

Statistical-Inference Project 1

Raghda

2024-05-11

This is an R Markdown document that include Statistical-Inference Project. The project has two parts that I will address bellow. # Part 1 ## Set parameters

```
lambda <- 0.2 # Rate parameter of exponential distribution
sample_size <- 40 # Number of samples
num_samples <- 10000 # Number of samples to generate
mu <- 1 / lambda # Mean of exponential distribution
sigma <- 1 / lambda # Standard deviation
set.seed(10000)
```

Generate random numbers from exponential distribution

```
n <- 1000 # Number of random numbers to generate
random_numbers <- rexp(n, rate = lambda)
```

Generate sample of averages

```
average_values <- replicate(num_samples, mean(rexp(sample_size, rate = lambda)))
```

Calculate sample mean and theoretical mean

```
sample_mean <- mean(average_values)
theoretical_mean <- mu
```

Calculate sample Variance and theoretical Variance

```
sample_Variance <- var(average_values)
theoretical_Variance <- (1/lambda)^2/sample_size # Variance of exponential distribution
```

Plot the results

```

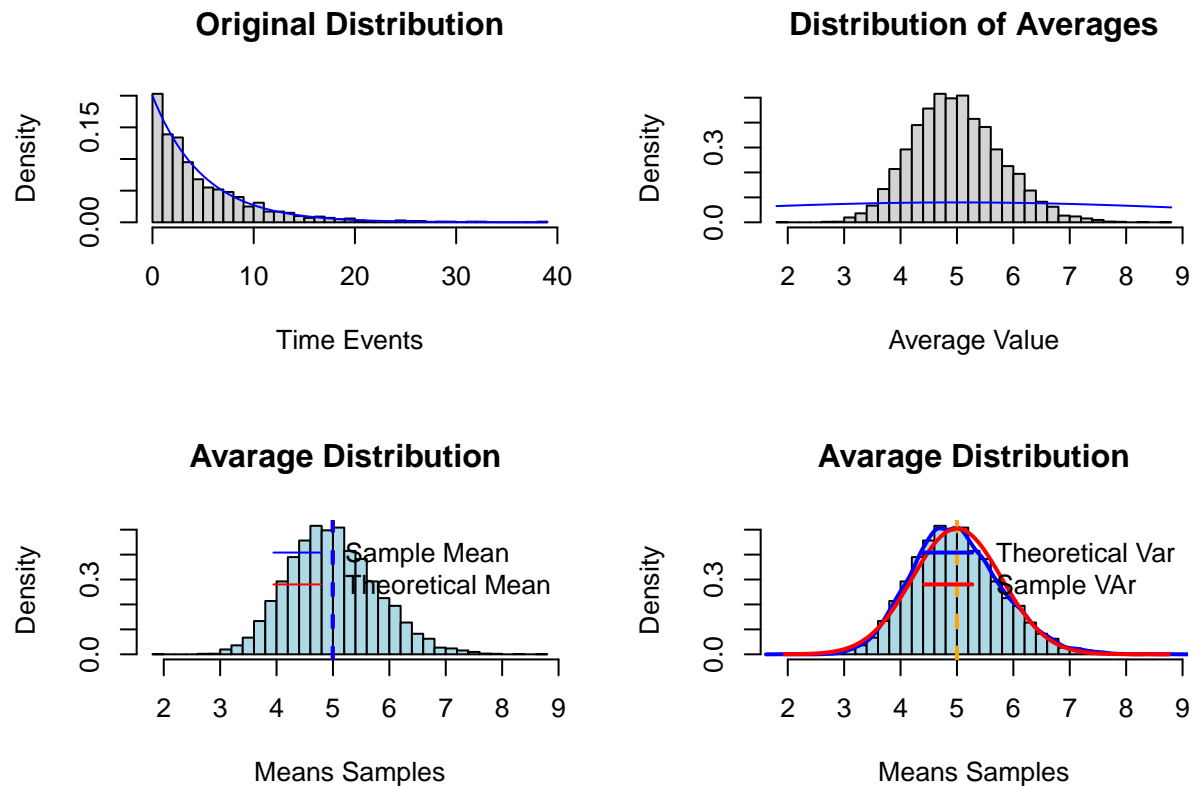
par(mfrow = c(2, 2))# Set up the plotting area with 2 rows and 2 columns
# Plot the original exponential distribution
hist(random_numbers, freq = FALSE,
      breaks = 30, main = "Original Distribution",
      xlab = "Time Events")
curve(dexp(x, rate = lambda),
      col = "blue", add = TRUE)

# Plot the distribution of averages of 40 exponentials
hist(average_values, freq = FALSE, breaks = 30,
      main = "Distribution of Averages",
      xlab = "Average Value")
curve(dnorm(x, mean = mu, sd = sigma),
      col = "blue", add = TRUE)

# Plot the distribution of sample means with sample and theoretical means
hist(average_values, breaks = 30, prob = TRUE,
      col = "lightblue", main = "Avarage Distribution ",
      xlab = "Means Samples", ylab = "Density")
abline(v = theoretical_mean, col = "red", lwd = 2, lty = 2)
abline(v = sample_mean, col = "blue", lwd = 2, lty = 2)
legend('topright', c("Sample Mean", "Theoretical Mean"),
      bty = "n", lty = c(1, 1), col = c("blue", "red"))

# Plot the distribution of sample means with density curve and theoretical variance
hist(average_values, breaks = 30, prob = TRUE,
      col = "lightblue", main = "Avarage Distribution",
      xlab = "Means Samples", ylab = "Density")
lines(density(average_values), col = "blue", lwd = 2)
abline(v = 1/lambda, col = "orange", lwd = 2, lty = 2)
xfit <- seq(min(average_values), max(average_values), length = 100)
yfit <- dnorm(xfit, mean = 1/lambda, sd = (1/lambda/sqrt(sample_size)))
lines(xfit, yfit, col = "red", lwd = 2)
legend('topright', c("Theoretical Var", "Sample VAr"),
      bty = "n", lwd = c(2, 2), col = c("blue", "red"))

