

Patel's Week 6 ePortfolio CKD Dataset

Harsh Patel

2024-10-04

Sampling Distribution of the sample mean

```
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
```

```
library(ggplot2)  
library(readxl)
```

```
Chronic_Kidney_Disease_data <- read_excel("Chronic_Kidney_Disease_data.xlsx")
```

I loaded the necessary libraries for my analysis. The dplyr package helps me manipulate data efficiently, ggplot2 is great for creating visualizations, and readxl allows me to import the Excel file containing the chronic kidney disease data.

```
ckd_patients <- Chronic_Kidney_Disease_data %>% filter(Diagnosis == 1)
```

I filtered the dataset to focus only on patients diagnosed with chronic kidney disease (CKD). This step is crucial as it ensures my analysis is specific to this patient group, allowing for accurate conclusions about CKD patients.

```
mean_bmi_ckd <- mean(ckd_patients$BMI, na.rm = TRUE)
mean_fasting_blood_sugar_ckd <- mean(ckd_patients$FastingBloodSugar, na.rm = TRUE)

print(paste("Mean BMI for CKD patients:", round(mean_bmi_ckd, 2)))
```

```
## [1] "Mean BMI for CKD patients: 27.74"
```

```
print(paste("Mean Fasting Blood Sugar for CKD patients:", round(mean_fasting_blood_sugar_ckd, 2)))
```

```
## [1] "Mean Fasting Blood Sugar for CKD patients: 133.57"
```

I calculated the average BMI and fasting blood sugar levels for CKD patients. The results show a mean BMI of 27.74, indicating that these patients are typically overweight. The mean fasting blood sugar of 133.57 mg/dL is elevated, suggesting potential insulin resistance or prediabetes. This wider spread indicates more variability in the BMI of CKD patients, which could be due to factors such as differences in body composition, age, or disease progression among the patients.

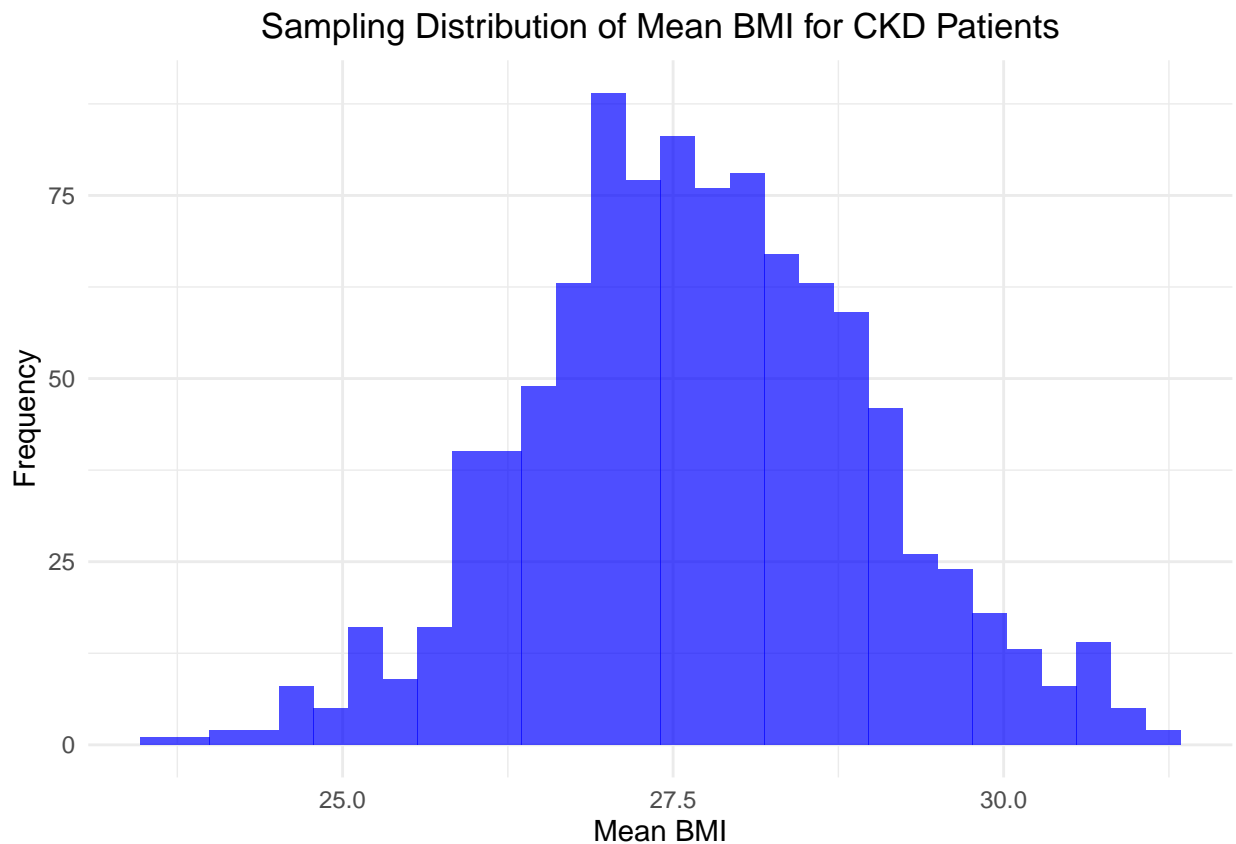
```
set.seed(2000)
sample_size <- 30
replicates <- 1000

sample_means_bmi <- numeric(replicates)
sample_means_fasting_blood_sugar <- numeric(replicates)

for (k in 1:replicates) {
  sample_data <- ckd_patients %>% sample_n(sample_size, replace = TRUE)
  sample_means_bmi[k] <- mean(sample_data$BMI, na.rm = TRUE)
  sample_means_fasting_blood_sugar[k] <- mean(sample_data$FastingBloodSugar, na.rm = TRUE)
}
```

