ToothGrowth Inferential Data Analysis

Elmar Langholz

April 22, 2015

Overview

The ToothGrowth data set explores the *The Effect of Vitamin C on Tooth Growth in Guinea Pigs* on *10 guinea pigs* (see Source). We will perform inferential data analysis on it with the intent of comparing the tooth growth with the dose levels and the delivery methods to identify the impact of these.

Exploratory data analysis

After loading the data, the first thing we should do is a get a sense of the structure of the data set.

```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

According the documentation we should expect the following varaibles defined in *Table 1* which holds true. In general, from this output, we can see that the observations are not directly tied to individual guinea pigs.

Table 1: Variables and description for the ToothGrowth data set

Variable Names	Description
supp	The supplement type or delivery method.
dose	The dose in milligrams.
len	The tooth (odontoblasts) length in micrometers.

The `supp` variable is categorical and can have only two values: **VC or OJ**. There are no missing values and are equally distributed.

Regarding the `dose` variable, it is also categorical and complete because it only encompasses the following numerical values **0.5, 1 or 2**. We convert it into a factor variable to facilitate grouping later on. These values are also equally distributed throughout the data set.

The `len` variable is the response. It is also valid since it does not contain any negative or and complete because all of its values are present. In summary, the mean is of **19.25** μm and can acquire values **from 4.2 to 33.9** μm **(inclusive)**.

Table 2: Summary of the exploration of individual variables

Column name	Type	Values
supp	Categorical	VC (ascorbin acid) or OJ (orange juice)
dose	Categorical	0.5 mg, 1 mg and 2 mg
len	Response	μ = 19.25 μm, range = [4.2, 33.9] μm

Comparing tooth growth by supplement type and dosage

Now that we have taken a look at the variables individually (see *Table 2*), lets look at how the categorical values interact with the response. The observations for the two categorical variables are evenly distributed and their variances are unequal (see Appendix 10, 11, 12). Since the amount of observations is relatively low we will use student's *t*-test to perform any *non-paired* independent group sample tests and will analyze the confidence intervals to validate the observations.

Impact of supplement type in tooth length

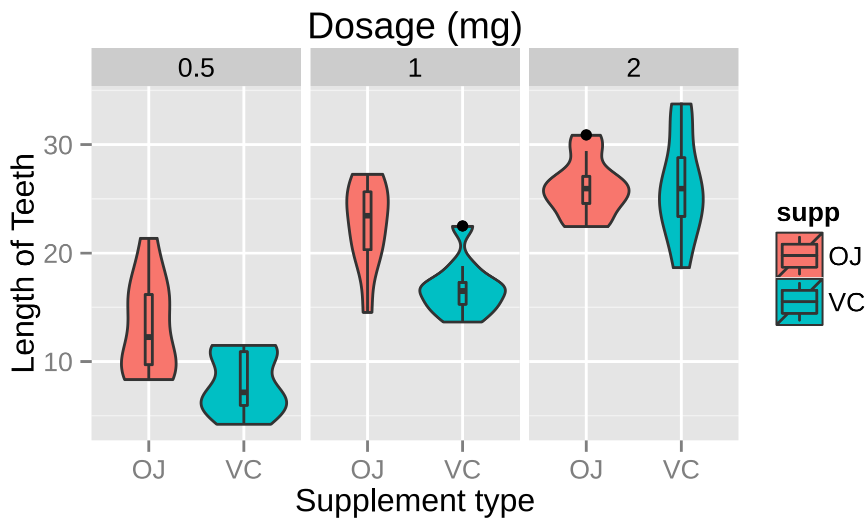

Figure 1: Effects of tooth length by supplement and dosage

Figure 1 shows that, on average, the orange juice delivery method has a higher length for dosages of 0.5 mg and 1.0 mg and for a dosage of 2.0 mg they are relatively the same. When we perform the group sample tests of 95% confidence level, the interval is (-0.17102, 7.57102). Since the interval covers zero, we can't reject the claim that the two groups are the same when not taking into account doses. Nonetheless, if we perform the supplement type tests with doses we are able to confirm through the confidence intervals (see *Table 3*) that the initial observation made is indeed correct because:

1. For the tests with 0.5 mg and 1.0 mg doses, the 95% confidence intervals lower and upper bounds (1.71906, 8.78094) and (2.80215, 9.05785) are positive and do not touch zero. This means that the influence of the orange juice supplement type has a higher impact than that of ascorbin acid.
2. For the test with the 2.0 mg dose, the 95% confidence intervals lower and upper bound (-3.79807, 3.63807) touches zero. This is an indication that we can't reject the claim that the two groups are the same.

Table 3: Summary of Supplement Type By Dose 95% Confidence Intervals

Group1	Group2	Lower	Upper
$OJ_{0.5}$	$VC_{0.5}$	1.719	8.781
OJ_1	VC_1	2.802	9.058
OJ_2	VC_2	-3.798	3.638

Impact of dose in tooth length

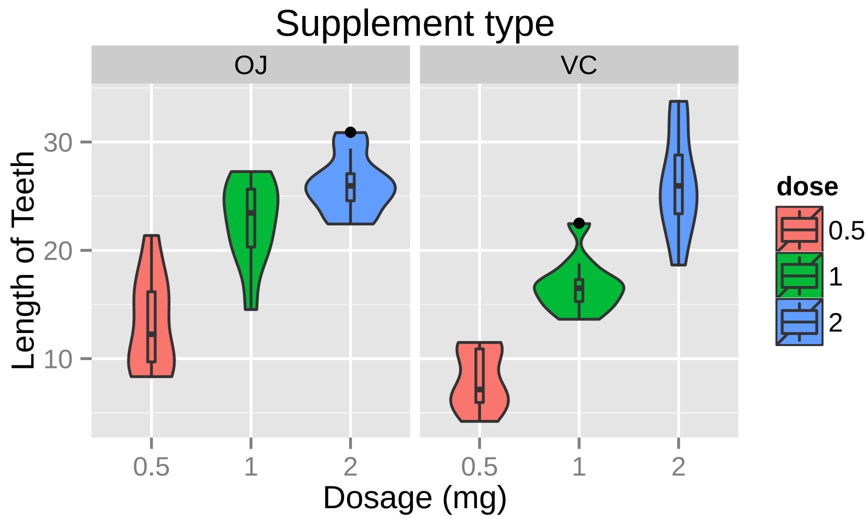

Figure 2: Effects of tooth length by dosage and supplement

Figure 2 shows that as we increase the dosage so does the odontoblasts length (independent of the supplement type). When we perform the group sample tests of 95% confidence level the intervals shown in *Table 4* confirm the previously stated observation because in all of the three tests the intervals as less than zero. This means that the group on the right side is the one that has a higher impact for each test and has a higher dosage than the one on the left.

Table 4: Summary of Dose 95% Confidence Intervals

Group1	Group2	Lower	Upper
0.5	1	-11.98	-6.276
0.5	2	-18.16	-12.83
1	2	-8.996	-3.734

Conclusion

Assuming that:

1. The guinea pigs were randomly selected and are impartial to any of the observed properties/attributes.
2. The sample is representative of the population and no bias was introduced.
3. There aren't any other properties/attributes influencing the length of the tooth besides the ones analyzed.
4. The dose and delivery methods are independent from each other and in turn each have a different variance within the data set.

We can conclude the following with a 0.95 probability:

1. On average, while the orange juice delivery method has a higher odontoblast length for dosages of 0.5 mgand 1.0 mg, for a dosage of 2.0 mg they are relatively the same and there is no clear influencer.
2. As the dose increases, independent of the delivery method (e.g. orange juice or ascorbic acid), so does the odontoblasts length.

Appendix

1. Loading the data

The ToothGrowth data set is provided by the datasets package. We must load the *datasets* package, load the *ToothGrowth* data set and attach to it with the purpose of accessing the variables directly instead of having to reference them through the loaded data frame.

