Statistical Inference Course Project Part 2: Tooth Growth Analysis

Hedley Stirrat

17 May 2020

Overview

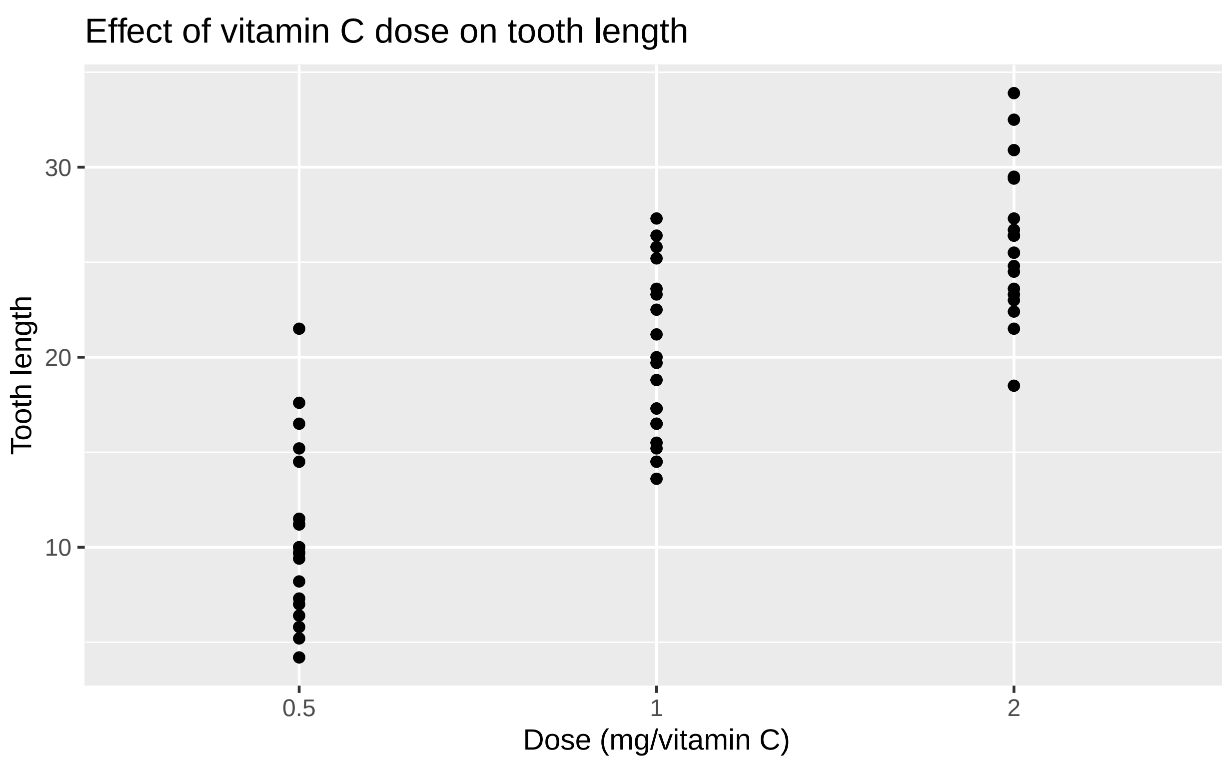
The ToothGrowth dataset in the R datasets package contains the results of an experiment that explored the effect of vitamin C on tooth growth in guinea pigs. Each guinea pig received one of three dose levels of vitamin C by one of two delivery methods.

The ToothGrowth dataframe consists of 60 observations of 3 variables. The variables are `len` (tooth length), `supp` (supplement type, VC or OJ for ascorbic acid and orange juice, respectively), and `dose` (dose in mg/day).

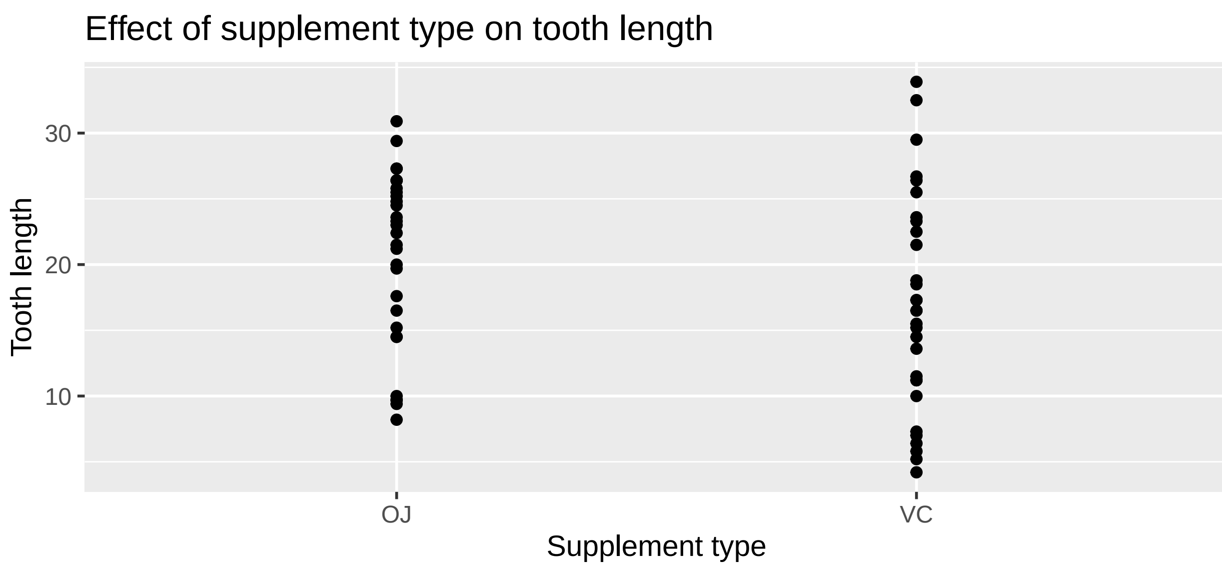
We will conduct an initial exploratory data analysis to generate hypotheses about the data, and then test those hypotheses in the final section.

