The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Aseem Rohatgi

The Data

The dataset we'll be using comes pre-built in R. It's stored as a data frame under the name `ToothGrowth`. The documentation for this dataset tells us that it is a collection of three variables-one response variable stating the length of odontoblasts in 60 guinea pigs, and two explanatory variables-one representing the method of delivery of vitamin C, and the other representing the dosage. First we look at its structure.

```
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 ...
```

We see that the variable `dose`, which represent the three levels of dosage is stored as a numeric variable. We should keep this in mind if we are to use it as a factor later on in our analysis. We should also check how the sample sizes of these groups differ.

```
ToothGrowth %>% count(supp, dose)
```

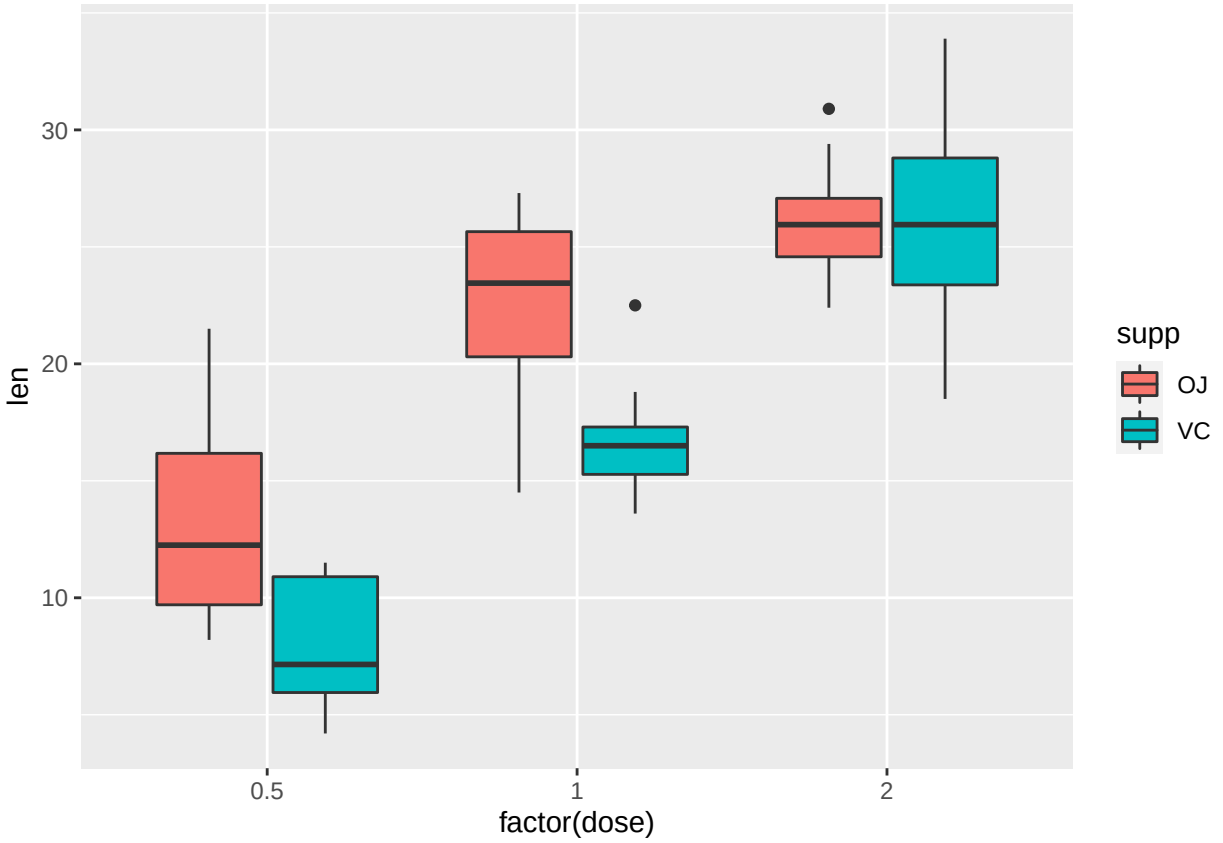
```
##   supp dose  n
## 1   OJ  0.5 10
## 2   OJ  1.0 10
## 3   OJ  2.0 10
## 4   VC  0.5 10
## 5   VC  1.0 10
## 6   VC  2.0 10
```

