Statistical Inference Course Project

Part 2: Basic inferential data analysis

Dennis van den Berg

25/07/2015

Introduction

This article analyses the effect of vitamin C on tooth growth in Guinea Pigs. Data was obtained from the *toothgrowth* dataset available from Base R.

This dataset contains 60 observations on 3 variables:

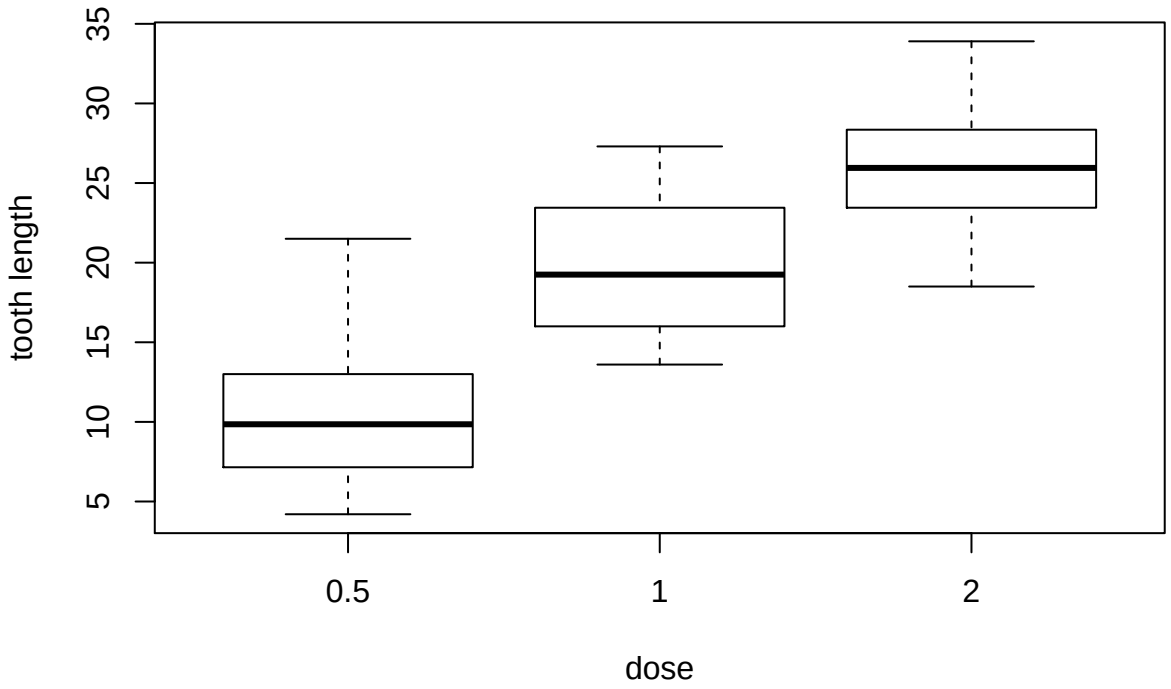
```
[,1]     len     numeric     Tooth length
[,2]     supp    factor      Supplement type (VC or OJ).
[,3]     dose    numeric     Dose in milligrams.
```

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 (supplement types OJ = orange juice or VC = ascorbic acid).

