Statistical inference course project part II

Halley Wang

October 25, 2015

This report explains how the hypothesis test (t-test in this case) shows the relationship between response variables (len) vs. predictor variables (supp and dose).

```
## Loading the necessary packages
library(ggplot2)
library(dplyr)
```

```
##
## 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
```

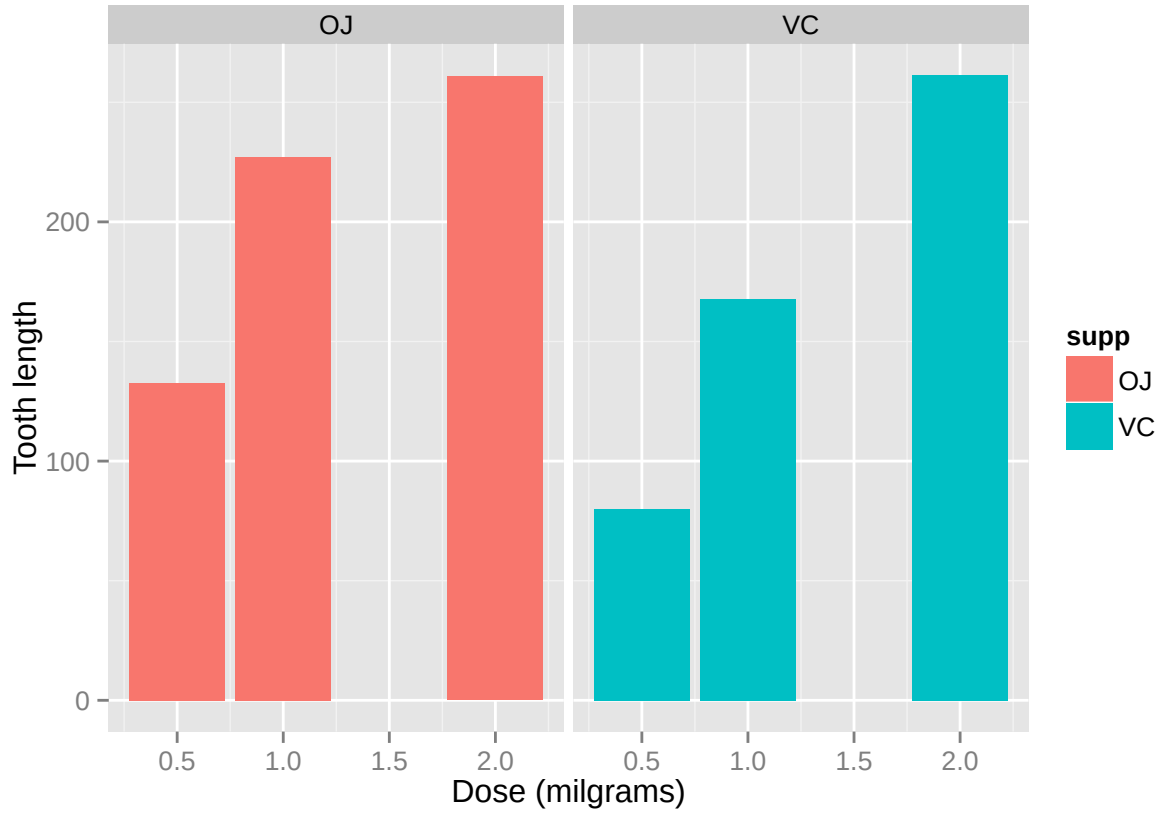
```
library(datasets)
```

```
## Summary of the data:
summary(ToothGrowth)
```

```
##       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
```

```
## make a plot of the data with respect to different supplement:
g2 = ggplot(data = ToothGrowth, aes(x = dose, y = len, fill = supp))


## Notice that a single predictor might correspond to many response values. In this case, we use the id
#g2 + geom_point() + facet_grid(. ~ supp)
g2 + geom_bar(stat = "identity") + facet_grid(. ~ supp) + labs(x = "Dose (milgrams)", y = "Tooth length
```



As we can see from the graph, the tooth length is positively related with the dose of supplement, d

```
## Plot the 2-variables linear regression model:
lmd = lm(len ~ dose + supp, data = ToothGrowth)
lmdSummary = summary(lmd)
lmdSummary
```

```
##
## Call:
## lm(formula = len ~ dose + supp, data = ToothGrowth)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.600 -3.700  0.373  2.116  8.800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.2725     1.2824   7.231 1.31e-09 ***
## dose          9.7636     0.8768  11.135 6.31e-16 ***
## suppVC       -3.7000     1.0936  -3.383   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.236 on 57 degrees of freedom
## Multiple R-squared:  0.7038, Adjusted R-squared:  0.6934
## F-statistic: 67.72 on 2 and 57 DF,  p-value: 8.716e-16
```

By assuming the true length of the teeth is linear related with the supplement and its dose, we mad
As we can see from the summary, the intercept = 9.2725 represents the teeth length at no supplement
When we are doing the hypothesis test. H0: the teeth length has no relation with the dose intake. v
Analogously for suppVC.

```
c.i = confint(lmd)
c.i
```

```
##                 2.5 %    97.5 %
## (Intercept)  6.704608 11.840392
## dose         8.007741 11.519402
## suppVC      -5.889905 -1.510095
```

At default (95%) confident interval, we say that the true intercept will fall in between 6.7046085 a