```



Conclusion

The distribution of averages of 40 exponentials matches a normal distribution, based on the visual inspection of the plot number two or (1,2). The histogram of the averages of 40 exponentials exhibits a bell-shaped curve, closely resembling the theoretical normal distribution curve overlaid on the plot. Also the mean and the Variance of the sampled data is almost perfectly estimating the mean and Variance of the Original exponential distribution.

Part 2

In this part we are supposed to analyze the ToothGrowth data in the R datasets package. `## Load the ToothGrowth data`

```
data(ToothGrowth)
# View the structure of the dataset
str(ToothGrowth)

## 'data.frame':  60 obs. of  3 variables:
## $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
## $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
## $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
# Summarize the dataset
summary(ToothGrowth)
```

```
##      len      supp      dose
## Min.   : 4.20   OJ:30   Min.    :0.500
## 1st Qu.:13.07   VC:30   1st Qu.:0.500
## Median :19.25                Median :1.000
## Mean   :18.81                Mean    :1.167
## 3rd Qu.:25.27                3rd Qu.:2.000
## Max.   :33.90                Max.    :2.000
```

Plotting the data in four different ways

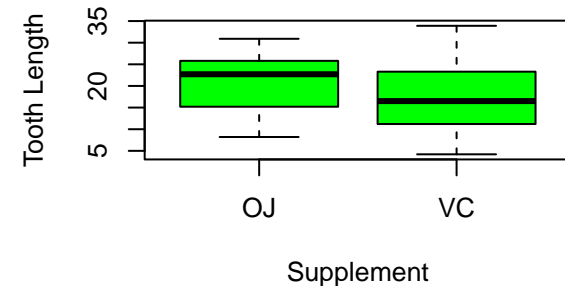
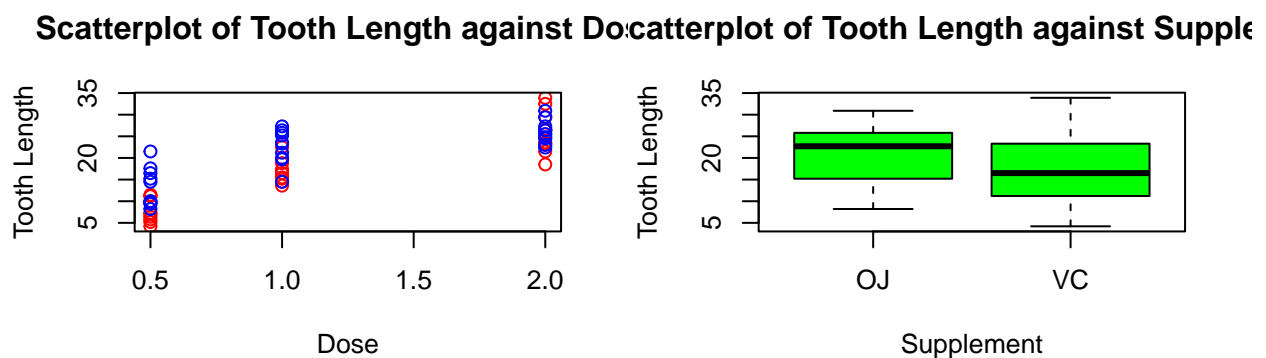
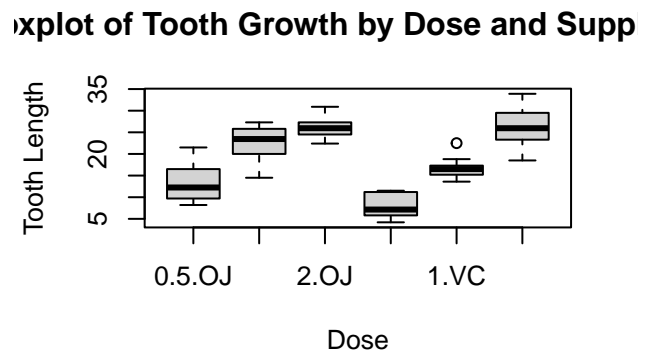
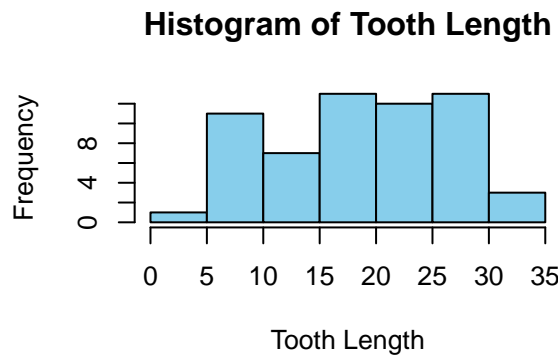
```
par(mfrow = c(2, 2))

# Visualize the distributions of variables
hist(ToothGrowth$len, main = "Histogram of Tooth Length",
     xlab = "Tooth Length",
     col = "skyblue")

# Boxplot of tooth growth by dose and supplement
boxplot(len ~ dose + supp, data = ToothGrowth,
        main = "Boxplot of Tooth Growth by Dose and Supplement",
        xlab = "Dose",
        ylab = "Tooth Length")

# Scatterplot of tooth length against dose
plot(len ~ dose, data = ToothGrowth,
     main = "Scatterplot of Tooth Length against Dose",
     xlab = "Dose",
     ylab = "Tooth Length",
     col = ifelse(ToothGrowth$supp == "VC", "red", "blue"))

# Scatterplot of tooth length against supplement
plot(len ~ supp, data = ToothGrowth,
