StatInference Part 2

2024-10-05

R Markdown

This analysis examines the ToothGrowth dataset to compare the effects of different supplements (Orange Juice and Ascorbic Acid) on guinea pig tooth growth at varying dosages (0.5, 1.0, and 2.0 mg/day). The study begins with exploratory data analysis (EDA) to visualize the relationship between supplement type, dosage, and tooth length using boxplots. The results indicate that orange juice significantly enhances tooth growth compared to ascorbic acid at dosages of 0.5 and 1.0 mg/day, while both supplements show similar effectiveness at a dosage of 2.0 mg/day. The analysis is grounded in the assumptions of normal distribution of tooth lengths and the absence of unmeasured factors affecting the results.

```
### Load the ToothGrowth Data and Perform Exploratory Data Analysis
# Load the necessary library and dataset
library(datasets)
data(ToothGrowth)

# Display the structure and summary of the dataset
str(ToothGrowth)  # Structure of the dataset
```

```
## '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 ...
```

```
head(ToothGrowth) # First few rows of the dataset
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
summary(ToothGrowth) # Summary statistics
```

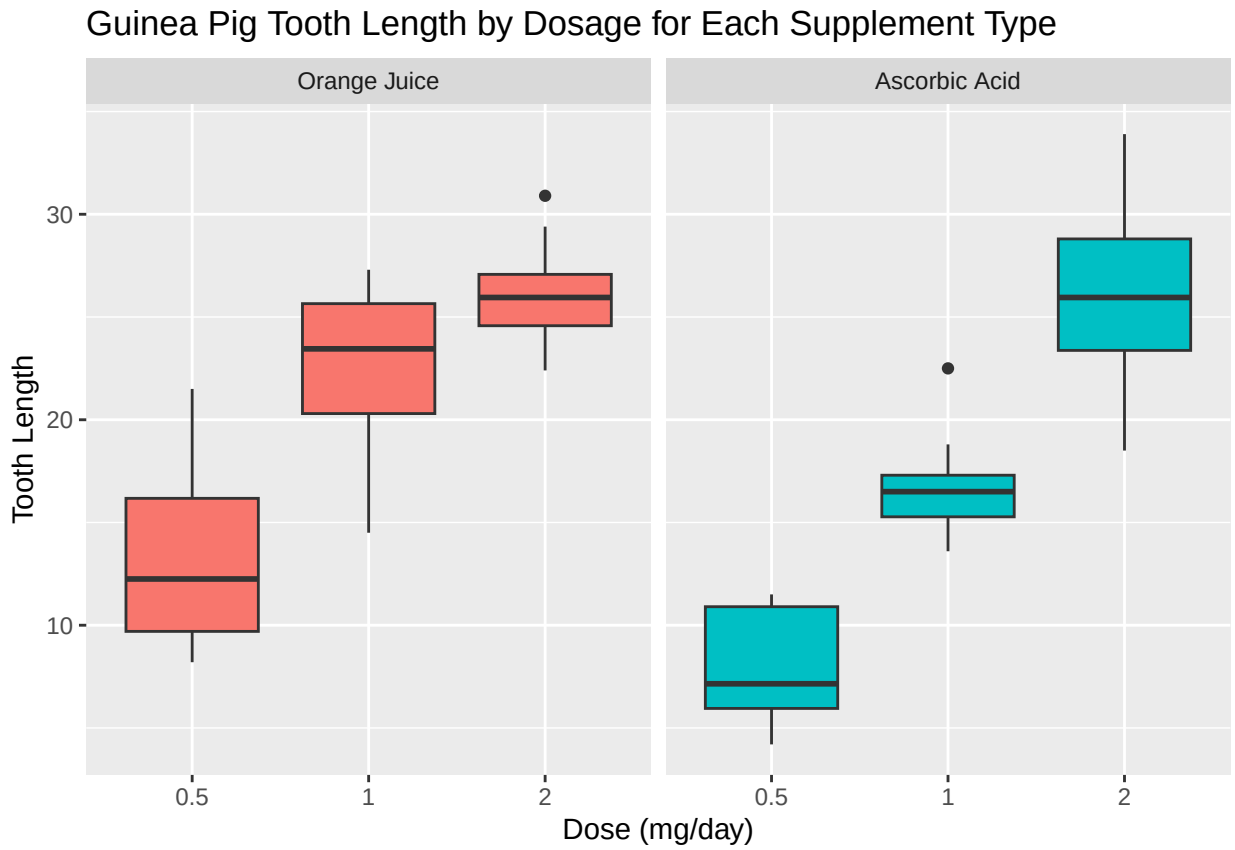
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```r
# Load ggplot2 for visualization
library(ggplot2)

