Basic Inferential Data Analysis

Bo Liu

March 27, 2017

Overview

This report explores the growth of the odontoblast, the cell responsible for tooth growth, of the incisor teeth as a criterion of vitamin C intake of a guinea pig. In this study, 60 guinea pigs each received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid.

Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(ggplot2);library(datasets);library(Hmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, round.POSIXt, trunc.POSIXt, units
```

```
data("ToothGrowth")
head(ToothGrowth)
```

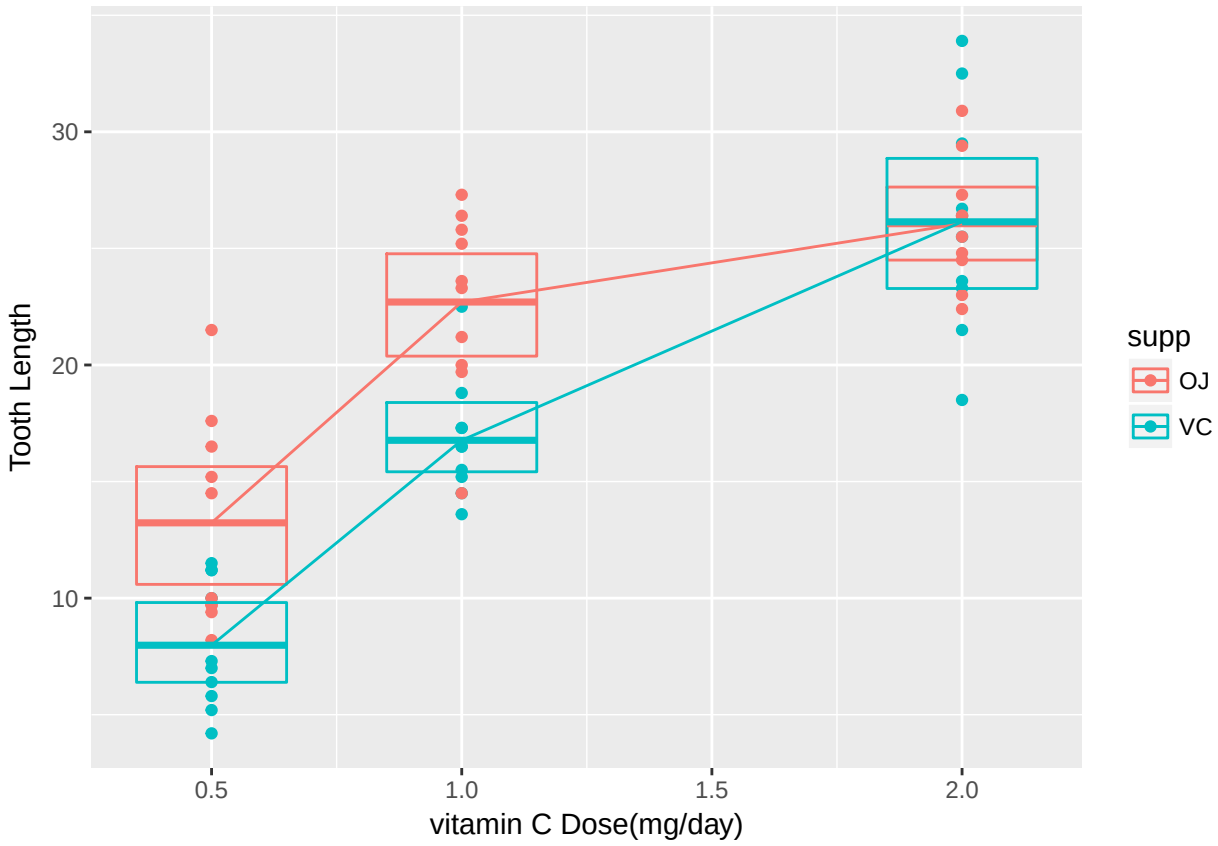
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
smry <- stat_summary(geom="crossbar", width=0.3, fun.data="mean_cl_boot")
meanline <- stat_summary(geom="line",fun="mean_se")
```

```
## Warning: Ignoring unknown parameters: fun
```

```
qplot(dose, len, data=ToothGrowth, colour=supp, xlab="vitamin C Dose(mg/day)", ylab="Tooth Length") + sm
```

```
## No summary function supplied, defaulting to `mean_se()
```



Provide a basic summary of the data.

From this figure, we can see the trend of len when supp and dose change. The len increasing as dose changing from 0.5 to 2 or supp status turn from VC to OJ.

Use confidence intervals to compare tooth growth by supp and dose.

First, calculate mu and sigma.

```
mu<-mean(ToothGrowth$len)
sigma<-var(ToothGrowth$len)
mu
```

```
## [1] 18.81333
```

```
sigma
```

```
## [1] 58.51202
```

Second, calculate the confidence intervals, say 95% intervals.

```
n <- length(ToothGrowth$len)
con_int<-mu + c(-1,1)*2*sigma/sqrt(n)
con_int
```

```
## [1]  3.705594 33.921073
```

We see that confidence intervals is [3.705594, 33.921073].
Third, let's calculate the mean by supp and dose.

```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ
## Levels: OJ VC
```

```
meansuppvc<-mean(ToothGrowth$len[ToothGrowth$supp=='VC'])
meansuppoj<-mean(ToothGrowth$len[ToothGrowth$supp=='OJ'])
meansuppvc
```

```
## [1] 16.96333
```

```
meansuppoj
```

```
## [1] 20.66333
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

```
meandose05<-mean(ToothGrowth$len[ToothGrowth$dose==.5])
meandose10<-mean(ToothGrowth$len[ToothGrowth$dose==1.0])
meandose20<-mean(ToothGrowth$len[ToothGrowth$dose==2.0])
meandose05
```

```
## [1] 10.605
```

```
meandose10
```

```
## [1] 19.735
```

```
meandose20
```

```
## [1] 26.1
```

Finally, compare those mean values to confidence intervals.

```
con_int-meansuppvc
```

```
## [1] -13.25774  16.95774
```

```
con_int-meansuppoj
```

```
## [1] -16.95774  13.25774
```

```
con_int-meandose05
```

```
## [1] -6.899406 23.316073
```

```
con_int-meandose10
```

```
## [1] -16.02941  14.18607
```

```
con_int-meandose20
```

```
## [1] -22.394406   7.821073
```

If the result values were both negtive or positive, means that the mean value is not in the confidence intervals. Otherwise, the mean value is in the range of confidence intervals.

Conclusions.

Since all the mean values are in the confidence intervals, we can conclude that all those results are believable.