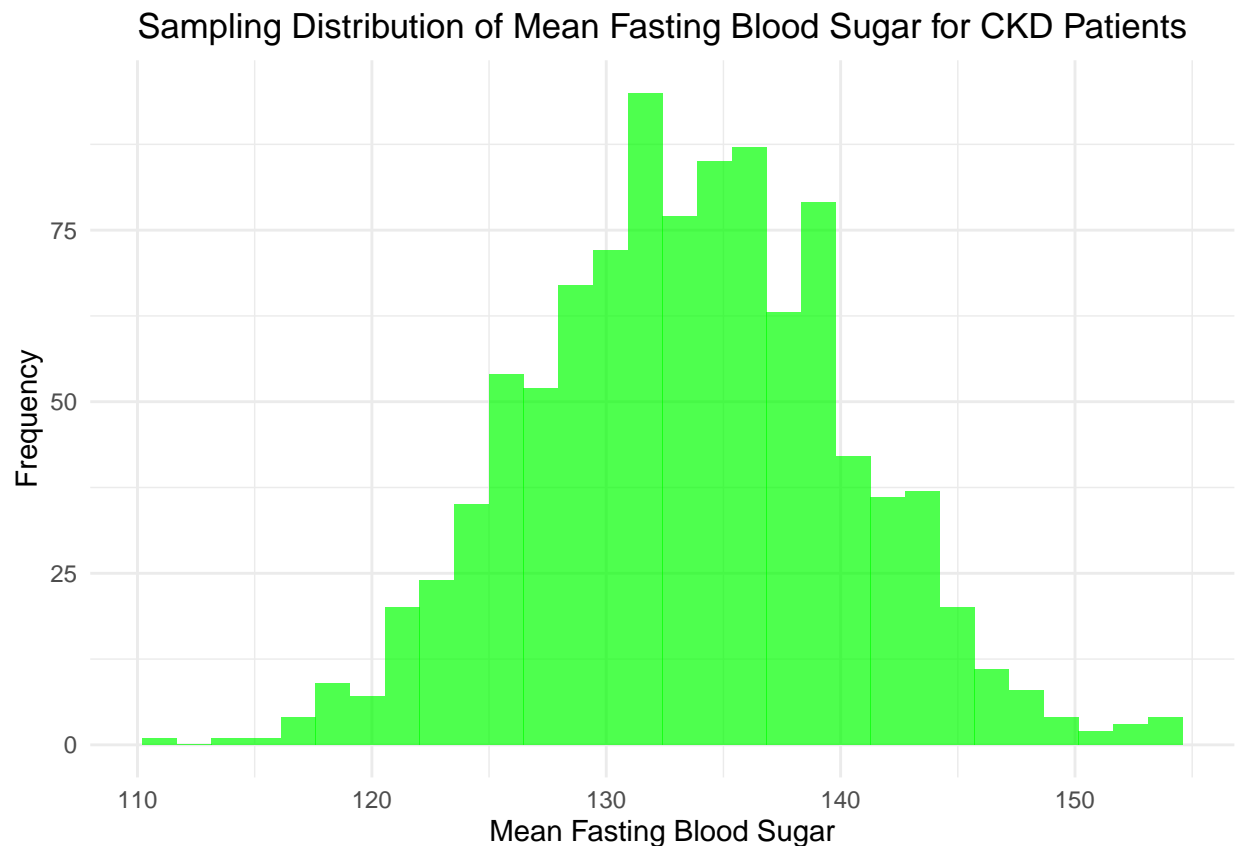
I set up a sampling process with a sample size of 30, repeating it 1,000 times. Using `set.seed(2000)` ensures my results are reproducible. For each repetition, I randomly selected patients and calculated their average BMI and fasting blood sugar. This method helps me understand how these averages might vary across different patient groups.

```
ggplot(data = data.frame(sample_means_bmi), aes(x = sample_means_bmi)) +  
  geom_histogram(bins = 30, fill = "blue", alpha = 0.7) +  
  labs(title = "Sampling Distribution of Mean BMI for CKD Patients",  
       x = "Mean BMI",  
       y = "Frequency") +  
  theme_minimal() +  
  theme(plot.title = element_text(hjust = 0.5))
```



