Module 3 Report: Emalee Schuler

Vitamin D Levels By Gender

Null Hypothesis H_0 : There is no significant difference in the mean vitamin D levels between females and males.

$H_0 : \mu_{females} = \mu_{males}$

Alternative Hypothesis H_a : The mean vitamin D levels between males and females are not equal.

$H_a : \mu_{females} \neq \mu_{males}$

Vitamin D Levels By Incidence of Illness:

Null Hypothesis H_0 : There is no significant difference in the mean vitamin D levels between healthy and diseased participants.

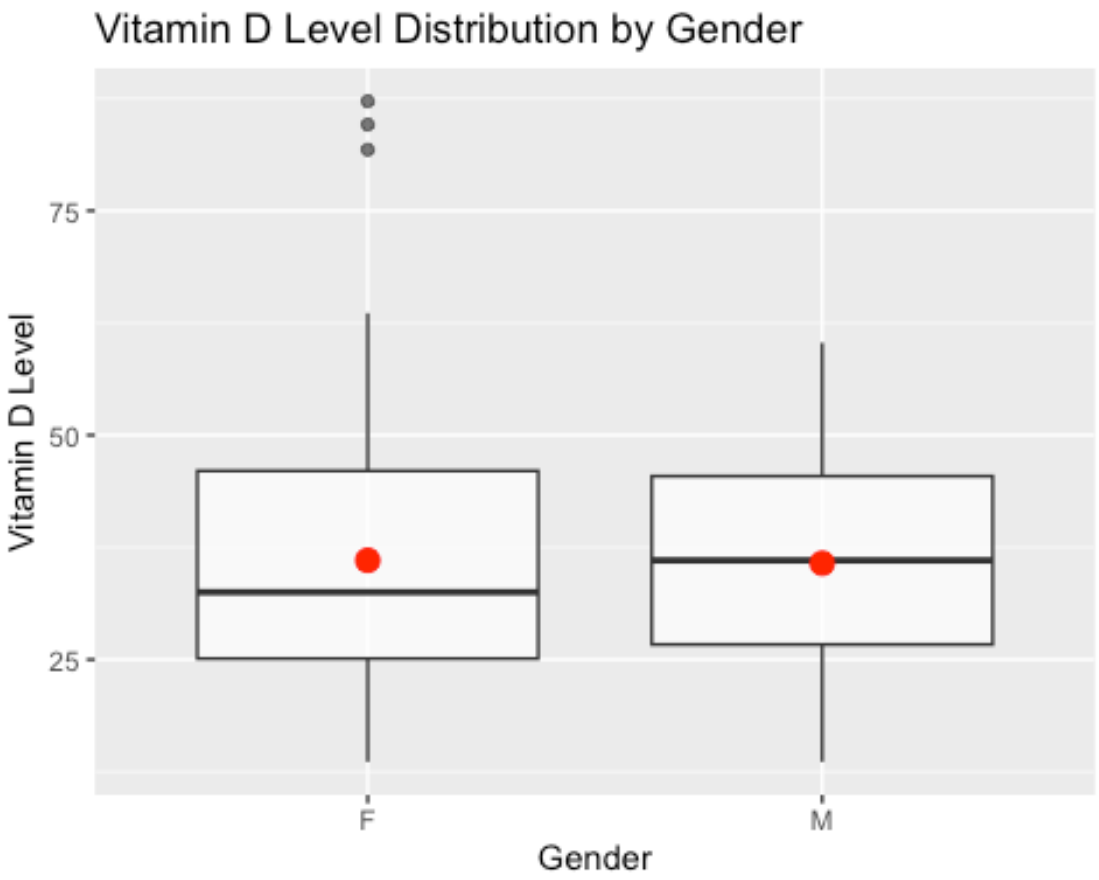
$H_0 : \mu_{healthy} = \mu_{diseased}$

Alternative Hypothesis H_a : The mean vitamin D levels between healthy and diseased participants.

