Basic Inferential Data Analysis

Vitalii Diakonov

2024-03-21

Overview:

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

- Load the ToothGrowth data and perform some basic exploratory data analyses
- Provide a basic summary of the data.
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
- State your conclusions and the assumptions needed for your conclusions.

Analysis:

Let's dive into the ToothGrowth dataset and explore its nuances. First off, I'll load the data and give it a once-over to see what we're working with:

```
# Load the ToothGrowth data
data("ToothGrowth")
tooth_growth <- ToothGrowth

# Perform basic exploratory data analysis
str(tooth_growth)
```

```
'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 ...
```

This dataset consists of 60 observations and 3 variables: 'len' representing the length of tooth growth, 'supp' indicating the supplement type ('OJ' or 'VC'), and 'dose' representing the dosage level. 'supp' is a factor variable with 2 levels: 'OJ' and 'VC'. 'dose' is a numerical variable representing the dosage level in milligrams.

```
# Check unique values in the 'dose' column
unique_doses <- unique(tooth_growth$dose)
unique_doses
```

```
[1] 0.5 1.0 2.0
```

I've explored the unique values in the 'dose' column to understand the dosage levels. The unique values are 0.5, 1, and 2, indicating three different dosage levels.

```
# Convert 'dose' to a factor variable
tooth_growth$dose <- factor(tooth_growth$dose)
str(tooth_growth)
```

```
'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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

I converted the 'dose' column to a factor variable to better represent the categorical nature of the dosage levels. Now, the 'dose' variable is a factor with 3 levels: 0.5, 1, and 2.

```
# Summary statistics of the ToothGrowth data
summary(tooth_growth)
```

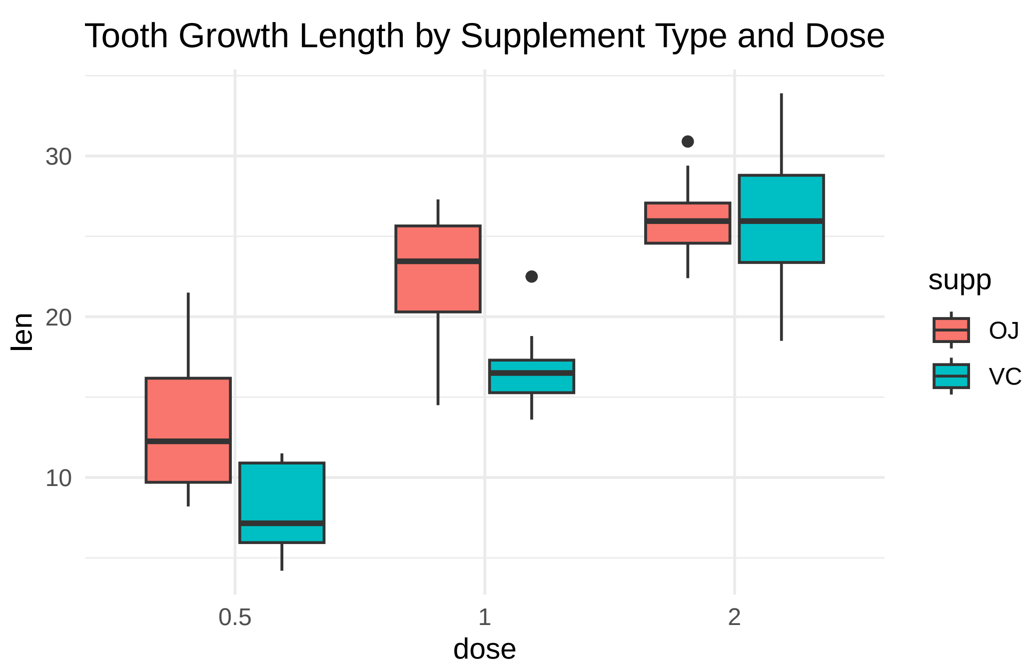
```
      len        supp     dose   
 Min.   : 4.20   OJ:30   0.5:20  
 1st Qu.:13.07   VC:30   1  :20  
 Median :19.25           2  :20  
 Mean   :18.81                   
 3rd Qu.:25.27                   
 Max.   :33.90                   
```

Here's a breakdown of the summary:

- **len (Length):** The minimum tooth growth length is 4.20, and the maximum length is 33.90. The median length is 19.25, with a mean of approximately 18.81. The first quartile (25th percentile) is 13.07, and the third quartile (75th percentile) is 25.27.
- **supp (Supplement Type):** There are two levels in the 'supp' variable: 'OJ' and 'VC'. Each level has 30 observations.
- **dose (Dosage Level):** The 'dose' variable represents the dosage level, with three levels: 0.5, 1, and 2. Each dosage level has 20 observations.

Next, I've created a boxplot to visualize the distribution of tooth growth lengths based on supplement type ('OJ' or 'VC') and dose levels (0.5, 1, and 2). This visualization helps in understanding the variation in tooth growth lengths across different supplement types and dosage levels.

