assignment_8_solution

Sahir Khan

November 11, 2024

```r
# Q1.A pipe manufacturing organization produces different kinds of pipes.
# We are given the monthly data of the wall thickness of certain types of pipes
# (data is available on LMS Clt-data.csv).
# The organization has an analysis to perform and one of the basic assumption
# of that analysis is that the data should be normally distributed.
# You have the following tasks to do:
# (a) Import the csv data file in R.
# (b) Validate data for correctness by counting number of rows and viewing the
# xtop ten rows of the dataset.
# (c) Calculate the population mean and plot the observations by making a histogram.
# (d) Mark the mean computed in last step by using the function abline.
data = read.csv("Clt-data.csv")
dim(data)
```

```
## [1] 9000    1
```

```r
head(data,10)
```

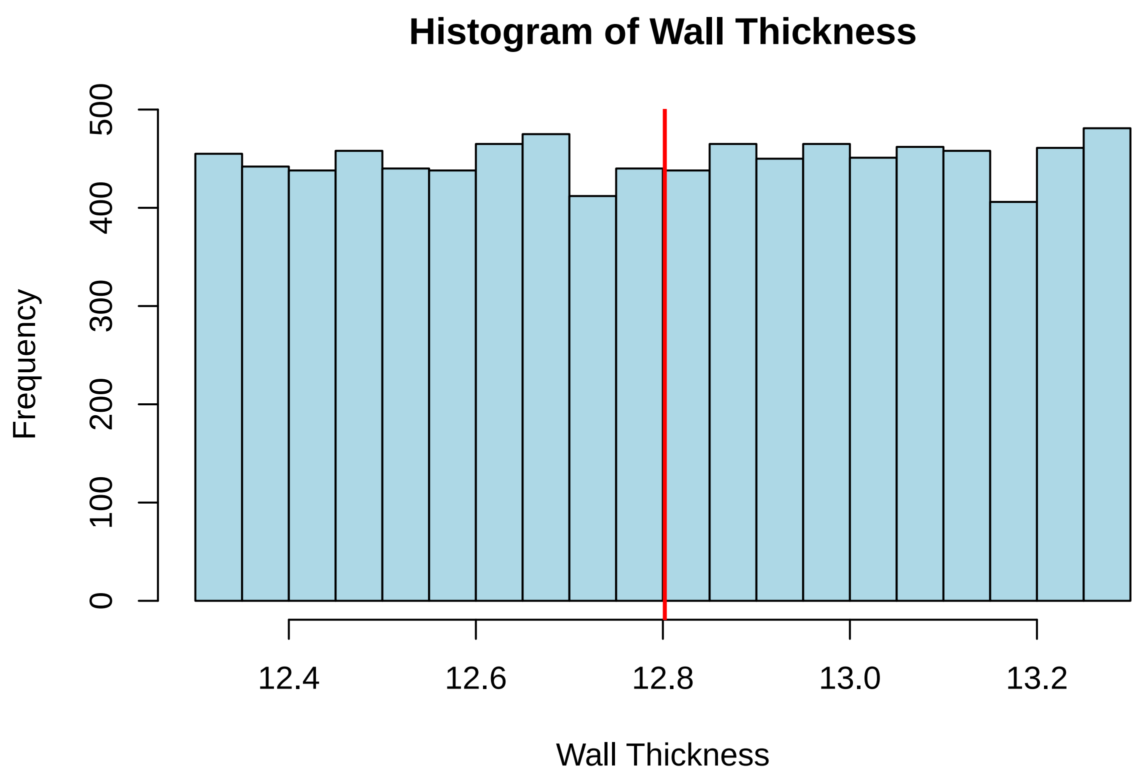
```
##    Wall.Thickness
## 1        12.35487
## 2        12.61742
## 3        12.36972
## 4        13.22335
## 5        13.15919
## 6        12.67549
## 7        12.36131
## 8        12.44468
## 9        12.62977
## 10       12.90381
```

```r
# Calculate the population mean
population_mean <- mean(data$Wall.Thickness)
print(paste("Population mean: ", population_mean))
```

```
## [1] "Population mean:  12.8020492455356"
```

