Data analysis on body fat

Ken Obata

April 3, 2020

Contents

0.1 General Instructions in RStudio

- Put your name in the author section above.
- Type your answers below the questions. **Do not change the R code**!!
- Save this .Rmd file to your computer and then knit the entire document to pdf.
- If you cannot knit to pdf on your computer, then knit to Word and save as a pdf file.
- Upload the pdf file to the Assignment 5 Activity in the Assignments section of CourseSpaces.

1 Body Fat Data Set

A new method of measuring the body fat percentge is investigated. The body fat percentage, age and gender (1=Male, 0=Female) of 18 normal adults is provided.

Reference: Mazess, Pepple, Gibbons. 1983. "Total Body composition by dualphoton absorptionmetry," *American Journal of Clinical Nutrition*, 40, 834-839.

2 Data preparation and descriptives

```
Bodyfat <- read.table("bodyfat.csv", sep=",", header=TRUE)   #read in data
print(nrow(Bodyfat))
```

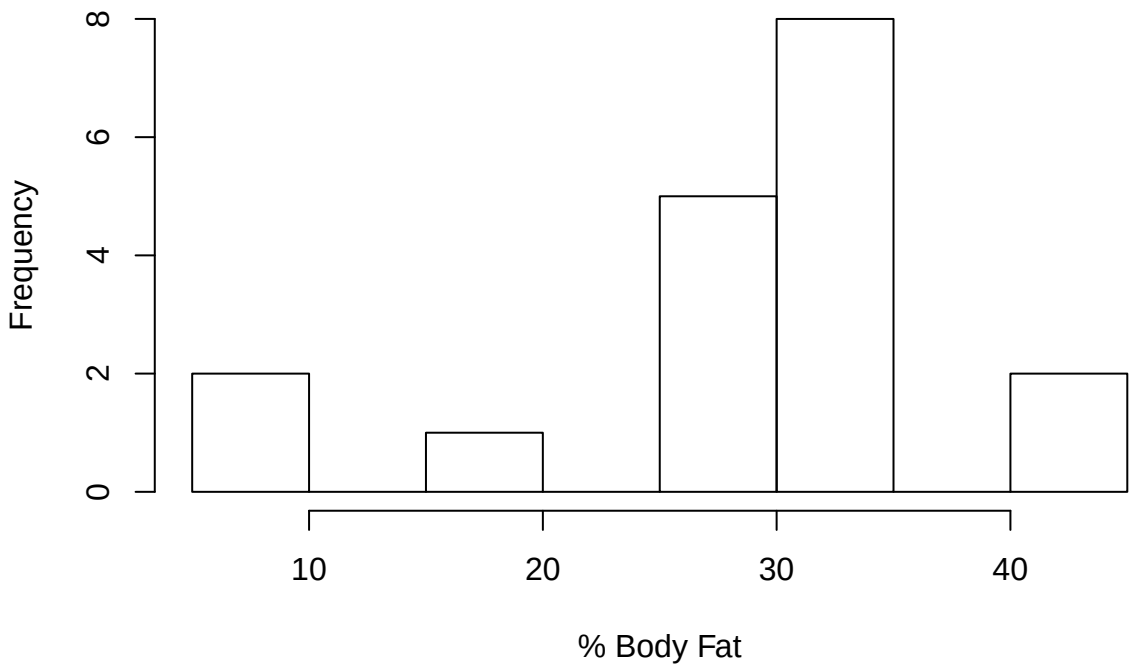
```
## [1] 18
```

```
Bodyfat$GenderF <- factor(Bodyfat$Gender, labels=c("F", "M")) #create factor Gender
summary(Bodyfat)
```

```
##      PerFat           Age            Gender       GenderF
##  Min.   : 7.80   Min.   :23.00   Min.   :0.0000   F:14
##  1st Qu.:26.27   1st Qu.:39.50   1st Qu.:0.0000   M: 4
##  Median :30.70   Median :51.50   Median :0.0000
##  Mean   :28.61   Mean   :46.33   Mean   :0.2222
##  3rd Qu.:33.60   3rd Qu.:56.75   3rd Qu.:0.0000
##  Max.   :42.00   Max.   :61.00   Max.   :1.0000
```