```r
# Visualize tooth growth based on supplement type and dose
library(ggplot2)
ggplot(tooth_growth, aes(x = dose, y = len, fill = supp)) +
    geom_boxplot() +
    labs(title = "Tooth Growth Length by Supplement Type and Dose") +
    theme_minimal() +
    theme(plot.title = element_text(hjust = 0.5))
```



The boxplot indicates that the disparities in tooth growth are more pronounced at the 0.5 and 1.0 mg/day dosage levels as opposed to the 2.0 mg/day dosage level. There's an observable trend where tooth length appears to increase with the dose. Additionally, the visual data suggests that for the 0.5 and 1.0 mg/day doses, the OJ supplement seems to lead to greater tooth growth, while at the 2.0 mg/day dose, the two supplements, OJ and VC, appear similarly effective.

T.test Hypothesis Testing

```
# Run t-tests and store full results
t_tests_full_results <- list()
means <- data.frame(dose = numeric(), supp = character(), mean_len = numeric(), stringsAsF

for(d in unique(tooth_growth$dose)) {
  # Conduct t-test
  t_test <- t.test(len ~ supp, data = tooth_growth, subset = dose == d)

  # Store the full t-test results
  t_tests_full_results[[as.character(d)]] <- t_test

  # Extract and store means
  mean_len_oj <- t_test$estimate["mean in group OJ"]
  mean_len_vc <- t_test$estimate["mean in group VC"]

  means <- rbind(means, data.frame(dose = d, supp = "OJ", mean_len = mean_len_oj))
  means <- rbind(means, data.frame(dose = d, supp = "VC", mean_len = mean_len_vc))
}

# Output the means for inspection
print(means)
```

```
                  dose supp mean_len
mean in group OJ   0.5   OJ    13.23
mean in group VC   0.5   VC     7.98
mean in group OJ1    1   OJ    22.70
mean in group VC1    1   VC    16.77
mean in group OJ2    2   OJ    26.06
mean in group VC2    2   VC    26.14
```

```r
# Output the full t-test results
print(t_tests_full_results)
```

```
$`0.5`

	Welch Two Sample t-test

data:  len by supp
t = 3.1697, df = 14.969, p-value = 0.006359
alternative hypothesis: true difference in means between group OJ and group VC is not equal t
95 percent confidence interval:
 1.719057 8.780943
sample estimates:
mean in group OJ mean in group VC 
           13.23             7.98 


$`1`

	Welch Two Sample t-test

data:  len by supp
t = 4.0328, df = 15.358, p-value = 0.001038
alternative hypothesis: true difference in means between group OJ and group VC is not equal t
95 percent confidence interval:
 2.802148 9.057852
sample estimates:
mean in group OJ mean in group VC 
           22.70            16.77 


$`2`

	Welch Two Sample t-test

data:  len by supp
t = -0.046136, df = 14.04, p-value = 0.9639
alternative hypothesis: true difference in means between group OJ and group VC is not equal t
95 percent confidence interval:
 -3.79807  3.63807
sample estimates:
mean in group OJ mean in group VC 
```

```
26.06            26.14
```

- **Dose 0.5 mg/day**: The mean tooth length for OJ is 13.23, and for VC is 7.98. Since the p-value is 0.00636, which is less than the alpha level of 0.05, this difference is statistically significant. Therefore, OJ is more effective than VC at this dosage.
- **Dose 1 mg/day**: The mean tooth length for OJ is 22.70, and for VC is 16.77. With a p-value of 0.00104, this significant difference also suggests that OJ is more effective than VC at this dosage.
- **Dose 2 mg/day**: The mean tooth lengths are 26.06 for OJ and 26.14 for VC, which are virtually identical. The p-value is 0.96385, indicating no significant difference between the supplements at this dose level. Therefore, both OJ and VC are equally effective at the 2 mg/day dosage.

Conclusion

In summary, OJ is more effective at promoting tooth growth at lower doses (0.5 and 1 mg/day), while at the highest dose tested (2 mg/day), there is no significant difference in efficacy between OJ and VC.