Statistical Inference Course Project

Jin-Hoon An

Friday, January 23, 2015

Part 2: Basic Inferential Data Analysis

Summary of ToothGrowth Dataset

```
##       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
```

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

The data allows us to compare a few different hypothesis to see if different doses of Vitamin c supplement and/or different supply methods support a hypothesis that they contributed to the a tooth's growth. *Please see the graphs of the distribution in the appendix of this report.*

Scenario #1

This scenario compares the impact of increasing doses of Vitamin C, independent of the delivery mechanism, to see if there is any statistical significant to tooth growth.

Test #1-1: *Compare the impact of doses of 0.5 units to 1.0 units of Vitamin C.*
$H_0 : \mu_2 - \mu_1 = 0$
$H_a : \mu_2 - \mu_1 > 0$
Where μ_1 and and μ_2 = average tooth growth for 0.5 and 1.0 units of Vitamin C respectively.

```
a <- 0.05
m1 <- ToothGrowth[ToothGrowth$dose==0.5,"len"]
m2 <- ToothGrowth[ToothGrowth$dose==1.0,"len"]
t1 <- t.test(m2,m1,paired = FALSE,var.equal = FALSE)
t1$conf
```

```
## [1]  6.276219 11.983781
## attr(,"conf.level")
## [1] 0.95
```

Reforming a t test for unequal sample variances, we see that the null hypothesis $\mu_0 = 0$ falls out of the range of the confidence interval. We therefore **reject** the null hypothesis. Alternatively, we can also reject the null hypothesis on the basis of the **t** statistic **6.477** exceeding **1.686**, which is the **95** percentile of the T distribution with **37.986** degrees of freedom.

Test #1-2: *Compare the impact of doses of 0.5 units to 1.0 units of Vitamin C.*
$H_0 : \mu_3 - \mu_2 = 0$
$H_a : \mu_3 - \mu_2 > 0$
Where μ_2 and and μ_3 = average tooth growth for 1.0 and 2.0 units of Vitamin C respectively.

```
m3 <- ToothGrowth[ToothGrowth$dose==2.0,"len"]
t1 <- t.test(m3,m2,paired = FALSE,var.equal = FALSE)
t1$conf
```

```
## [1] 3.733519 8.996481
## attr(,"conf.level")
## [1] 0.95
```

Peforming a similar t test, we see that the null hypothesis $\mu_0 = 0$ falls out of the range of the confidence interval. We therefore **reject** the null hypothesis. Alternatively, we can also reject the null hypothesis on the basis of the **t** statistic **4.9** exceeding **1.687**, which is the **95** percentile of the T distribution with **37.101** degrees of freedom.
The results of tests 1-1 and 1-2 support the alternative hypothesis that vitamin C in increased doses does increase average tooth growth in guinea pigs.

Scenario #2

This scenario compares the impact of the Vitamin C supply method, **OJ** vs **VC**, for common dosage levels to see if there is any statistical significant to tooth growth.

Test #2-1: *Compare the impact of supply method when dosing 0.5 units of Vitamin C.*
$H_0 : \mu_2 - \mu_1 = 0$
$H_a : \mu_2 - \mu_1 \neq 0$
Where μ_1 and μ_2 = average tooth growth for 0.5 units of Vitamin C when dosing with VC and OJ respectively.

```
a  <- 0.05
m1 <- ToothGrowth[ToothGrowth$dose==0.5 & ToothGrowth$supp=="VC","len"]
m2 <- ToothGrowth[ToothGrowth$dose==0.5 & ToothGrowth$supp=="OJ","len"]
t1 <- t.test(m2,m1,paired = FALSE,var.equal = FALSE,conf.level=(1-a/2))
t1$conf
```

```
## [1] 1.125088 9.374912
## attr(,"conf.level")
## [1] 0.975
```

We see that the null hypothesis $\mu_0 = 0$ falls out of the range of the confidence interval. We therefore **reject** the null hypothesis. Alternatively, we can also reject the null hypothesis on the basis of the $|t|$ statistic **3.17** exceeding **1.753**, which is the **97.5** percentile of the T distribution with **14.969** degrees of freedom.

