Statistical Inference Course Project Pt.2

Omer Yavin

March 3, 2018

Basic inferential data analysis

In this part the focus will be on analyzing actual data found in the R database. Namely - ToothGrowth data set. From the description of the database:

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

The following code loads the data set and uses some known 'tricks' to review how it is built and how it behaves.

```
tg_data<-ToothGrowth

#Dataset dimentions
dim(tg_data)
```

```
## [1] 60  3
```

```
#First few lines
head(tg_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
```

```
#General review of the data
summary(tg_data)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   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
```

All this allows some questions to be raised regarding the data at hand.

Questions

1. How does vitamin C dosages affect tooth growth?

2. Does the method of delivery of vitamin C affect tooth growth?
3. Does supplement type affect the answer to Q1? Does dose level affect the answer to Q2?

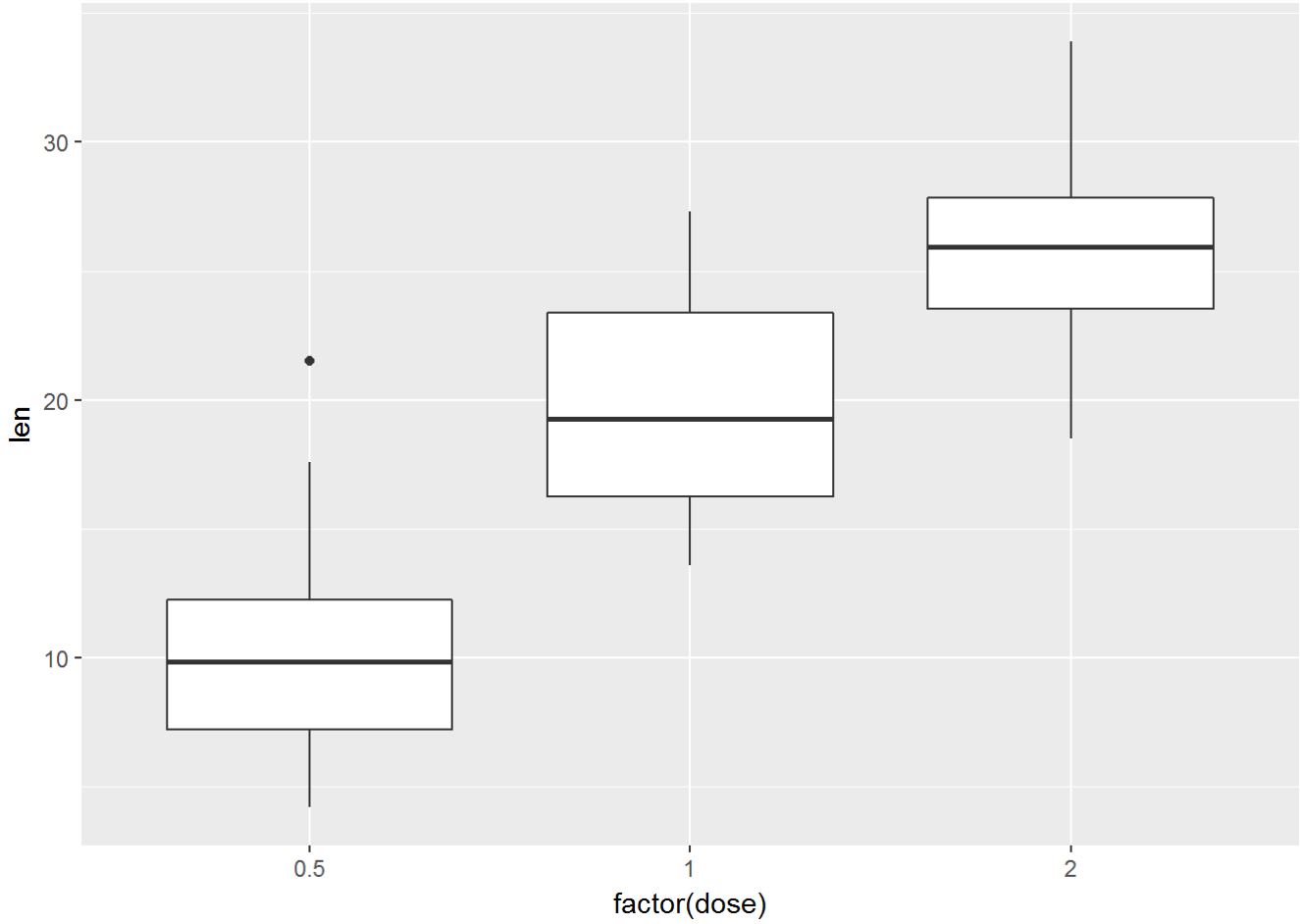
There are probably more questions that can be asked. The following analysis will focus on these questions and try to give some answers.

