inferential data analysis

wenlarry

October 15, 2016

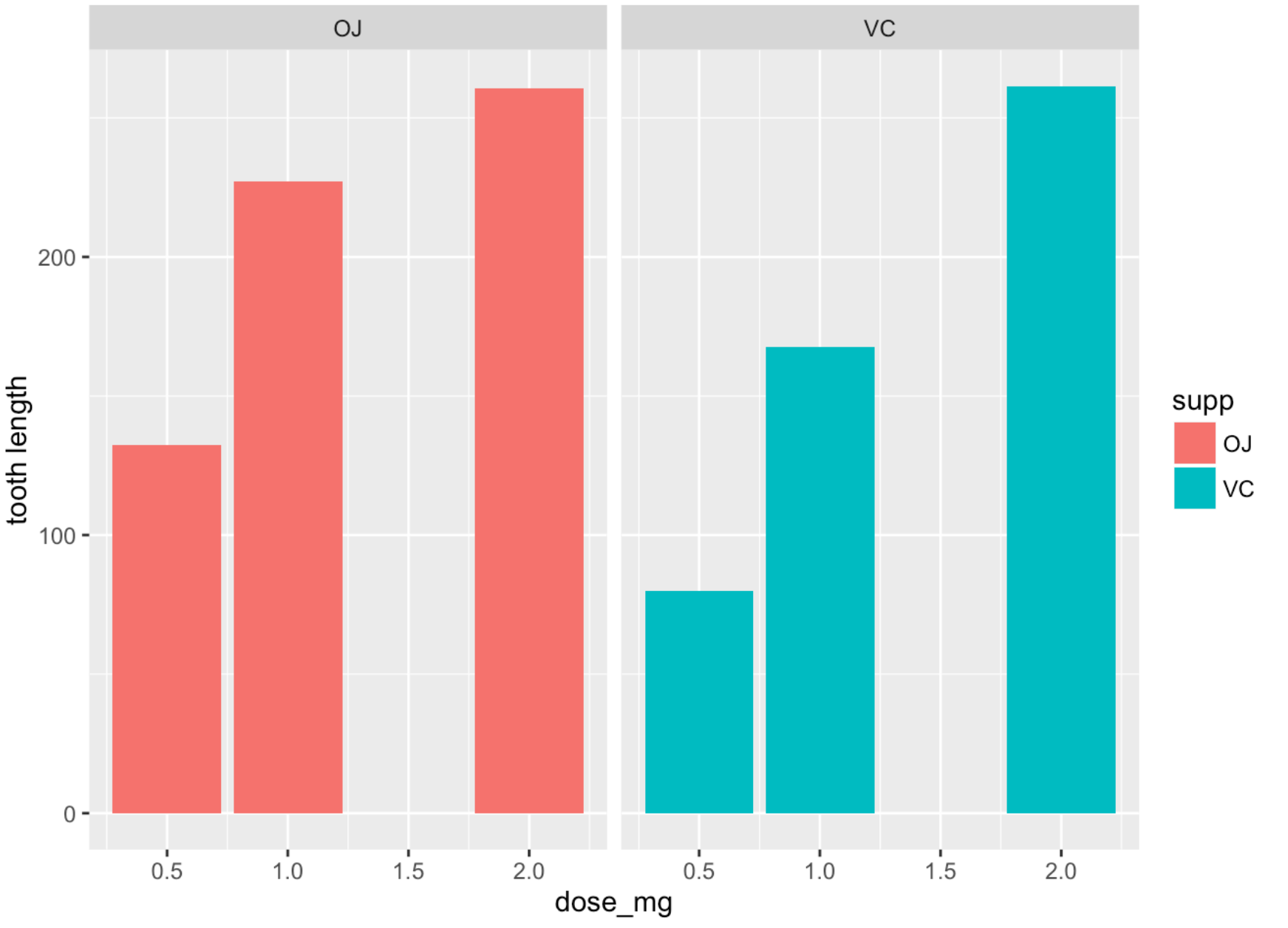
OVERVIEW

An analysis on the effects of various dosages of 2 supplements (OJ & VC) on tooth growth. This analysis uses the ToothGrowth data in R and commences with an exploratory data analysis that is followed by hypothesis tests. Confidence intervals and p values are used in the hypothesis tests.

EXPLORATORY DATA ANALYSIS

data(ToothGrowth) as.factor(ToothGrowth$dose)

```
library(ggplot2)
ggplot(ToothGrowth,aes(dose,len,fill=supp))+
geom_bar(stat="identity" )+  facet_wrap(~supp )+
ylab("tooth length") +xlab("dose_mg")
```



DATA SUMMARY

We can infer from the plot that tooth lengths positively correlate with both dosages of OJ and VC.

When the dosages are at 0.5 to 1.0mg, OJ is more effective for tooth growth.

CONFIDENCE INTERVALS AND HYPOTHESIS TESTS

Hypothesis#1 - OJ & VC deliver tooth growth

```
data<-ToothGrowth

H1<-t.test(len~supp   ,data)
H1$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

```
H1$p.value
```

```
## [1] 0.06063451
```

Confidence interval includes 0 with p value more than 0.05. No rejection of the null hypothesis.

Hypothesis#2- OJ & VC deliver tooth growth at 1mg dosage

```
d1<-subset(data,dose==1)
H2<-t.test(len~supp  ,d1)
H2$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
H2$p.value
```

```
## [1] 0.001038376
```

Confidence interval excludes 0 with p value less than 0.05. Reject the null hypothesis. Accept the alternative hypothesis that 1mg of OJ delivers more tooth growth.

Hypothesis#3 - OJ & VC deliver the same length of tooth growth at 2mg dosage

```
d2<-subset(data,dose==2)
H3<-t.test(len~supp   , d2)
H3$conf.int
```

```
## [1] -3.79807  3.63807
## attr(,"conf.level")
## [1] 0.95
```

```
H3$p.value
```

```
## [1] 0.9638516
```

Confidence level includes 0 with p value more than 0.05. No rejection of the null hypothesis.

CONCLUSION

OJ delivers more tooth growth upto 1 mg of dosage. At 2 mg of dosage, OJ and VC deliver the same level of growth.

The assumptions are that the data are close to the standard normal distribution and the samples represent the population.