Research of confidence intervals in effect of Vitamin C supplement on guinea pig tooth growth

Alex Gaggin

Overview

Confidence intervals in ToothGrowth dataset show that Vitamin C supplement increases length of guinea pigs' tooth, and orange juice should be preferrable way of delivering supplement at lower doses.

Basic exploratory data analyses

```
data(ToothGrowth)
# ?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
```

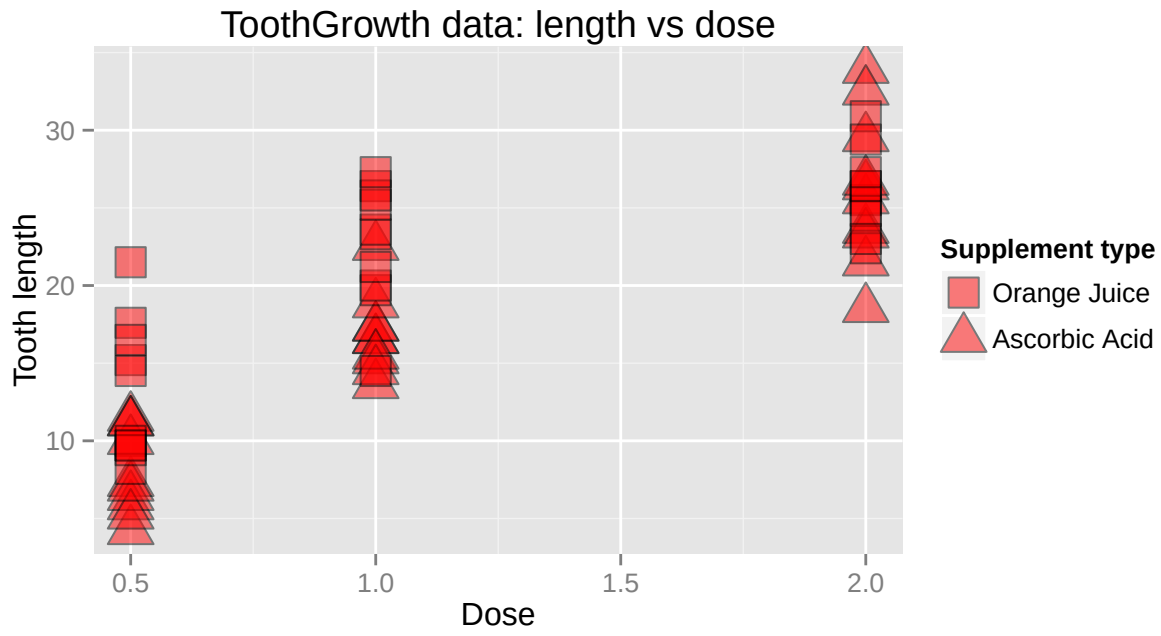
```
summary(ToothGrowth)
```

```
##       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
```

```
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 ...
```

```
print(ggplot(data=ToothGrowth, aes(x=dose, y=len, shape=supp)) +
              xlab("Dose") +
              ylab("Tooth length") +
       ggtitle("ToothGrowth data: length vs dose") +
       geom_point(size=6, alpha=0.5, fill="red") + #guides(shape=FALSE) +
         scale_shape_manual(values=c(22, 24),
                            name="Supplement type",
                         breaks=c("OJ", "VC"),
                         labels=c("Orange Juice", "Ascorbic Acid"))
               )
```



ToothGrowth dataset describes response in the length of teeth of 10 guinea pigs treated with Vitamin C. There are six groups produced for three dose levels (0.5, 1 and 2 mg) and two supplement delivery methods (orange juice and ascorbic acid).

Confidence intervals

Intuitive interpretation of the data on the chart is that supplement does affects tooth length, but delivery type efficiency depends on dose. Let's verify by calculating confidence intervals.

It's not completely clear if 10 guinea pigs described in the help page for ToothGrowth dataset are the same animals in each of groups, or different ones. Even if they are the same, there's no clear indication in the dataset of particular rodents, so we are going to treat them as if they were different, and data points aren't paired. If we will still see that changes are signigicant, then it will be true even if pigs were the same and data points paired, because paired data is less variable, and paired confidence intervals are narrower.