The histogram I created shows how the average BMI is distributed across my samples. It looks roughly normal, centered around 27.5-28, which aligns with my earlier calculations. The distribution is fairly symmetrical, indicating a balanced spread of BMI values. There's some variability in BMI, possibly due to factors like age or disease progression.

```
ggplot(data = data.frame(sample_means_fasting_blood_sugar), aes(x = sample_means_fasting_blood_sugar)) +  
  geom_histogram(bins = 30, fill = "green", alpha = 0.7) +  
  labs(title = "Sampling Distribution of Mean Fasting Blood Sugar for CKD Patients",  
        x = "Mean Fasting Blood Sugar",  
        y = "Frequency") +  
  theme_minimal() +  
  theme(plot.title = element_text(hjust = 0.5))
```



The second histogram displays the distribution of average fasting blood sugar levels. It's also approximately normal, centered around 133-134 mg/dL, matching my earlier calculations. The distribution is slightly skewed to the right, suggesting some patients have higher than average blood sugar levels. The range is relatively narrow, from about 120 to 150 mg/dL.

Interval estimates for the mean

```
mean_bmi_ckd <- mean(ckd_patients$BMI, na.rm = TRUE)
sd_bmi_ckd <- sd(ckd_patients$BMI, na.rm = TRUE)

mean_fasting_blood_sugar_ckd <- mean(ckd_patients$FastingBloodSugar, na.rm = TRUE)
sd_fasting_blood_sugar_ckd <- sd(ckd_patients$FastingBloodSugar, na.rm = TRUE)
```

I calculated the mean and standard deviation for BMI and fasting blood sugar in CKD patients. These statistics help me understand the typical values and variability of these health metrics in this patient group.

```
print(paste("Mean BMI for CKD patients:", round(mean_bmi_ckd, 2)))
```

```
## [1] "Mean BMI for CKD patients: 27.74"
```

```
print(paste("Standard Deviation of BMI for CKD patients:", round(sd_bmi_ckd, 2)))
```

```
## [1] "Standard Deviation of BMI for CKD patients: 7.32"
```

```
print(paste("Mean Fasting Blood Sugar for CKD patients:", round(mean_fasting_blood_sugar_ckd, 2)))
```

```
## [1] "Mean Fasting Blood Sugar for CKD patients: 133.57"
```

```
print(paste("Standard Deviation of Fasting Blood Sugar for CKD patients:", round(sd_fasting_blood_sugar_ckd, 2)))
```

```
## [1] "Standard Deviation of Fasting Blood Sugar for CKD patients: 36.48"
```

I displayed the calculated means and standard deviations for both BMI and fasting blood sugar levels among CKD patients. I found that the mean BMI is 27.74 (indicating an overweight status) with a standard deviation of 7.32, suggesting variability in BMI within the patient population. The mean fasting blood sugar level is 133.57 mg/dL with a standard deviation of 36.48, indicating elevated blood sugar levels that could pose significant health risks.

```
n <- nrow(ckd_patients)
```

