

Statistical Inference - Project Part 2

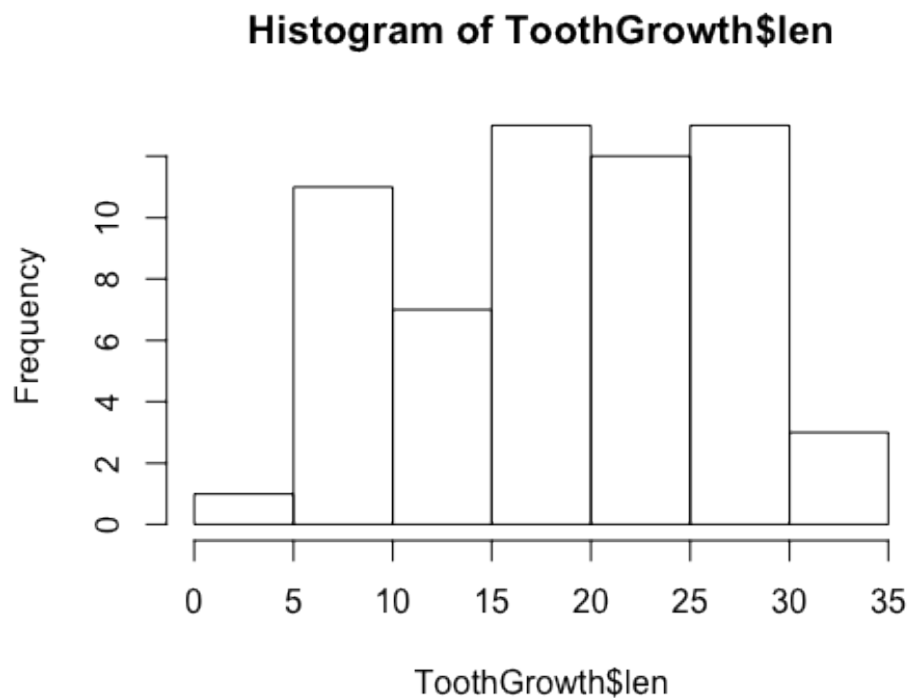
Kari Innes
September 15, 2014

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

```
## '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 ...
```

```
## len supp dose  
## 1 4.2 VC 0.5  
## 2 11.5 VC 0.5  
## 3 7.3 VC 0.5  
## 4 5.8 VC 0.5  
## 5 6.4 VC 0.5  
## 6 10.0 VC 0.5
```

Histogram of tooth length



Please see **Appendix** for additional exploratory graphs.

2. Provide a basic summary of the data

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.  
## 4.2 13.1 19.2 18.8 25.3 33.9
```

Calculate mean length and standard deviation for each supp

```
## supp len  
## 1 OJ 20.66  
## 2 VC 16.96
```

```
## supp len  
## 1 OJ 6.606  
## 2 VC 8.266
```

Calculate mean length and standard deviation for each dose

```
## dose len
## 1 0.5 10.61
## 2 1.0 19.73
## 3 2.0 26.10
```

```
## dose len
## 1 0.5 4.500
## 2 1.0 4.415
## 3 2.0 3.774
```

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

Compare tooth growth by supplement

Null hypotheses: There is no difference between tooth growth under the two supplements

H0: $\mu(OJ) = \mu(VC)$

Ha: $\mu(VC) \neq \mu(OJ)$

```
## [1] -8.1511 0.7511
## attr(,"conf.level")
## [1] 0.975
```

Compare tooth growth by dose

Null hypotheses: There is no difference between tooth growth under the **doses 0.5 and 1.0**

H0: $\mu(0.5) = \mu(1.0)$

Ha: $\mu(1.0) > \mu(0.5)$

```
## [1] 6.276 11.984
## attr(,"conf.level")
## [1] 0.95
```

Null hypotheses: There is no difference between tooth growth under the **doses 0.5 and 2.0**

H0: $\mu(0.5) = \mu(2.0)$

Ha: $\mu(2.0) > \mu(0.5)$

```
## [1] -18.16 -12.83
## attr(,"conf.level")
## [1] 0.95
```

Null hypotheses: There is no difference between tooth growth under the **doses 1.0 and 2.0**

H0: $\mu(1.0) = \mu(2.0)$

Ha: $\mu(2.0) > \mu(1.0)$

