An Exploration of Tooth Growth

Randall

January 25, 2015

The data for this exploration is documented:

> The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

The variables are `len`, tooth length in mm, `supp`, whether vitamin C was administered as orange juice (OJ) or as ascorbic acid (VC), and `dose` in mg of vitamin C.

For the purposes of this project I will treat tooth length as a dependent variable of dose and supplement type.

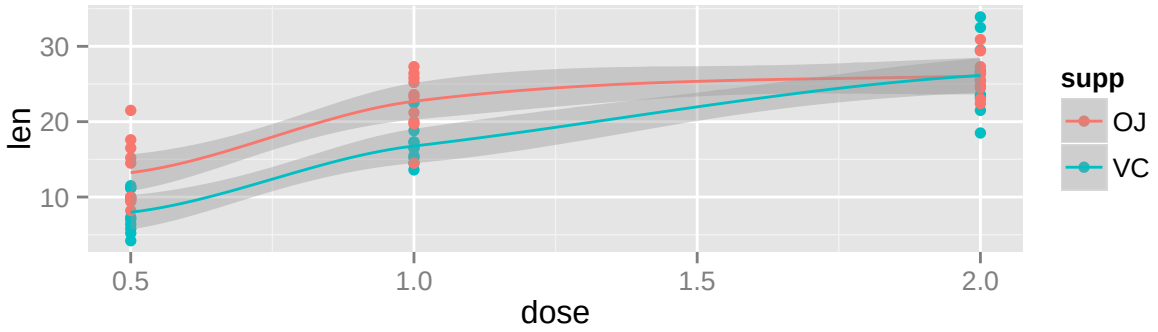
```
library("datasets"); library("ggplot2"); library("UsingR")
library("tidyr") # change shape of data
library("dplyr") # data manipulation and loads magrittr
library("magrittr")
library("xtable") # formats tables in knitr
data("ToothGrowth")
```

```
ToothGrowth %<>% tbl_df # convert ToothGrowth promise into dplyr::tbl_df object
ToothGrowth[sample(x = 60, size = 5), ] %>%
    xtable %>% print(comment=FALSE) #Show random 5 values
```

	len	supp	dose
1	16.50	OJ	0.50
2	32.50	VC	2.00
3	14.50	OJ	1.00
4	19.70	OJ	1.00
5	13.60	VC	1.00

A quick exploratory plot of the data give an indication what is good to study further:

```
qplot(data = ToothGrowth, x = dose, y = len, color = supp) + geom_smooth()
```



Notably, it looks like increased Vitamin C is associated with longer tooth length, and at least at lower doses, that orange juice is a more effective delivery method than ascorbic acid by itself.

Association Between Vitamin C and Tooth Length

Let's hypothesize[1] that a dose of 1mg of vitamin C is associated with longer teeth in guinea pigs than a vitmain C dose of 0.5mg.

- $H_0 : \mu_{length,dose=1mg} \leq \mu_{length,dose=0.5mg}$
- $H_a : \mu_{length,dose=1mg} > \mu_{length,dose=0.5mg}$

```
ToothGrowth %<>% group_by(dose)
len_dose_0.5 <- ToothGrowth %>% filter(dose == 0.5)
len_dose_1 <- ToothGrowth %>% filter(dose == 1)

t.test(len_dose_1$len - len_dose_0.5$len)
```

```
##
##  One Sample t-test
##
## data:  len_dose_1$len - len_dose_0.5$len
## t = 6.9669, df = 19, p-value = 1.225e-06
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   6.387121 11.872879
## sample estimates:
## mean of x
##      9.13
```

Yes, it is very safe to reject the null hypothesis that 1mg of vitamin C is not associated with longer tooth length than 0.5mg.

The 95% confidence interval says that the increased dosage of vitamin C corresponds with between 6.387 and 11.873 mm longer teeth. Furthermore the null hypothesis is rejected with a lot of power, it is rejected even with a p-value as low as 0.000001225.

Note that this assumes some additional properties. For instance that both populations have similar variance:

```
var(len_dose_1$len)
```

```
## [1] 19.49608
```

```
var(len_dose_0.5$len)
```

```
## [1] 20.24787
```

It also assumes that doses were chosen randomly, without any confounding reasons. For example, if guinea pigs with longer teeth were larger at the beginning and had a larger appetite, they might consume more of the supplements. Though we should investigate the methodology, it seems unlikely that the researchers make this type of absurd mistake.

[1] This is not a legitimate way to conduct science. A hypothesis should be chosen before an experiment is performed, and then the experiment is used to find a weight of evidence. Choosing a hypothesis to match the data is *not* science, but it is an okay way to practice statistics for the sake of an assignment.