```
library(datasets); data(ToothGrowth); attach(ToothGrowth)
```

2. Supplement type validation

```
suppValues <- unique(supp); suppValues
```

```
## [1] VC OJ
## Levels: OJ VC
```

3. Frequency of the supplement types variable

Table 5: Frequency of the values of supp

supp	Freq
OJ	30
VC	30

4. Dosage validation

```
doseValues <- unique(dose); doseValues
```

```
## [1] 0.5 1   2
## Levels: 0.5 1 2
```

5. Converting the dosage to a factor variable

```
ToothGrowth$dose <- dose <- as.factor(dose)
```

6. Frequency of the dosage variable

Table 6: Frequency of the values of dose

dose	Freq
0.5	20
1	20
2	20

7. Length validation

```
data.frame(hasNegativeValues = sum(len < 0) > 0, hasMissingValues = sum(is.na(len)) > 0)
```

```
##   hasNegativeValues hasMissingValues
## 1             FALSE            FALSE
```

8. Length summary

```
lenSummary <- summary(len); lenSummary
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20   13.08   19.25   18.81   25.28   33.90
```

9. Frequency of the supplement type and dosage variables

Table 7: Frequency of the values of supp and dose

supp	dose	Freq
OJ	0.5	10
VC	0.5	10
OJ	1	10
VC	1	10
OJ	2	10
VC	2	10

10. Summary of length by supplement type

Table 8: Summary of Tooth Length by Supplement Type

supp	mean	sd	var
OJ	20.66	6.606	43.63
VC	16.96	8.266	68.33

11. Summary of length by dosage

Table 9: Summary of Tooth Length by Dose

dose	mean	sd	var
0.5	10.61	4.5	20.25
1	19.73	4.415	19.5
2	26.1	3.774	14.24

12. Summary of length by dosage and supplement type

Table 10: Summary of Tooth Length by Dosage and Supplement

dose	supp	mean	sd	var
0.5	OJ	13.23	4.46	19.89
0.5	VC	7.98	2.747	7.544
1	OJ	22.7	3.911	15.3
1	VC	16.77	2.515	6.327
2	OJ	26.06	2.655	7.049
2	VC	26.14	4.798	23.02

13. Split data by supplement type

```
suppOJ <- subset(ToothGrowth, supp == "OJ"); suppVC <- subset(ToothGrowth, supp == "VC")
```

14. Supplement type testing

```
suppConfInt <- t.test(suppOJ$len, suppVC$len, paired = FALSE, var.equal = FALSE)$conf.int
```

15. Supplement type testing with doses

```
suppByDoseConfInts <- vapply(levels(dose),
                             function (d) {
                                 confInt <- t.test(subset(suppOJ, dose == d)$len,
                                                   subset(suppVC, dose == d)$len,
                                                   paired = FALSE,
                                                   var.equal = FALSE)$conf.int
                                 c(confInt[1], confInt[2])
                             },
                             FUN.VALUE = numeric(2))
```

15. Dose testing

```
doseComparers <- combn(levels(dose), 2)
doseConfInts <- apply(doseComparers,
                      2,
                      function (doseSet) {
                          confInt <- t.test(subset(ToothGrowth, dose == doseSet[1])$len,
                                            subset(ToothGrowth, dose == doseSet[2])$len,
                                            paired = FALSE,
                                            var.equal = FALSE)$conf.int
                          c(confInt[1], confInt[2])
                      })
```