The sample sizes are all the same, albeit small. Let's now look at some basic summary statistics for this data set, with and without grouping.

```
## # A tibble: 6 x 7
##    dose supp  quartile1 median quartile2   iqr  mean
##   <dbl> <fct>     <dbl>  <dbl>     <dbl> <dbl> <dbl>
## 1   0.5 OJ         9.7   12.2       16.2  6.48 13.2
## 2   0.5 VC         5.95   7.15      10.9  4.95  7.98
## 3   1   OJ        20.3   23.5       25.6  5.35 22.7
## 4   1   VC        15.3   16.5       17.3  2.03 16.8
## 5   2   OJ        24.6   26.0       27.1  2.5  26.1
## 6   2   VC        23.4   26.0       28.8  5.43 26.1
```

```
##   quartile1 median quartile2  iqr     mean
## 1    13.075  19.25    25.275 12.2 18.81333
```

Plotting boxplots for the different dosages in the two delivery methods would allow us to understand this data better.



Is there a difference in tooth growth among the two delivery methods?

We will consider each dosage level separately and see if there is a difference between the results of the two delivery methods using an unpaired t-test. We use an unpaired t-test since the documentation states that "Each animal received one of three dose levels of vitamin C by one of two delivery methods". So the data are not paired, and each observation comes from a different guinea pig, the numbers of which are quite small in our sample.

Dosage of 0.5 mg/day

Hypotheses

Our null hypothesis is that there is no difference between the mean growth in length of odontoblasts in guinea pigs by the two delivery methods with a dosage of 0.5 mg/day. Our alternative hypothesis states that such a difference exists.

Test

A simple `t.test` command will give us required results.

```
##
##  Welch Two Sample t-test
##
## data:  OJ_0.5 and VC_0.5
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

The p-value is 0.0063586. Since this is much less than 0.05, we reject our null hypothesis at 95% significance.

Conclusion

We can conclude that orange juice seems to be better than ascorbic acid in inducing growth of odontoblasts in guinea pigs when given a dosage of 0.5 mg/day.

Dosage of 1 mg/day

Hypotheses

Our null hypothesis is that there is no difference between the mean growth in length of odontoblasts in guinea pigs by the two delivery methods with a dosage of 1 mg/day. Our alternative hypothesis states that such a difference exists.

Test

A simple `t.test` command will give us required results.

```
##
##  Welch Two Sample t-test
##
## data:  OJ_1 and VC_1
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

The p-value is 0.0010384. Since this is much less than 0.05, we reject our null hypothesis at 95% significance.

Conclusion

We can conclude that orange juice seems to be better than ascorbic acid in inducing growth of odontoblasts in guinea pigs when given a dosage of 1 mg/day.

Dosage of 2 mg/day

Hypotheses

Our null hypothesis is that there is no difference between the mean growth in length of odontoblasts in guinea pigs by the two delivery methods with a dosage of 2 mg/day. Our alternative hypothesis states that such a difference exists.

