Title: statistical_inference_project_part2

Author: Derek Chang

Overview

This part is a statistical analysis based on the dataset "ToothGrowth". In this report, a basic summary of the data and the statistical analysis are provided.

Data description

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).

Data format

A data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length

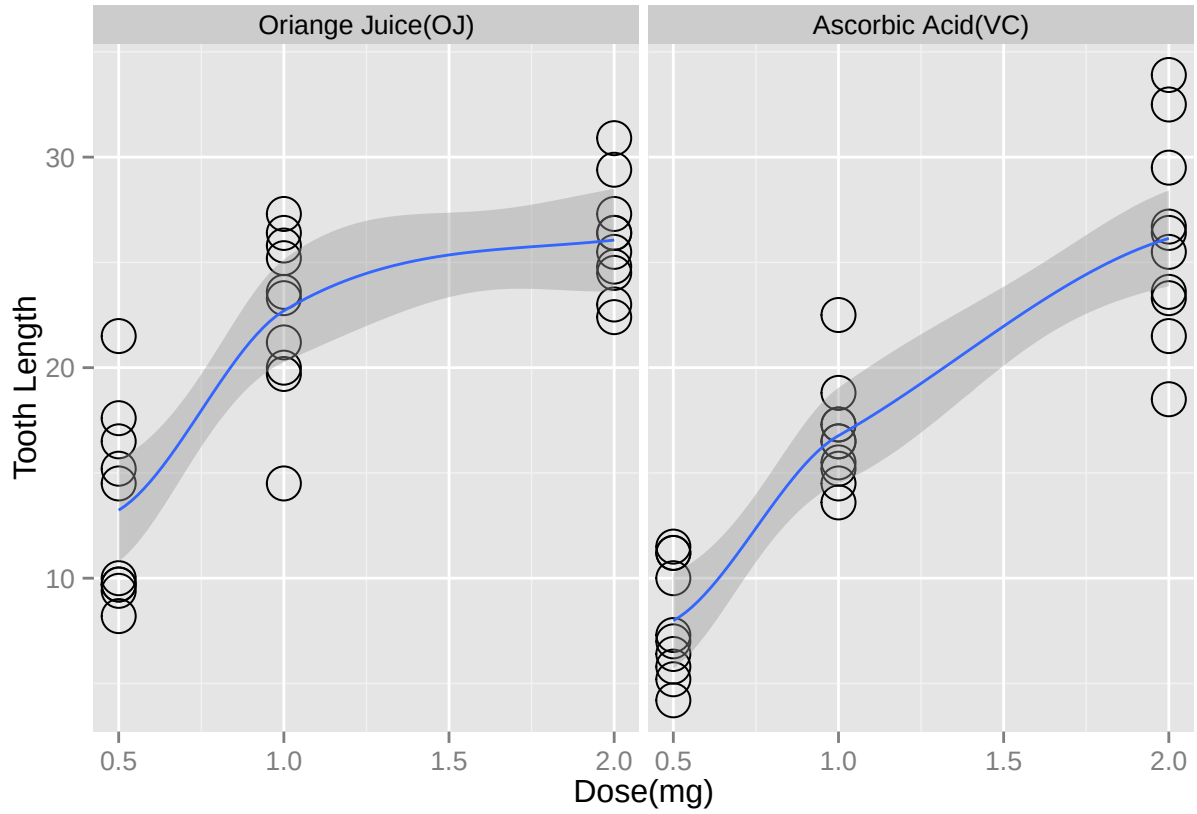
[,2] supp factor Supplement type (VC or OJ).

[,3] dose numeric Dose in milligrams.

Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
library(ggplot2)
# Load the ToothGrowth data
toothgrowth = ToothGrowth
# perform some basic exploratory data analyses
toothgrowth$supp <- as.factor(toothgrowth$supp)
levels(toothgrowth$supp) = c("Oriange Juice(OJ)","Ascorbic Acid(VC)")
p1 <- ggplot(toothgrowth,aes(dose,len))+
  geom_point(pch=1,cex=6)+
  facet_wrap(~supp)+
  geom_smooth()+
  labs(x="Dose(mg)",y="Tooth Length")
p1
```

```
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to cl
```



In the figure, the trend of tooth length is upward after increasing the dose levels of Vitamin C.

Provide a basic summary of the data

```
summary(toothgrowth)
```

```
##       len                        supp         dose
##  Min.   : 4.20   Oriange Juice(OJ):30   Min.   :0.500
##  1st Qu.:13.07   Ascorbic Acid(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
```

```
# summary of the delivery method, orange juice
summary(subset(toothgrowth,supp=="Oriange Juice(OJ)"))
```

```
##       len                        supp         dose
##  Min.   : 8.20   Oriange Juice(OJ):30   Min.   :0.500
##  1st Qu.:15.53   Ascorbic Acid(VC): 0   1st Qu.:0.500
##  Median :22.70                          Median :1.000
##  Mean   :20.66                          Mean   :1.167
##  3rd Qu.:25.73                          3rd Qu.:2.000
##  Max.   :30.90                          Max.   :2.000
```

```r
# summary of the delivery method, ascorbic acid
summary(subset(toothgrowth,supp=="Ascorbic Acid(VC)"))
```

```
##       len                       supp         dose
##  Min.   : 4.20   Oriange Juice(OJ): 0   Min.   :0.500
##  1st Qu.:11.20   Ascorbic Acid(VC):30   1st Qu.:0.500
##  Median :16.50                          Median :1.000
##  Mean   :16.96                          Mean   :1.167
##  3rd Qu.:23.10                          3rd Qu.:2.000
##  Max.   :33.90                          Max.   :2.000
```

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

```r
# two sample t-test
t.test(len~supp, toothgrowth[toothgrowth$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 Oriange Juice(OJ) mean in group Ascorbic Acid(VC)
##                           13.23                            7.98
```

```r
t.test(len~supp, toothgrowth[toothgrowth$dose==1.0,])
```

```
##
##  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 Oriange Juice(OJ) mean in group Ascorbic Acid(VC)
##                           22.70                            16.77
```

```r
t.test(len~supp, toothgrowth[toothgrowth$dose==2.0,])
```

```
##
##  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 Oriange Juice(OJ) mean in group Ascorbic Acid(VC)
##                          26.06                           26.14
```

Based on two sample t-test of different delivery methods, when the dose levels equal to 0.5 and 1.0, null hypothesis can be rejected, the true difference in means is not equal to 0. However, when the dose level is 2.0, the 95 percent confidence interval covers zero, so we cannot reject null hypothesis.

State your conclusions and the assumptions needed for your conclusions

In this analysis, we assume the subsets of different delivery methods are independent two groups. From exploratory data analyses, no matter what delivery method is chosen, when the dose level of Vitamin C is increased, the length of odontoblasts is also increased. In addition, based on the two sample t-test, for dosage of 0.5 mg and 1.0 mg, orange juice can produce a stronger effect on the tooth growth than ascorbic acid.