

Patel's Week 7 Dataset Two Sample Tests

Harsh Patel

2024-10-11

```
library(readxl)

Chronic_Kidney_Disease_data <-
read_excel("C:/Users/hpate/Downloads/Chronic_Kidney_Disease_data.xlsx")
```

I started by loading the Chronic Kidney Disease dataset from an Excel file. This dataset contains various health metrics related to chronic kidney disease, which I will analyze further.

```
hemoglobin_levels <- Chronic_Kidney_Disease_data$HemoglobinLevels
gender <- Chronic_Kidney_Disease_data$Gender
```

Here, I extracted the hemoglobin levels and gender variables from the dataset. I plan to analyze how hemoglobin levels differ between males and females in this dataset.

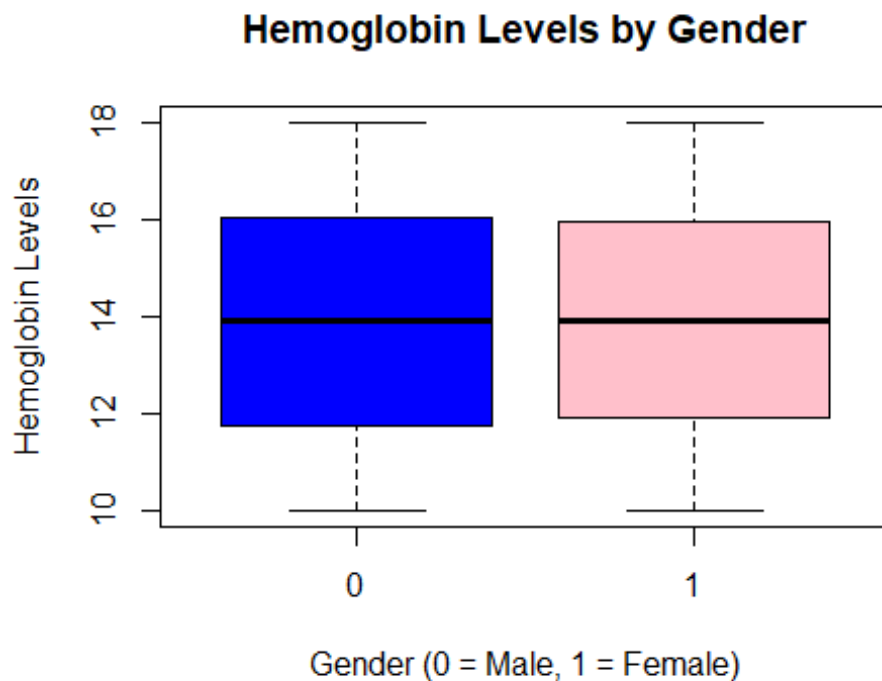
```
summary(hemoglobin_levels)

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  10.00   11.84   13.91   13.93   16.01   18.00

summary(gender)

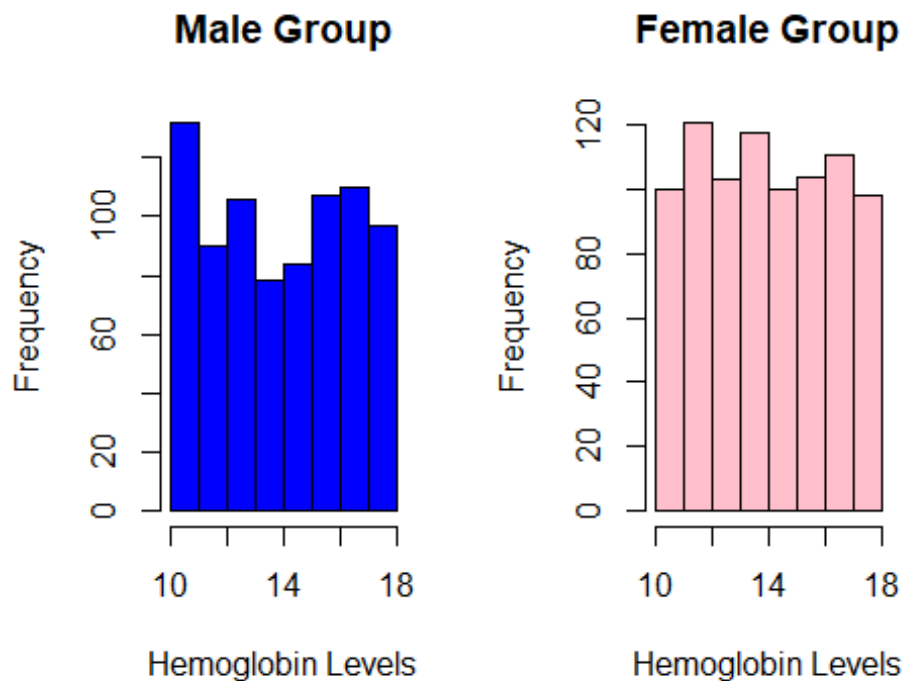
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000   0.0000   1.0000   0.5154   1.0000   1.0000

boxplot(hemoglobin_levels ~ gender,
        xlab = "Gender (0 = Male, 1 = Female)",
        ylab = "Hemoglobin Levels",
        main = "Hemoglobin Levels by Gender",
        col = c("blue", "pink"))
```



