Statistical Inference Course Project Part 2

Katya Demidova

10 December 2015

Overview

In this part of the course project we're going to analyze the ToothGrowth data from the R `datasets` package. Description of this data set: 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).

Loading the data; basic summary

Let us load the data set and perform some basic exploratory data analyses.

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## '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 ...
```

The ToothGrowth data is a data frame with 60 observations of 3 variables:

- `len`: tooth length (numeric),
- `supp`: supplement type / delivery method (factor with 2 levels `OJ` and `VC`, orange juice and ascorbic acid respectively),
- `dose`: dose of Vitamin C in miligrams (numeric).

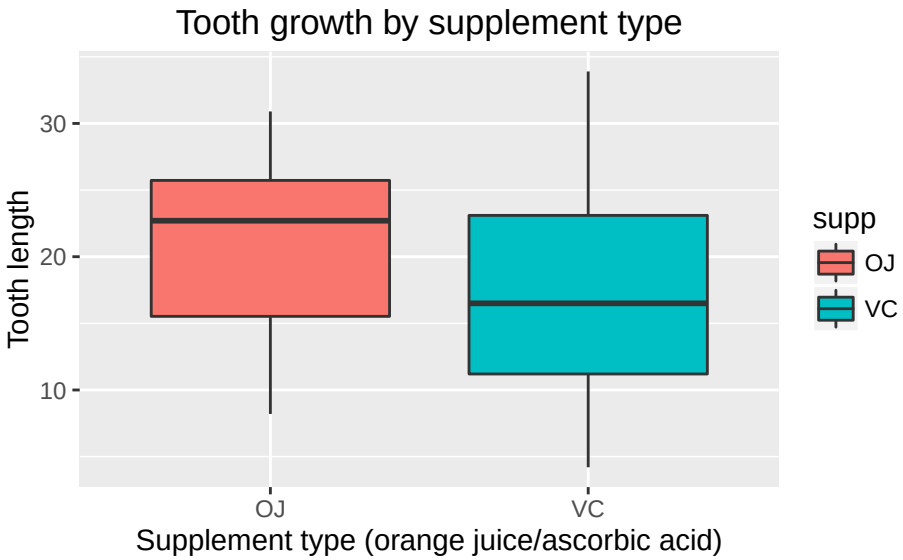
Basic summary of data:

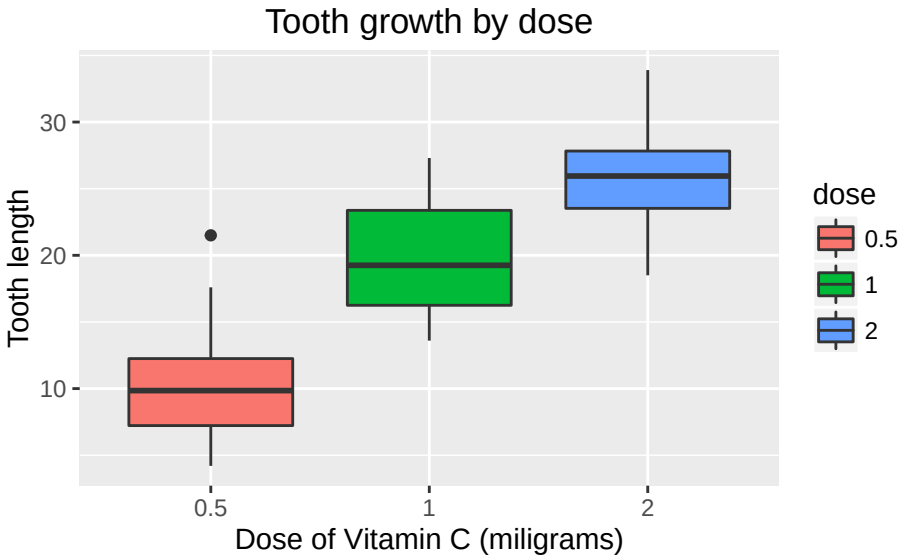
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

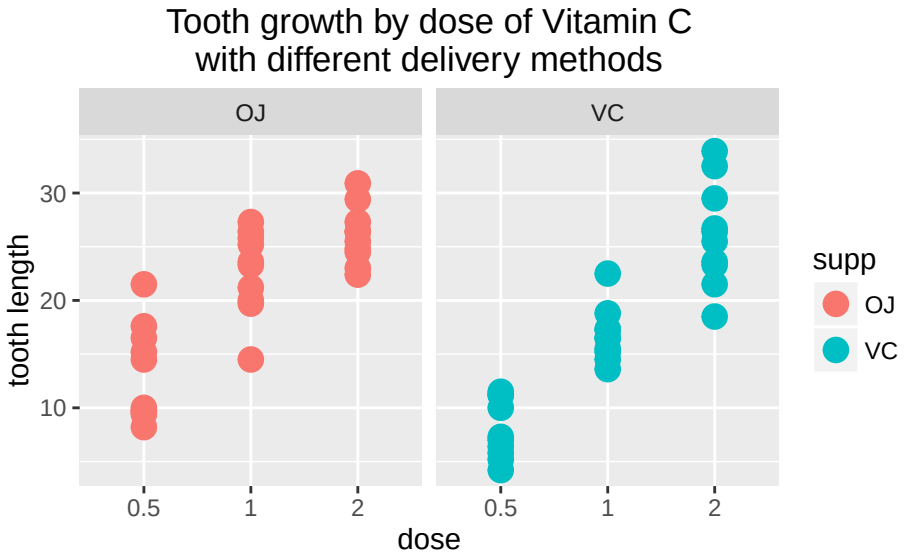
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20   13.08   19.25   18.81   25.28   33.90
```

Though the `dose` variable is numeric in the original data set, we think it would be more convenient if we coerced this variable to a factor.

