Effect of vitamin C on tooth growth in guinea pigs

Overview

We apply basic data exploratory analysis techniques to the ToothGrowth data from the R datasets package. The length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs was observed after applying feeding methods. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, namely orange juice (OJ) or ascorbic acid (VC).

Basic exploratory analysis

```
library("ggplot2")
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data("ToothGrowth")

dim(ToothGrowth)
```

```
## [1] 60  3
```

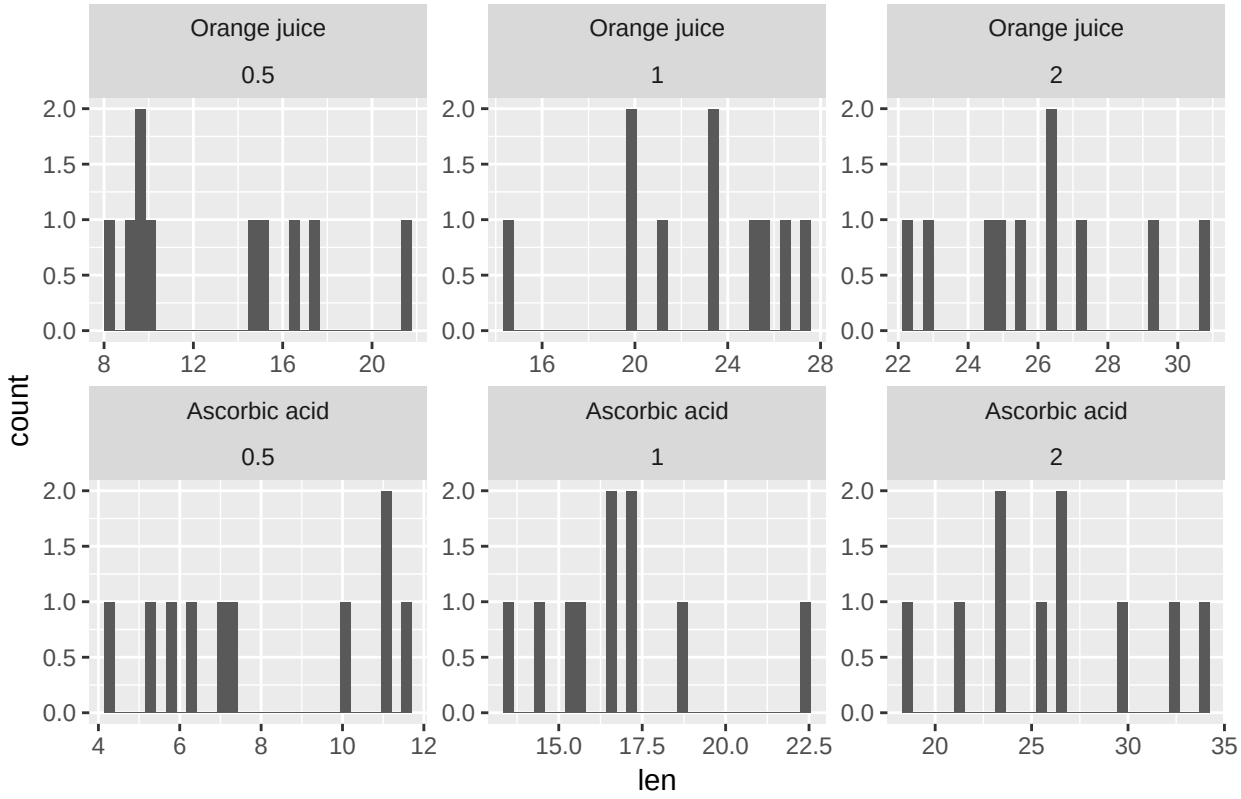
```
summary(ToothGrowth)
```

```
##       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
```

