Statistical Inference: Course project 2

Rachel Smith

8/4/2020

Overview

The following report is in completion of part 2 of the Coursera Statistical Inference Course Project. The following criteria are required:

1. Load the ToothGrowth data and perform some basic exploratory analysis
2. Provide a basic summary of the data
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose
4. State conclusions and assumptions needed for these conclusions

Load and summarize data

```
data("ToothGrowth")

df_toothgrowth <- as_tibble(ToothGrowth)

head(df_toothgrowth)
```

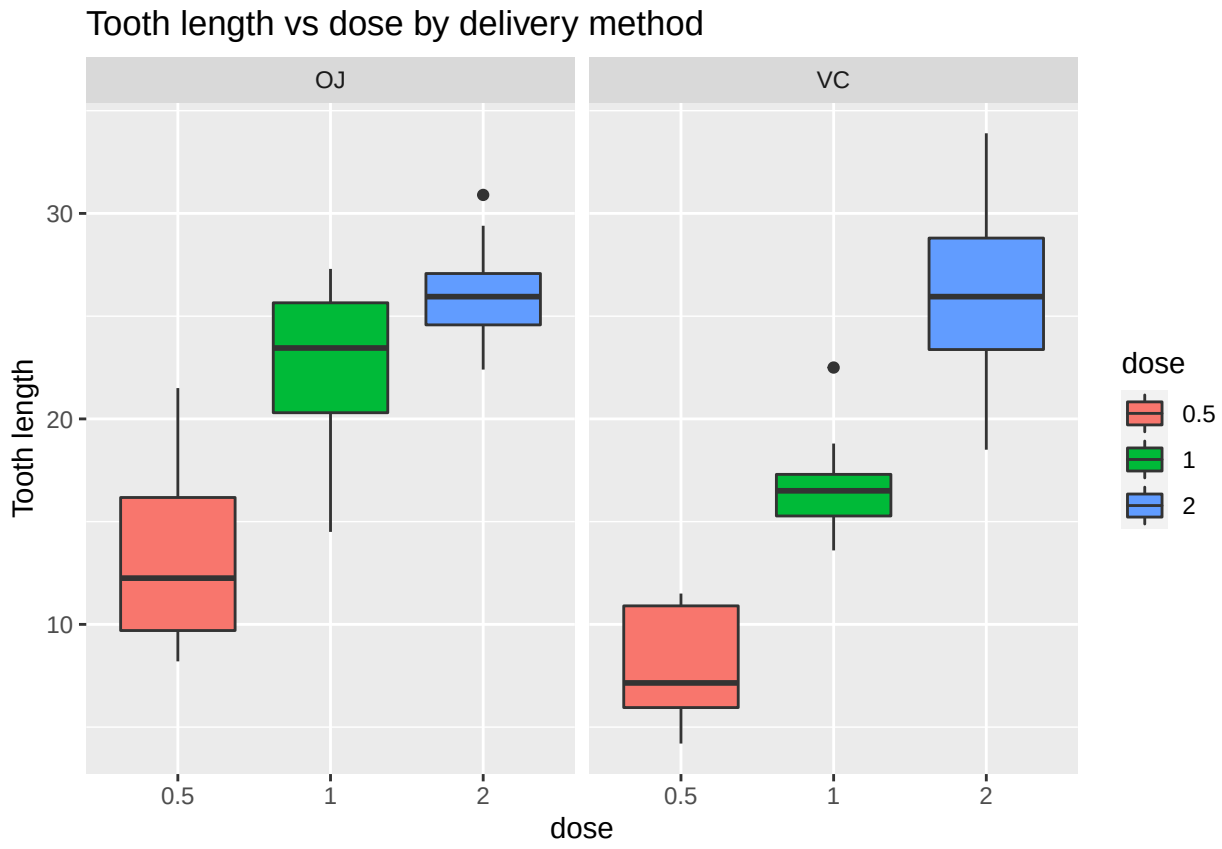
```
## # A tibble: 6 x 3
##     len supp   dose
##   <dbl> <fct> <dbl>
## 1   4.2 VC      0.5
## 2  11.5 VC      0.5
## 3   7.3 VC      0.5
## 4   5.8 VC      0.5
## 5   6.4 VC      0.5
## 6  10   VC      0.5
```

```
summary(ToothGrowth)
```

