Inferential data analysis of tooth growth data

Izaak Jephson

20/01/2020

Overview

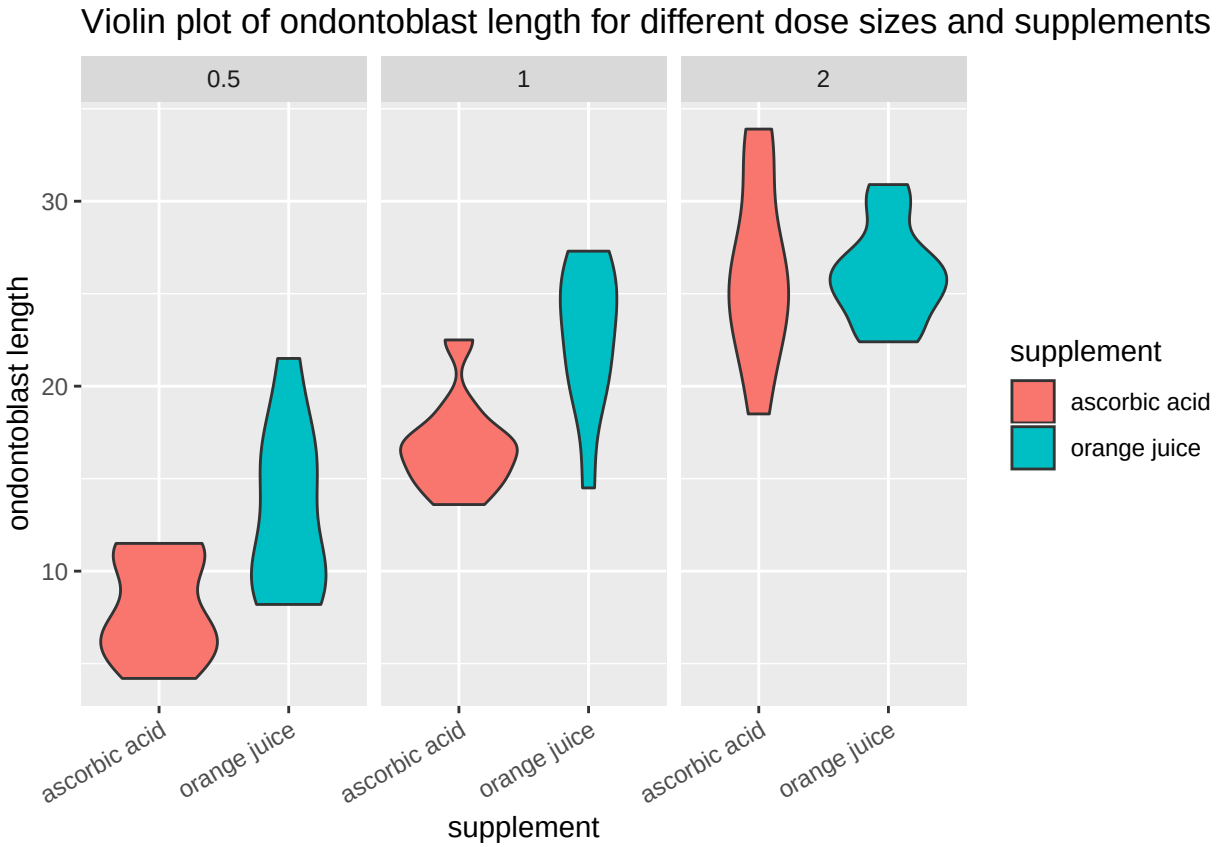
This analysis looks at the length of odontoblasts in 60 guinea pigs, given different doses of of vitamin C by one of two delivery methods: orange juice or ascorbic acid. The data indicate that generally, odontoblast length increases with increased dose for both delivery methods and that odontblast length is greater in the guinea pigs given orange juice than in those given ascorbic acid, although this effect is only present at low doses.

Exploratory data analysis

```
# Create a violin plot of ondontoblast length for diffent dose sizes and supplements

data <- ToothGrowth %>% mutate(supp = str_replace(supp, "OJ","orange juice"))
data <- data %>% mutate(supp = str_replace(supp, "VC","ascorbic acid"))

g <- ggplot(data, mapping = aes(supp,len, fill = supp)) + geom_violin() +
        facet_grid(cols=vars(dose)) +
        labs(title = "Violin plot of ondontoblast length for different dose sizes and supplements", x =
        scale_fill_discrete(name = "supplement") +
        theme(axis.text.x = element_text(angle = 30, hjust = 1))
g
```



From the plot above, the most obvious relationship is that for both supplements, ondontoblast length increases with dose. Comparing the relative effect of the two supplements, it appears that at low doses (0.5 mg/day), the orange juice supplements are associated with higher ondontoblast length than the ascorbic acid. However this effect appears to be lessened in medium doses (1 mg/day) and it is unclear if there is any difference at higher doses (2 mg/day).