I determined the sample size (n) by counting the number of rows in the dataset, which represents the total number of CKD patients. Knowing the sample size is essential for subsequent calculations, as it affects the precision of the confidence intervals for the means.

```
z_star <- qnorm(0.975)
```

I calculated the z-value (z-star) for a 95% confidence level using the qnorm function. This value is critical for determining the margin of error in the confidence interval calculations, indicating how many standard deviations from the mean correspond to the desired confidence level.

```
margin_of_error_bmi <- z_star * (sd_bmi_ckd / sqrt(n))  
ci_bmi_lower <- mean_bmi_ckd - margin_of_error_bmi  
ci_bmi_upper <- mean_bmi_ckd + margin_of_error_bmi
```

I calculated the margin of error for the mean BMI using the previously calculated z-star and the standard deviation. I then developed the lower and upper bounds of the confidence interval for BMI. This interval gives me a range within which I can be 95% confident that the true mean BMI of CKD patients lies.

```
margin_of_error_fbs <- z_star * (sd_fasting_blood_sugar_ckd / sqrt(n))  
ci_fbs_lower <- mean_fasting_blood_sugar_ckd - margin_of_error_fbs  
ci_fbs_upper <- mean_fasting_blood_sugar_ckd + margin_of_error_fbs
```

I calculated the margin of error for the mean fasting blood sugar levels and subsequently obtained the confidence interval for fasting blood sugar. This interval provides a range where I can confidently assert that the true mean fasting blood sugar level of CKD patients is likely to fall.

```
print(paste("95% Confidence Interval for BMI:", round(ci_bmi_lower, 2), "to", round(ci_bmi_upper, 2)))

## [1] "95% Confidence Interval for BMI: 27.37 to 28.11"

print(paste("95% Confidence Interval for Fasting Blood Sugar:", round(ci_fbs_lower, 2), "to", round(ci_fbs_upper, 2)))

## [1] "95% Confidence Interval for Fasting Blood Sugar: 131.74 to 135.4"
```

I printed the calculated 95% confidence intervals for both BMI and fasting blood sugar levels. The BMI confidence interval ranges from 27.37 to 28.11, while the fasting blood sugar confidence interval ranges from 131.74 to 135.4 mg/dL. These intervals indicate where I expect the true population means to lie with 95% certainty, highlighting potential health concerns related to overweight status and elevated blood sugar levels in CKD patients.

Hypothesis Testing

```
alpha <- 0.05

mean_bmi_ckd <- mean(ckd_patients$BMI, na.rm = TRUE)
sd_bmi_ckd <- sd(ckd_patients$BMI, na.rm = TRUE)
n <- nrow(ckd_patients)

mu_0 <- 25
t_statistic <- (mean_bmi_ckd - mu_0) / (sd_bmi_ckd / sqrt(n))
t_statistic

## [1] 14.61818
```

The output 14.61818 represents the t-statistic I calculated from my sample data. This value shows me how many standard deviations the sample mean BMI of CKD patients is from the hypothesized mean of 25. A t-statistic of 14.61818 is quite large, which suggests that there is a substantial difference between my sample mean and the hypothesized value. This indicates that the average BMI of CKD patients is far from 25.

```
p_value <- 2 * pt(-abs(t_statistic), df = n - 1)
p_value
```

```
## [1] 2.191174e-45
```

The output 2.191174e-45 is the p-value associated with my t-statistic. This extremely small p-value tells me that there is a very low probability of observing a sample mean as extreme as mine if the null hypothesis (that the mean BMI is 25) were true. Essentially, this p-value suggests that my observed data is highly unlikely under the null hypothesis.

```
if (p_value < alpha) {
  conclusion <- "Reject the null hypothesis: There is significant evidence that the average BMI of CKD patients is not 25."
} else {
  conclusion <- "Fail to reject the null hypothesis: There is insufficient evidence to suggest that the average BMI of CKD patients is not 25."
}

print(conclusion)
```

```
## [1] "Reject the null hypothesis: There is significant evidence that the average BMI of CKD patients is not 25."
```

The conclusion “Reject the null hypothesis: There is significant evidence that the average BMI of CKD patients is not 25.” summarizes what I found from my hypothesis test. Since the p-value is far less than the significance level of 0.05, I reject the null hypothesis. This provides strong evidence that the average BMI of CKD patients in my sample is statistically significantly different from the value of 25. This finding emphasizes the importance of considering the average BMI when evaluating the health status and management of CKD patients.