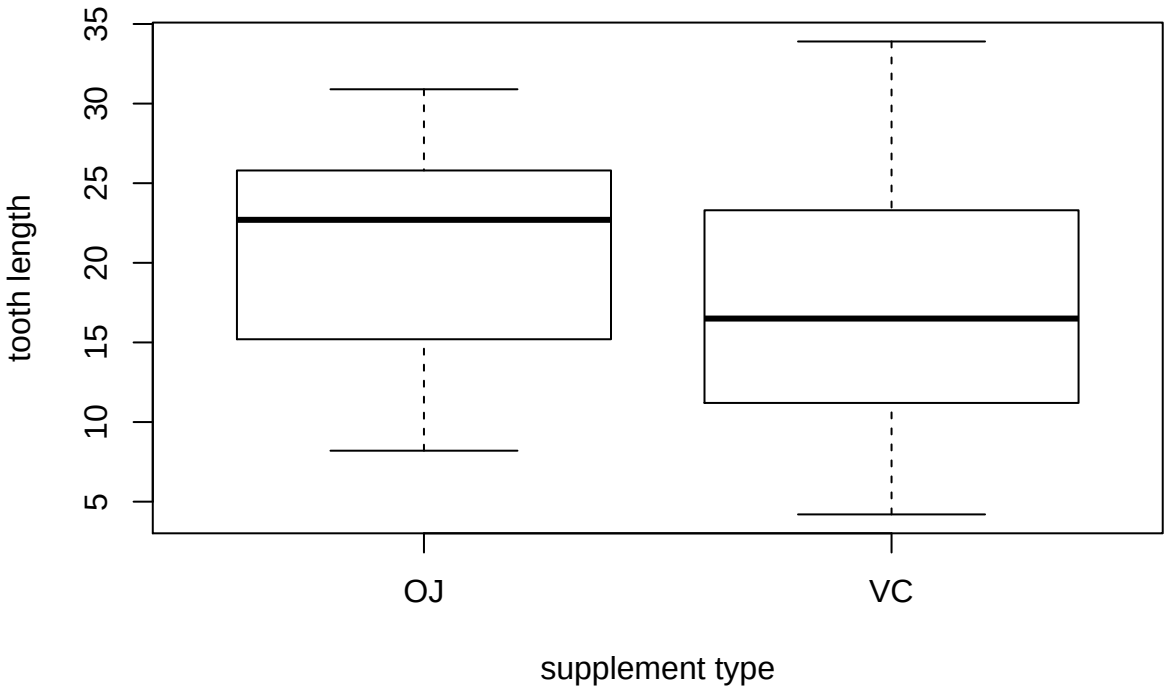
Exploratory Analysis

In this section we perform a quick exploratory analysis of the factors influencing tooth length, in order to generate hypotheses.

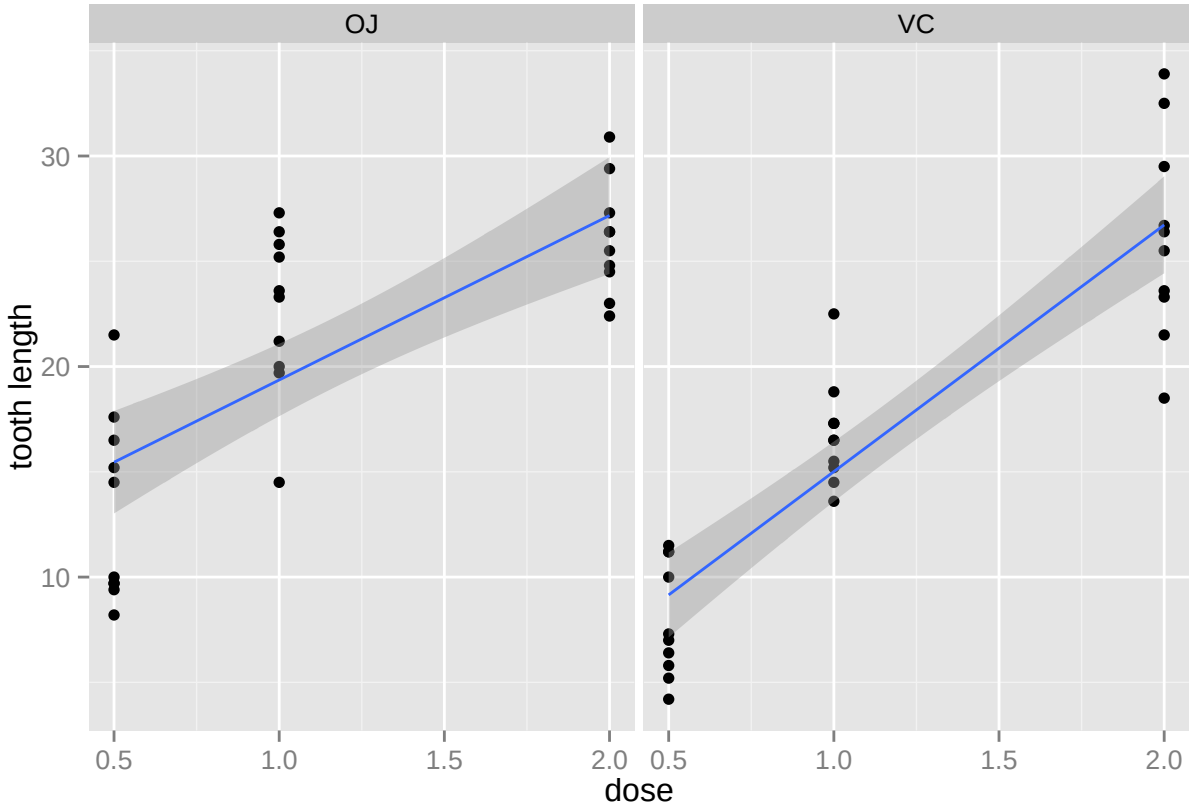
Dose vs tooth length



Supplement type vs tooth length



Supplement type and dose vs tooth length (multivariate)



Mean tooth lengths for different supplement types and doses:

```
##     dose
## supp   0.5     1     2
##   OJ 13.23 22.70 26.06
##   VC  7.98 16.77 26.14
```

Preliminary Summary and Hypotheses

This exploratory tables and plots seem to point out an increasing tooth length for larger doses. Furthermore, supplement type Orange Juice (OJ) seems to give larger tooth lengths than Ascorbic Acid (VC).