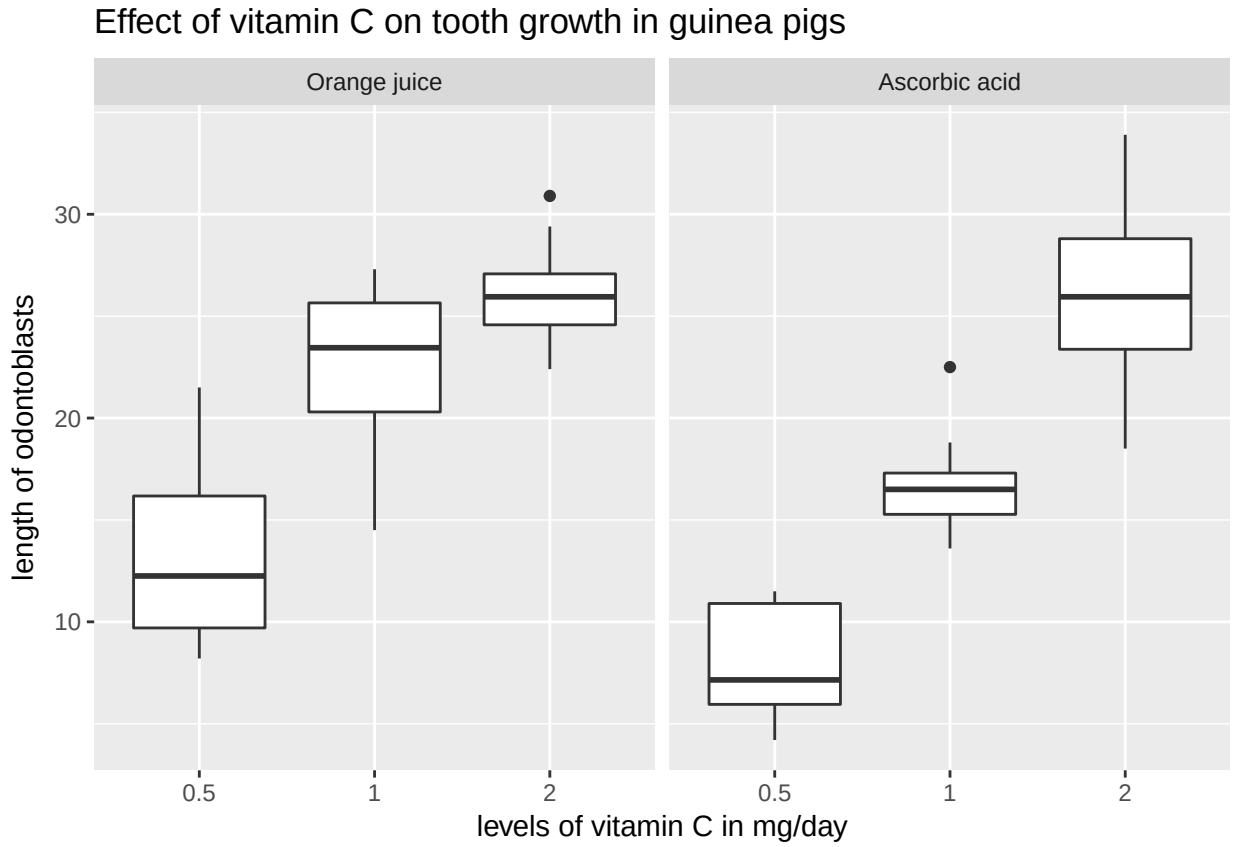
```
ToothGrowth$supp <- factor(ToothGrowth$supp, levels = c("OJ", "VC"), labels = c("Orange juice", "Ascorbi

ggplot(ToothGrowth, aes(x = len)) +
  geom_histogram() +
  facet_wrap(~ supp + dose, scales = "free") +
  labs(title = "Empirical distributions for each supplement and dose combination.")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Empirical distributions for each supplement and dose combination.




```
ggplot(ToothGrowth, aes(x = factor(dose), y = len)) +
  geom_boxplot() +
  facet_wrap(~supp) +
  labs(x = "levels of vitamin C in mg/day",
       y = "length of odontoblasts",
       title = "Effect of vitamin C on tooth growth in guinea pigs")
```




```
data_summary <- group_by(ToothGrowth, supp, dose) %>%
  summarise(len_mean = mean(len), sd_len = sd(len))

data_summary
```

```
## Source: local data frame [6 x 4]
## Groups: supp [?]
##
##            supp  dose len_mean   sd_len
##          <fctr> <dbl>    <dbl>    <dbl>
## 1  Orange juice   0.5    13.23 4.459709
## 2  Orange juice   1.0    22.70 3.910953
## 3  Orange juice   2.0    26.06 2.655058
## 4 Ascorbic acid   0.5     7.98 2.746634
## 5 Ascorbic acid   1.0    16.77 2.515309
## 6 Ascorbic acid   2.0    26.14 4.797731
```

Confidence intervals

We calculate confidence bounds for each supplement and dose combination resulting in a total of 6 combinations. The confifence bounds are calculated by resampling with replacement of the original data n times (default = 1000) due to the irregular distribution pattern of the data. We assume that the resampled distribution is representative of the true population distribution. The mean value is calculated for each bootstrap sample and the confidence bounds are selected from the ordered mean values. By default 95% confidence bonds are calculated.

```r
conf_resample <- function(data, n = 1000, prop = 0.95) {
  # Resample data with replacement n times and calculate the mean for each sample.
  means <- vector(mode = "numeric", length = n)
  for (i in seq_along(means)) {
    sam <- sample(data, replace = TRUE)
    means[i] <- mean(sam)
  }