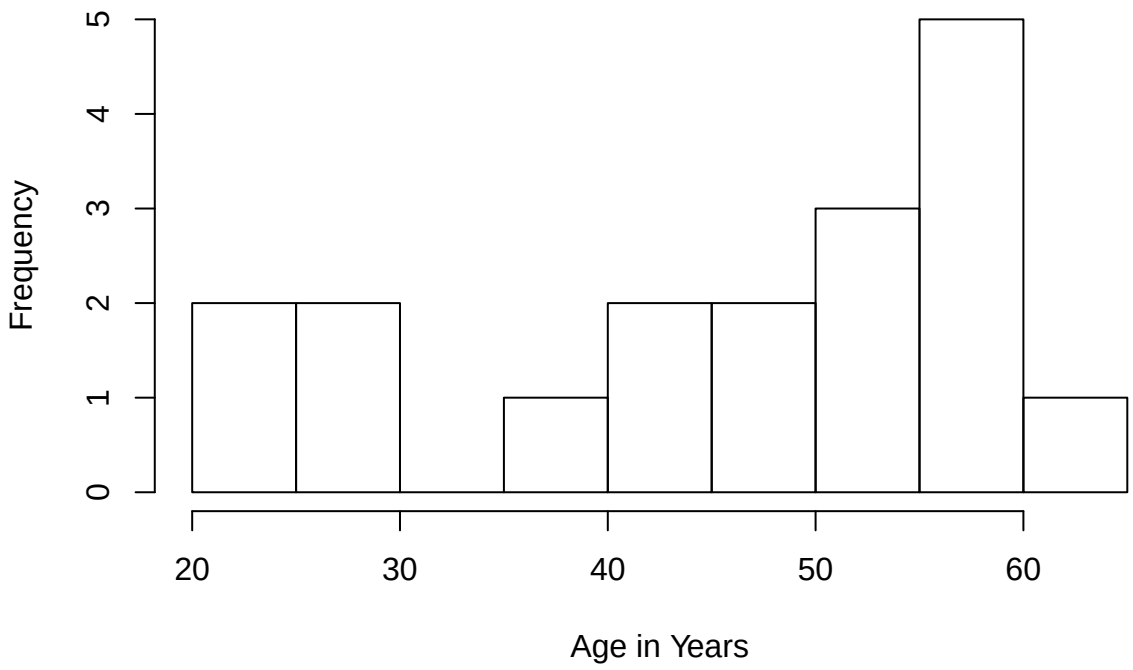
```
hist(Bodyfat$PerFat, main='Figure 1: Histogram of percentage body fat', xlab='% Body Fat')
```

Figure 1: Histogram of percentage body fat




```
hist(Bodyfat$Age, main='Figure 2: Histogram of Age', xlab='Age in Years')
```

Figure 2: Histogram of Age



---

3 Analyze the percentage fat for females

3.1 Inferences for the mean, μ

```
y <- Bodyfat$PerFat[Bodyfat$GenderF=='F']  #choose females
y
```

```
##  [1] 27.9 31.4 25.9 25.2 31.1 34.7 42.0 29.1 32.5 30.3 33.0 33.8 41.1 34.5
```

```
#mean of PerFat for females
mean(y)
```