Exploratory data analysis







The plots above could help us with comparing tooth growth by `supp` and `dose`:

- it is clear from the plots that the higher the dose, the larger the teeth, in other words, it looks like there is a positive correlation between dose level and tooth length in guinea pigs;
- we might also make an assumption that orange juice causes larger tooth growth than ascorbic acid.

Statistical Inference

Tooth growth by dose level

As mentioned above, there seems to be a positive correlation between dose level and tooth length in guinea pigs. Mean values of tooth length by dose:

```
## Source: local data frame [3 x 2]
##
##     dose mean(len)
##   (fctr)     (dbl)
## 1    0.5    10.605
## 2      1    19.735
## 3      2    26.100
```

To compare tooth growth by `dose` and check if there is a statistically significant difference between the means, we will use t-tests:

```
with(tg, t.test(len[dose=="0.5"], len[dose=="1"]))
```

```
##
##  Welch Two Sample t-test
##
## data:  len[dose == "0.5"] and len[dose == "1"]
## t = -6.4766, df = 37.986, p-value = 0.0000001268
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
```

```
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

```
with(tg, t.test(len[dose=="1"], len[dose=="2"]))
```

```
##
##  Welch Two Sample t-test
##
## data:  len[dose == "1"] and len[dose == "2"]
## t = -4.9005, df = 37.101, p-value = 0.00001906
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

```
with(tg, t.test(len[dose=="0.5"], len[dose=="2"]))
```

```
##
##  Welch Two Sample t-test
##
## data:  len[dose == "0.5"] and len[dose == "2"]
## t = -11.799, df = 36.883, p-value = 0.00000000000004398
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

P-values of these t-tests, respectively:

```
## [1] 0.0000001268301
```

```
## [1] 0.0000190643
```

```
## [1] 0.00000000000004397525
```

From these numbers, we can make a conclusion that **dosage of Vitamin C affects tooth growth in guinea pigs**.

Tooth growth by supplement type

We have made an assumption that that orange juice causes larger tooth growth than ascorbic acid. Mean values of tooth length by supplement type:

```
## Source: local data frame [2 x 2]
##
##     supp mean(len)
##   (fctr)     (dbl)
## 1     OJ  20.66333
## 2     VC  16.96333
```

To compare tooth growth by `supp` and check if there is a statistically significant difference between the means, we will use a t-test:

```
with(tg, t.test(len[supp=="OJ"], len[supp=="VC"]))
```

```
##
##  Welch Two Sample t-test
##
## data:  len[supp == "OJ"] and len[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
```

The resulting p-value is `0.06063`. Though it is very close to `0.05` (most common alpha-level), the difference between the means is **not** significant. So, **delivery method doesn't affect tooth growth in guinea pigs**.

Tooth growth by supplement type and dosage

Now, let us compare tooth growth by delivery method (`supp`) in groups with different dosage:

```
## [1] 0.006358607
```

```
## [1] 0.001038376
```

```
## [1] 0.9638516
```

As we can see, delivery method **does affect** tooth growth in guinea pigs when dosage is 0.5 or 1, with **orange juice causing more tooth growth**. However, **when dosage is 2 miligrams, orange juice does not significantly differ from ascorbic acid**. If we take into consideration the fact that this dosage causes largest tooth growth, these two supplement types are equally successful.

Appendix

```
library(ggplot2)
library(dplyr)
options(scipen=99)
library(datasets)
tg <- ToothGrowth
str(tg)
```

```
ggplot(data = tg, aes(supp, len)) + geom_boxplot(aes(fill = supp)) + ggtitle("Tooth growth by supplement
```

```
ggplot(data = tg, aes(dose, len)) + geom_boxplot(aes(fill = dose)) + ggtitle("Tooth growth by dose") + y
```

```
ggplot(data = tg, aes(dose, len)) + geom_point(aes(colour = supp), size=4) + ggtitle("Tooth growth by do
```

```
with(tg, t.test(len[dose=="0.5"], len[dose=="1"]))$p.value
with(tg, t.test(len[dose=="1"], len[dose=="2"]))$p.value
with(tg, t.test(len[dose=="0.5"], len[dose=="2"]))$p.value
```

```
with(tg, t.test(len[supp=="OJ" & dose=="0.5"], len[supp=="VC" & dose=="0.5"]))$p.value
with(tg, t.test(len[supp=="OJ" & dose=="1"], len[supp=="VC" & dose=="1"]))$p.value
with(tg, t.test(len[supp=="OJ" & dose=="2"], len[supp=="VC" & dose=="2"]))$p.value
```