Statistical inference project part 2: data analysis

Nick Plummer

27/07/2018

Overview

For the second half of the project we are going to analyze the `TootGrowth` dataset, from the `datasets` library, observations of tooth growth of Guinea pigs and the effect of dose of vitamin C on this growth.

The `ToothGrowth` dataset consists of 60 observations of 3 variables:

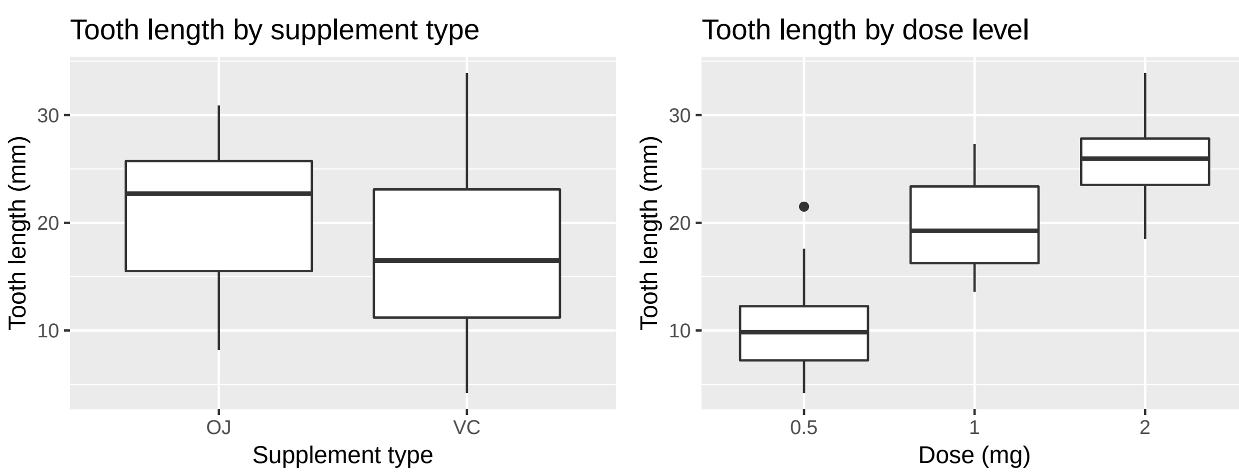
- `len`: Tooth length in millimeters (numeric variable)
- `supp`: Supplement type (factor variable with levels VC - ascorbic acid - and OJ - orange juice)
- `dose`: Dose in milligrams (numeric variable, either 0.5, 1, or 2 mcg/day)

We will treat `dose` as a factor variable in the subsequent analysis. This will leave six subgroups of 10 Guinea pigs recieving a specific dose via specific supplement:

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

Question 1 - Exploratory data analysis and summary

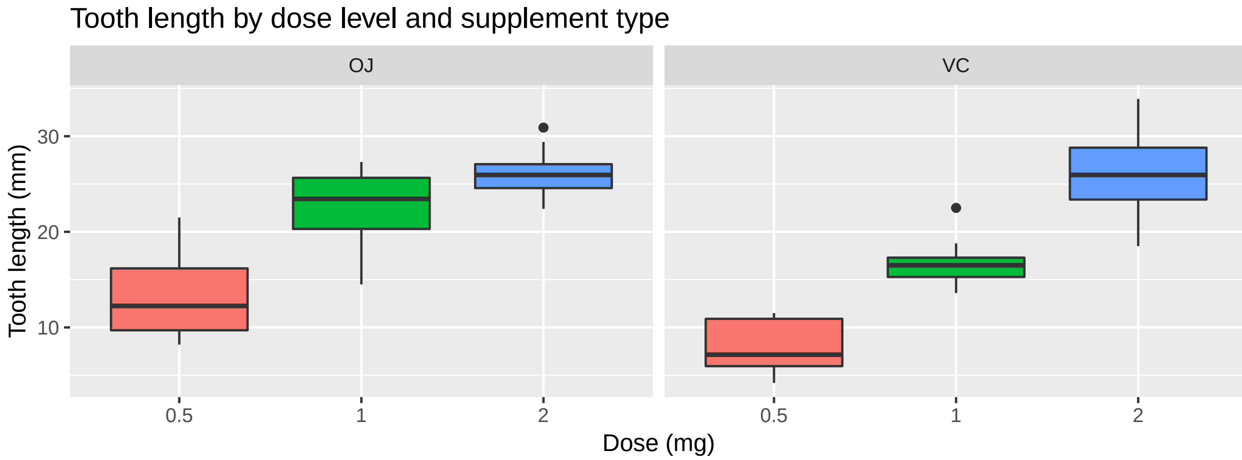
The average guinea pig tooth length is 18.813 with a standard deviation of 7.649. However we are more interested in exploring differences in tooth length across different combinations of supplement types and dosages.



The left plot demonstates that guinea pigs recieving orange juice have a greater median tooth length than those recieving ascorbic acid directly. The right plot shows a strong correlation between doseage and tooth length. This can be summaried in a table:

```
##   dose    q25 median    q75
## 1  0.5  7.225   9.85 12.250
## 2  1.0 16.250  19.25 23.375
## 3  2.0 23.525  25.95 27.825
```

These differences remain when comparing dosages within groups with the same mode of vitamin C delivery, although this appears to be minimal at the highest doses.



Question 2 - Data exploration using hypothesis testing

We will assume that the guinea pigs were randomly assigned to one of the groups (independence of observations) and that they were sampled from a nearly normal population (normality) allowing us to use the t distribution for hypothesis testing.