     main = "Scatterplot of Tooth Length against Supplement",
     xlab = "Supplement",
     ylab = "Tooth Length",
     col = ifelse(ToothGrowth$dose == 0.5, "green", "orange"))
```



Using T-test to compare tooth growth by supp and dose

```
t05 <- t.test(len ~ supp,
  data = rbind(ToothGrowth[(ToothGrowth$dose == 0.5) &
    (ToothGrowth$supp == "OJ"),],
    ToothGrowth[(ToothGrowth$dose == 0.5) &
    (ToothGrowth$supp == "VC"),]),
  var.equal = FALSE)

t1 <- t.test(len ~ supp,
  data = rbind(ToothGrowth[(ToothGrowth$dose == 1) &
    (ToothGrowth$supp == "OJ"),],
    ToothGrowth[(ToothGrowth$dose == 1) &
    (ToothGrowth$supp == "VC"),]),
  var.equal = FALSE)

t2 <- t.test(len ~ supp,
  data = rbind(ToothGrowth[(ToothGrowth$dose == 2) &
    (ToothGrowth$supp == "OJ"),],
    ToothGrowth[(ToothGrowth$dose == 2) &
    (ToothGrowth$supp == "VC"),]),
  var.equal = FALSE)

# Make summary of the conducted t.tests, which compare the delivery methods by dosage
```

```

# take p-values and CI
summaryBYsupp <- data.frame(
  "p-value" = c(t05$p.value, t1$p.value, t2$p.value),
  "Conf.Low" = c(t05$conf.int[1], t1$conf.int[1], t2$conf.int[1]),
  "Conf.High" = c(t05$conf.int[2], t1$conf.int[2], t2$conf.int[2]),
  row.names = c("Dosage .05", "Dosage 1", "Dosage 2"))
# Show the data table
summaryBYsupp

```

```

##           p.value  Conf.Low Conf.High
## Dosage .05 0.006358607  1.719057  8.780943
## Dosage 1   0.001038376  2.802148  9.057852
## Dosage 2   0.963851589 -3.798070  3.638070

```

Conclusion

We reject the null hypothesis, which states that there is no difference in tooth growth between the .5 and 1 milligrams/day, with a 95% confidence level. We find p-values below the .05 threshold, and there is no zero in the confidence levels. So, the mode of distribution matters for dosages of .5 and 1 mg each day. We are unable to reject the null hypothesis, which states that there is no difference in tooth growth depending on the delivery strategy for two milligrams per day, with a 95% confidence level. We see p-values that are higher than the .05 threshold, and the confidence intervals include 0. Therefore, the mode of delivery is irrelevant for a dosage of two milligrams per day.