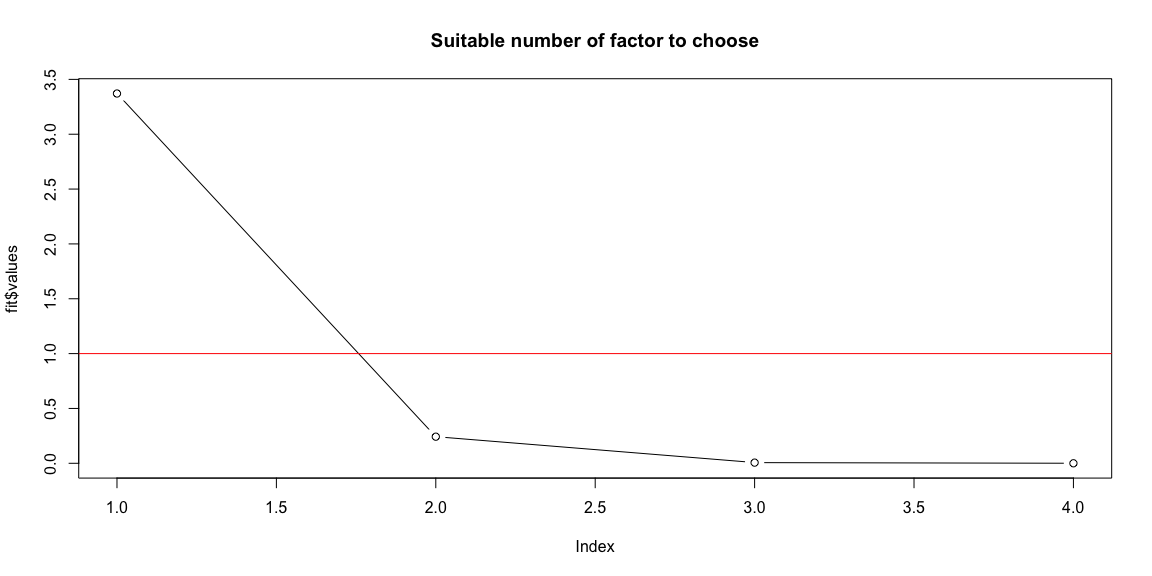
## R-square

Feature R\_square  
1 Age 0.087054954  
2 Weight 0.381058597  
3 Height 0.000863748  
4 Neck 0.238655937  
5 Chest 0.490937911  
6 Abdomen 0.678456403  
7 Hip 0.400277245  
8 Thigh 0.300906684  
9 Knee 0.242366841  
10 Ankle 0.059807419  
11 Biceps 0.231879038  
12 Forearm 0.133012552  
13 Wrist 0.114926297

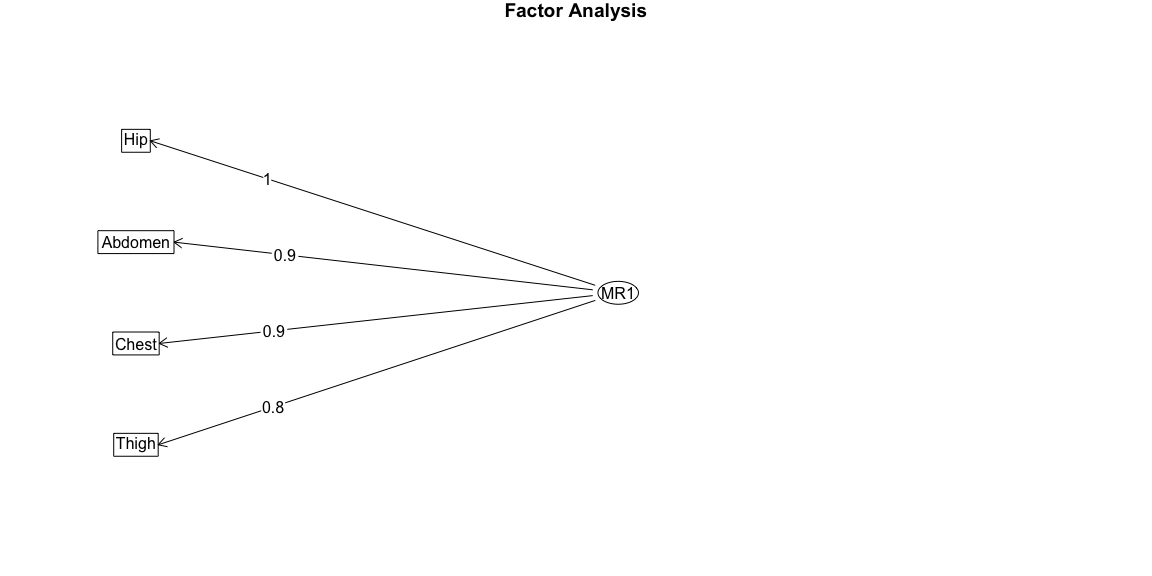
According to the Scatter plots and R-square values above, among the Features, , , and , show stronger linear association with . To minimize the numbers of features used for linear model, and test the correlation among , , and , factor analysis was conducted.