Length per dose

We calculate confidence intervals for tooth length change for four pair of groups: dose change from 0.5 to 1 and dose change from 1 to 2 - for each type of supplement. Notice that t.test()'s default value for paired is FALSE, and this is what we want. We also presume that variability in groups isn't equal and this is also the default setting for the t test.

```
CheckIntervalByDose <- function(before,after,supp) {
      if(missing(supp)) suppC <- TRUE else suppC <- ToothGrowth$supp == supp
      t.test(ToothGrowth$len[ToothGrowth$dose == after & suppC],
       ToothGrowth$len[ToothGrowth$dose == before & suppC]
       )$conf.int %>% cat
}
CheckIntervalByDose(0.5, 1, "OJ")
```

```
## 5.524366 13.41563
```

```
CheckIntervalByDose(1,   2, "OJ")
```

```
## 0.1885575 6.531443
```

```
CheckIntervalByDose(0.5, 1, "VC")
```

```
## 6.314288 11.26571
```

```
CheckIntervalByDose(1,   2, "VC")
```

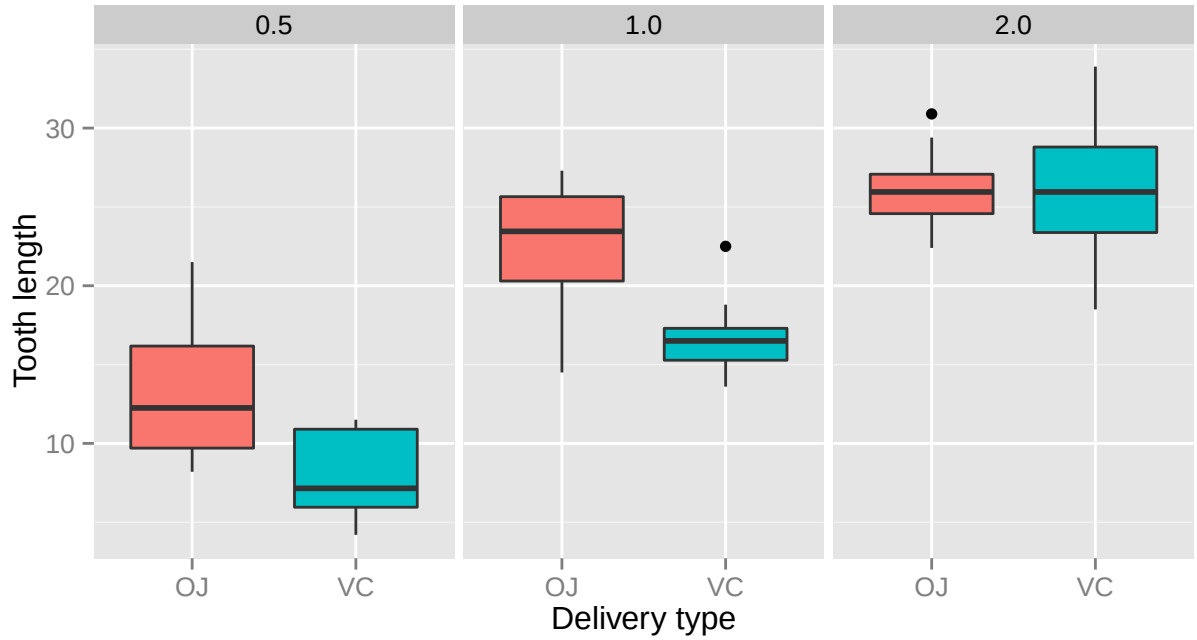
```
## 5.685733 13.05427
```

We see that the confidence intervals lie entirely above zero in every case, so it's clear that increasing the dosage of the supplement does indeed increase tooth length in guinea pigs with 95% confidence level.

Length per supplement

Does supplement delivery affect tooth length?

```
print(ggplot(ToothGrowth, aes(x=supp, y=len, fill=supp)) + geom_boxplot() +
        ylab("Tooth length") + guides(fill=FALSE) +xlab("Delivery type") +
        facet_wrap(~dose))
```



It looks like for lower doses orange juice is more efficient delivery mechanism, while for 2mg dose it doesn't matter anymore. Let's check these assumptions by calculating confidence intervals, assuming data isn't paired and variance can be different for each group.

```
CheckIntervalBySupp <- function(dose) {
      t.test(ToothGrowth$len[ToothGrowth$dose == dose & ToothGrowth$supp == "OJ"],
       ToothGrowth$len[ToothGrowth$dose == dose & ToothGrowth$supp == "VC"]
       )$conf.int %>% cat
}
CheckIntervalBySupp(0.5)
```

```
## 1.719057 8.780943
```

```
CheckIntervalBySupp(1)
```

```
## 2.802148 9.057852
```

```
CheckIntervalBySupp(2)
```

```
## -3.79807 3.63807
```

Indeed, we see that it's only at higher dosage of 2mg we can't say if vitamin C delivery method makes a difference. But for 0.5mg and 1mg orange juice is better, as confidence intervals in these cases lie entirely above zero. Efficiency doesn't seem to be dependant on the dose in a linear way though, as confidence for 0.5mg and 1mg aren't very different.