# Rename supplement levels for clarity
t = ToothGrowth
levels(t$supp) <- c("Orange Juice", "Ascorbic Acid")

# Create boxplots to visualize tooth length by dosage and supplement type
ggplot(t, aes(x=factor(dose), y=len)) +
  facet_grid(.~supp) +  # Separate plots for each supplement
  geom_boxplot(aes(fill = supp), show_guide = FALSE) +  # Boxplot for tooth length
  labs(title="Guinea Pig Tooth Length by Dosage for Each Supplement Type",
       x="Dose (mg/day)",
       y="Tooth Length")  # Axis labels
```

```
## Warning: The 'show_guide' argument of 'layer()' is deprecated as of ggplot2 2.0.0.
## i Please use the 'show.legend' argument instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```




```r
### Basic Summary of the Data
# The boxplots indicate that increasing dosage correlates with increased tooth growth.
# Notably, orange juice appears more effective than ascorbic acid for dosages of 0.5 to 1.0 mg/day,
# while both supplements show similar effectiveness at 2.0 mg/day.
```

```r
### Confidence Intervals & Hypothesis Tests

#### Hypothesis #1
# Hypothesis: Orange juice and ascorbic acid provide similar tooth growth across the dataset.
hypoth1 <- t.test(len ~ supp, data = t)
hypoth1$conf.int  # Confidence interval
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```r
hypoth1$p.value   # P-value
```

```
## [1] 0.06063451
```

```r
# Interpretation
# The confidence interval includes 0, and the p-value exceeds 0.05.
# Thus, we cannot reject the null hypothesis.

#### Hypothesis #2
# Hypothesis: For a dosage of 0.5 mg/day, both supplements deliver similar tooth growth.
hypoth2 <- t.test(len ~ supp, data = subset(t, dose == 0.5))
hypoth2$conf.int  # Confidence interval
```

```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```r
hypoth2$p.value   # P-value
```

```
## [1] 0.006358607
```

```r
# Interpretation
# The confidence interval does not include 0, and the p-value is below 0.05.
# We can reject the null hypothesis, supporting the claim that orange juice provides
# greater tooth growth at this dosage.

#### Hypothesis #3
# Hypothesis: For a dosage of 1 mg/day, both supplements deliver similar tooth growth.
hypoth3 <- t.test(len ~ supp, data = subset(t, dose == 1))
hypoth3$conf.int  # Confidence interval
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```r
hypoth3$p.value   # P-value
```

```
## [1] 0.001038376
```

```r
# Interpretation
# The confidence interval does not include 0, and the p-value is less than 0.05.
# We can reject the null hypothesis, indicating that orange juice leads to
# greater tooth growth at this dosage as well.

#### Hypothesis #4
# Hypothesis: For a dosage of 2 mg/day, both supplements deliver similar tooth growth.
hypoth4 <- t.test(len ~ supp, data = subset(t, dose == 2))
hypoth4$conf.int  # Confidence interval
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```r
hypoth4$p.value   # P-value
```

```
## [1] 0.9638516
```

```r
# Interpretation
# The confidence interval includes 0, and the p-value is greater than 0.05.
# Thus, we cannot reject the null hypothesis.

### Conclusions & Assumptions
# Orange juice promotes more tooth growth than ascorbic acid for dosages 0.5 and 1.0 mg/day.
# However, for a dosage of 2.0 mg/day, both supplements are equally effective.
# Therefore, we cannot conclude that orange juice is more effective than ascorbic acid for the entire d

# Assumptions:
# - Normal distribution of tooth lengths
# - No other unmeasured factors affecting tooth length
```