Statistical Inference Course Project[Part 2]

Sriram Vadlamani

In the second part of the project, we analyze the `ToothGrowth` data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
#Load the libraries
library(datasets)
library(ggplot2)

#Explore the data with str and summary
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

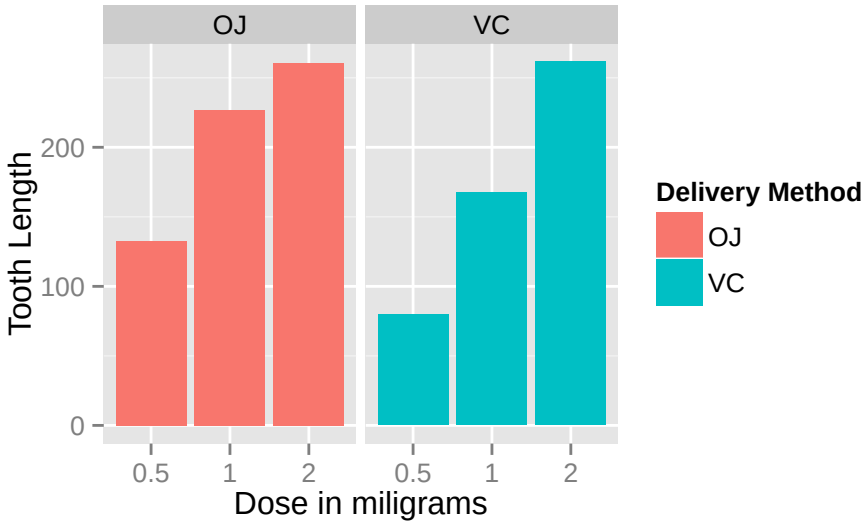
```
summary(ToothGrowth)
```

```
##       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
```

2. Provide a basic summary of the data.

ToothGrowth dataset has 60 observations for 3 categories - length, supplement and dose. Possible dose level values are 0.5, 1, 2. Supplement or the delivery methods are orange juice or ascorbic acid.

```
#Plot the data
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
    geom_bar(stat="identity",) +
    facet_grid(. ~ supp) +
    xlab("Dose in miligrams") +
    ylab("Tooth Length") +
    guides(fill=guide_legend(title="Delivery Method"))
```



From the picture above, there is a clear correlation between the tooth length and the dose levels of Vitamin C, for both delivery methods - Orange Juice and Ascorbic Acide.

The effect of the dose can also be identified using regression analysis.

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

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

```
##
## 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
```

The model explains 70.38% of the variance in the data.

1. The intercept is 9.2725, which means with no supplement of Vitamin C, the average tooth length would be 9.2725 units.

2. The coefficient of `dose` is 9.7635714. Increasing the delievered dose by 1 mg, everything else being equal, would increase the tooth length 9.7635714 units.

3. The last coefficient is for the supplement type. Since the supplement type is a categorical variable, dummy variables are used.

4. The computed coefficient is for `suppVC` and the value is -3.7 which means delivering a given dose as ascorbic acid, without changing the dosage, would result in 3.7 units of decrease in the tooth length. Since there are only two categories, we can also conclude that on average, delivering the dosage as orange juice would increase the tooth length by 3.7 units.

4. Conclusions and the assumptions needed for conclusions

95% confidence intervals for two variables and the intercept are as follows.

```
confint(fit)
```

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

From the above results, it's clear that - the three coefficients intercept, dosage and delivery type - have an effect on growth of teeth in guinea pigs as the 95% confidence interval results shows.

Assumptions

1. The distribution is normal
2. The pigs provided is a random sample but is representative of the population.