2A - Difference between supplement types

First, we will check if the observed difference in tooth length means between pigs who received their dose using orange juice and those who received their dose via ascorbic acid is statistically significant, via a t-test with unequal variances of the two samples:

```
tidy(t_diff_supp <- t.test(len ~ supp, ToothGrowth, var.equal = FALSE))
```

```
## # A tibble: 1 x 10
##   estimate estimate1 estimate2 statistic p.value parameter conf.low
##      <dbl>     <dbl>     <dbl>     <dbl>   <dbl>     <dbl>    <dbl>
## 1     3.70      20.7      17.0      1.92  0.0606      55.3   -0.171
## # ... with 3 more variables: conf.high <dbl>, method <chr>,
## #   alternative <chr>
```

p=0.061 (>0.05) so we fail to reject the null hypothesis, i.e. there is no significant difference in tooth length across the two supplement types (95% confidence intervals -0.171 to 7.571)

2B - Differences in dose levels

Since we are dealing with three different dose levels we need to apply three pairwise comparison t-tests to cover all factor level combinations.

The table below summarizes the results of the three 3 pairwise comparisons. `comparison` shows which subsets were used for the respective t-test, `statistic` the mean difference with 95% confidence intervals `conf.low` and `conf.high`, and the p value `p.value`.

```
##      comparison statistic  conf.low conf.high      p.value
## 1   µ2mg - µ1mg  4.900484  3.733519  8.996481 1.906430e-05
## 2 µ2mg - µ0.5mg 11.799046 12.833833 18.156167 4.397525e-14
```

```
## 3 µ1mg - µ0.5mg  6.476648  6.276219 11.983781 1.268301e-07
```

In all three cases the p-value is <0.05, meaning there is a significant difference in the effect the dose of vitamin C has on the average tooth length of guinea pigs.

Conclusions

We can raise two conclusions from this simplistic analysis: - Dose of vitamin C supplementation is strongly linked with tooth length - Form of supplementation (ascorbic acid vs orange use) is not significantly linked with tooth length

This is caviat to a number of assumptions: - The experiment was done with random assignment of guinea pigs to different dose level categories and supplement type (control for confounders) - Members of the sample population, i.e. the 60 guinea pigs, are representative of the entire population of guinea pigs (generalisability) - Assumptions of normality and independence for the use of t-tests, with variances assumed to be different for the two groups being compared

Further investigation into the interaction between dose and mode of delivery of vitamin C is likely warrented.

Appendix (Full source code)

This section includes the source code which was used for generating the results of the analysis.

```r
# Set up
knitr::opts_chunk$set(echo = TRUE)
library(datasets)
library(ggplot2)
library(dplyr)
library(broom)
library(grid)


# Subgroups
with(ToothGrowth, table(dose, supp))


# Basic descriptives
length_mean <- mean(ToothGrowth$len)
length_sd <- sd(ToothGrowth$len)


# Plot length vs type and length vs dose
g_len_by_supp <- ggplot(ToothGrowth, aes(supp, len)) +
  geom_boxplot() +
  xlab('Supplement type') +
  ylab('Tooth length (mm)') +
  ggtitle('Tooth length by supplement type')

g_len_by_dose <- ggplot(ToothGrowth, aes(as.factor(dose), len)) +
  geom_boxplot() +
  xlab('Dose (mg)') +
  ylab('Tooth length (mm)') +
  ggtitle('Tooth length by dose level')
```

```r
grid.newpage()
pushViewport(viewport(layout = grid.layout(1, 2, widths = c(0.5, 0.5))))
print(g_len_by_supp, vp = viewport(layout.pos.row = 1, layout.pos.col = 1))
print(g_len_by_dose, vp = viewport(layout.pos.row = 1, layout.pos.col = 2))


# Plot length vs dose by type
ToothGrowth %>% group_by(dose) %>%
  summarize(
    q25 = quantile(len, 0.25),
    median = quantile(len, 0.5),
    q75 = quantile(len, 0.75)
  ) %>%
  as.data.frame

ggplot(ToothGrowth, aes(as.factor(dose), len)) +
  geom_boxplot(aes(fill = as.factor(dose))) +
  facet_grid(. ~ supp) +
  xlab('Dose (mg)') +
  ylab('Tooth length (mm)') +
  ggtitle('Tooth length by dose level and supplement type') +
  theme(legend.position = "none")


# Effect of supplement type
tidy(t_diff_supp <- t.test(len ~ supp, ToothGrowth, var.equal = FALSE))


# Multiple t-test
pairwise_results <- t.test(ToothGrowth$len[ToothGrowth$dose == 2],
       ToothGrowth$len[ToothGrowth$dose == 1]) %>%
  tidy %>%
  mutate(
    comparison = "µ2mg - µ1mg"
  ) %>%
  select(11,4,7,8,5)

pairwise_results <- t.test(ToothGrowth$len[ToothGrowth$dose == 2],
       ToothGrowth$len[ToothGrowth$dose == 0.5]) %>%
  tidy %>%
  mutate(
    comparison = 'µ2mg - µ0.5mg'
  ) %>%
  select(11,4,7,8,5) %>%
  bind_rows(pairwise_results, .)

pairwise_results <- t.test(ToothGrowth$len[ToothGrowth$dose == 1],
       ToothGrowth$len[ToothGrowth$dose == 0.5]) %>%
  tidy %>%
  mutate(
    comparison = 'µ1mg - µ0.5mg'
  ) %>%
  select(11,4,7,8,5) %>%
```

```
  bind_rows(pairwise_results, .)
```