

Toothgrowth Analysis

Vadzim Dabravolski

July 26, 2015

This is a report generated for the Statistic Inference course. The sources of calculations are located in [author's git repo](#).

TASK 1: “Do some exploratory data analysis”

First of all, it's important to learn a bit about the data sample. For this one can run following simple command in R:

```
summary(ToothGrowth)
head(ToothGrowth)
```

Let's try to do some manipulations with data and group it by **dose** and **delivery method**:

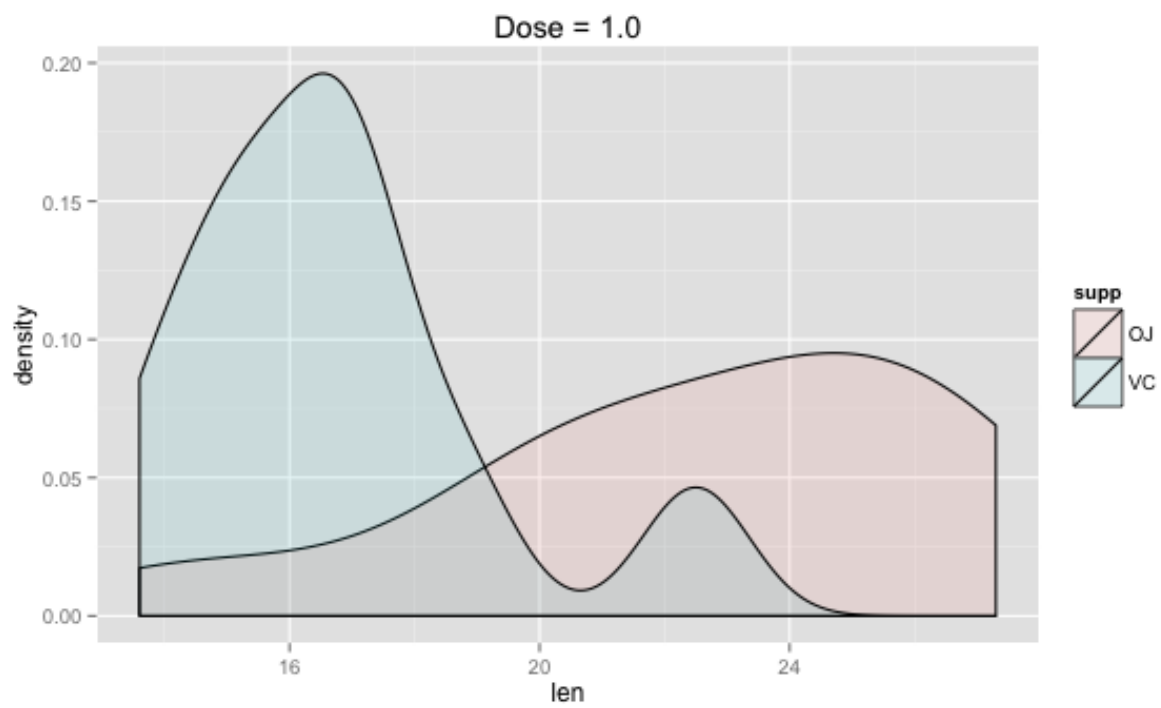
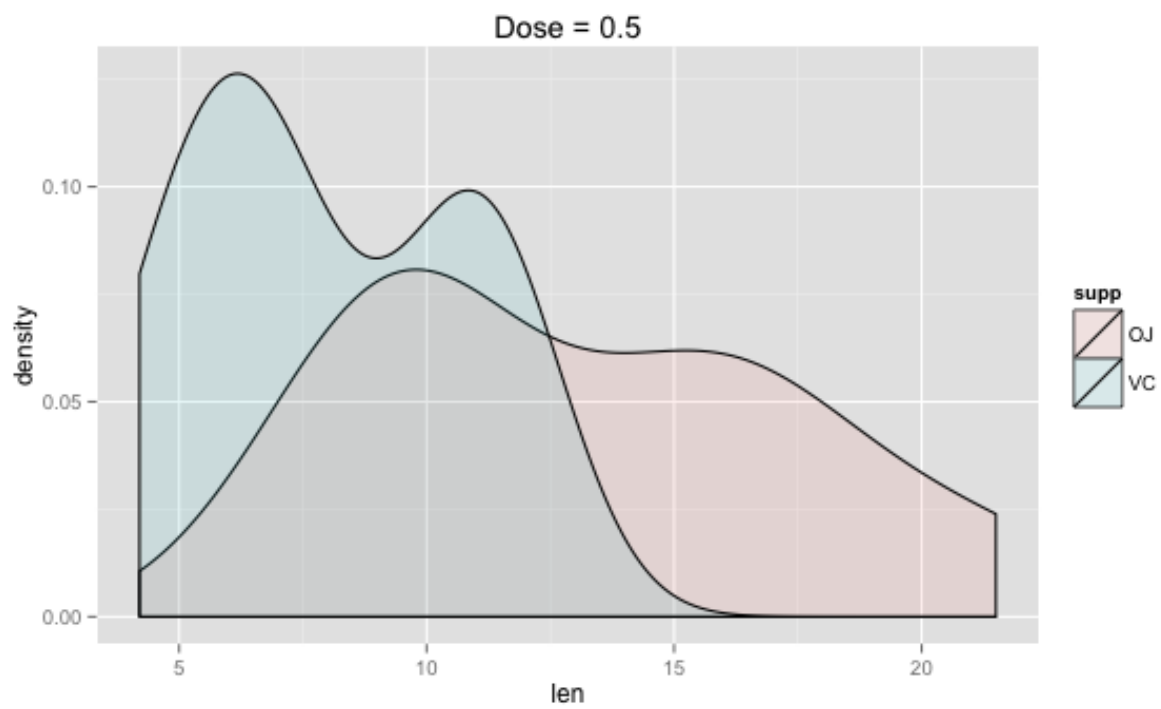
```
tg <- data.frame(ToothGrowth)
#Group by delivery method
tg_oj <- filter(tg, supp=="OJ")
tg_vc <- filter(tg, supp=="VC")

