

Effects of delivery method of vitamic C on the length of odontoblasts in guinea pigs

Nathalie Descusse-Brown

January 31, 2019

Overview

The work presented here investigates the characteristics of the dataset ‘Toothgrowth’ in the R datasets package. The dataset corresponds to the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice (“OJ”) or ascorbic acid (a form of vitamin C and coded as VC). We sought with this work to identify some statistically significant conclusions from the dataset, namely whether the vitamin C delivery method of either orange juice or ascorbic acid is associated with a larger odontoblasts length.

Data Processing

First we start by downloading the ToothGrowth dataset from the R dataset package.

```
library("datasets")
toothgrowth <- ToothGrowth
names(toothgrowth) <- c("toothlength", "supplementtype", "doseinmgperday")
head(toothgrowth)
```

```
##   toothlength supplementtype doseinmgperday
## 1         4.2             VC           0.5
## 2        11.5             VC           0.5
## 3         7.3             VC           0.5
## 4         5.8             VC           0.5
## 5         6.4             VC           0.5
## 6        10.0             VC           0.5
```

```
summary(toothgrowth)
```

```
##   toothlength   supplementtype doseinmgperday
## 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
```

```
dim(toothgrowth)
```

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

We can see from the above dataset information that there is one data point per guinea pig. We want to investigate whether a specific vitamic C delivery method (either via orange juice or ascorbic acid) is associated with a larger odontoblasts length. We start by performing some exploratory analysis to identify some potential trends.

```
library(reshape2)
library(ggplot2)
```

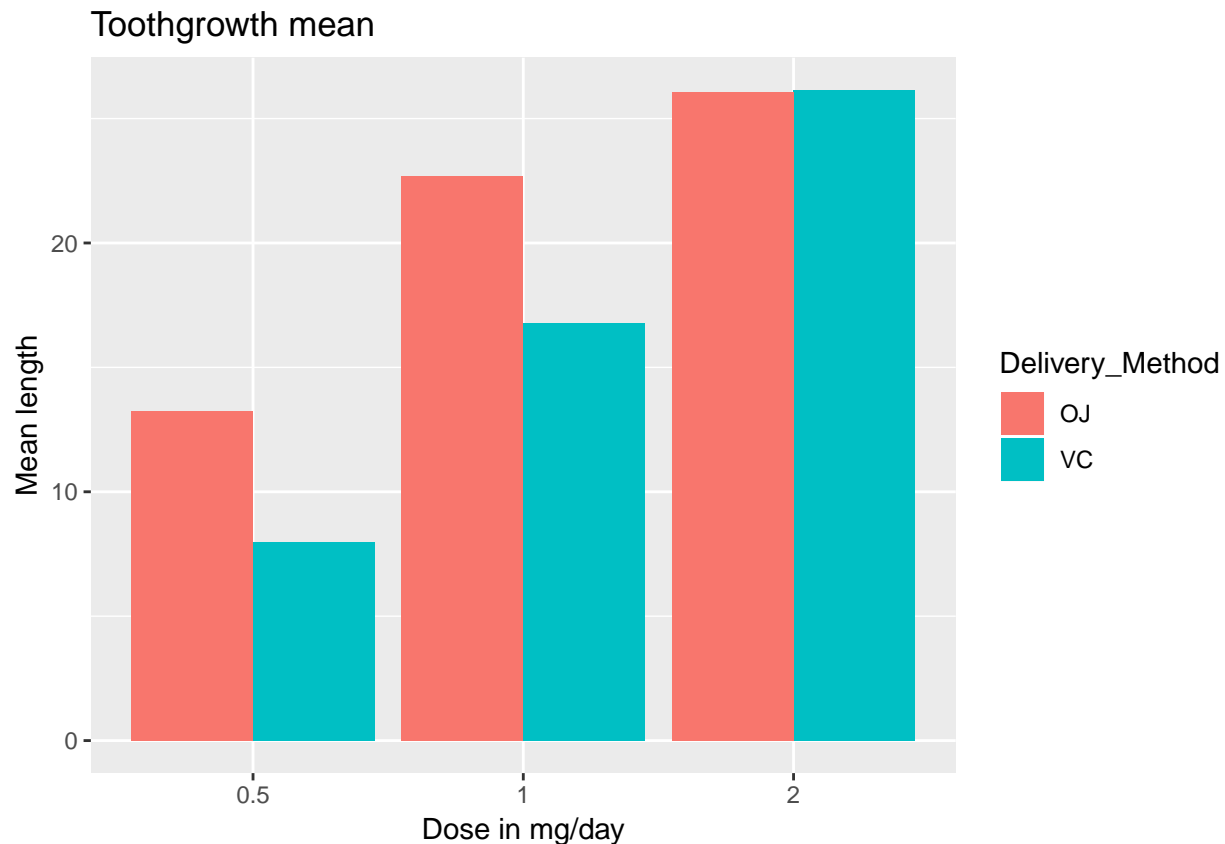
```

meanlength <- tapply(toothgrowth$toothlength,list(toothgrowth$supplementtype,toothgrowth$doseinmgperday),
meanlength2 <- melt(meanlength,id.vars = "dose")

names(meanlength2)[1] <- "Delivery_Method"

ggplot(meanlength2,arr.ind=TRUE, aes(fill=Delivery_Method, y=value, x=as.character(Var2))) +
  geom_bar( stat="identity", position=position_dodge()) + ggtitle("Toothgrowth mean") +
  xlab("Dose in mg/day") + ylab("Mean length")