Confidence intervals and hypothesis testing

Assuming that the ondontoblast length is a normal random variable. We can find a confidence interval for the mean in each case tested (i.e. for all three dose levels and both supplements, 6 cases in total). We can then compare these to determine whether differences in ondontoblast length are statistically significant or merely the result of random chance. In each case, the distribution of sample means was modelled as a t distribution, owing to the small sample size for each case.

```
# Calculate t confidence intervals for mean of each case

groups <- data %>% group_by(supp,dose) %>% summarise(mean = mean(len))

lower <- data %>%
                group_by(supp,dose) %>%
                        summarise(lower = t.test(len)$conf.int[1])

upper <- data %>%
                group_by(supp,dose) %>%
                        summarise(upper = t.test(len)$conf.int[2])

groups <- groups %>%
```

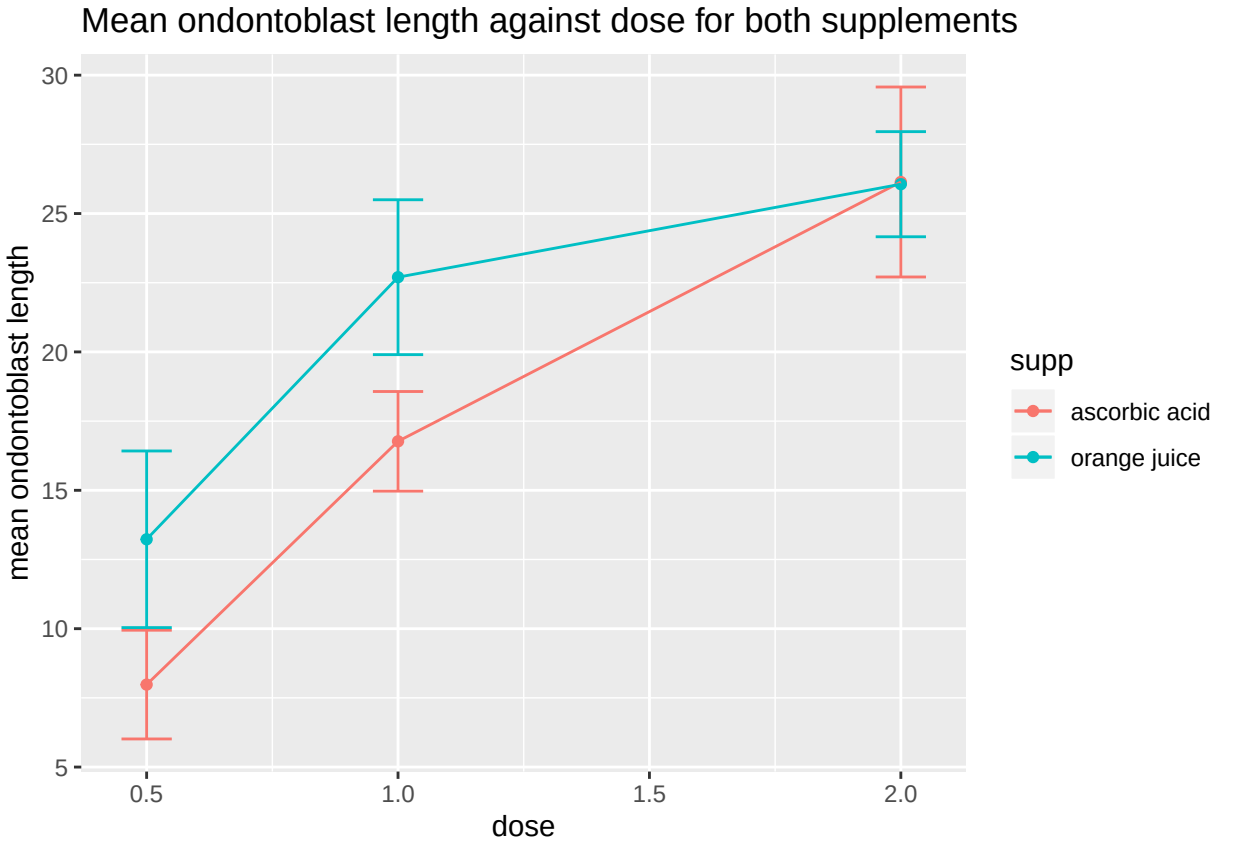
```
                left_join(lower) %>%
                left_join(upper)
```

```
## Joining, by = c("supp", "dose")
## Joining, by = c("supp", "dose")
```

The means and associated confidence intervals for each case were then plotted.

```
# Plot means and confidence intervals

g <- ggplot(groups, mapping = aes(x = dose, y = mean, colour = supp)) +
        geom_errorbar(mapping = aes(ymin = lower, ymax = upper), width = 0.1) +
        geom_point() +
        geom_line() +
        labs(title = "Mean ondontoblast length against dose for both supplements", y = "mean ondontoblas
        scale_fill_discrete(name = "supplement", labels = c("orange juice", "ascorbic acid"))
g
```



Conclusions

As we saw in the previous plot, the line grpah above shows that mean ondontoblast length does indeed increase with dose for both supplements. The error bars on the plot represent 95% confidence intervals. For each test consider H_0 to be the null hypothesis, that there is no difference between the two groups being compared. From this plot (and the associated confidence intervals), we can conclude that, at the 95% confidence level, orange juice supplements are associated with longer ondontoblast lengths than ascorbic acid supplements for doses of 0.5 mg/day and 1 mg/day (i.e. for these two tests we can reject H_0) as the confidence intervals do not overlap. However, when comparing doses of 2 mg/day of orange juice and ascorbic

acid, we cannot reject H_0 as the confidence intervals for the sample means overlap. Thus, we have no evidence that doses of 2 mg/day of orange juice are associated with greater ondontoblast growth than ascorbic acid supplements of the same dose.

The confidence intervals for the mean ondontoblast length for different doses of ascorbic acid are all distinct and do not overlap. Therefore we can reject H_0 at the 95% confidence level for each of these comparisons. The same can be said when comparing the results associated with doses of orange juice at 0.5 mg/day and 1 mg/day. However, the confidence intervals for mean ondontoblast length of guinea pigs given orange juice supplements of 1 and 2 mg/day overlap. Therefore, we cannot declare them distinct at the 95% confidence level.

All of these conclusions assume a t distribution for the samples, and by extension, a normal underlying distribution for the population.