## Factor analysis

### Determine the number of factors



### Extract the factors and Interprest the factors

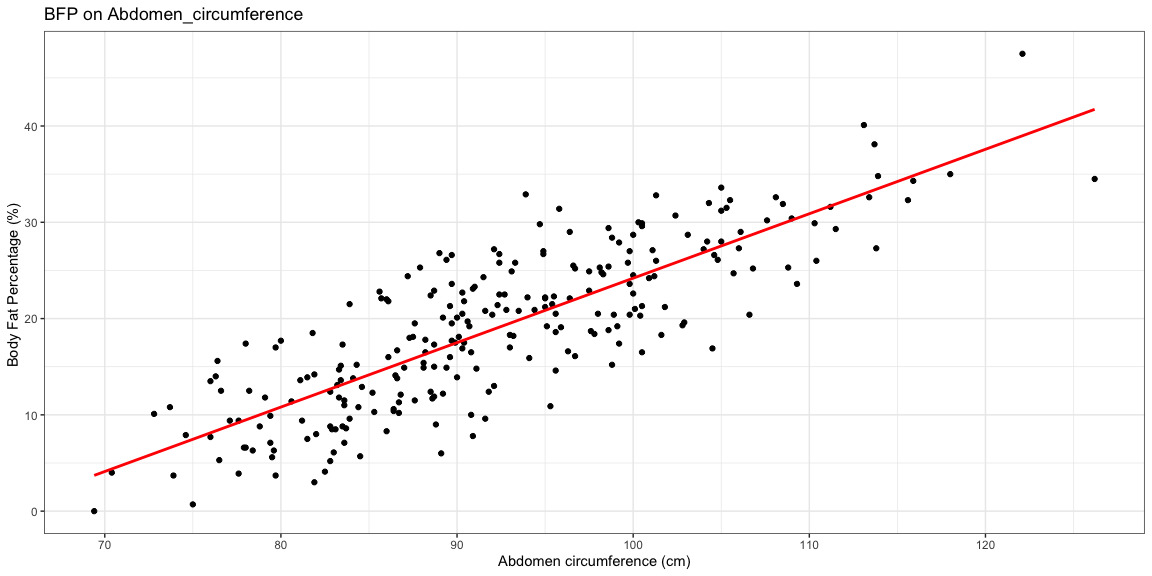


Chest Abdomen Hip Thigh  
MR1 0.9107751 0.9459778 0.9764733 0.8485894

However, only one factor was found among these four features and , , , is high that suggests that common factor has strong collinearity with other variables, thereby, reducing the unique explanatory capability of . This may affect the validity of the common factor in further analysis. To avoid this, was replaced with that showed strongest linear association with for building the linear model.