Result

Assuming data is normally distributed, and using t distribution for safer approximation for small given number of observations, the confidence intervals show that

- change in teeth length is statistically significant for each level of dose increase,
- orange juice is more effective delivery mechanism for vitamic c for lower doses of the supplement.

Appendices

Appendix 1. Effect of ignoring data structure aspects

It probably wouldn't make much sense for a real research to lose information we have about the experiment, but just for the training purposes, let's see if the result changes when we do that.

First, let's ignore supplement delivery type when checking length dependance on the dose.

```
CheckIntervalByDose(0.5, 1)
```

```
## 6.276219 11.98378
```

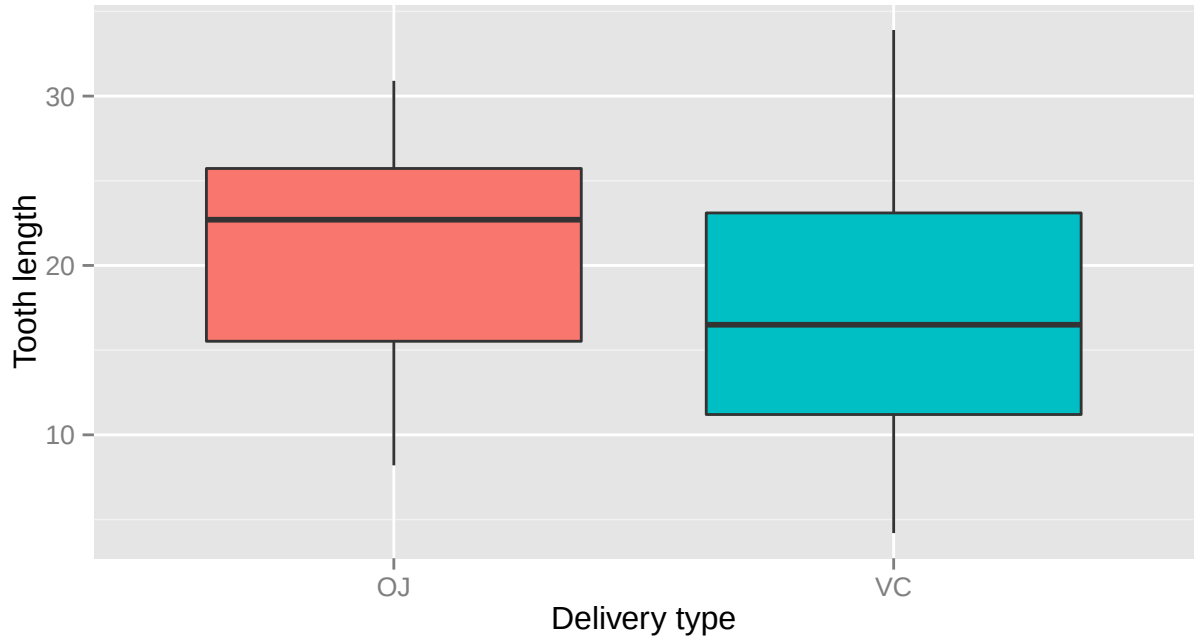
```
CheckIntervalByDose(1,   2)
```

```
## 3.733519 8.996481
```

Intervals are still entirely above zero, and the distances aren't dramatically different than before - it's probably because increased variability was compensated by having twice as much data points, and having more points increases confidence.

Now, let's try to see delivery type effect without separation by dose.

```
print(ggplot(ToothGrowth, aes(x=supp, y=len, fill=supp)) + geom_boxplot() +
        ylab("Tooth length") + guides(fill=FALSE) +xlab("Delivery type"))
```



The plot looks like orange juice is better, but is it?

```
t.test(ToothGrowth$len[ToothGrowth$supp == "OJ"],
       ToothGrowth$len[ToothGrowth$supp == "VC"]
       )$conf.int %>% cat
```

```
## -0.1710156 7.571016
```

We see that if variance isn't the same for each group, and rodents were different in each group, then we can't say that in general orange juice is more effective, because our confidence interval touches zero.

Appendix 2. Effect of presumed equal variance

By the way, what if we are to presume equal variance here, will this assumption tip the scales?

```
t.test(ToothGrowth$len[ToothGrowth$supp == "OJ"],
       ToothGrowth$len[ToothGrowth$supp == "VC"],
       var.equal=TRUE
       )$conf.int %>% cat
```

```
## -0.1670064 7.567006
```

Well, interval became a bit narrower but not narrow enough to miss zero. So in general, orange juice isn't more effective.