

Statistical Inference Final Project (Part 2)

MP

Sunday, September 27, 2015

Introduction

The ToothGrowth in datasets library shows the effect of vitamin C on Tooth Growth in Guinea Pigs. 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). Here there is a data frame with 60 observations on 3 variables: [1] len numeric Tooth length [2] supp factor Supplement type (VC or OJ) [3] dose numeric Dose in milligrams

```
library(ggplot2)
qplot(len, dose, data = ToothGrowth, color = supp)
```

as shown in figure above there exists high correlation between Dose in milligram and Tooth Length. The summarized characteristic of dataset is as below:

```
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

DataSet could be fitted with GLM (Generalized Linear Model) function. in the Coefficients *Supplementtype* has lower significance against the *Doseinmilligrams* according to P-Value and because of that P-Values are below 0.05 both features are important for prediction of *Toothlength*.

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

##
## Call:
## glm(formula = len ~ dose + supp, data = ToothGrowth)
##
## Deviance Residuals:
```

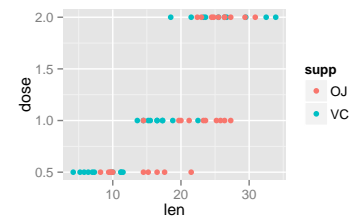


Figure 1: len vs. dose, colored by supp

```
##      Min      1Q  Median      3Q      Max
## -6.600 -3.700  0.373  2.116  8.800
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)   9.2725     1.2824   7.231
## dose          9.7636     0.8768  11.135
## suppVC        -3.7000     1.0936  -3.383
##              Pr(>|t|)
## (Intercept) 1.31e-09 ***
## dose        6.31e-16 ***
## suppVC      0.0013 **
## ---
## Signif. codes:
##  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 17.93956)
##
##      Null deviance: 3452.2  on 59  degrees of freedom
## Residual deviance: 1022.6  on 57  degrees of freedom
## AIC: 348.42
##
## Number of Fisher Scoring iterations: 2

confint(fit)

## Waiting for profiling to be done...

##              2.5 %      97.5 %
## (Intercept)  6.759111 11.785889
## dose         8.045008 11.482135
## suppVC       -5.843425 -1.556575
```

In this study number of observations are less than previous one and so there need t-statistic in this case. Because of uncorrelated variables in Confidence Interval The Paired Argument is FALSE and because of same number of observations the var.equal is TRUE.

```
t.test(ToothGrowth$len, ToothGrowth$dose, paired = FALSE,
       var.equal = TRUE)$conf

## [1] 15.68451 19.60883
## attr(,"conf.level")
## [1] 0.95
```

Conclusion

The goal of this study was Tooth length Prediction with two Supplement type (VC or OJ) and Dose in milligrams features. the P_value on t-statistic show high significance of Dose in milligrams in prediction of Tooth length.