Statistical Inference Analysis on ToothGrowth Dataset

Dhruv Mishra

The `ToothGrowth` data set contains "*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).*" [Ref.; C. I. Bliss (1952) The Statistics of Bioassay. Academic Press]

Exploratory data analyses

The values which the dose variable can take are 0.5, 1.0 and 2.0 and thus we will convert it into a factor. The relevant plot is in the appendix section.

```
library(ggplot2)
data(ToothGrowth)
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

The distributions of tooth length by delivery method and dose levels seem to indicate that for 0.5 and 1 mg, the mean tooth length is greater for those where orange juice (OJ) was used. For the 2 mg level, no noticeable difference in the mean is observed, but the interquartile range for the OJ group seems to be smaller.

Basic Summary of data

```
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Confidence intervals and hypothesis tests

See if the difference if mean equal to zero is present in the 95 percent confidence interval or t statistic is significantly larger or p value is very large than 0.05. The relevant codes are in the appendix section.

Supp OJ vs VC

The confidence interval contains zero so we can't reject the null hypothesis which means that we can't say that the delivery method has an effect on the tooth length.

Dose 0.5 vs 1

The confidence interval of -11.98 to -6.27 means that the null hypothesis is rejected and the growth caused by 0.5 is less that 1. The p value is also very small implying the rejection of null hypothesis.

Dose 1 vs 2

The confidence interval of -8.99 to -3.73 means that the null hypothesis is rejected and the growth caused by 1 is less that 2. The p value is also very small implying the rejection of null hypothesis.

The above test implies the order 2>1>0.5 for dose variable.

Assumptions

For this experiment, we've assumed there's a common variance in the guinea pigs population (var.equal=TRUE) while performing tests on the dose variable. The experiment design suggest that each Guinea pig was randomly assigned to a combination of dosage and supplement type, so the test that we performed used the independent samples methodology.

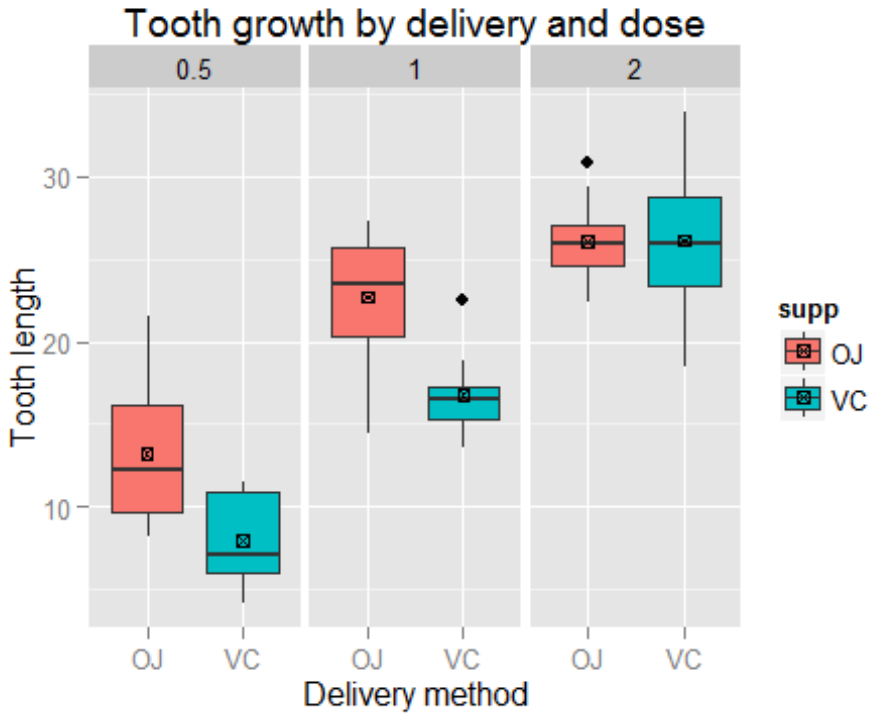
Conclusion

The supplement type has no effect on tooth growth. However increasing the dose level leads to increased tooth growth.

Appendix

Exploratory data analysis

```
ggplot(data=ToothGrowth, aes(y=len, x=supp, fill=supp)) +
geom_boxplot() + facet_wrap(~ dose, ncol=3) + ylab("Tooth length") +
xlab("Delivery method") + ggtitle("Tooth growth by delivery and dose")
+ stat_summary(fun.y=mean, geom="point", shape=7, size=2)
```



The square dot represents the respective mean.

Hypothesis Tests

Supp OJ vs VC

```
t.test(ToothGrowth$len[ToothGrowth$supp=="OJ"],
ToothGrowth$len[ToothGrowth$supp=="VC"], paired = FALSE, var.equal =
FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and
ToothGrowth$len[ToothGrowth$supp == "VC"]
## 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 of x mean of y
##  20.66333  16.96333
```

Dose 0.5 vs 1

```
t.test(ToothGrowth$len[ToothGrowth$dose==0.5],
ToothGrowth$len[ToothGrowth$dose==1], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 0.5] and
ToothGrowth$len[ToothGrowth$dose == 1]
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

Dose 1 vs 2

```
t.test(ToothGrowth$len[ToothGrowth$dose==1],
ToothGrowth$len[ToothGrowth$dose==2], paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and
ToothGrowth$len[ToothGrowth$dose == 2]
## t = -4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.994387 -3.735613
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```