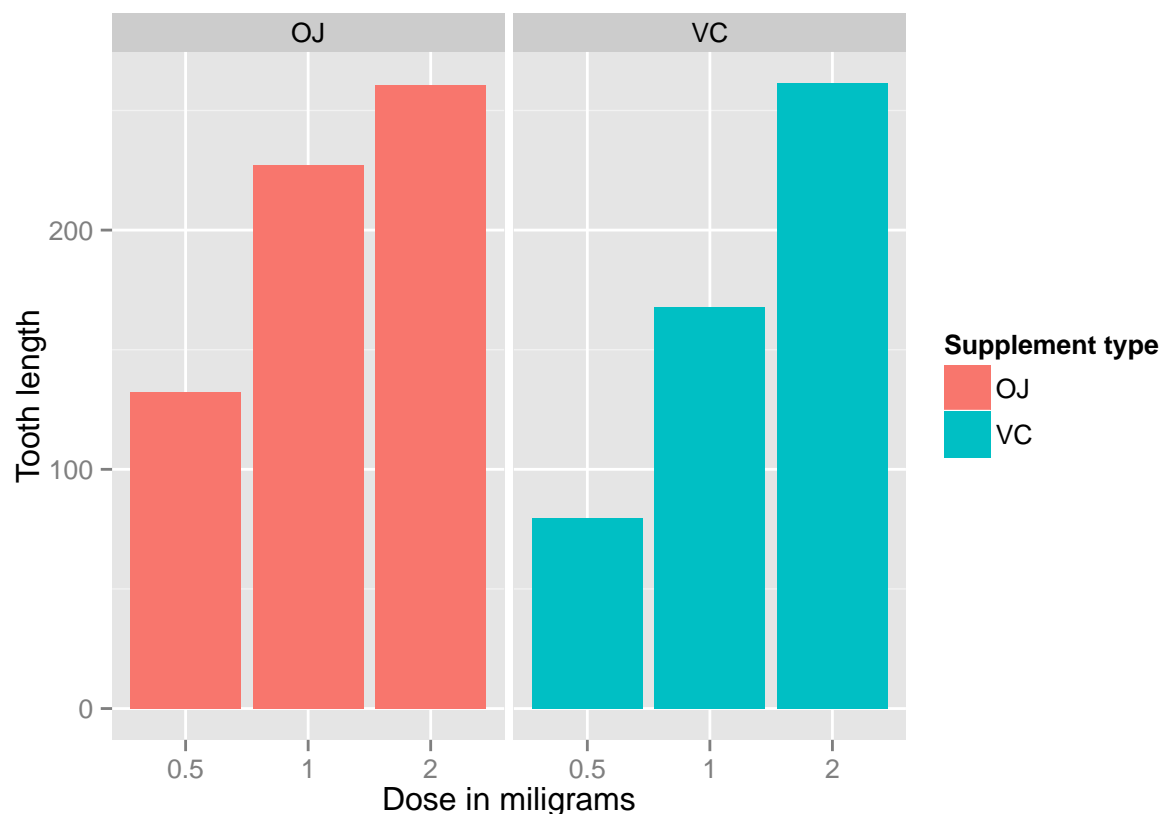


Statistical Inference Course Project, Part 2: Basic Inferential Data Analysis

Sefa Kilic

In the second part of the project, we analyze the `ToothGrowth` data in the R datasets package. The data is set of 60 observations, 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).

```
library(datasets)
library(ggplot2)
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
  geom_bar(stat="identity",) +
  facet_grid(. ~ supp) +
  xlab("Dose in milligrams") +
  ylab("Tooth length") +
  guides(fill=guide_legend(title="Supplement type"))
```



As can be seen above, there is a clear positive correlation between the tooth length and the dose levels of Vitamin C, for both delivery methods.

The effect of the dose can also be identified using regression analysis. One interesting question that can also be addressed is whether the supplement type (i.e. orange juice or ascorbic acid) has any effect on the tooth length. In other words, how much of the variance in tooth length, if any, can be explained by the supplement type?

```
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)    9.273      1.282    7.23  1.3e-09 ***
## dose          9.764      0.877   11.14  6.3e-16 ***
## suppVC        -3.700      1.094   -3.38  0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.24 on 57 degrees of freedom
## Multiple R-squared:  0.704, Adjusted R-squared:  0.693
## F-statistic: 67.7 on 2 and 57 DF, p-value: 8.72e-16
```

The intercept is 9.2725, meaning that with no supplement of Vitamin C, the average tooth length is 9.2725 units. The coefficient of `dose` is 9.7636. It can be interpreted as increasing the delivered dose 1 mg, all else equal (i.e. no change in the supplement type), would increase the tooth length 9.7636 units. The last coefficient is for the supplement type. Since the supplement type is a categorical variable, dummy variables are used. The computed coefficient is for `suppVC` and the value is -3.7 meaning that delivering a given dose as ascorbic acid, without changing the dose, 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.

confidence here