Coursera: Statistical Inference - Course Project

Konstantin Greger

2015/01/24

Part 2: Analysis of the “ToothGrowth”" Data in the R “datasets”" Package

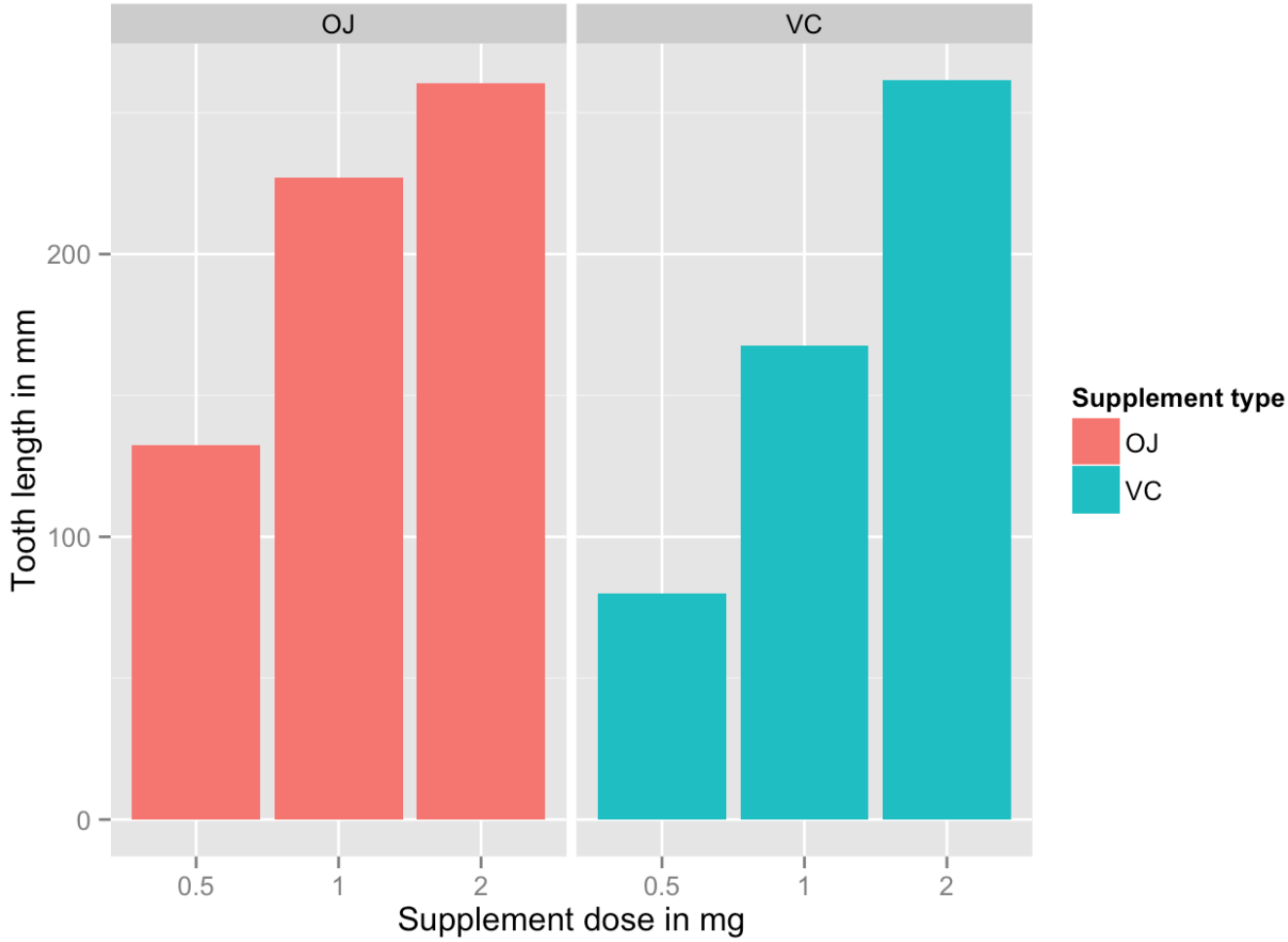
```
# load the data
library(datasets)
tooth = ToothGrowth
```

The “ToothGrowth” data in the R “datasets” package consists of 60 observations. According to the official description of the dataset (https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html) it provides “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).”

Tooth length varies between 8.2 and 30.9 mm for the celivery using orange juice, and between 4.2 and 33.9 mm for the delivery using ascorbic acid.

The following plot shows the data distribution over delivery methods and supplement doses:

```
# plot the distributions
library(ggplot2)
ggplot(data = tooth,
       aes(x = as.factor(dose),
           y = len,
           fill = supp)) +
    geom_bar(stat = "identity") +
    facet_grid(. ~ supp) +
    xlab("Supplement dose in mg") +
    ylab("Tooth length in mm") +
    guides(fill = guide_legend(title = "Supplement type"))
```



The data show a clear positive correlation between tooth length and the dose levels of Vitamin C for both delivery methods. The supplement dose appears to be of greater importance when the supplement is delivered as ascorbic acid as compared to orange juice.

The null hypothesis \(H_0\) is that the supplement type of vitamin C generally does not have an effect on the tooth length.

```
# check for differences due to different supplement type
# assumption: unequal variances between the two groups
t.test(len ~ supp,
       data = tooth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean in group OJ mean in group VC
##            20.66            16.96
```

With a p-value of 0.06 and a confidence interval containing zero we can not reject the null hypothesis \(H_0\) that the different supplement types have no effect on tooth length.

It is worthwhile to perform pairwise comparisons between the different supplement doses:

```
# create subsets for the supplement dose level pairs
tooth_0.5_1.0 <- subset(tooth, dose %in% c(0.5, 1.0))
tooth_0.5_2.0 <- subset(tooth, dose %in% c(0.5, 2.0))
tooth_1.0_2.0 <- subset(tooth, dose %in% c(1.0, 2.0))
```

```
# check for differences due to different dose levels: 0.5, 1.0
# assumption: unequal variances between the two groups
t.test(len ~ dose,
       data = tooth_0.5_1.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.477, df = 37.99, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.984  -6.276
## sample estimates:
## mean in group 0.5   mean in group 1
##             10.61             19.73
```

```
# check for differences due to different dose levels: 0.5, 2.0
# assumption: unequal variances between the two groups
t.test(len ~ dose,
       data = tooth_0.5_2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.8, df = 36.88, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.16 -12.83
## sample estimates:
## mean in group 0.5   mean in group 2
##             10.61             26.10
```

```
# check for differences due to different dose levels: 1.0, 2.0
# assumption: unequal variances between the two groups
t.test(len ~ dose,
       data = tooth_1.0_2.0)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.901, df = 37.1, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996 -3.734
## sample estimates:
## mean in group 1 mean in group 2
##           19.73           26.10
```

For all three dose level pairs (0.5 and 1.0 mg, 0.5 and 2.0 mg, 1.0 and 2.0 mg), the p-values are smaller than 0.05 (0, 0.36, and 0.08, respectively), and none of the confidence intervals does contain zero. Therefire we can reject the null hypothesis \(H_0\) and establish that increasing the dose level leads to an increase in tooth length.

All analyses were performed under the assumption of a random assignment of guinea pigs to different dose level categories and supplement types, and also the representativeness of the samples.