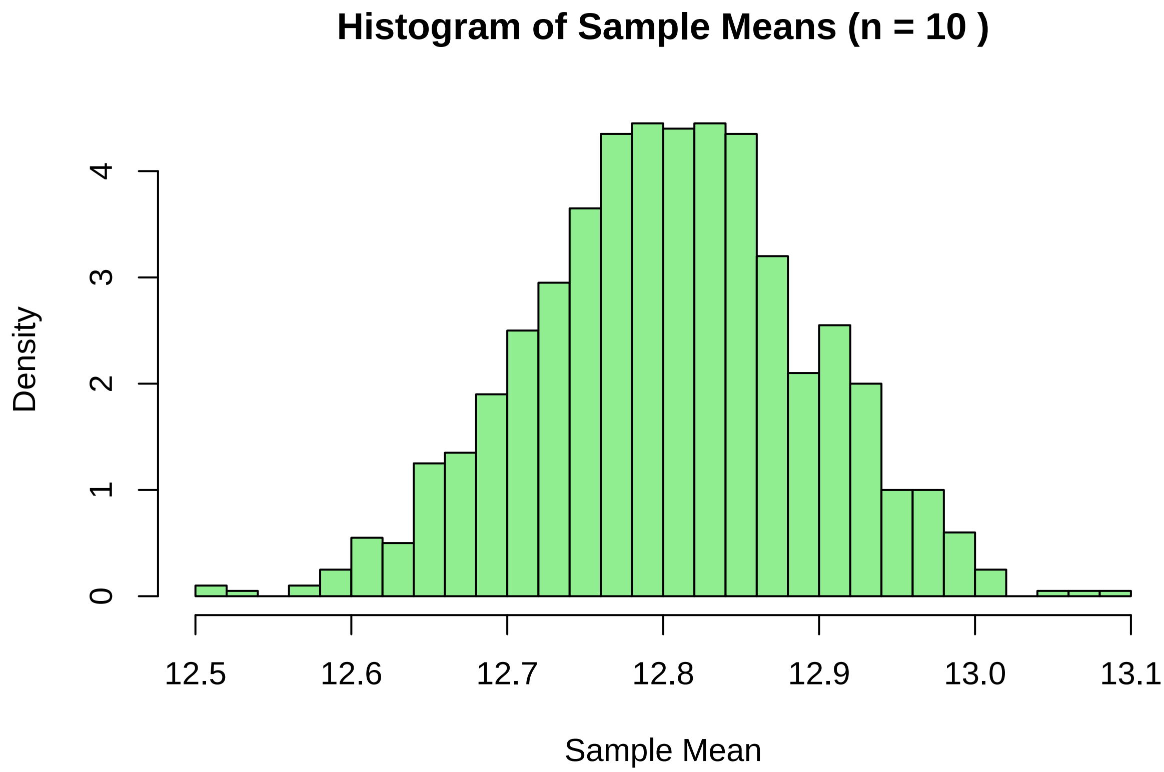
```r
# Plot a histogram of the data
hist(data$Wall.Thickness, breaks = 30,
     main = "Histogram of Wall Thickness",
     xlab = "Wall Thickness", col = "lightblue", border = "black")
```

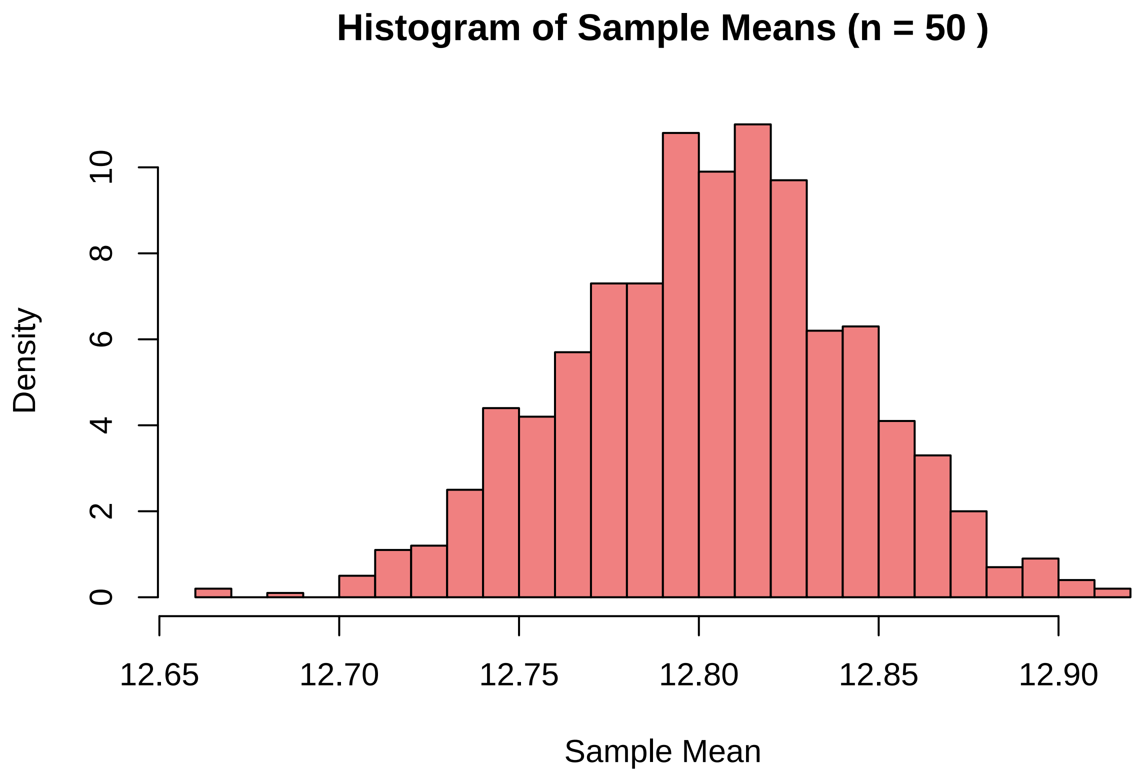
```r
# Mark the population mean on the histogram
abline(v = population_mean, col = "red", lwd = 2)
```