## Linear model and visualization

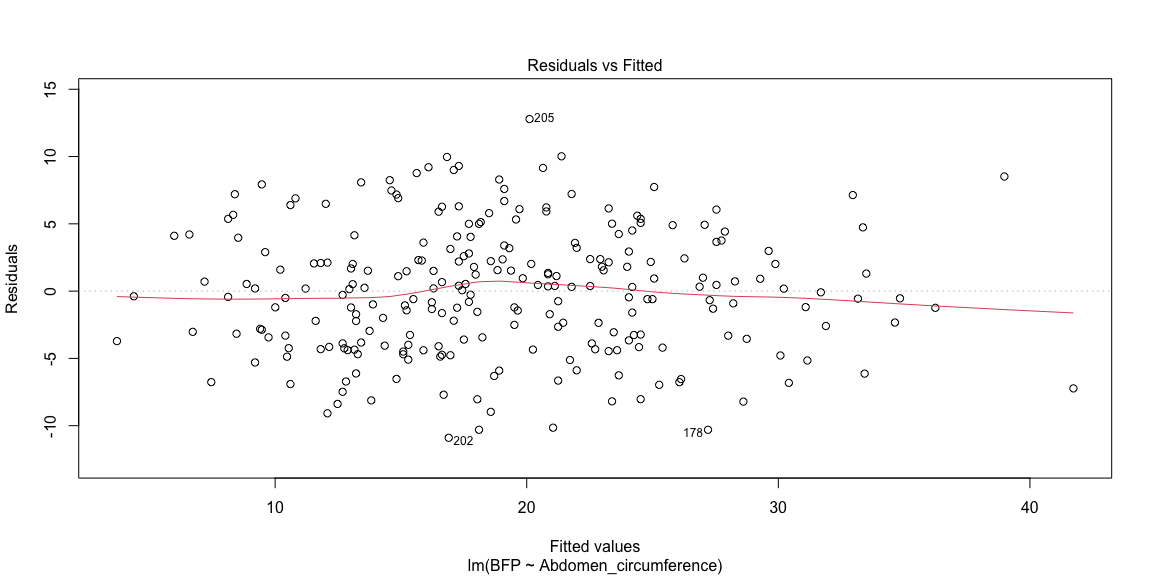
Estimate Std. Error t value Pr(>|t|)  
(Intercept) -42.7341340 2.71650562 -15.73129 3.826304e-39  
Abdomen\_circumference 0.6692803 0.02925776 22.87530 4.846625e-63



Based on the calculation above, the linear model:

BFP = 0.67 \*Abdomen -42.73

## Model Diagnostics



The smoothing curve in the residual vs fitting plot is close to the Residuals=0 line and does not have strong pattern, which suggests this model is a good fit linear regression model.

# Conclusion

There is a linear relationship between BFP and abdomen circumference.A coefficient of 0.63, suggesting a positive association between the two variables.The fitted model has a very low p-value of 2e-16, indicating strong evidence to reject the null hypothesis that the true slope is 0.

# Citation

Behnke, A.R. and Wilmore, J.H. (1974). Evaluation and Regulation of Body Build and Composition. Prentice-Hall, Englewood Cliffs, N.J.

Katch, Frank and McArdle, William (1977). Nutrition, Weight Control, and Exercise. Houghton Mifflin Co., Boston.

Siri, W.E. (1956), Gross composition of the body. Advances in Biological and Medical Physics, vol. IV, edited by J.H. Lawrence and C.A.Tobias, Academic Press, Inc., New York.

Deurenberg, P. (1999) The assessment of obesity: methods for measuring body fat and global prevalence of obesity. Bailliere’s Clinical Endocrinology and Metabolism 13(1):1-11

# APPENDIX: The Code

knitr::opts\_chunk$set(fig.align="center",  
 fig.height=6,  
 fig.width=12,  
 warning = FALSE,  
 message = FALSE,  
 comment = NA,  
 echo=FALSE)  
library(tidyverse)# dplyr::tibble  
library(knitr) #allows you to create Appendix with all\_labels()  
library(tidyverse)  
library(dplyr)  
library(psych)   
theme\_set(theme\_bw()) #sets default ggplot output style  
#Load Data  
dat<-read.csv("bodyfat\_raw\_data.csv")  
#View summary statistics  
summary(dat)  
  
dat<-filter(dat, Height >29.5 & Weight < 300)  
  