In order to believe in the ability to answer these questions with the data provided, some assumptions are made. These will be discussed in the end.

Dosage

A box plot can visualize the effect of vitamin C dosage on tooth growth.

```
#Convert dose to factor
#Plot len vs dose
p<-ggplot(data=tg_data,aes(x=factor(dose),y=len))
p+geom_boxplot()
```



An upwards trend is clearly visible here, with the increase of vitamin C dosage. This will be further examined by setting two separate null hypotheses and testing them each against their respective alternative hypotheses.

Some definitions to start:

- $\mu_{0.5}$ - the mean length of odontoblasts for a dose of 0.5 milligrams/day of vitamin C.
- $\mu_{1.0}$ - the mean length of odontoblasts for a dose of 1.0 milligrams/day of vitamin C.
- $\mu_{2.0}$ - the mean length of odontoblasts for a dose of 2.0 milligrams/day of vitamin C.

The null hypotheses:

1. $H_{0_1} : \mu_{1.0} = \mu_{0.5}$

2. $H_{0_2} : \mu_{1.0} = \mu_{2.0}$

The respective alternative hypotheses defined below.

1. $H_{1_1} : \mu_{1.0} > \mu_{0.5}$
2. $H_{1_2} : \mu_{1.0} < \mu_{2.0}$

Tests will utilize the one sided t-test.

Test 1:

```
##
##  Welch Two Sample t-test
##
## data:  c(tg_data$len[tg_data$dose == 1]) and c(tg_data$len[tg_data$dose == 0.5])
## t = 6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

Test 2:

```
##
##  Welch Two Sample t-test
##
## data:  c(tg_data$len[tg_data$dose == 1]) and c(tg_data$len[tg_data$dose == 2])
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

In both cases the null hypothesis is rejected. This means there is enough evidence to believe with at least 95% certainty that $\mu_{1.0} > \mu_{0.5}$ and that $\mu_{1.0} < \mu_{2.0}$. Combined together there is a clear and proven upwards trend in odontoblast length with the increase* of vitamin C dosage.

Method of delivery

Now the general affect of just the type of supplement the vitamin C was delivered by will be examined.

First it's important to show that tests are even, meaning - each supplement type was tested equal number of times for each dosage. Otherwise results could be skewed.

We could count, but we're far too sophisticated (and lazy) for that… :)