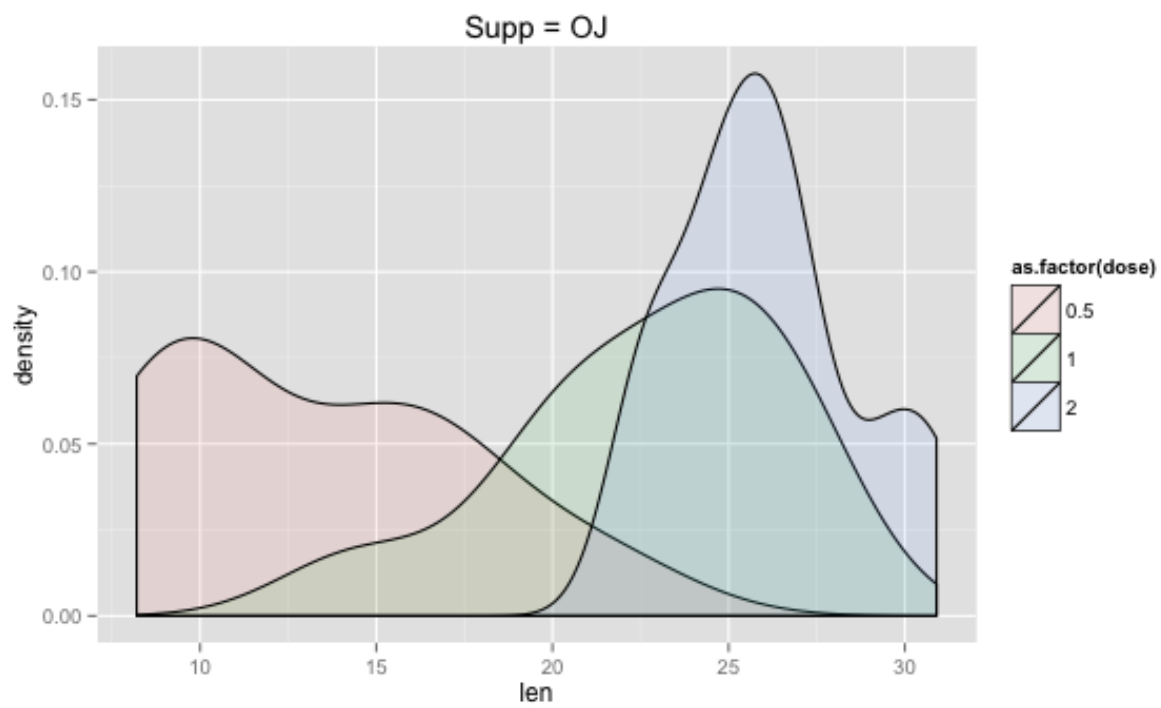
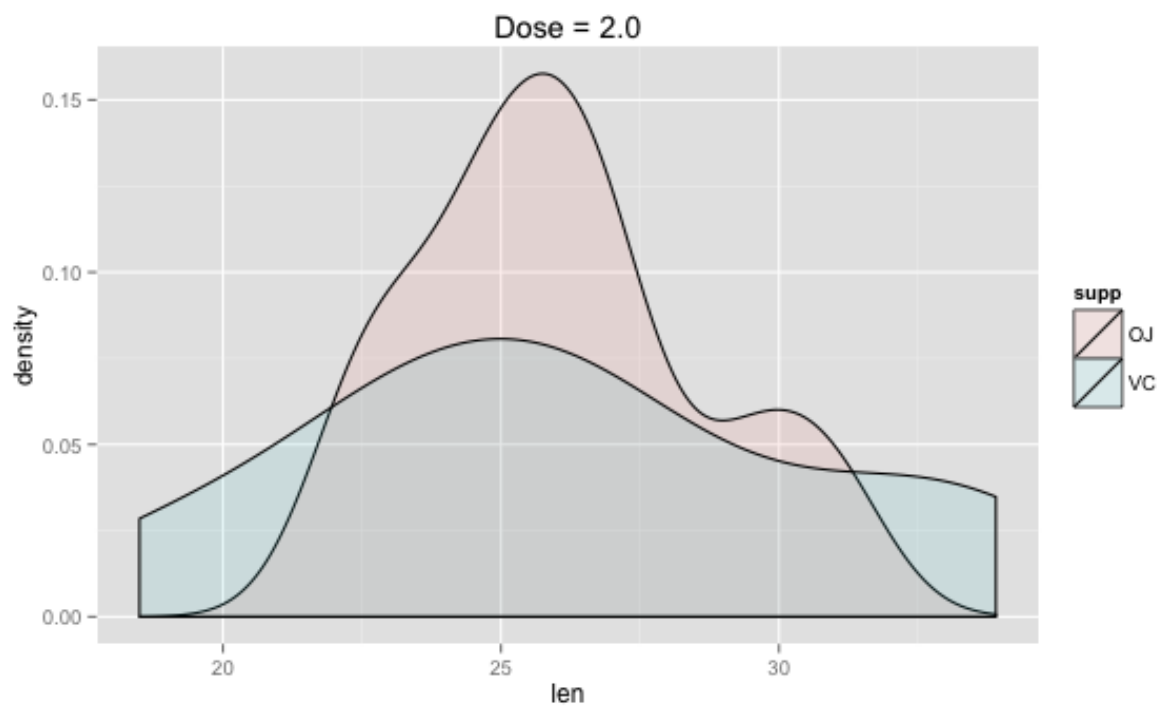
#Group into subsets with equal dose
tg_equal_dose_1 <- rbind(filter(tg_oj,dose==0.5),filter(tg_vc,dose==0.5))
tg_equal_dose_2 <- rbind(filter(tg_oj,dose==1.0),filter(tg_vc,dose==1.0))
tg_equal_dose_3 <- rbind(filter(tg_oj,dose==2.0),filter(tg_vc,dose==2.0))
```