Test

A simple `t.test` command will give us required results.

```
##
##  Welch Two Sample t-test
##
## data:  OJ_2 and VC_2
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

The p-value is 0.9638516. Since this is much greater than 0.05, we do not sufficient evidence to reject our null hypothesis at 95% significance.

Conclusion

We can conclude that orange juice seems to be no better than ascorbic acid in inducing growth of odontoblasts in guinea pigs when given a dosage of 2 mg/day.

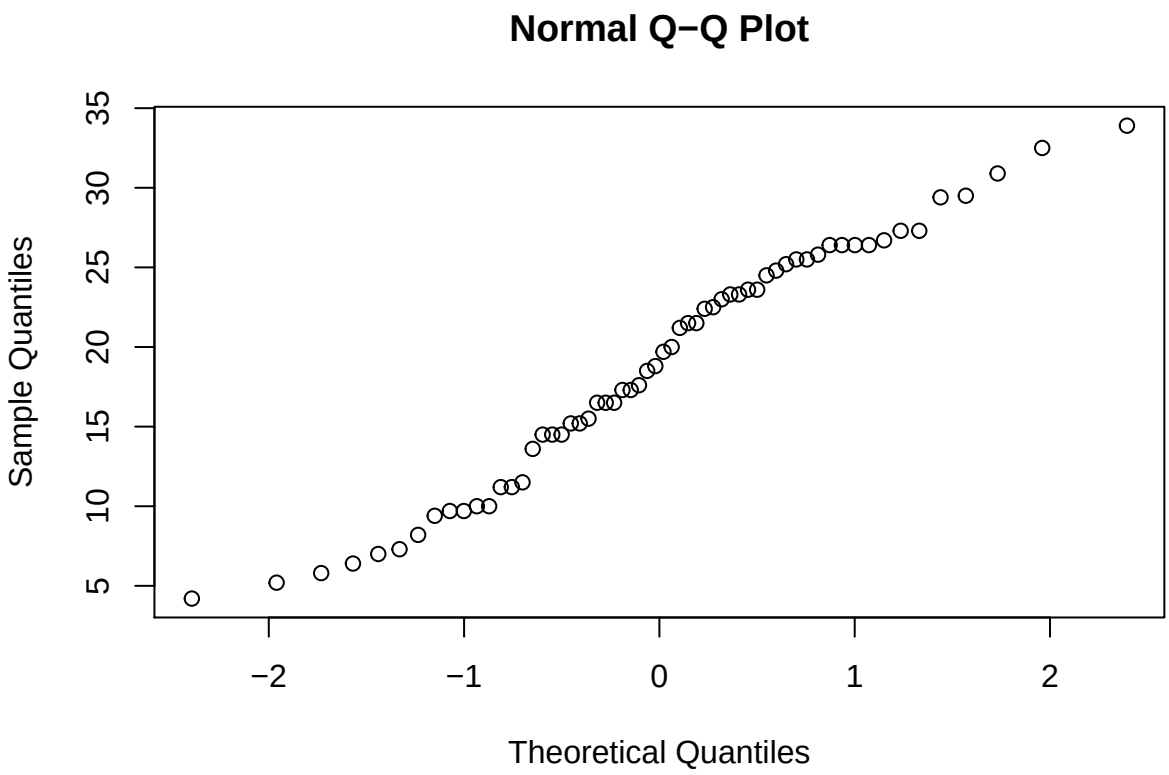
Conclusion

The tests indicate that at lower dosages, orange juice seems to be acting better than ascorbic acid in increasing the length of odontoblasts. However, this difference in their impact on mean growth of odontoblasts diminishes as dosage increases to 2 mg/day.

Also, notice that the confidence interval for the difference in mean growth under a dose of 0.5 mg/day is wider and "lower in value", compared to that under a dose of 1 mg/day. But increasing the dose to 2 mg/day results in both delivery methods becoming equivalent. This suggests that there must an optimal dosage between 1 and 2 mg/day where the difference in mean growth is highest, and we might want to investigate this further.

Assumptions

We assume that our response variable, `len`, is Gaussian. This is a reasonable assumption to make as can be seen from the normal Q-Q plot below.



The observation can be assumed to be independent as well, as long as the guinea pigs used for the study were genetically diverse. Since we have no reason to believe that the guinea pigs belonged to the same family, we can be safe with our assumption of independence and hence, of a simple random sample.