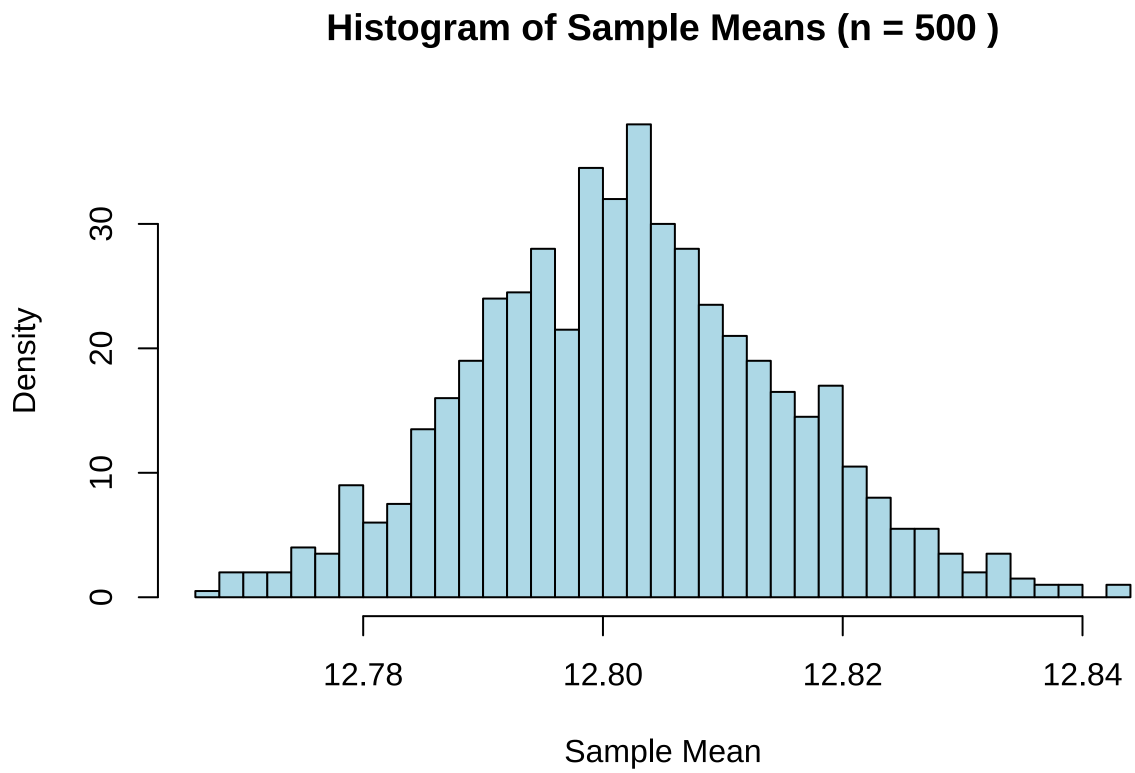


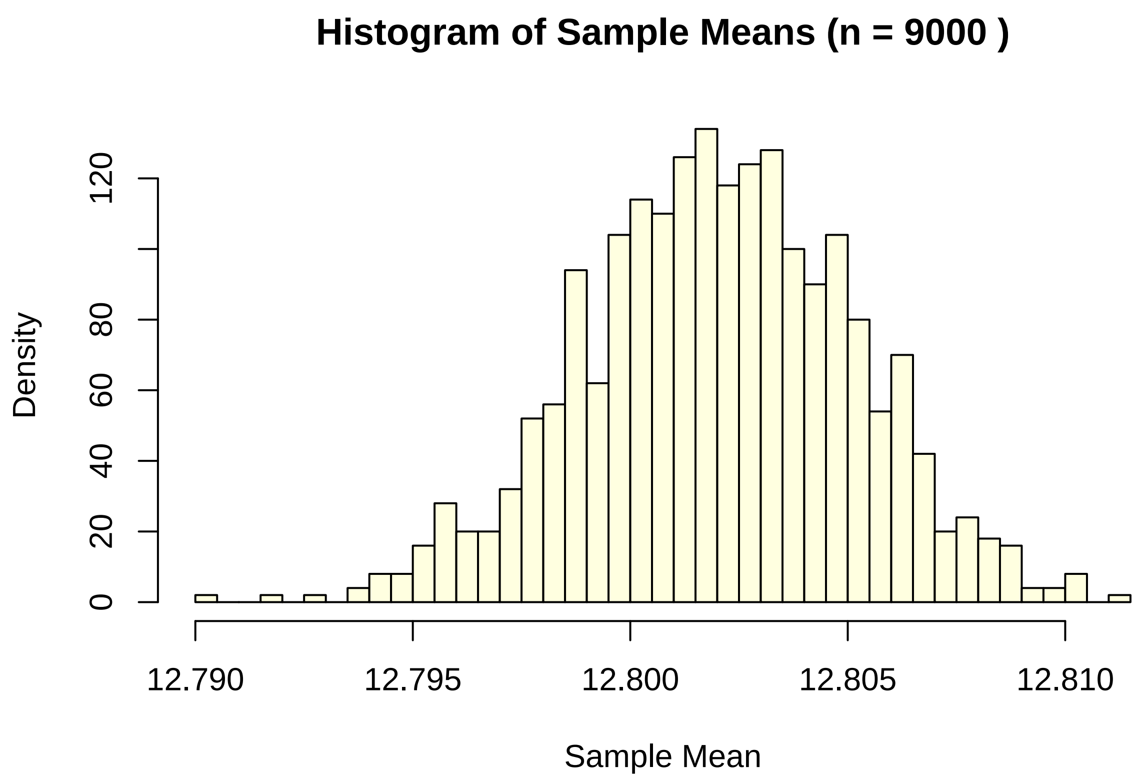

```r
# Function to draw samples and calculate mean
draw_samples <- function(sample_size, sample_colour, num_samples = 1000) {
  sample_means <- numeric(num_samples)

  for (i in 1:num_samples) {
    sample_means[i] <- mean(sample(data$Wall.Thickness, sample_size, replace = TRUE))
  }

  # Plot the histogram of sample means
  hist(sample_means, breaks = 30,
       main = paste("Histogram of Sample Means (n =", sample_size, ")"),
       xlab = "Sample Mean", col = sample_colour, freq = FALSE)
  abline(v = mean(sample_size), col = "red", lwd = 2)
}

# Sample sizes
sample_sizes <- c(10, 50, 500, 9000)
sample_colours <- c("lightgreen", "lightcoral","lightblue","lightyellow")
n=4

# Plot sample means for each sample size
for (i in 1:n) {
```

```
  draw_samples(sample_sizes[i],sample_colours[i])
}
```

