Statistical Inference Course Project-2 (ToothGrowth data)

Chakra C

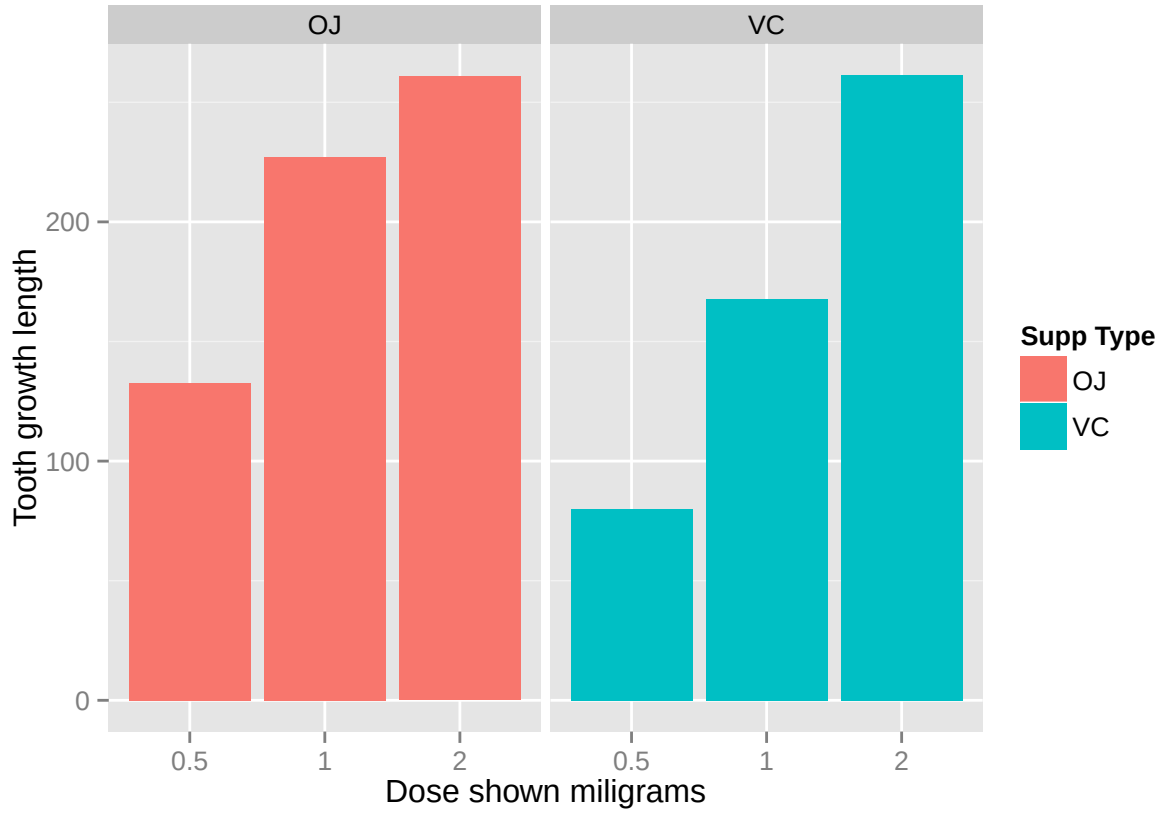
Sunday, May 24, 2015

Statistical Inference Course second Project talks about the ToothGrowth data in the R datasets package. The data anlysis is on 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), and set of 60 observations.

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

```
## Warning: package 'ggplot2' was built under R version 3.1.2
```

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
    geom_bar(stat="Identity",) +
    facet_grid(. ~ supp) +
    xlab("Dose shown miligrams") +
    ylab("Tooth growth length") +
    guides(fill=guide_legend(title="Supp Type"))
```



From the above Graph, in both delivery methods it is showing clear correlation between the Dose levels of Vitamin C and Tooth growth length. Using regression analysis, the effect of dose can be identified. And Graph showing whether the Supplement type (Orange Juice or ascorbic acid) has effect on the Tooth growth length. Similarly the how much variance in Tooth Growth length is also 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)   9.2725     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
```

Which explains 70% of the variance in the data. And the intercept is 9.2725, that means with no supplement type of Vitamin C, the average tooth growth length is 9.2725 units. The coefficient of `Dose` is 9.7635714. This can be interpreted as growing the delievered dose 1 mg, all else equal. which is no change in the supplement type. This will increase the tooth growth length 9.7635714 units. The last coefficient is for the supplement type. It is the supplement type is a categorical variable, empty variables. The calculated coefficient is for `suppVC` and the value is -3.7 that delivering a given dose as ascorbic acid, there is no changing the dose, which result in 3.7 units of reduce in the tooth growth length. But there are only two categories, we can conclude that on average, delivering the dosage as orange juice would increase the tooth growth length by 3.7 units.

So, 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
```

Overall, the confidence intervals mean if we are collecting a different set of data and many time we estimate parameters of the linear model. And 95% of the time, the coefficient estimations will be in these ranges. In each coefficient (intercept, `dose` and `suppVC`), the null hypothesis is that the coefficients are zero, meaning that no tooth growth length variation is explained by that variable. All p-values are less than 0.05, rejecting the null hypothesis and suggesting that each variable explains a significant portion of variability in tooth growth length, assuming the significance level is 5%.