```
## [1] -8.996 -3.734
## attr(,"conf.level")
## [1] 0.95
```

4. State your conclusions and the assumptions needed for your conclusions.

Assumptions:

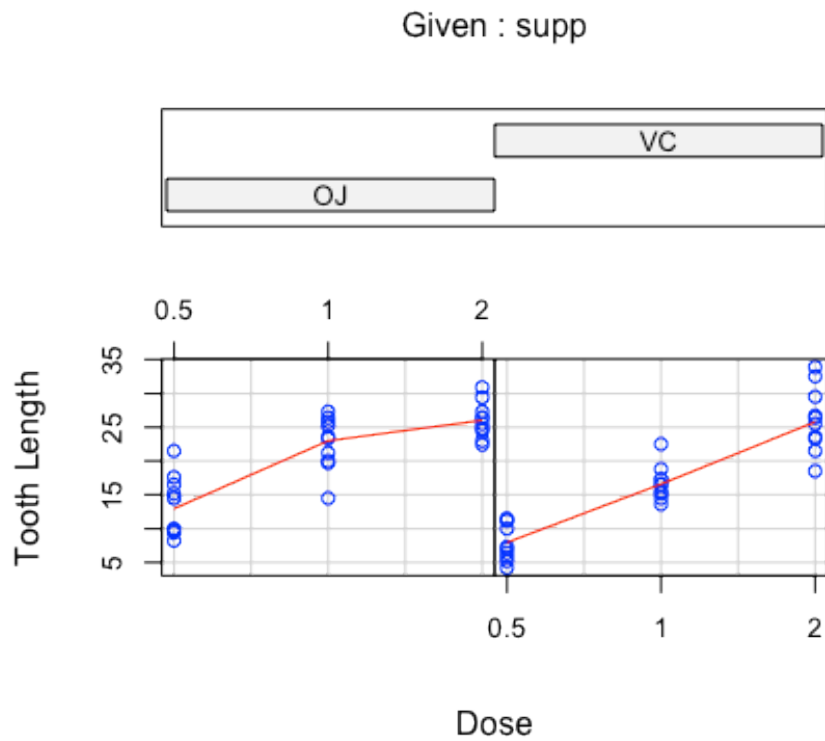
1. The different doses and supplements were given to different guinea pigs (and not the same guinea pigs at different time periods). Therefore I did not use a paired test.
2. The tooth growth data is assumed to be iid normal. This assumption is supported by the histogram which shows that the data is roughly symmetric and mound shaped and is not skewed.
3. The alternate hypothesis for the supplement assume a two sided t-test, simply testing that the supplements are not equal. The alternate hypotheses for the dose tests assume a one-sided t-test, testing that the higher dose results in longer length (as suspected by looking at the exploratory graphs).

Conclusions:

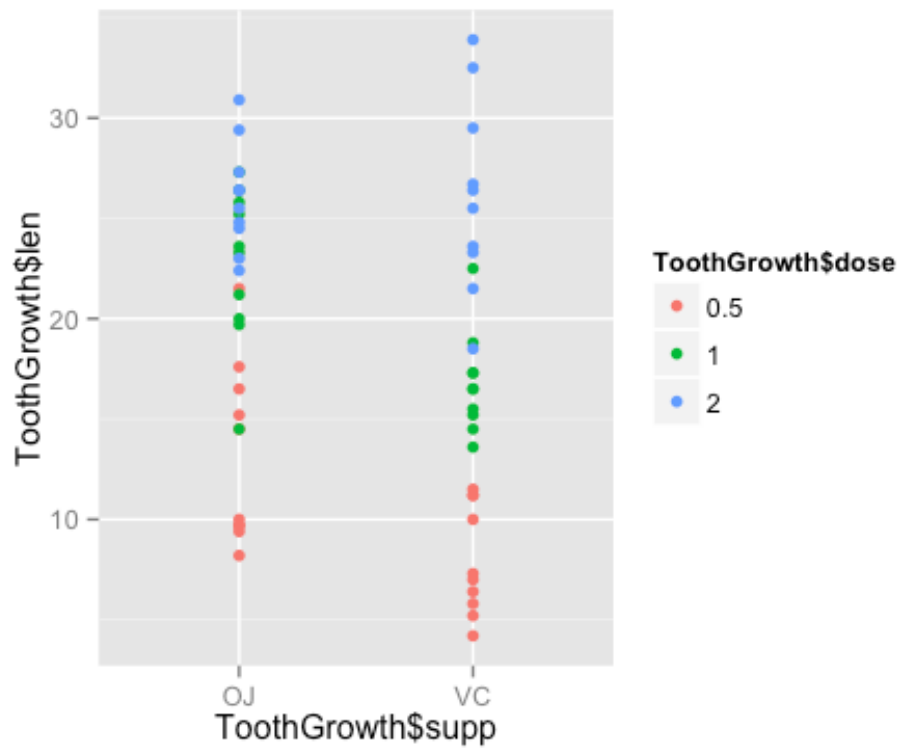
1. Fail to reject the null hypothesis that the means of the different supplements are equal. The confidence interval for the associated t-test contains $\mu = 0$.
2. Reject the null hypothesis for each of the tests of various doses. None of the tests (0.5 vs. 1.0, 0.5 vs. 2.0, 1.0 vs. 2.0) results in a 95% confidence interval that includes $\mu = 0$.

Appendix

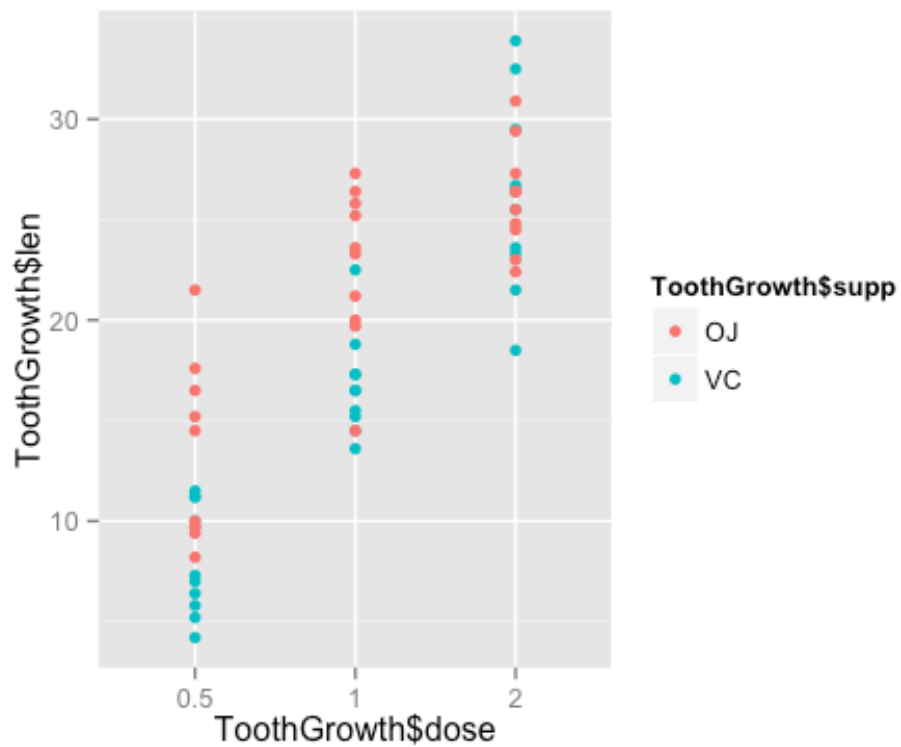
Plot tooth length by both dose and supplement



Plot tooth length by dose - color coded by supplement



Plot tooth length by supplement - color coded by dose



Code

Load Data