  # Sort mean and select confidence bounds.
  means <- sort(means)
  min_id <- which.min(abs(seq(0, 1, length.out = n) - (1 - prop)/2))
  max_id <- which.min(abs(seq(0, 1, length.out = n) - ((1 - prop)/2 + prop)))

  c(means[min_id], mean(means), means[max_id])
}
```

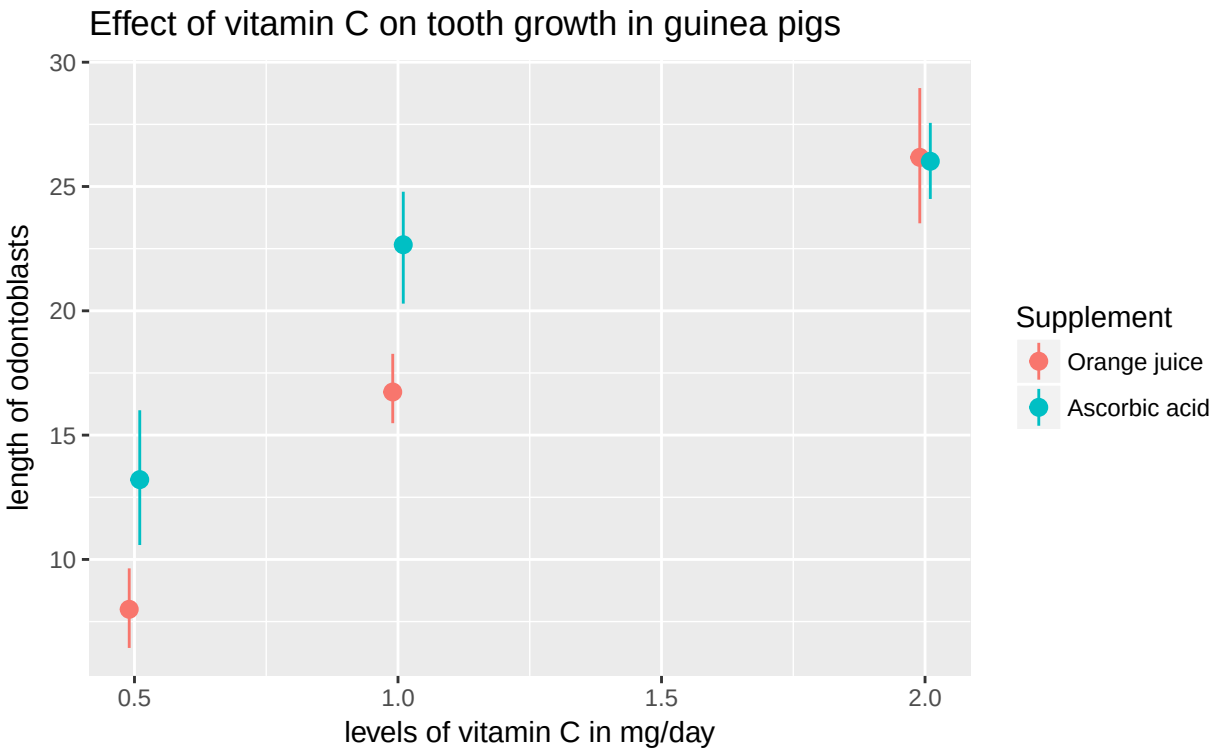
Let's calculate the confidence intervals for the different supplement and dose combinations. First we split the data into different data subsets, one for each combination of supplement and dose. Then we apply the the `conf_resample` function to each subsample and store the minimum and maximum confidence interval bonds in two seperate columns and add the mean value.

```r
ToothGrowth$groups <- with(ToothGrowth, paste(supp, dose))
data_summary$data <- split(ToothGrowth$len, ToothGrowth$groups)

data_summary$conf <- lapply(data_summary$data, conf_resample)

data_summary$min <- sapply(data_summary$conf, function(x) x[1])
data_summary$mean <- sapply(data_summary$conf, function(x) x[2])
data_summary$max <- sapply(data_summary$conf, function(x) x[3])
```

Conclusions

For dosages of 0.5 and 1 mg vitamin C per day we observed significant differences in tooth growth in guinea pigs. Guinea pigs feed with Ascorbic acid showed an increased tooth growth compared to orange juice. However, no differences in tooth growth were observed for guinea pigs feed with the highest dose of 2 mg vitamin C per day. Overall the tooth growth increased with increased doses of vitamin C. However guinea pigs fed with 1 mg vitamin C per day from ascorbic acid showed similar growth rates compared to guinea pigs feed with 2 mg vitamin C per day (either from ascorbic acid or orange juice).

```r
# Slightly shift the data on the x axis
data_summary$dose[data_summary$supp == "Orange juice"] <- data_summary$dose[data_summary$supp == "Orange
data_summary$dose[data_summary$supp != "Orange juice"] <- data_summary$dose[data_summary$supp != "Orange

ggplot(data_summary, aes(x = dose, y = mean, ymin = min, ymax = max, col = supp)) +
  geom_point() +
  geom_pointrange() +
  scale_color_discrete("Supplement") +
  labs(x = "levels of vitamin C in mg/day",
       y = "length of odontoblasts",
       caption = "Mean values and confidence intervals for each combination of dose and
supplement. Note that the actual vitamn C dosages of 0.5, 1 and 2 mg/day have
been slightly shifted to prevent overlapping.",
title = "Effect of vitamin C on tooth growth in guinea pigs")
```



Mean values and confidence intervals for each combination of dose and supplement. Note that the actual vitamn C dosages of 0.5, 1 and 2 mg/day have been slightly shifted to prevent overlapping.