```



From the above plot we make the null hypothesis that for any of the three doses administered (0.5mg/day, 1mg/day and 2mg/day) there is no difference in odontoblasts length between the ascorbic acid (VC) or orange juice (OJ) so that the difference of the mean between both data sets (orange juice - ascorbic acid) is zero. We want to test this hypothesis versus an alternative hypothesis, which is that the difference is greater than zero, with statistical significance, which would indicate that orange juice is associated with a greater odontoblasts length than for ascorbic acid.

Results

Permutations

As the number of data points is small for each dose, we use permutations to obtain more meaningful statistics for each dose. We choose a very high number of permutations to ensure that the p value of the hypothesis testing is not affected by the actual permutations chosen at random.

0.5mg/day dose

```

set.seed(2)
toothgrowthdose1 <- subset(toothgrowth,doseinmgperday==0.5)
dose1growth <- toothgrowthdose1$toothlength
group1 <- as.character(toothgrowthdose1$supplementtype)
testStat1 <- function(v,g) mean(v[g == "OJ"]) - mean(v[g == "VC"])
observedStat1 <- testStat1(dose1growth,group1)
permutations1 <- sapply(1:100000,function(i) testStat1(dose1growth,sample(group1)))
prob1 <- mean(permutations1>observedStat1)

```

We calculated the probability of the permutations being above the observed difference in mean odontoblasts length between the group administered 0.5mg/day of orange juice and the group administered 0.5mg/day of ascorbic acid to be 0.00234, which means the null hypothesis can be rejected in favour of the alternative hypothesis, i.e. the difference appears significant so we can conclude that for a dose of 0.5mg/day, vitamin C delivered via orange juice is found to be associated with a larger odontoblasts length than for delivery via ascorbic acid.

1mg/day dose

```

set.seed(2)
options(scipen = 999)
toothgrowthdose2 <- subset(toothgrowth,doseinmgperday==1.0)
dose2growth <- toothgrowthdose2$toothlength
group2 <- as.character(toothgrowthdose2$supplementtype)
testStat2 <- function(v,g) mean(v[g == "OJ"]) - mean(v[g == "VC"])
observedStat2 <- testStat2(dose2growth,group2)
permutations2 <- sapply(1:100000,function(i) testStat2(dose2growth,sample(group2)))
prob2 <- mean(permutations2>observedStat2)

```

We calculated the probability of the permutations being above the observed difference in mean odontoblasts length between the group administered 1.0mg/day of orange juice and the group administered 1.0mg/day of ascorbic acid to be 0.00067, which means the null hypothesis can be rejected in favour of the alternative hypothesis, i.e. the difference appears significant so we can conclude for a dose of 1.0mg/day, vitamin C delivered via orange juice is found to be associated with a larger odontoblasts length than for delivery via ascorbic acid.

