Title Analysis of the ToothGrowth Data from R Datasets Package

Overview

This project explores the *ToothGrowth* data from the R *datasets* library by performing some preliminary analysis on the data and drawing some initial conclusions.

Exploratory Data Analysis

```
library(dplyr)      # Data manipulation (filter, mutate, group_by, etc.)
library(ggplot2)    # Plotting (qplot, ggplot etc.)
library(knitr)      # Dynamic Report Creation
library(datasets)   # R included sample data sets
data(ToothGrowth)
```

To setup the analysis we load the required libraries and the *ToothGrowth* data set. Per the R documentation for the *datasets* package [1], *ToothGrowth* is described as follows:

> The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

```
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 ...
```

```
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
```

```
table(ToothGrowth$supp)
```

```
##
## OJ VC
## 30 30
```

```
table(ToothGrowth$dose)
```

```
##
## 0.5   1   2
##  20  20  20
```

```
table(ToothGrowth %>% select(supp, dose))
```

```
##     dose
## supp 0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

The exploratory analysis of the *ToothGrowth* data frame demonstrates that the data consist of 30 observations each of *supp* OJ (Orange Juice) and VC (Ascorbic Acid); the 30 observations of each *supp* consist of 10 observations each of the three *dose* levels (0.5, 1 and 2).

Data Summary

```
tgBySuppDose <- ToothGrowth %>%
    group_by(Suppliment = supp, Dose = dose) %>%
    summarize(
        Length = mean(len)
        ,CI = 2 * sd(len) / sqrt(n())
    ) %>%
    arrange(Suppliment, Dose, desc(Length))

kable(tgBySuppDose
      ,format = "markdown"
      ,caption = "ToothGrowth by Suppliment and Dose"
)
```

Suppliment	Dose	Length	CI
OJ	0.5	13.23	2.820567
OJ	1.0	22.70	2.473504
OJ	2.0	26.06	1.679206
VC	0.5	7.98	1.737124
VC	1.0	16.77	1.590821
VC	2.0	26.14	3.034352

Create a data frame grouped by *Suppliment* and *Dose*. Derived summary data for each group include the mean length (*Length*) and the 95% confidence interval (*CI*). The confidence interval is derived by multiplying the standard deviation (σ) by two and dividing by the square root of the number of observations (n) as: $2\sigma/\sqrt{n}$.

Based on the summarized data we find that in general increasing the dosage of Vitamin C leads to an increase in the mean of Tooth Length via both delivery methods (orange juice and ascorbic acid). While the highest supplied dosage (2.0 mg) of each produce similar means (26.06 and 26.14, respectively) further analysis for variability and confidence intervals is required.

Confidence Intervals and Hypothesis Test

```
# plot1 <- ggplot(
#     ToothGrowth
#     ,aes(
#         x = Dose
#         ,y = Length
#         ,fill = Dose
#     )
# ) + geom_boxplot() +
#     facet_grid(.~Suppliment) +
#     labs(
#         title = "Analysis of ToothGrowth Data by Suppliment and Dose"
#         ,x = "Dose (mg)"
#         ,y = "Tooth Length"
#     )
# print(plot1)
#
# plot2 <- ggplot(
#     ToothGrowth
#     ,aes(
#         x = Dose
#         ,y = Length
#         ,color = Suppliment
#         ,group = Suppliment
#     )
# ) + geom_errorbar(
#     aes(
#         ymin = Length + se
#         ,ymax = Length + se
#     )
#     ,color = "black"
#     ,width=.1
#     #,position = pd
# )
#
# print(plot2)
```

Outliers OJ 2 and Ascorbic Acid 1

Conclusions

Reference

[1] https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html