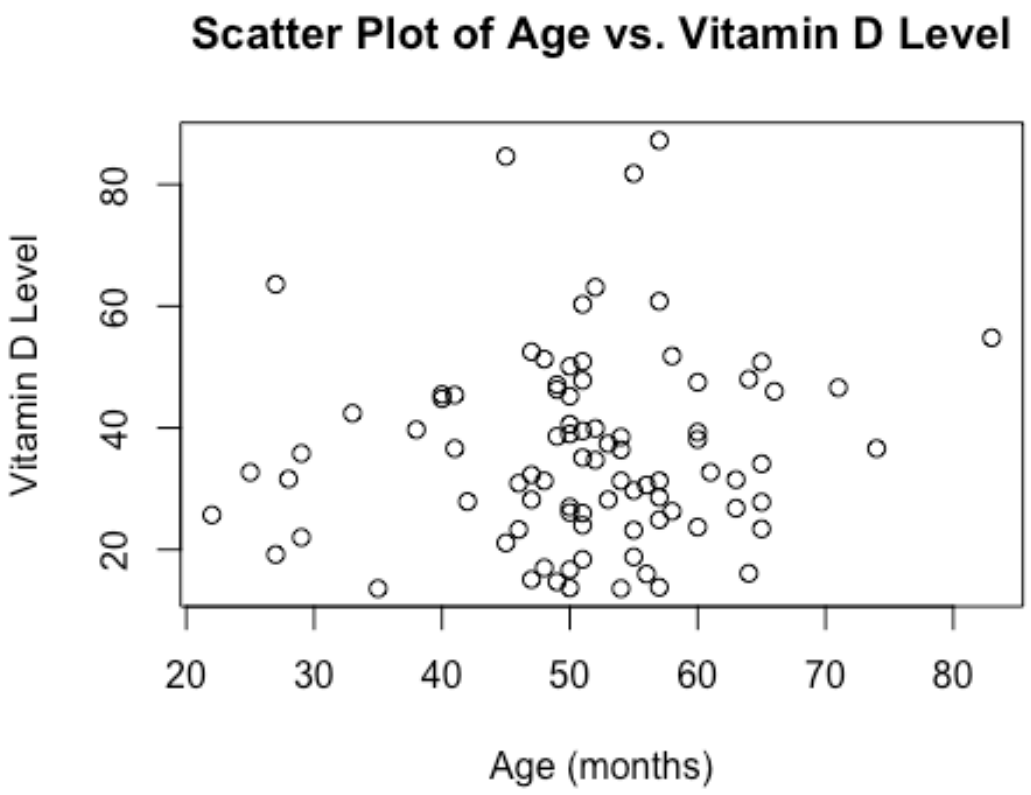
$H_a : \mu_{healthy} \neq \mu_{diseased}$

```
library(ggplot2)
library(dplyr)


Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union

data1_LSC598 <- read.csv("~/Downloads/data1_LSC598.txt", sep="")

data1_LSC598 <- data1_LSC598 %>% filter(!is.na(vitD_level))


ggplot(data1_LSC598, aes(x = gender, y = vitD_level)) +
    geom_boxplot(alpha = 0.7) +
  stat_summary(fun = mean, geom = "point", shape = 20, size = 5, color =
"red", fill = "red") +
```

```
    theme(legend.position = "none") +
     scale_fill_brewer(palette = "Set1") + ggtitle("Vitamin D Level
Distribution by Gender") +
  xlab("Gender") +
    ylab("Vitamin D Level")
```




```
plot(data1_LSC598$age_month, data1_LSC598$vitD_level, xlab = "Age (months)",
ylab = "Vitamin D Level", main = "Scatter Plot of Age vs. Vitamin D Level")
```




```
cor(data1_LSC598$age_month, data1_LSC598$vitD_level)

[1] 0.08259572
```

Interpretation of side-by-side boxplot comparing the vitamin D levels of females and males:

The boxes for females and males are similar in height with the female box being slightly taller. This indicates that females have a larger spread of data. This makes sense because the females have more outliers in their vitamin D results. The median for females is slightly lower than that of males. However, they do have very similar means. All in all, the boxplots are very similar with a few minute differences and it can reasonably be concluded that there are not major differences in vitamin D levels between males and females.

Interpretation of the correlation coefficient for age and vitamin D level:

A correlation coefficient of 0.08259572 for age and vitamin D level indicates a very weak and positive relationship between age and vitamin D level in this data set. This means that there is little to no meaningful association between age and vitamin D level in this data set. This is illustrated in the scatter-plot because there is no clear linear relationship shown by the points. While there is a positive association, it is weak. The data points are widely dispersed, and the correlation is not practically significant.

T-Tests

```
#T-Test to examine whether the vitamin D levels differ between females and
males
vitD_females <- data1_LSC598[data1_LSC598$gender == "F", "vitD_level"]

vitD_males <- data1_LSC598[data1_LSC598$gender == "M", "vitD_level"]

t.test(vitD_females, vitD_males)


    Welch Two Sample t-test

data:  vitD_females and vitD_males
t = 0.10163, df = 58.166, p-value = 0.9194
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.196153  6.859056
sample estimates:
mean of x mean of y
 36.05645  35.72500

#T-Test to examine whether the vitamin D levels differ between healthy and
disease participants

vitD_healthy <- data1_LSC598[data1_LSC598$group == 0, "vitD_level"]

vitD_disease <- data1_LSC598[data1_LSC598$group == 1, "vitD_level"]

t.test(vitD_healthy, vitD_disease)


    Welch Two Sample t-test

data:  vitD_healthy and vitD_disease
t = 0.89425, df = 64.73, p-value = 0.3745
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.896267 10.213690
sample estimates:
mean of x mean of y
 37.83714  34.67843
```

Interpretation of T-Test to examine whether the vitamin D levels differ between females and males:

The p-value (0.9194) is much larger than the typical significance level of 0.05. Therefore, we fail to reject the null hypothesis. This means that there is no statistically significant difference in the mean vitamin D levels between females and males.

The confidence interval also includes zero which supports the idea that there is no significant difference. In other words, the data does not provide sufficient evidence to conclude that there is a meaningful difference in vitamin D levels between females and males.

Interpretation of T-Test to examine whether the vitamin D levels differ between healthy and disease participants:

The p-value (0.3745) is greater than the typical significance level of 0.5. Therefore, we fail to reject the null hypothesis. Therefore, there is no statistically significant difference in the mean vitamin D levels between healthy and diseased participants.

The confidence interval also includes zero, which further supports the idea that there is no significant difference. In other words, the data does not provide sufficient evidence to conclude that there is a meaningful difference in vitamin D levels between the healthy and diseased participants.