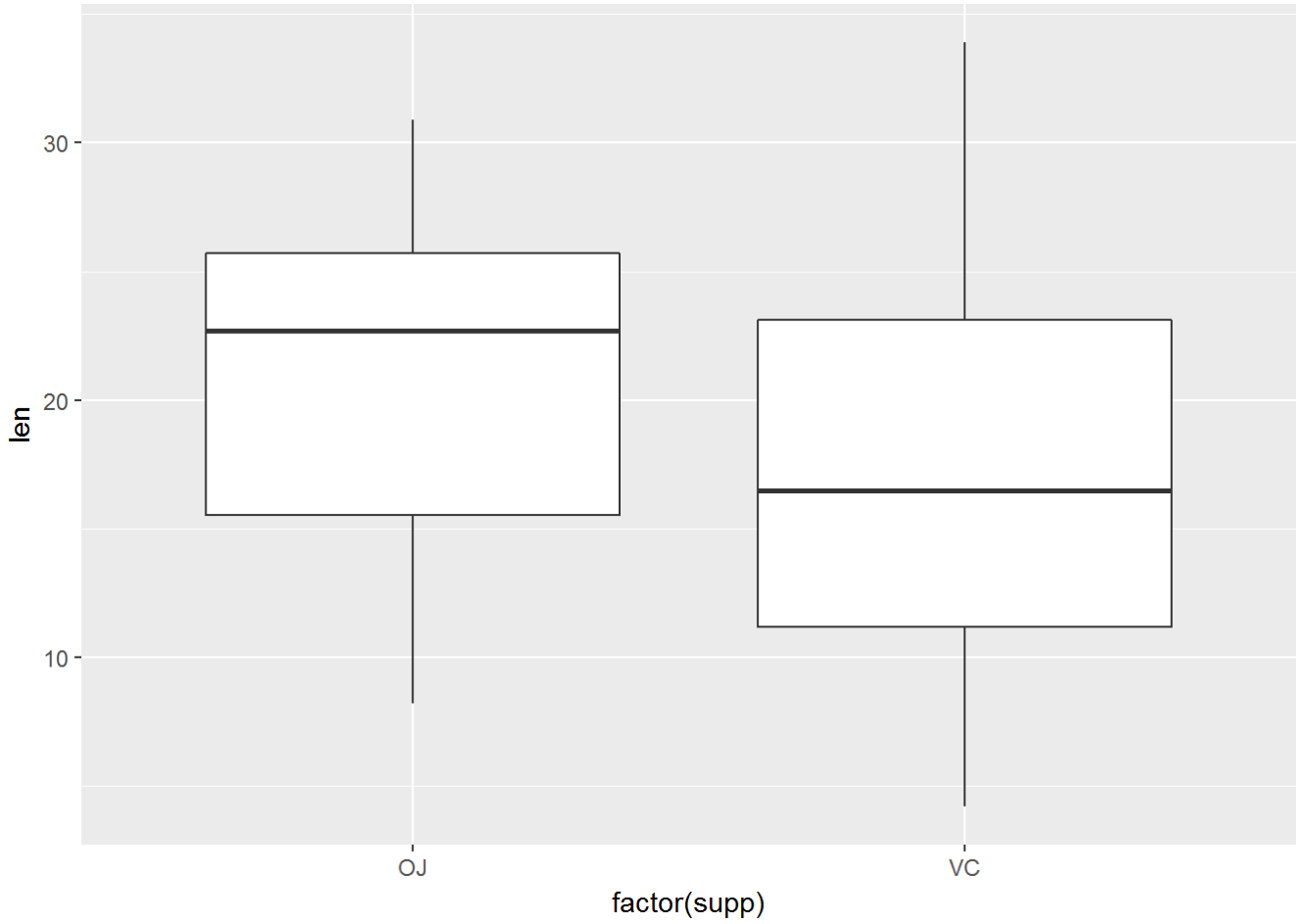
```
tg_temp<-tg_data
tg_temp$ones <- rep(1,nrow(tg_temp))
dcast(tg_temp,supp~dose,value.var = "ones",fun.aggregate = sum)
```

```
##   supp 0.5  1  2
## 1   OJ  10 10 10
## 2   VC  10 10 10
```

Indeed each supplement type (OJ/VC) was tested for each level of dosage an equal number of 10 times. Now for the question…

Does the supplement type affect the level of tooth growth for any dosage of vitamin C?

```
p<-ggplot(data=tg_data,aes(x=factor(supp),y=len))
p+geom_boxplot()
```



Odontoblast length is slightly bigger for OJ when compared to ascorbic acid as a form of delivery for vitamin C, but results seem to overlap. Tests to follow.