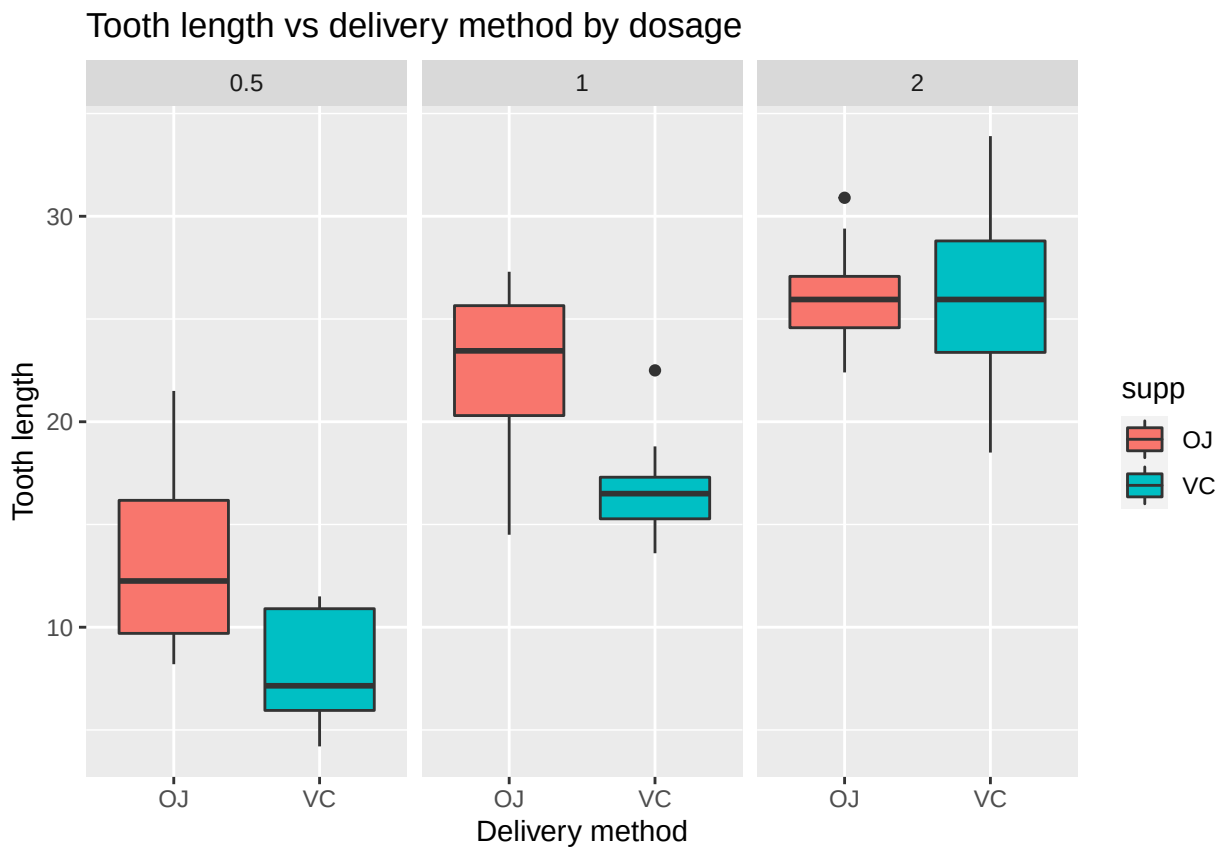
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Exploratory data analysis

```
# convert dose to factor

df_toothgrowth <- df_toothgrowth %>% mutate(dose = as.factor(dose))

# plot tooth length as a function of dose, by supplement

df_toothgrowth %>% ggplot(aes(x = dose, y = len, fill = dose)) +

       geom_boxplot() +
       facet_grid(~supp) +
       ggtitle("Tooth length vs dose by delivery method") +
       ylab("Tooth length")
```




```
# plot tooth length as function of delivery method by dose

df_toothgrowth %>% ggplot(aes(x = supp, y = len, fill = supp)) +

       geom_boxplot() +
       facet_grid(~dose) +
       ggtitle("Tooth length vs delivery method by dosage") +
       xlab("Delivery method") +
       ylab("Tooth length")
```



What affects tooth growth?

```
t.test(len ~ supp, data = df_toothgrowth)
```

Does tooth length depend on delivery method?

```
##
##  Welch Two Sample t-test
##
## data:  len by 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
```

The p-value is 0.06 and confidence interval contains zero, therefore we cannot reject the null and must assume that supplement delivery method doesn't affect tooth length.

Does tooth length depend on dose? Compare 0.5 to 1.0

```
t.test(len ~ dose, data = df_toothgrowth %>% filter(dose == c("0.5","1")))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.4725, df = 17.976, p-value = 0.0002952
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -14.43327  -5.20673
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.63             20.45
```

Compare 0.5 to 2

```
t.test(len ~ dose, data = df_toothgrowth %>% filter(dose == c("0.5","2")))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -7.3335, df = 17.635, p-value = 9.362e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -19.72833 -10.93167
## sample estimates:
## mean in group 0.5   mean in group 2
##             10.63             25.96
```

Compare 1 to 2

```
t.test(len ~ dose, data = df_toothgrowth %>% filter(dose == c("1","2")))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -3.6827, df = 17.949, p-value = 0.00171
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -10.899993  -2.980007
## sample estimates:
## mean in group 1 mean in group 2
##           19.02           25.96
```

All p-values are less than 0.05 and no CIs cross 0; therefore we can reject the null hypothesis and assume that dose affects tooth length.

Conlusions

Under the assumptions that the sample is representative of the population, and the distribution of the sample means follow the Central Limit Theorem, we can conclude that the supplement delivery method does not affect tooth growth, but dosage size does affect tooth length.