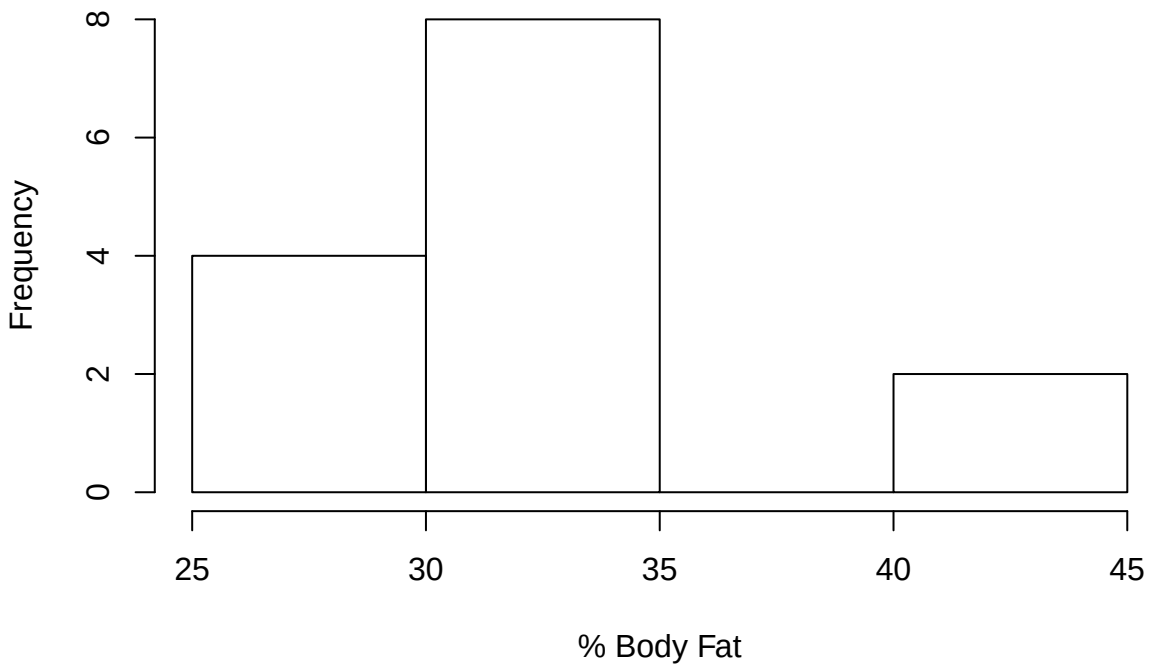
```
## [1] 32.32143
```

```
#sd of PerFat for females
sd(y)
```

```
## [1] 4.89995
```

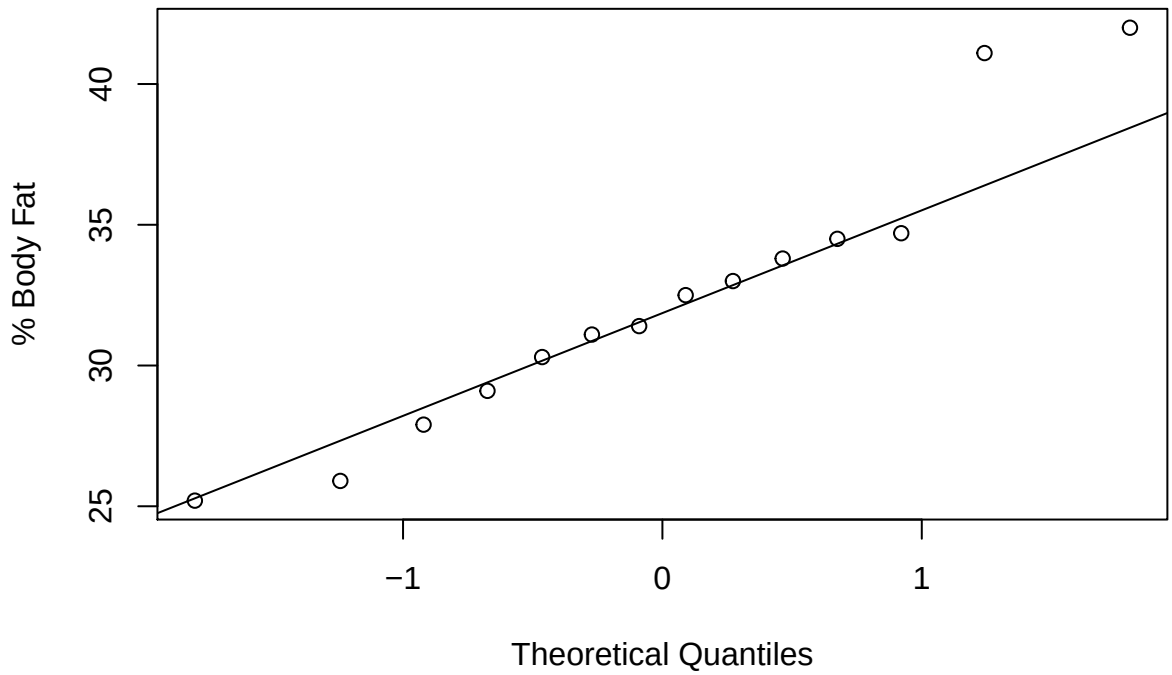
```
hist(y, main='Figure 3: Histogram of percentage body fat, females', xlab='% Body Fat')
```