ggplot(dat, aes(x= Age,y=BFP))+geom\_point(color = "blue")  
ggplot(dat, aes(x= Weight,y=BFP))+geom\_point(color = "chocolate4")  
ggplot(dat, aes(x= Height,y=BFP))+geom\_point(color = "deeppink")  
ggplot(dat, aes(x= Neck,y=BFP))+geom\_point(color = "orange")  
ggplot(dat, aes(x= Chest,y=BFP))+geom\_point(color = "cadetblue")  
ggplot(dat, aes(x= Abdomen,y=BFP))+geom\_point(color = "purple")  
ggplot(dat, aes(x= Hip,y=BFP))+geom\_point(color = "yellowgreen")  
ggplot(dat, aes(x= Thigh,y=BFP))+geom\_point(color = "lightblue")  
ggplot(dat, aes(x= Knee,y=BFP))+geom\_point(color = "darkgreen")  
ggplot(dat, aes(x= Ankle,y=BFP))+geom\_point(color = "brown1")  
ggplot(dat, aes(x= Biceps,y=BFP))+geom\_point(color = "skyblue")  
ggplot(dat, aes(x= Forearm,y=BFP))+geom\_point(color = "darksalmon")  
ggplot(dat, aes(x= Wrist,y=BFP))+geom\_point(color = "darkgrey")  
  
  
  
Age<- lm(dat$BFP ~ dat$Age)  
r\_squared\_Age <-summary(Age)$r.squared  
  
Weight<- lm(dat$BFP ~ dat$Weight)  
r\_squared\_Weight <-summary(Weight)$r.squared  
  
Height<- lm(dat$BFP ~ dat$Height)  
r\_squared\_Height <-summary(Height)$r.squared  
  
Neck<- lm(dat$BFP ~ dat$Neck)  
r\_squared\_Neck <-summary(Neck)$r.squared  
  
Chest<- lm(dat$BFP ~ dat$Chest)  
r\_squared\_Chest <-summary(Chest)$r.squared  
  
Abdomen<- lm(dat$BFP ~ dat$Abdomen)  
r\_squared\_Abdomen <-summary(Abdomen)$r.squared  
  
Hip<- lm(dat$BFP ~ dat$Hip)  
r\_squared\_Hip <-summary(Hip)$r.squared  
  
Thigh<- lm(dat$BFP ~ dat$Thigh)  
r\_squared\_Thigh <-summary(Thigh)$r.squared  
  
Knee<- lm(dat$BFP ~ dat$Knee)  
r\_squared\_Knee <-summary(Knee)$r.squared  
  
Ankle<- lm(dat$BFP ~ dat$Ankle)  
r\_squared\_Ankle <-summary(Ankle)$r.squared  
  
Biceps<- lm(dat$BFP ~ dat$Biceps)  
r\_squared\_Biceps <-summary(Biceps)$r.squared  
  
Forearm<- lm(dat$BFP ~ dat$Forearm)  
r\_squared\_Forearm <-summary(Forearm)$r.squared  
  
Wrist<- lm(dat$BFP ~ dat$Wrist)  
r\_squared\_Wrist <-summary(Wrist)$r.squared  
  
R\_square<-c(r\_squared\_Age, r\_squared\_Weight, r\_squared\_Height, r\_squared\_Neck, r\_squared\_Chest, r\_squared\_Abdomen, r\_squared\_Hip, r\_squared\_Thigh, r\_squared\_Knee, r\_squared\_Ankle, r\_squared\_Biceps, r\_squared\_Forearm, r\_squared\_Wrist)  
Feature <-c("Age", "Weight", "Height", "Neck", "Chest", "Abdomen", "Hip", "Thigh", "Knee", "Ankle", "Biceps", "Forearm", "Wrist")  
df<-data.frame(Feature, R\_square)  
df  
  
# keep feature "Chest, Abdomen, Hip and Thigh" for anaylysis   
data<-dat[, c(7:10)]  
fit<-fa(data, nfactors = 4, rotate = "varimax")  
plot(fit$values, type = "b", main = "Suitable number of factor to choose")  
abline(h =1, col = "red")  
fa\_result<-fa(data)  
fa.diagram(fa\_result, simple = TRUE)  
factor\_scores <-as.data.frame(fa\_result$scores)  
cor(factor\_scores, data)  
BFP<-dat$BFP  
Abdomen\_circumference<-dat$Abdomen  
  
# Linear model  
model<- lm(BFP ~ Abdomen\_circumference)  
  
# Coefficients and p-value  
summary(model)$coefficients  
  
# visualization  
ggplot(dat, aes(x = Abdomen\_circumference, y = BFP)) +  
 geom\_point() +  
 geom\_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "red")+  
 labs(title="BFP on Abdomen\_circumference",x="Abdomen circumference (cm)",y="Body Fat Percentage (%) ")  
  
plot(model,which = 1)