Test #2-2: *Compare the impact of supply method when dosing 1.0 units of Vitamin C.*
$H_0 : \mu_2 - \mu_1 = 0$
$H_a : \mu_2 - \mu_1 \neq 0$ Where μ_1 and μ_2 = average tooth growth for 1.0 units of Vitamin C when dosing with VC and OJ respectively.

```
a  <- 0.05
m1 <- ToothGrowth[ToothGrowth$dose==1.0 & ToothGrowth$supp=="VC","len"]
m2 <- ToothGrowth[ToothGrowth$dose==1.0 & ToothGrowth$supp=="OJ","len"]
t1 <- t.test(m2,m1,paired = FALSE,var.equal = FALSE,conf.level=(1-a/2))
t1$conf
```

```
## [1] 2.2781 9.5819
## attr(,"conf.level")
## [1] 0.975
```

We see that the null hypothesis $\mu_0 = 0$ clearly falls out of the range of the confidence interval. We therefore **reject** the null hypothesis. Alternatively, we can also reject the null hypothesis on the basis of the $|t|$ statistic **4.033** exceeding **1.75**, which is the **97.5** percentile of the T distribution with **15.358** degrees of freedom.

Test #2-3: *Compare the impact of supply method when dosing 2.0 units of Vitamin C.*
$H_0 : \mu_2 - \mu_1 = 0$
$H_a : \mu_2 - \mu_1 \neq 0$ Where μ_1 and μ_2 = average tooth growth for 2.0 units of Vitamin C when dosing with VC and OJ respectively.

```
a  <- 0.05
m1 <- ToothGrowth[ToothGrowth$dose==2.0 & ToothGrowth$supp=="VC","len"]
m2 <- ToothGrowth[ToothGrowth$dose==2.0 & ToothGrowth$supp=="OJ","len"]
t1 <- t.test(m2,m1,paired = FALSE,var.equal = FALSE,conf.level=(1-a/2))
t1$conf
```

```
## [1] -4.430131  4.270131
## attr(,"conf.level")
## [1] 0.975
```

Unlike all the previous cases, in this test, the null hypothesis $\mu_0 = 0$ falls within the range of the confidence interval. We therefore **accept** the null hypothesis. Alternatively, we can also **accept** the null hypothesis on the basis of the $|t|$ statistic **0.046** not exceeding **1.761**, which is the **97.5** percentile of the T distribution with **14.04** degrees of freedom.

The results of tests 2-1 and 2-1 support the alternative hypothesis that there is a difference in tooth growth depending on the supply mechanism of Vitamin C for doses of 0.5 and 1.0 units. In these cases, utilzing OJ vs VC appears to result in longer tooth growth. However, with an increaed dose of 2 units, the difference between the two supply method does not appear to be statistically significant. Further tests should be conducted to see if this is the case for other dosages above 2.0 units.

Conclusion

This results of this study indicate that Vitamin C in increased doses has a statistically significant impact on tooth growth. In addition, for at least moderate doses of Vitamin C (in this case 1.0 units or less), the choice of supply method, OJ vs VC, also has an impact on tooth growth, with OJ resulting in longer teeth than VC. However that distinction did not hold true when dosing 2.0 units of Vitamin C, providing some support of the conclusion that supply method has an impact when administering lower doses of Vitamin C, but not necessarily with larger doses.