```
data(ToothGrowth) str(ToothGrowth) head(ToothGrowth)
```

```
library(ggplot2)
```

Histogram

```
hist(ToothGrowth$len)
```

Summary of Data

```
summary(ToothGrowth$len)
```

Mean and sd

```
mean_supp <- aggregate(len ~ supp, ToothGrowth, FUN = "mean") sd_supp <-  
aggregate(len ~ supp, ToothGrowth, FUN = "sd") mean_supp sd_supp
```

```
mean_dose<- aggregate(len ~ dose, ToothGrowth, FUN = "mean") sd_dose <- aggregate(len  
~ dose, ToothGrowth, FUN = "sd") mean_dose sd_dose
```

T Tests

```
t.test(len ~ I(relevel(supp, 2)), paired = FALSE, conf.level = 0.975, data =  
ToothGrowth)$conf
```

```
ToothGrowthdose) ToothGrowth51 <- subset(ToothGrowth, dose %in% c(0.5, 1.0))  
t.test(len ~ I(relevel(dose, 2)), paired = FALSE, data = ToothGrowth51)$conf
```

```
ToothGrowthdose) ToothGrowth51 <- subset(ToothGrowth, dose %in% c(0.5, 2.0))  
t.test(len ~ I(relevel(dose, 2)), paired = FALSE, data = ToothGrowth51)$conf
```

```
ToothGrowthdose) ToothGrowth51 <- subset(ToothGrowth, dose %in% c(1.0, 2.0))  
t.test(len ~ I(relevel(dose, 2)), paired = FALSE, data = ToothGrowth51)$conf
```