Some definitions:

- μ_{OJ} - the mean length of odontoblasts given orange juice as the form of delivery for vitamin C.
- μ_{VC} - the mean length of odontoblasts given ascorbic acid as the form of delivery for vitamin C.
- $H_0 : \mu_{OJ} = \mu_{VC}$
- $H_1 : \mu_{OJ} > \mu_{VC}$

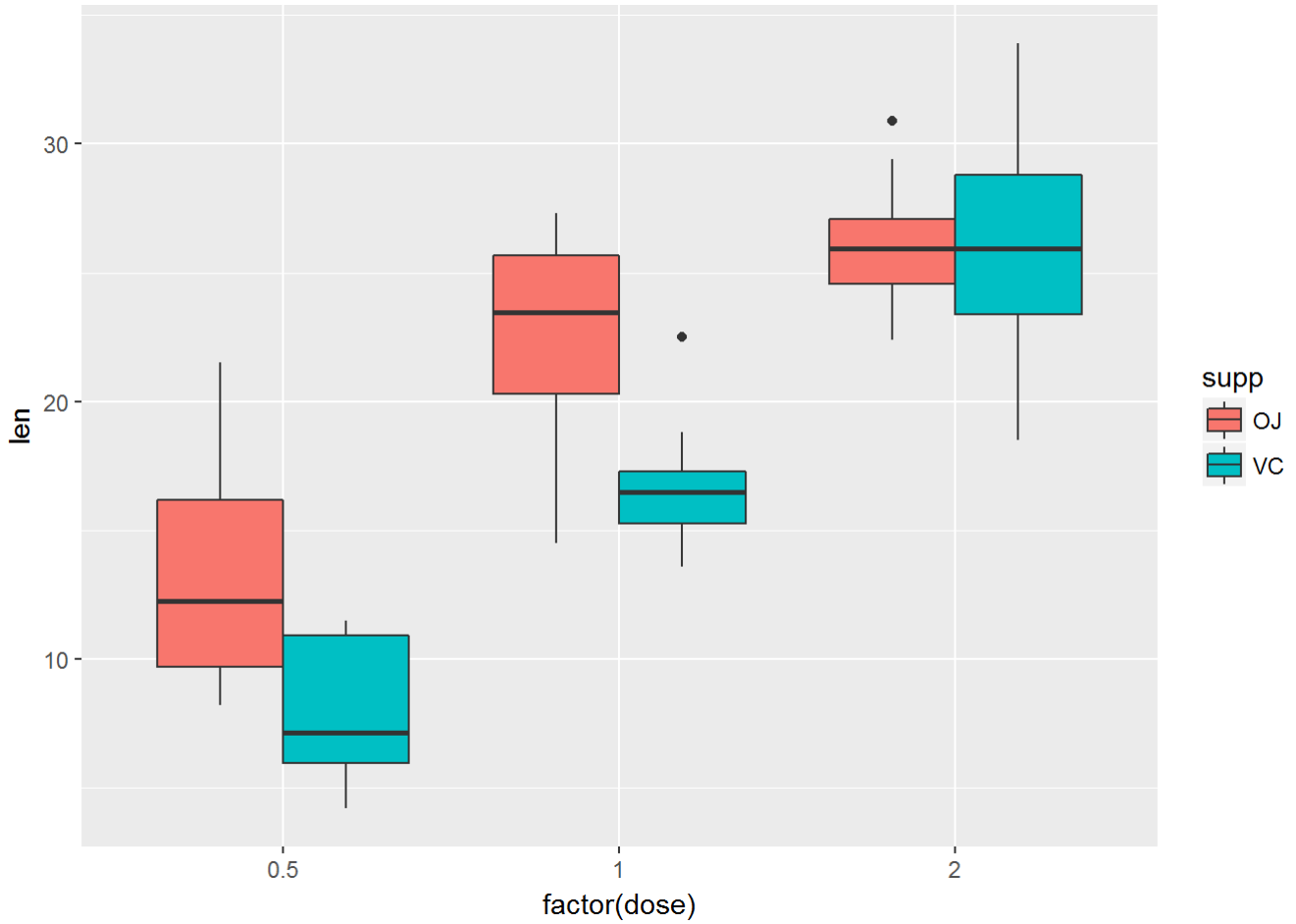
```
##
##  Welch Two Sample t-test
##
## data:  c(tg_data$len[tg_data$supp == "OJ"]) and c(tg_data$len[tg_data$supp == "VC"])
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

From the actual t.test it can be concluded that indeed there is evidence to support, with at least 95% confidence that the effect of delivering vitamin C through orange juice is more significant to teeth growth than using ascorbic acid as a supplement. Note that $p - value \approx 0.05$, but it is still smaller.

Dose and supplement type seperate examination

To see the separate affect of supplement type for each dosage and vise versa, a slightly more complex box plot will be formed by plotting for dose levels, then separating each plot by supplement type.