Figure 3: Histogram of percentage body fat, females




```
qqnorm(y, main='Figure 4: Normal QQ plot of percentage body fat, females', ylab='% Body Fat')
qqline(y, lty=1)
```

Figure 4: Normal QQ plot of percentage body fat, females



For Figure3, since there are only 14 observations, it is hard to interpret but among 14 female sample data with range of body fat% from 25% to 45%, the range from 30% to 35% has the highest frequency. For Figure

4, the normal QQ plot (Fig.4) suggests that the sample is consistent with the Normal distribution, except for two outliers whose body fat% exceeds 40%.

Since σ^2 is unknown, we need to esimate it and use the t-distribution to compute a confidence interval for the mean percentage body fat for females.

```
# 95% confidence interval
n <- length(y)
mean(y) + c(-1,1) * qt(.975, n-1) * sd(y) / sqrt(n)
```

```
## [1] 29.49228 35.15058
```

```
qt(.975, n-1)
```

```
## [1] 2.160369
```

```
t.test(y, mu=35)
```

```
##
##  One Sample t-test
##
## data:  y
## t = -2.0454, df = 13, p-value = 0.06161
## alternative hypothesis: true mean is not equal to 35
## 95 percent confidence interval:
##  29.49228 35.15058
## sample estimates:
## mean of x
##  32.32143
```

95 percent confidence interval is [29.49228, 35.15058]

$H_0 : \mu = 35$. After performing the hypothesis test, we found that p-value = 0.06161 which is less than 0.1 and bigger than 0.05. So there is a moderately strong evidence against the null hypothesis of a mean percentage body fat of 35% (95% CI=29.49228, 35.15058).

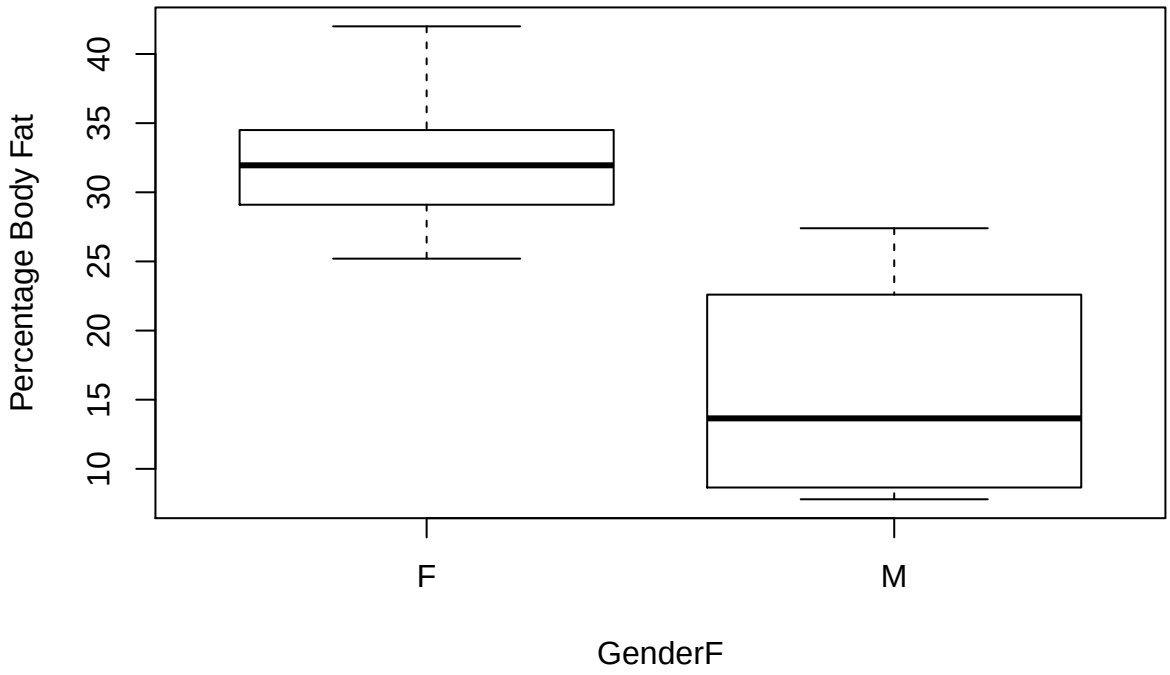
---

4 Compare percentage body fat for males and females

4.1 Inferences about the differences in the means, $\mu_F - \mu_M$

```
#Graph the data; side-by-side boxplots
boxplot(PerFat ~ GenderF, data=Bodyfat, ylab='Percentage Body Fat',
        main='Figure 5: Percentage Body Fat for Males and Females')
```