2mg/day dose

```

set.seed(2)
toothgrowthdose3 <- subset(toothgrowth,doseinmgperday==2.0)
dose3growth <- toothgrowthdose3$toothlength
group3 <- as.character(toothgrowthdose3$supplementtype)
testStat3 <- function(v,g) mean(v[g == "OJ"]) - mean(v[g == "VC"])
observedStat3 <- testStat3(dose3growth,group3)
permutations3 <- sapply(1:100000,function(i) testStat3(dose3growth,sample(group3)))
prob3 <- mean(permutations3>observedStat3)

```

We calculated the probability of the permutations being above the observed difference in mean odontoblasts length between the group administered 2.0mg/day of orange juice and the group administered 2.0mg/day of ascorbic acid to be 0.51444, which means the null hypothesis cannot be rejected in favour of the alternative hypothesis, i.e. that for this dose the difference between delivery via orange juice and delivery via ascorbic acid doesn't appear to be statistically significant.

The results of the permutation analyses confirm the trends observed in the exploratory histogram.

Bootstrapping

We also perform bootstrapping to check the results we obtained from the permutations, and we expect to reach exactly the same conclusions.

0.5mg/day dose

```
B <- 100000
oj1 <- toothgrowthdose1[toothgrowthdose1$supplementtype=="OJ",]
vc1 <- toothgrowthdose1[toothgrowthdose1$supplementtype=="VC",]
diff1 <- oj1$toothlength - vc1$toothlength
resamples1 <- matrix(sample(diff1,40*B,replace=TRUE),B,40)
medians1 <- apply(resamples1,1,median)
sd(medians1)
```

```
## [1] 2.397045
```

```
quantile(medians1,c(0.025,0.975))
```

```
## 2.5% 97.5%
```

```
## 2.7 10.0
```

1.0mg/day dose

```
B <- 100000
oj2 <- toothgrowthdose2[toothgrowthdose2$supplementtype=="OJ",]
vc2 <- toothgrowthdose2[toothgrowthdose2$supplementtype=="VC",]
diff2 <- oj2$toothlength - vc2$toothlength
resamples2 <- matrix(sample(diff2,40*B,replace=TRUE),B,40)
medians2 <- apply(resamples2,1,median)
sd(medians2)
```

```
## [1] 0.6508012
```

```
quantile(medians2,c(0.025,0.975))
```

```
## 2.5% 97.5%
```

```
## 6.7 8.4
```

2.0mg/day dose

```
B <- 100000
oj3 <- toothgrowthdose3[toothgrowthdose3$supplementtype=="OJ",]
vc3 <- toothgrowthdose3[toothgrowthdose3$supplementtype=="VC",]
diff3 <- oj3$toothlength - vc3$toothlength
resamples3 <- matrix(sample(diff3,40*B,replace=TRUE),B,40)
medians3 <- apply(resamples3,1,median)
sd(medians3)
```

```
## [1] 0.9278679
```

```
quantile(medians3,c(0.025,0.975))
```

```
## 2.5% 97.5%
```

```
## -1.6 1.9
```

As expected the bootstrapping analysis confirmed the results from the permutations: 1. For dose of 0.5mg/day we see the bootstrapping confidence 95% interval for the mean of the difference between the 'orange juice' group and the 'ascorbic acid' group is strictly above zero, indicating that we can reject the null hypothesis

and that there is a statistically significant difference between the two delivery method, with orange juice delivery being associated with larger odontoblasts length. 2. For dose of 1mg/day we observe a similar outcome to the results for 0.5mg/day dose, indicating that we can reject the null hypothesis and that there is a statistically significant difference between the two delivery method, with orange juice delivery being associated with larger odontoblasts length. 3. However, for a dose of 2.0mg/day we see the bootstrapping confidence 95% interval for the mean of the difference between the 'orange juice' group and the 'ascorbic acid' group includes zero, indicating that we cannot reject the null hypothesis and that there is no statistically significant difference between the two delivery methods.

We can also see that for the 0.5mg/day dose vs 1.0mg/day there is a larger spread of results, indicated by the larger population standard deviation, which explains why the permutations results identified a smaller probability of exceeding the observed statistics. This is also aligned with the wider confidence interval observed for the 0.5mg/day dose vs 1.0mg/day dose.

It should be noted that the above conclusions make the assumptions that the sample means give a good estimate of the general population mean, i.e. that the dataset is not simply a collection of very rare results amongst the population.