Initial exploratory analysis

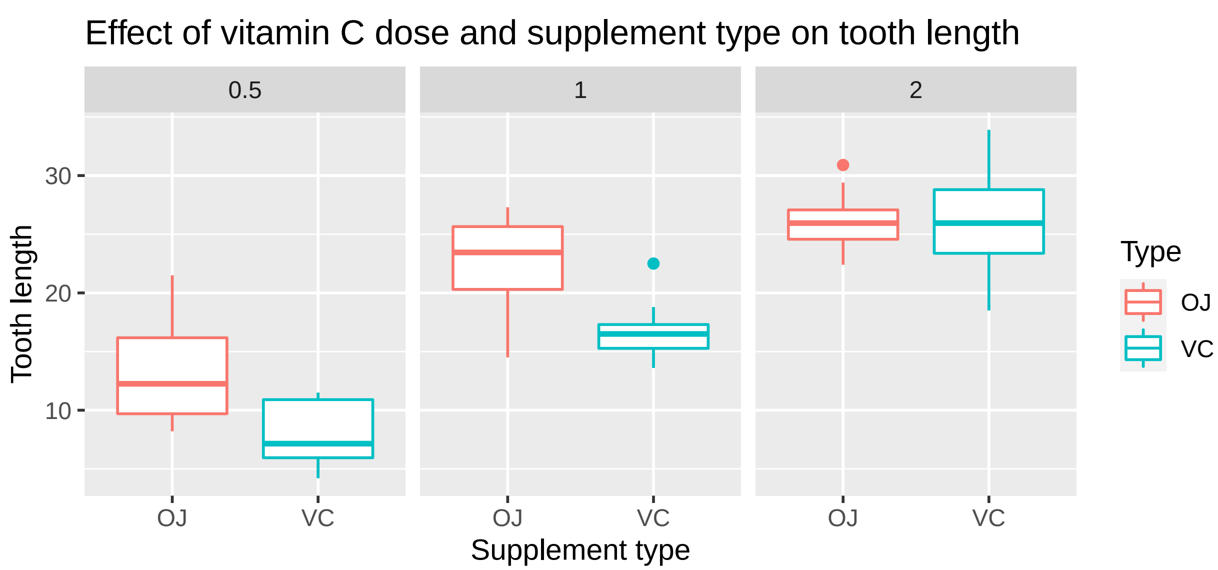
Simply plotting dose vs length suggests that higher vitamin C doses may lead to longer tooth lengths:



It's not clear that the supplement type affects tooth length in an obvious way:



We can look at these two variables (`supp` and `dose`) and their combined effect on tooth length. Splitting up the data this way appears to suggest that at lower dosages (0.5 mg and 1.0 mg), the OJ supplement type leads to better tooth growth than VC, while at the highest dosage (2.0 mg), supplement type does not have an effect on tooth growth.



Hypothesis testing

Assumptions

- We have left the default `t.test` parameter of `paired = FALSE` as the data do not appear to be paired
- We have left the default `t.test` parameter of `conf.level = 0.95` to test for significance at the 95% confidence level
- We have left the default `t.test` parameter of `alternative = "two.sided"` to conduct a two-sided test (although one-sided may be more appropriate here)
- We have left the default `t.test` parameter of `var.equal = FALSE` with the assumption that the variances between groups are not equal

- We also make the assumption that the variables are independent and identically distributed

Dosage vs tooth length

The first plot above suggested that higher vitamin C doses may be associated with longer tooth lengths. Let's test this hypothesis by running t-tests on unique dosage combinations (0.5 mg and 1.0 mg; 0.5 mg and 2.0 mg; 1.0 mg and 2.0 mg).

```
##                 p.value lower.confidence upper.confidence
## 0.5-1.0mg 1.268301e-07           -11.98            -6.28
## 0.5-2.0mg 4.397525e-14           -18.16           -12.83
## 1.0-2.0mg 1.906430e-05            -9.00            -3.73
```

In all three comparisons, we have identified that higher doses are associated with longer tooth lengths, and these differences are significant at the 95% confidence level. Therefore, we reject the null hypothesis that vitamin C dosage levels are not associated with a difference in tooth lengths.