Figure 5: Percentage Body Fat for Males and Females



The box plots (Fig.5) indicates that the median percentage body fat for female is higher than that for males. Also, the variation in percentage body fat of male is higher than that for females.(Plot for males has more variation in 50 percentile range.)

4.2 Inferences for the differences: Assume variances equal and unknown

```
#this uses pooled estimate of variance for test that H0: mu_F - mu_M = 0
t.test(PerFat ~ GenderF, data=Bodyfat, var.equal=TRUE)
```

```
##
##  Two Sample t-test
##
## data:  PerFat by GenderF
## t = 5.0037, df = 16, p-value = 0.0001299
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   9.622642 23.770216
## sample estimates:
## mean in group F mean in group M
##        32.32143        15.62500
```

$H_0 : \mu_f - \mu_m = 0$ where μ_f is mean percentage body fat for females, and μ_m is mean for males. Since p-value = 0.0001299<0.01 , the mean percentage body fat for females is significantly higher than that for males, with an avarage difference of 32.32143 - 15.62500. 0.95 percent confidence interval is: [9.622642, 23.770216].

4.3 Inferences for the differences: Do not assume variances are equal

```
#this does NOT use pooled estimate of variance for test that HO: mu_F - mu_M = 0
t.test(PerFat ~ GenderF, data=Bodyfat, var.equal=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  PerFat by GenderF
## t = 3.5684, df = 3.5258, p-value = 0.02888
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   2.986674 30.406183
## sample estimates:
## mean in group F mean in group M
##        32.32143        15.62500
```

$H_0 : \mu_f - \mu_m = 0$ where μ_f is mean percentage body fat for females, and μ_m is mean for males. Since p-value = 0.02888<0.05, there is evidence against the null hypothesis. In other words, the mean percentage body fat for females is significantly higher than that for males, with an avarage difference of 32.32143 - 15.62500. 0.95 percent confidence interval is: [2.986674, 30.406183]. ———————

5 Regression analysis to investigate relationship, Percentage body fat as a function of Age

```
#Graph the data; Scatterplot
plot(PerFat ~ Age, data=Bodyfat, pch=as.character(GenderF),
     xlab='Age in years', ylab='Percentage body fat',
     main='Figure 6: Percentage body fat versus Age')
legend('bottomright', c('F=Female','M=Male'), bty='n')
```