We will investigate these relationships in the following section. To do this, we formulate a number of hypotheses:

1. Supplement type Orange Juice gives significantly larger tooth length than Ascorbic Acid
2. Dose level gives significant difference in tooth length. We will check this for the following cases:
    a. Comparison dose = 0.5 vs 1.0
    b. Comparison dose = 0.5 vs 2.0
    c. Comparison dose = 1.0 vs 2.0

Hypothesis Testing

Supplement type versus tooth length

Determining an Independent Group T Interval for the difference in mean tooth length of supplement type Orange Juice and Ascorbic Acid:

```
len_OJ <- subset(ToothGrowth, supp=="OJ", len)
len_VC <- subset(ToothGrowth, supp=="VC", len)
t.test(len_VC, len_OJ, paired = FALSE, var.equal = FALSE)$conf
```

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

Since the 95% confidence interval contains 0 our hypothesis that supplement type Orange Juice gives significantly larger tooth length than Ascorbic Acid is NOT VALIDATED.

Dose vs tooth length

Determining an Independent Group T Interval for difference in mean tooth length for different doses:

```
len_dose05 <- subset(ToothGrowth, dose==.5, len)
len_dose1 <- subset(ToothGrowth, dose==1, len)
len_dose2 <- subset(ToothGrowth, dose==2, len)
t.test(len_dose1, len_dose05, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1]  6.276219 11.983781
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(len_dose2, len_dose05, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] 12.83383 18.15617
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(len_dose2, len_dose1, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] 3.733519 8.996481
## attr(,"conf.level")
## [1] 0.95
```

We see that for all 3 tests the 95% confidence interval is strictly positive, so we conclude that our hypotheses that dose level gives significant difference in tooth length is VALIDATED for all 3 cases:

```
a. Comparison dose = 0.5 vs 1.0
b. Comparison dose = 0.5 vs 2.0
c. Comparison dose = 1.0 vs 2.0
```

Conclusions

We conclude the following:

- Supplement type Orange Juice does not give significantly larger tooth length than Ascorbic Acid
- A dose level of 1.0 mg gives a significant larger tooth length than dose level of 0.5 mg
- A dose level of 2.0 mg gives a significant larger tooth length than dose level of 1.0 mg
- A dose level of 2.0 mg gives a significant larger tooth length than dose level of 0.5 mg

In all our 4 hypothesis tests we used the following assumptions:

- Groups are independent
- Unpaired results
- Unequal variance
- Confidence level of 95%

Data Sources

C. I. Bliss (1952) The Statistics of Bioassay. Academic Press.

References

McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley.

Appendix: Complete R Code

```r
library(datasets)
library(ggplot2)
```

```r
boxplot(len ~ dose, data=ToothGrowth, xlab="dose", ylab="tooth length")
```

```r
boxplot(len ~ supp, data=ToothGrowth, xlab="supplement type", ylab="tooth length")
```

```r
qplot(dose, len, data=ToothGrowth, xlab="dose", ylab="tooth length", facets = . ~ supp) + geom_smooth(m
```

```r
tapply(ToothGrowth$len, list(supp=ToothGrowth$supp, dose=ToothGrowth$dose), mean)
```

```r
len_OJ <- subset(ToothGrowth, supp=="OJ", len)
len_VC <- subset(ToothGrowth, supp=="VC", len)
t.test(len_VC, len_OJ, paired = FALSE, var.equal = FALSE)$conf
```

```r
len_dose05 <- subset(ToothGrowth, dose==.5, len)
len_dose1 <- subset(ToothGrowth, dose==1, len)
len_dose2 <- subset(ToothGrowth, dose==2, len)
t.test(len_dose1, len_dose05, paired = FALSE, var.equal = FALSE)$conf
t.test(len_dose2, len_dose05, paired = FALSE, var.equal = FALSE)$conf
t.test(len_dose2, len_dose1, paired = FALSE, var.equal = FALSE)$conf
```