```
p<-ggplot(data=tg_data,aes(x=factor(dose),y=len,fill=supp))
p+geom_boxplot()
```



For doses of 0.5mg/day and 1.0mg/day results are as can be expected after answering Q 1 and 2 - as dose goes up, odontoblasts length goes up; per given dosage OJ grants higher values of odontoblasts length. For a vitamin C dose of 2.0mg/day it seems the effect of supplement type is eliminated. For ascorbic acid there is a higher variance in results, but the means look about the same.

To verify this theory a hypothesis will be defined, then a t.test performed over appropriately filtered data.

- $\mu_{2.0_{OJ}}$ - the mean length of odontoblasts given orange juice as the form of delivery for vitamin C.
- $\mu_{2.0_{VC}}$ - the mean length of odontoblasts given ascorbic acid as the form of delivery for vitamin C.
- $H_0 : \mu_{2.0_{OJ}} = \mu_{2.0_{VC}}$
- $H_1 : \mu_{2.0_{OJ}} \neq \mu_{2.0_{VC}}$

```
##
##  Welch Two Sample t-test
##
## data:  c(tg_data_2.0$len[tg_data_2.0$supp == "OJ"]) and c(tg_data_2.0$len[tg_data_2.0$supp
== "VC"])
## 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 of x mean of y
##     26.06     26.14
```

Zero (and the actual diff in means) is well within the 95% confidence interval, and also $p - value \approx 1$. We fail to reject the null hypothesis, meaning for a dose of 2.0 mg/day, the type of supplement used to deliver vitamin C is insignificant.

Conclusions

To conclude an answer for each question presented:

1. For the 3 dosages measured we found a positive correlation between a larger dosage of vitamin C and the length of odontoblasts, signifying elevated tooth growth.
2. It was found that generally, using orange juice as the supplement to deliver vitamin C gives stronger results in terms of tooth growth when compared to ascorbic acid.
3. It was found that for a dose of 2.0 there is no evidence that OJ is preferable to ascorbic acid in terms of the mean length of odontoblasts**. If results for higher dosages were available, it would be interesting to see if the effect of the the supplement reverses or remains equal (or, in fact, if a higher dosage has a reversed effect on tooth growth all together).

Assumptions

When answering these questions it was assumed that 10 tests of each type is a significant enough amount to make any conclusions (the sample is representative of the entire population), and that they are independent samples.. It was also assumed that, between the various dosage levels and types of supplements, the guinea pigs tested were similar enough in terms of the factors contributing to odontoblast length (size, weight, possibly age) and, of course, that they had about the same length of adontoblasts at the beginning of the experiment.

Of course, link between odobtoblast length and tooth growth is taken for granted but there was an effort made to refer to the prior rather than the latter thus disregarding any uncertainties in terms of this link.

(*) - This is true strictly within the values tested. Maybe for some higher dosage there will be a decline but we haven't the data to support (or test) this.

(**) - There is is still some upside to using OJ as the supplement since it gives less varying results, but the mean result is the same.