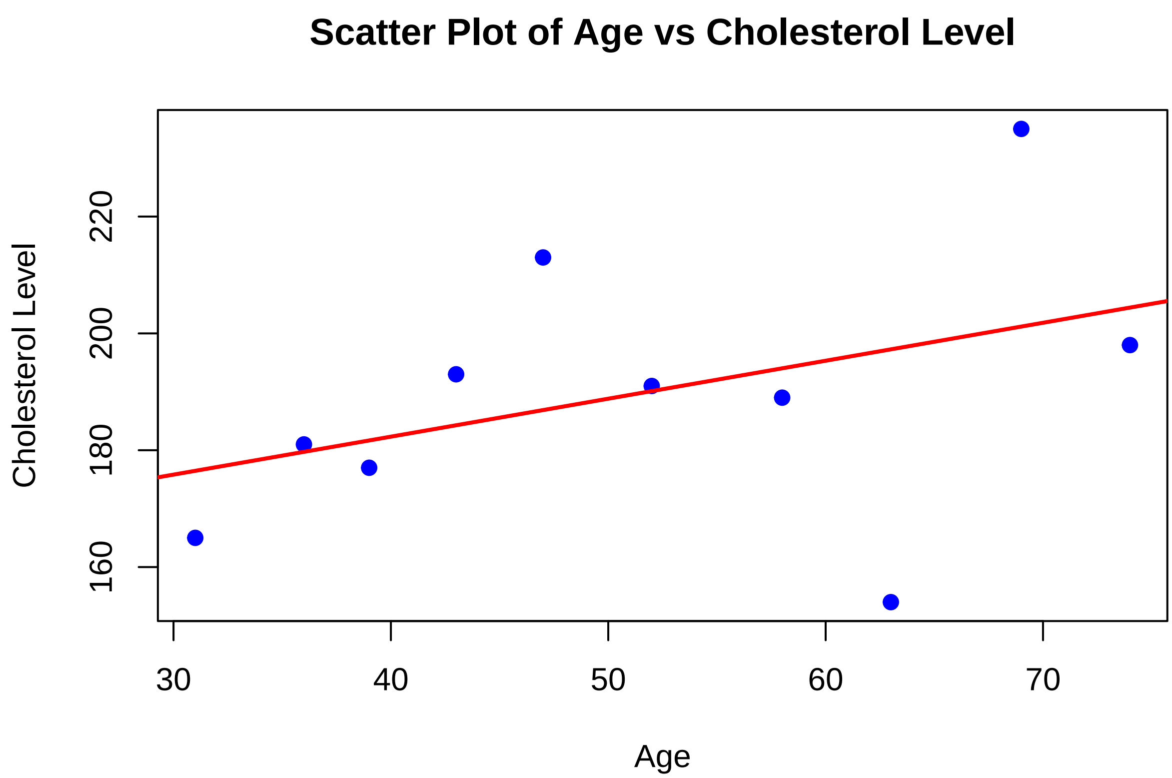



Histogram of Sample Means (n = 50)
Density
0
2
4
6
8
10
12.65
12.70
12.75
12.80
12.85
12.90
Sample Mean


Histogram of Sample Means (n = 500)
Density
0
10
20
30
12.78
12.80
12.82
12.84
Sample Mean




```
# Q2.The following table gives information on ages and cholesterol levels for a
# random sample of 10 men. Plot the scatter diagram and a regression line that
# will enable us to predict Cholesterol level on age. Further, estimate the
# cholesterol level of a 60 year-old man.
Age <- c(58, 69, 43, 39, 63, 52, 47, 31, 74, 36)
Cholesterol <- c(189, 235, 193, 177, 154, 191, 213, 165, 198, 181)

# Scatter plot
plot(Age, Cholesterol, main = "Scatter Plot of Age vs Cholesterol Level",
     xlab = "Age", ylab = "Cholesterol Level", pch = 19, col = "blue")
# Linear regression model
model <- lm(Cholesterol ~ Age)

# Add regression line to the plot
abline(model, col = "red", lwd = 2)
```




```r
# Predict cholesterol for Age = 60
predicted_cholesterol <- predict(model, data.frame(Age = 60))
print(paste("Estimated cholesterol level for a 60-year-old man:", round(predicted_cholesterol, 2)))
```

```
## [1] "Estimated cholesterol level for a 60-year-old man: 195.32"
```

```r
# Q3.A research methodology course has recently been added to the PhD curriculum
# at the Thapar Institute of Engineering and Technology, Patiala. To evaluate
# its effec- tiveness, students take a test on formulating research problems and
# writing research papers both before and after completing the course. Assume
# that the differences between the pre-course and post-course test scores are
# normally distributed, and a high score on the test indicates a strong level of
# assertive- ness. Do the collected data, at 5% level of significance, provide
# enough evidence to conclude that research scholars become more assertive after
# completing the course?
before_test <- c(145, 173, 158, 141, 167, 159, 154, 167, 145, 153)
after_test <- c(155, 167, 156, 149, 168, 162, 158, 169, 157, 161)

# Differences between before and after scores
differences <- after_test - before_test
print(differences)
```

```
##  [1] 10 -6 -2  8  1  3  4  2 12  8
```

```r
# Paired t-test
t_test_result <- t.test(after_test, before_test, paired = TRUE, alternative = "greater")
print(t_test_result)
```

```
##
##  Paired t-test
##
## data:  after_test and before_test
## t = 2.2597, df = 9, p-value = 0.0251
## alternative hypothesis: true mean difference is greater than 0
## 95 percent confidence interval:
##  0.7551657       Inf
## sample estimates:
## mean difference
##               4
```