Statistical Inference Peer Assessment - Part 2

Paulo Viana

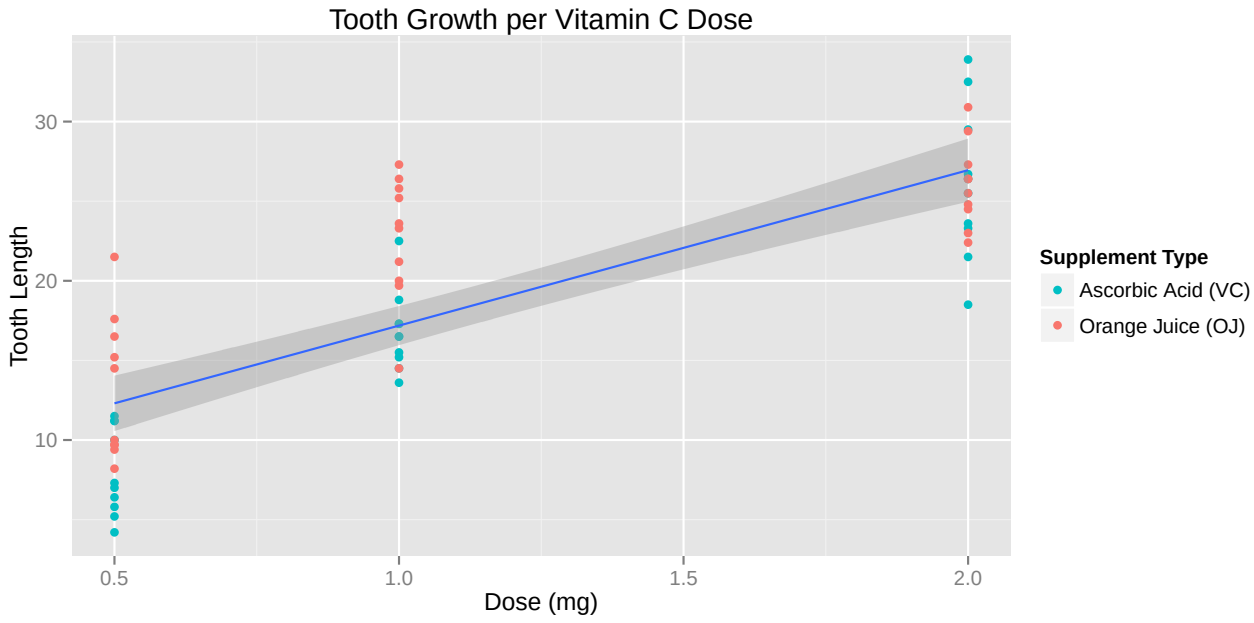
In the second part of the assessment we will analyze the ToothGrowth dataset. The dataset show the effect of vitamin C on tooth growth in guinea pigs, and is described as "the response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid)".

The first step is to load the dataset and do some exploratory analysis.

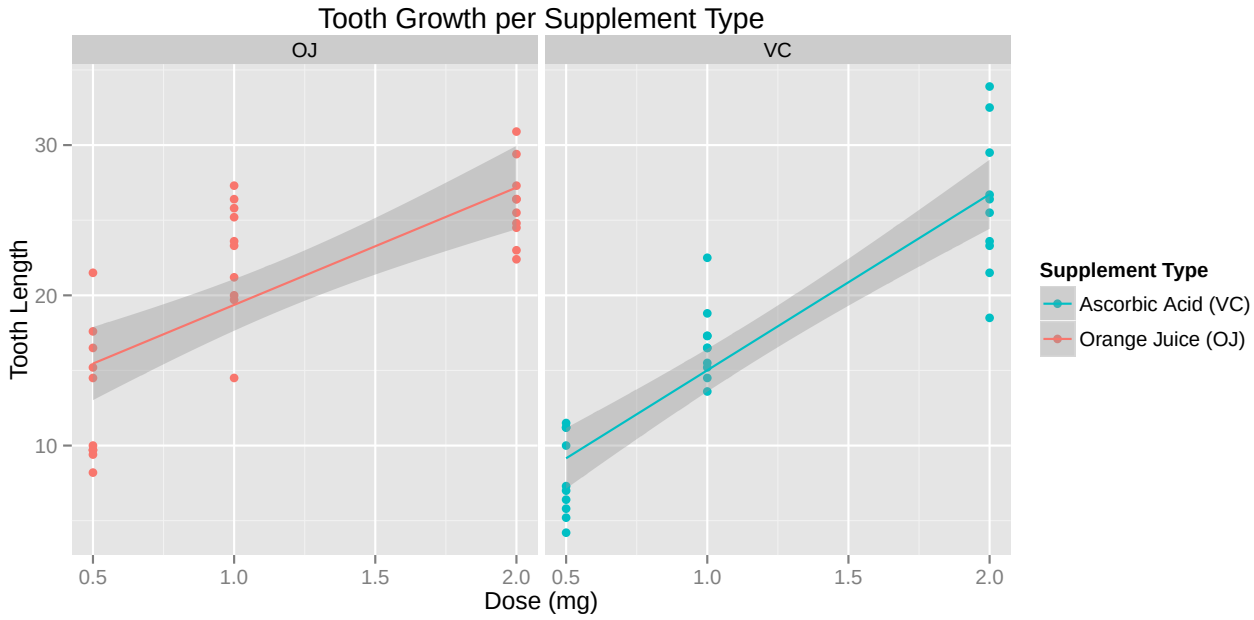
```
library(ggplot2)
data(ToothGrowth)

g <- ggplot(ToothGrowth, aes(x = dose, y = len)) +
    xlab('Dose (mg)') + ylab('Tooth Length') +
    geom_point(aes(colour = supp)) +
    scale_colour_discrete(name = 'Supplement Type',
                          breaks = c('VC', 'OJ'),
                          labels = c('Ascorbic Acid (VC)', 'Orange Juice (OJ)'))


g + stat_smooth(method='lm') +
    ggtitle('Tooth Growth per Vitamin C Dose')
```




```
g + stat_smooth(aes(group = supp, colour = supp), method='lm') +
    facet_grid(. ~ supp) + ggtitle('Tooth Growth per Supplement Type')
```



Both graphs show a clear positive correlation between the vitamin C intake and the tooth length, for both methods. Let's examine this relation more closely.

```
fit.supp <- lm(len ~ dose + supp, data = ToothGrowth)

summary(fit.supp)
```

```
##
## 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
```

We'll begin taking a look at the coefficients. The intercept states that the average value of tooth length for pigs not taking any form of vitamin C is 9.2725. The second, dose, points that the average increase in tooth length per increased *mg* of vitamin intake, *ceteris paribus*, is 9.7635714 units. The last, suppVC, relates to the difference of the tooth length averages for each supplement. This means that, all else equal, the average tooth length for a pig receiving ascorbic acid will be 3.7 units smaller than the average length for a pig receiving the same dosage of orange juice.

The coefficient of determination R^2 is equal to 0.7037969 indicating that our model explains approximately 70% of the variation of the data. The p-values are < 0.05, so we can accept the hypothesis that the supplements are reponsible for the tooth growth of the pigs.

Our last step is to check our confidence intervals:

```
confint(fit.supp)
```

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

If we collect new data and repeat the simulations, the coefficients will be within these bounds 95% of the time.