Figure 6: Percentage body fat versus Age

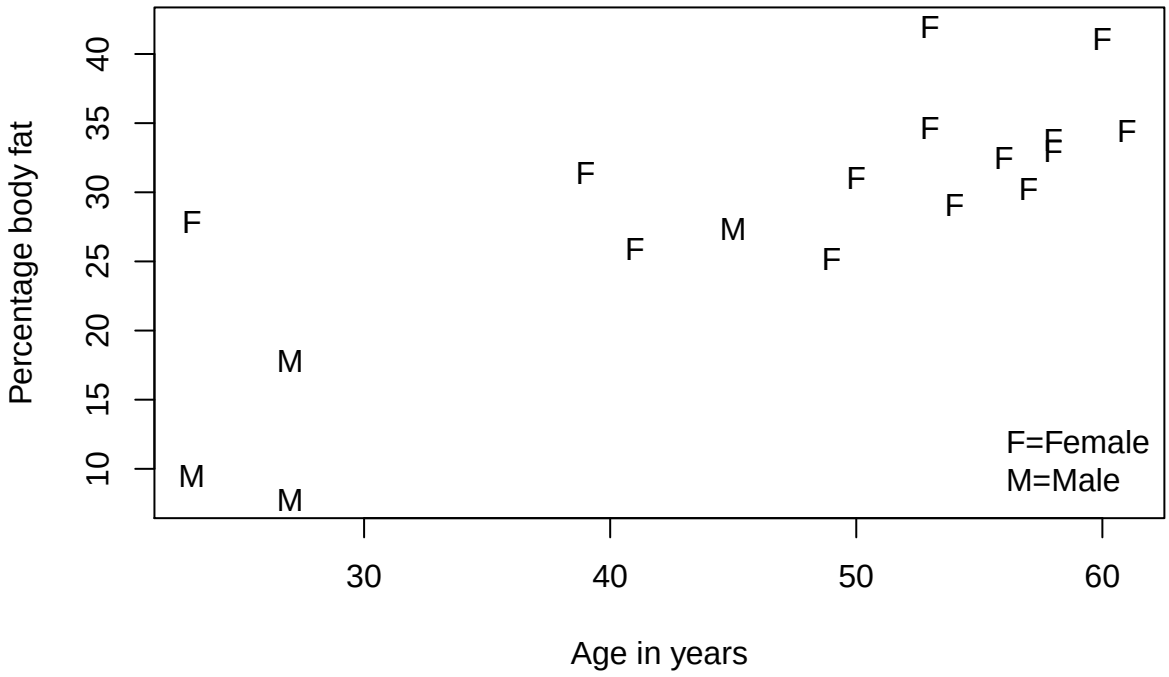


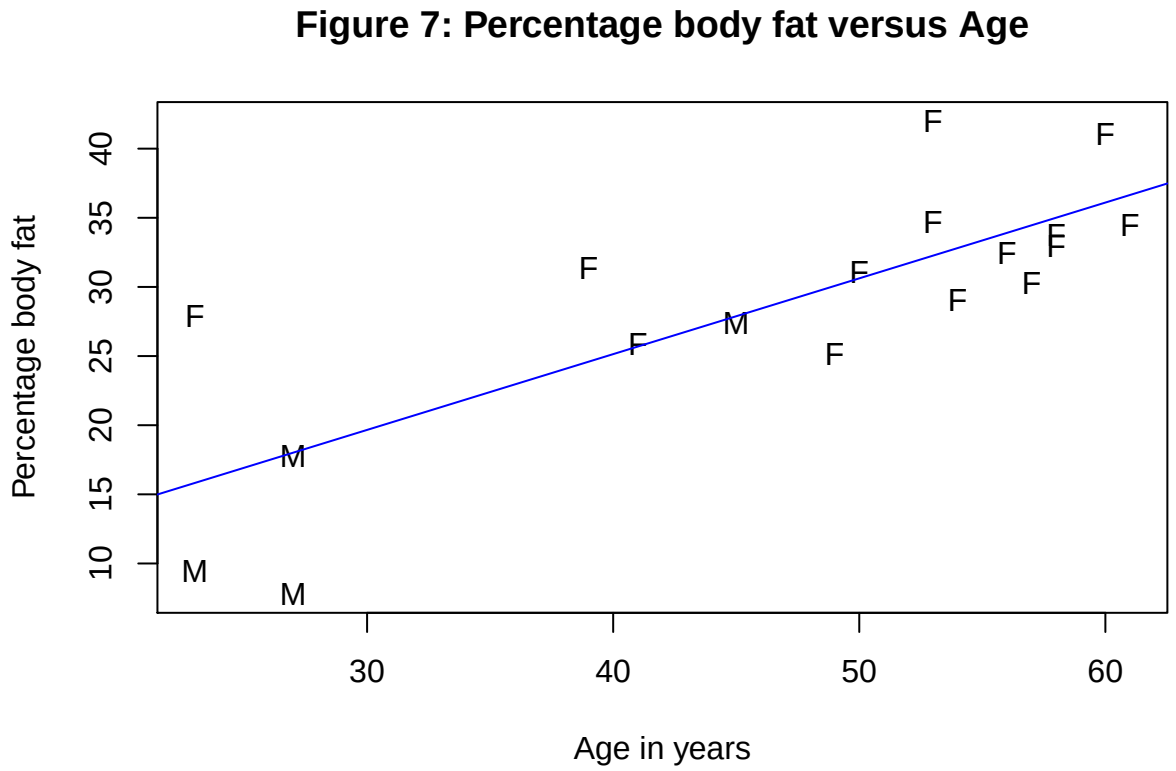
Figure6 suggests that there is a relationship between age and percentage body fat. As age increases, percentage body fat also increases for both male and female.

```
#Fit the regression model
Bodyfat.lm <- lm(PerFat ~ Age, data=Bodyfat)
summary(Bodyfat.lm)
```

```
##
## Call:
## lm(formula = PerFat ~ Age, data = Bodyfat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.2166  -3.3214  -0.8424   1.9466  12.0753
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.2209     5.0762   0.635    0.535
## Age           0.5480     0.1056   5.191 8.93e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.754 on 16 degrees of freedom
## Multiple R-squared:  0.6274, Adjusted R-squared:  0.6041
## F-statistic: 26.94 on 1 and 16 DF,  p-value: 8.93e-05
```

```
plot(PerFat ~ Age, data=Bodyfat, pch=as.character(GenderF),
     xlab='Age in years', ylab='Percentage body fat',
```

```
     main='Figure 7: Percentage body fat versus Age')
abline(Bodyfat.lm, col=4)
```



