project_tooth

lipnerova

December 21, 2015

Info on data:

The response is 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 or ascorbic acid (a form of vitamin C and coded as VC).

Zadání:

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

- Load the ToothGrowth data and perform some basic exploratory data analyses
- Provide a basic summary of the data.
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
- State your conclusions and the assumptions needed for your conclusions.

Some criteria that you will be evaluated on

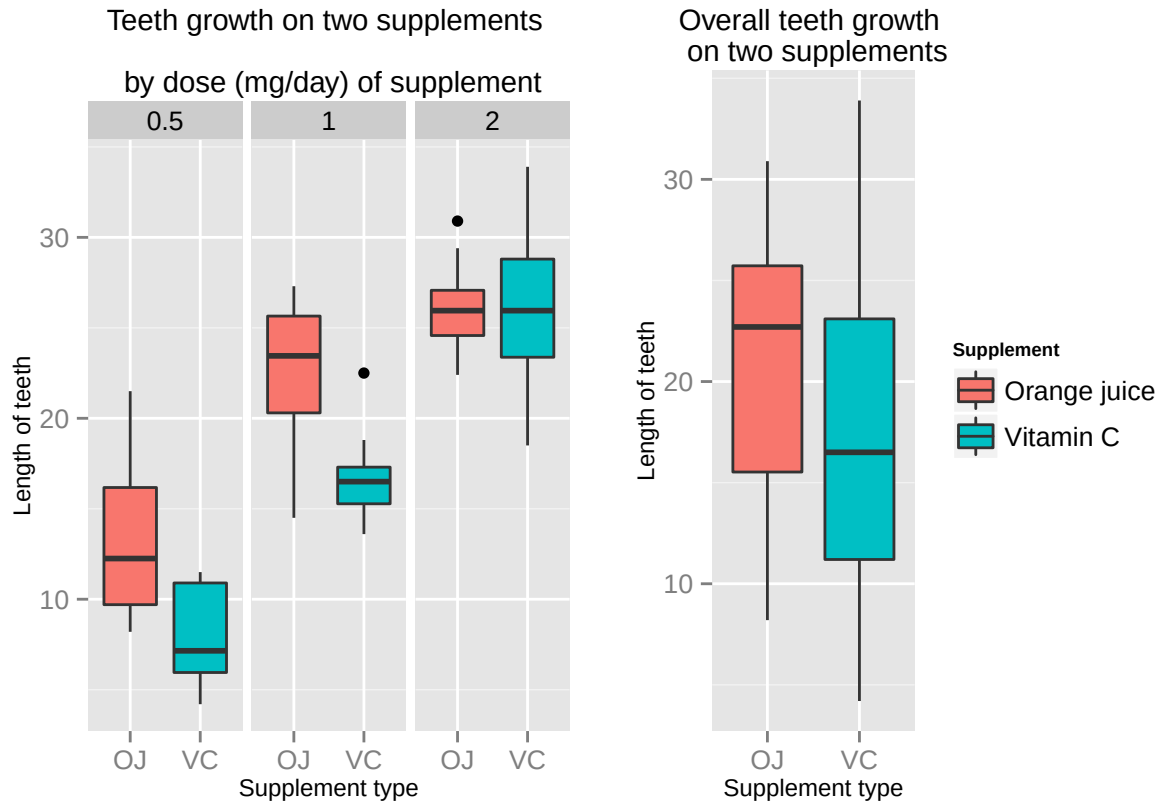
- Did you perform an exploratory data analysis of at least a single plot or table highlighting basic features of the data?
- Did the student perform some relevant confidence intervals and/or tests?
- Were the results of the tests and/or intervals interpreted in the context of the problem correctly?
- Did the student describe the assumptions needed for their conclusions?

Loading the data

```
# Load libraries
library(dplyr)
library(ggplot2)
library(gridExtra)
library(knitr)

# Load dataset
library(datasets)
data(ToothGrowth)
a<-(ToothGrowth) #give it easily typed name
```

EDA

```
p1 <- ggplot(a, aes(y=len, x=as.factor(supp))) +
  geom_boxplot(aes(fill=supp)) +
  guides(fill=FALSE) +
  facet_grid(.~dose) +
  labs(title="Teeth growth on two supplements
       \n by dose (mg/day) of supplement",
       x="Supplement type",
       y="Length of teeth") +
  theme(title = element_text(size = rel(0.75), hjust = 0.5))


p2 <- ggplot(a, aes(x=supp, y=len)) + geom_boxplot(aes(fill=supp)) +
  labs(title="Overall teeth growth \n on two supplements",
       x="Supplement type",
       y="Length of teeth") +
  scale_fill_discrete(name="Supplement",
                      labels=c("Orange juice", "Vitamin C")) +
  theme(title = element_text(size = rel(0.75), hjust = 0.5))

# Arrange them:
grid.arrange(p1, p2, ncol=2, nrow=1)
```




```
a_tab<-group_by(a, supp, dose) %>% summarise(mean(len))
kable(x=a_tab, digits=1, align="c", caption="Summary table of data")
```

Table 1: Summary table of data

supp	dose	mean(len)
OJ	0.5	13.2
OJ	1.0	22.7
OJ	2.0	26.1
VC	0.5	8.0
VC	1.0	16.8
VC	2.0	26.1

Testing

Test hypothesis, that there is a difference between orange juice and vitamin C (all doses together. Test each dose separately for above stated difference => together 4 tests.

Hypotheses

Overall test: H0 there is no difference between OJ and VC H1: the OJ has bigger effect than VC

```
t.test(len ~ factor(supp), data=a, paired=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by factor(supp)
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

PArtial tests - dose 0.5

```
# Dose of 0.5 mg/day:
t.test(len ~ supp, data=a[a$dose==0.5,])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```r
# Dose of 1 mg/day:
t.test(len ~ supp, data=a[a$dose==1,])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```r
# Dose of 2 mg/day:
t.test(len ~ supp, data=a[a$dose==2,])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```