The summary statistics reveal that hemoglobin levels range from 10.00 to 18.00, with a mean close to 13.93. The gender summary shows a slight predominance of females (mean of 0.5154) in the dataset. The boxplot illustrates the distribution of hemoglobin levels by gender, indicating that males tend to have a slightly wider range compared to females.

```
par(mfrow = c(1, 2))  
  
hist(hemoglobin_levels[gender == 0], main = "Male Group", xlab = "Hemoglobin  
Levels", col = "blue")  
hist(hemoglobin_levels[gender == 1], main = "Female Group", xlab =  
"Hemoglobin Levels", col = "pink")
```



In the histograms, I observed the distribution of hemoglobin levels for both genders. The male group has a broader distribution, while the female group's distribution is more concentrated around the mean, suggesting potential differences in hemoglobin levels between the two groups.

```
t_test_result <- t.test(hemoglobin_levels ~ gender,
                        alternative = "two.sided",
                        var.equal = FALSE)
```

I conducted a t-test to assess whether there is a significant difference in hemoglobin levels between males and females. This statistical test will help determine if the observed differences are meaningful or likely due to random variation.

```
t_stat <- t_test_result$statistic
p_value <- t_test_result$p.value
conf_int <- t_test_result$conf.int

print(paste("Test Statistic (t):", t_stat))

## [1] "Test Statistic (t): -0.269031666466681"
```

```
print(paste("p-value:", p_value))
## [1] "p-value: 0.787939265732962"
print(paste("95% Confidence Interval:", conf_int[1], "to", conf_int[2]))
## [1] "95% Confidence Interval: -0.25673255445924 to 0.194799536231591"
```

The t-test results indicate a test statistic of approximately -0.269 and a p-value of 0.788. The 95% confidence interval ranges from -0.257 to 0.195. These results suggest that the difference in hemoglobin levels between genders is not statistically significant.

```
if (p_value < 0.05) {
  print("Conclusion: We reject the null hypothesis. There is a significant
  difference in Hemoglobin Levels between males and females.")
} else {
  print("Conclusion: We fail to reject the null hypothesis. There is no
  significant difference in Hemoglobin Levels between males and females.")
}
## [1] "Conclusion: We fail to reject the null hypothesis. There is no
  significant difference in Hemoglobin Levels between males and females."
```

Based on the p-value of 0.788, I fail to reject the null hypothesis. This conclusion indicates that there is no significant difference in hemoglobin levels between males and females in this dataset, suggesting that gender does not impact hemoglobin levels in this particular group.