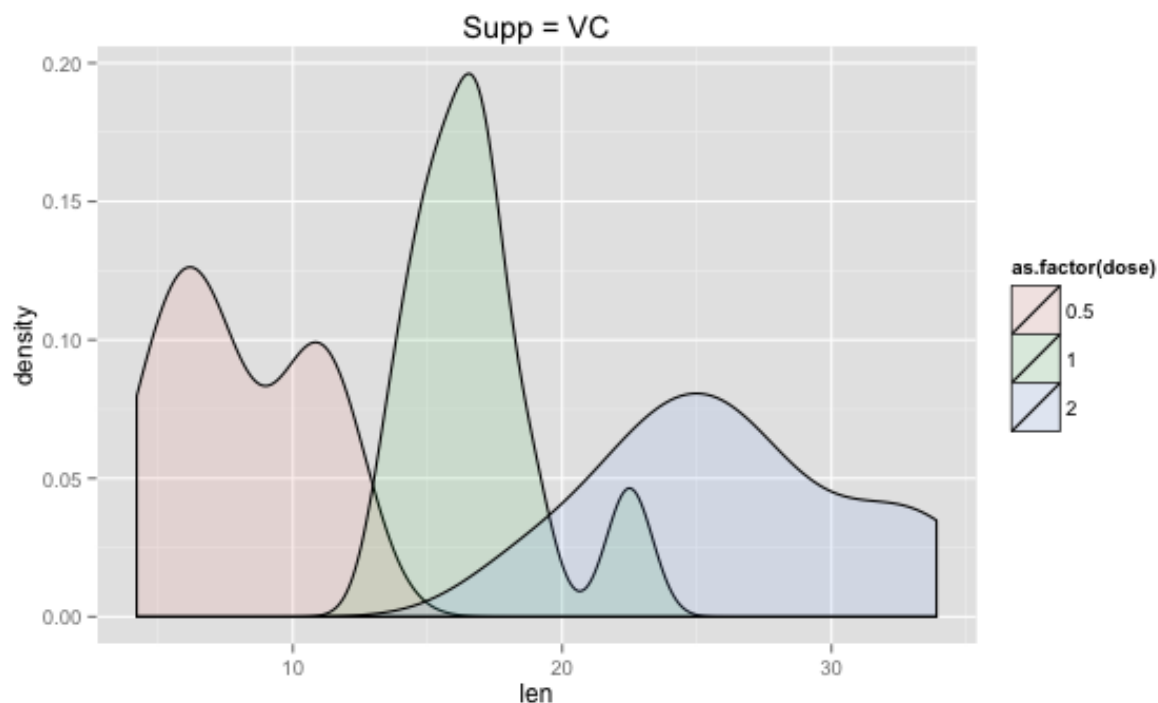
Now let's try to plot some data using following code:

```
ggplot(tg_oj, aes(len,fill=as.factor(dose))) + geom_density(alpha = 0.1) +ggtitle("Supp = OJ")
ggplot(tg_vc, aes(len,fill=as.factor(dose))) + geom_density(alpha = 0.1)+ggtitle("Supp = VC")
ggplot(tg_equal_dose_1, aes(len,fill=supp)) + geom_density(alpha = 0.1)+ggtitle("Supp = OJ") +ggtitle("Dose = 0.5")
ggplot(tg_equal_dose_2, aes(len,fill=supp)) + geom_density(alpha = 0.1)+ggtitle("Supp = OJ") +ggtitle("Dose = 1.0")
ggplot(tg_equal_dose_3, aes(len,fill=supp)) + geom_density(alpha = 0.1)+ggtitle("Supp = OJ") +ggtitle("Dose = 2.0")
```

Let's take a look on the plots.







TASK 2: “Calculate confident intervals”

To calculate intervals of confidence I made a hypothesis that distribution of tooth for two different delivery methods (“OJ” and “VC”) for the equal doses (0.5, 1.0 and 2.0). Let’s plug it into R code:

```
# confidence interval for dose=0.5 for two distribution of length grouped by delivery method
conf_1 <- t.test(len ~ supp, paired=FALSE, var.equal = TRUE, data=tg_equal_dose_1)$conf
# similar computation as above but for dose=1.0
conf_2 <- t.test(len ~ supp, paired=FALSE, var.equal = TRUE, data=tg_equal_dose_2)$conf
# similar computation as above but for dose=2.0
conf_3 <- t.test(len ~ supp, paired=FALSE, var.equal = TRUE, data=tg_equal_dose_3)$conf
```

The following confidence levels for original hypothesis are calculated:

- for dose=0.5 confidence interval is [1.770262 8.729738] with probability=95%
- for dose=1.0 confidence interval is [2.840692 9.019308] with probability=95%
- for dose=2.0 confidence interval is [-3.722999 3.562999] with probability=95%

ASSUMPTIONS

In the series of calculations I assumed that:

- the distributions for dose and delivery method (supp) are normal and independent.