Estimated model: percentage body fat=3.2209 + 0.5480*Age

```
qqnorm(resid(Bodyfat.lm), main="Figure 8: QQ plot of residuals from regression", ylab="residuals")
qqline(resid(Bodyfat.lm), lty=1)
```

Figure 8: QQ plot of residuals from regression

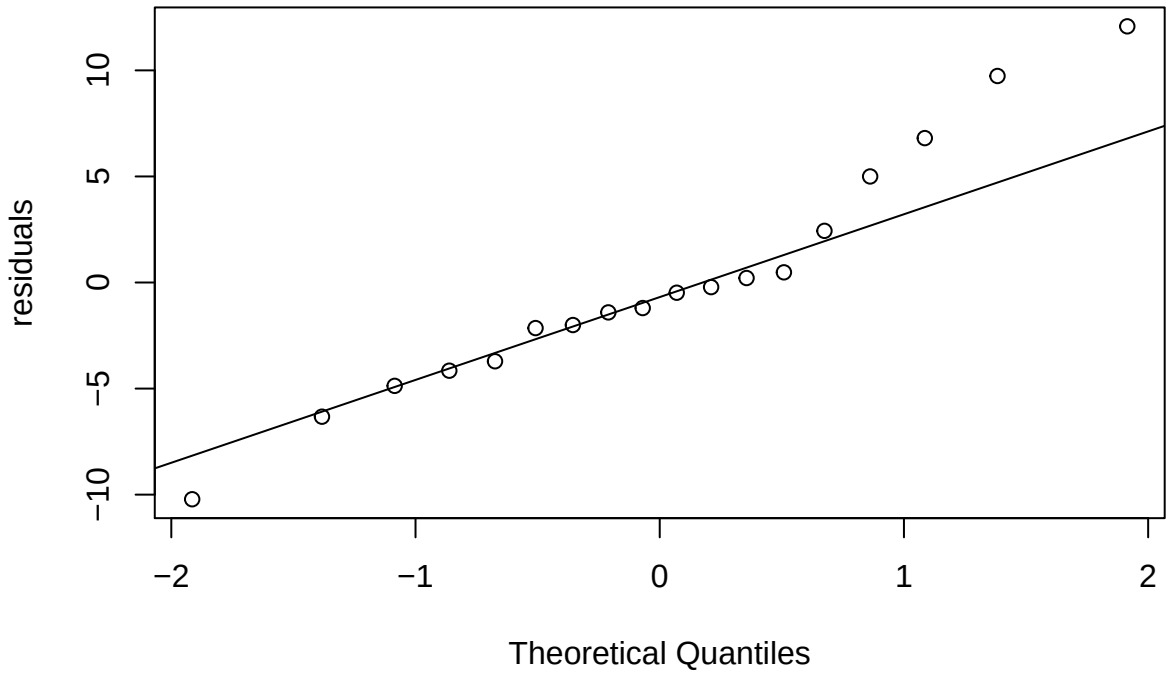


Figure8. shows that the distribution of the residuals appear to be Normal.