Supplement type vs tooth length

As described above, it appears as though the supplement type may have an effect on tooth length at low dosages of vitamic C, but not at the higher 2.0 mg dose.

Before we test this hypothesis, let's first test whether the different supplement types are associated with different tooth lengths regardless of dosage level:

```r
generate_summary(t.test(len ~ supp, data = ToothGrowth))
```

```
##      p.value lower.confidence upper.confidence
## 1 0.06063451            -0.17             7.57
```

With a p-value of 0.06 and our 95% confidence interval containing zero, we cannot reject the null hypothesis that the two different supplement types do not lead to different tooth lengths when we look at data containing all dosage levels.

However, let's see whether we can identify differences in tooth lengths caused by the different supplement types if we control for dosage levels:

```
##           p.value lower.confidence upper.confidence
## 0.5mg 0.006358607             1.72             8.78
## 1.0mg 0.001038376             2.80             9.06
## 2.0mg 0.963851589            -3.80             3.64
```

The t-tests support our initial hypotheses; we've shown that at dosages of 0.5 or 1.0 mg, the OJ supplement type is associated with a longer tooth length with $p < 0.01$ at the 95% confidence level. Additionally, we found no significant difference between the supplement types at the 2 mg dosage level, as the 95% confidence interval included zero with a p-value of 0.96.

Conclusion

We've shown that at low vitamin C dosages of 0.5 and 1.0 mg, the OC supplement method was associated with longer tooth lengths, while at 2.0 mg dosage, there was no significant difference in tooth lengths between the supplement types

We've also shown that, regardless of supplement type, higher vitamin C dosages are associated with longer tooth lengths.

Appendix

Code used to set up the analysis:

```
library(ggplot2)
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
```

Code used in the exploratory data analysis:

```
ggplot(ToothGrowth, aes(x = as.factor(dose), y = len)) +
    geom_point() +
    ggtitle("Effect of vitamin C dose on tooth length") +
    labs(y = "Tooth length", x = "Dose (mg/vitamin C)")
```

```
ggplot(ToothGrowth, aes(x = supp, y = len)) +
    geom_point() +
    ggtitle("Effect of supplement type on tooth length") +
    labs(y = "Tooth length", x = "Supplement type")
```

```
ggplot(ToothGrowth, aes(x = supp, y = len, colour = supp)) +
    geom_boxplot() +
    facet_wrap(~ dose) +
    ggtitle("Effect of vitamin C dose and supplement type on tooth length") +
    labs(y = "Tooth length", x = "Supplement type", colour = "Type")
```

Code used in the hypothesis testing:

```
dose_vs_len <- list()
dose_vs_len[['0.5-1.0']] <- t.test(len ~ dose, data = ToothGrowth[ToothGrowth$dose %in% c(0.5,1.0),])
dose_vs_len[['0.5-2.0']] <- t.test(len ~ dose, data = ToothGrowth[ToothGrowth$dose %in% c(0.5,2.0),])
dose_vs_len[['1.0-2.0']] <- t.test(len ~ dose, data = ToothGrowth[ToothGrowth$dose %in% c(1.0,2.0),])

generate_summary <- function(result_list) {

    if (length(result_list) == 10) {
        `p-value` <- result_list[['p.value']]
        `lower confidence` <- round(result_list[['conf.int']][[1]], 2)
        `upper confidence` <- round(result_list[['conf.int']][[2]], 2)
    } else {
        `p-value` <- unlist(lapply(result_list, function(x) x[['p.value']]))
        `lower confidence` <- round(unlist(lapply(result_list, function(x) x[['conf.int']][[1]])),2)
        `upper confidence` <- round(unlist(lapply(result_list, function(x) x[['conf.int']][[2]])),2)
    }
    data.frame(`p-value`, `lower confidence`, `upper confidence`)
```

```
}

generate_summary(dose_vs_len)
```

```
##              p.value lower.confidence upper.confidence
## 0.5-1.0 1.268301e-07           -11.98            -6.28
## 0.5-2.0 4.397525e-14           -18.16           -12.83
## 1.0-2.0 1.906430e-05            -9.00            -3.73
```

```
generate_summary(t.test(len ~ supp, data = ToothGrowth))
```

```
##      p.value lower.confidence upper.confidence
## 1 0.06063451            -0.17             7.57
```

```
supp_vs_len <- list()
supp_vs_len[['0.5mg']] <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose %in% c(0.5),])
supp_vs_len[['1.0mg']] <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose %in% c(1.0),])
supp_vs_len[['2.0mg']] <- t.test(len ~ supp, data = ToothGrowth[ToothGrowth$dose %in% c(2.0),])

generate_summary(supp_vs_len)
```

```
##           p.value lower.confidence upper.confidence
## 0.5mg 0.006358607             1.72             8.78
## 1.0mg 0.001038376             2.80             9.06
## 2.0mg 0.963851589            -3.80             3.64
```