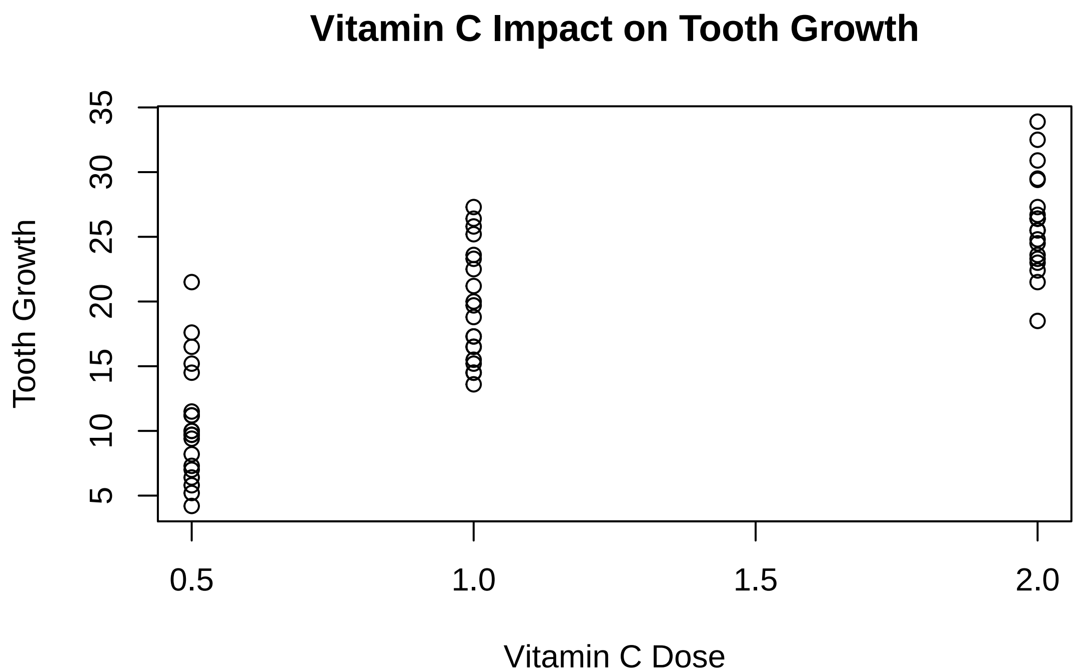
Appendix

Assumptions

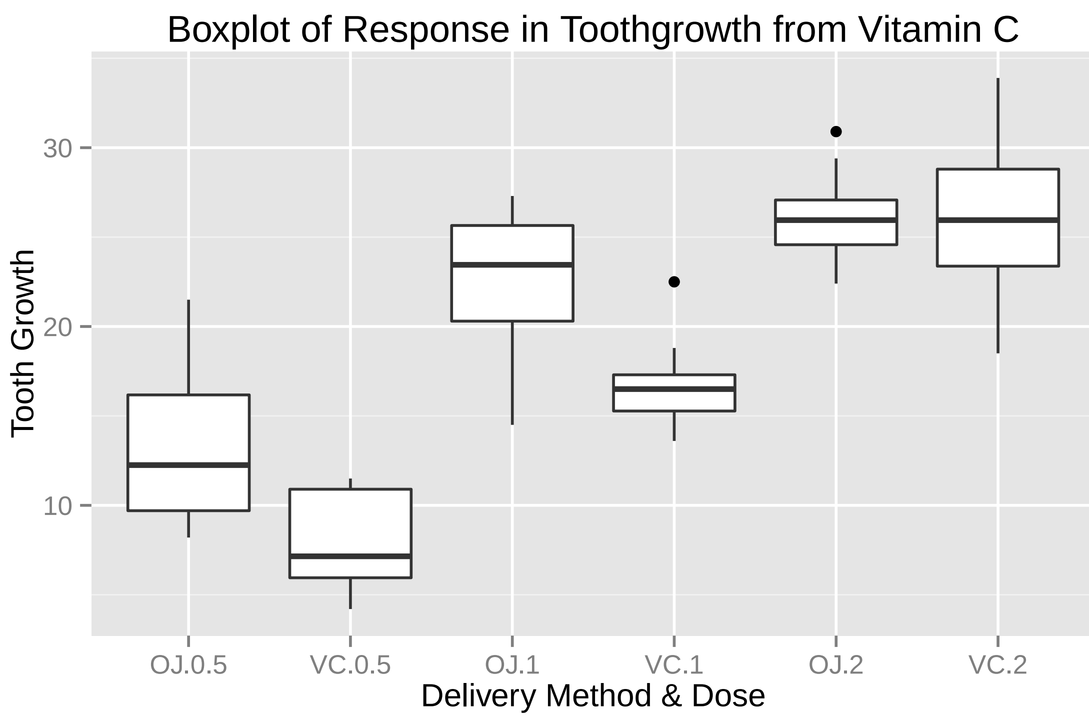
This analysis assumes that the guinea pigs were randomly assigned to a set of predefined combinations of doses and supply methods of Vitamin C. In addition, it also assumes that there were no other uncontrolled factors that could independently lead to tooth growth on individual subjects that might confound the overall results.

Distribution Graphs

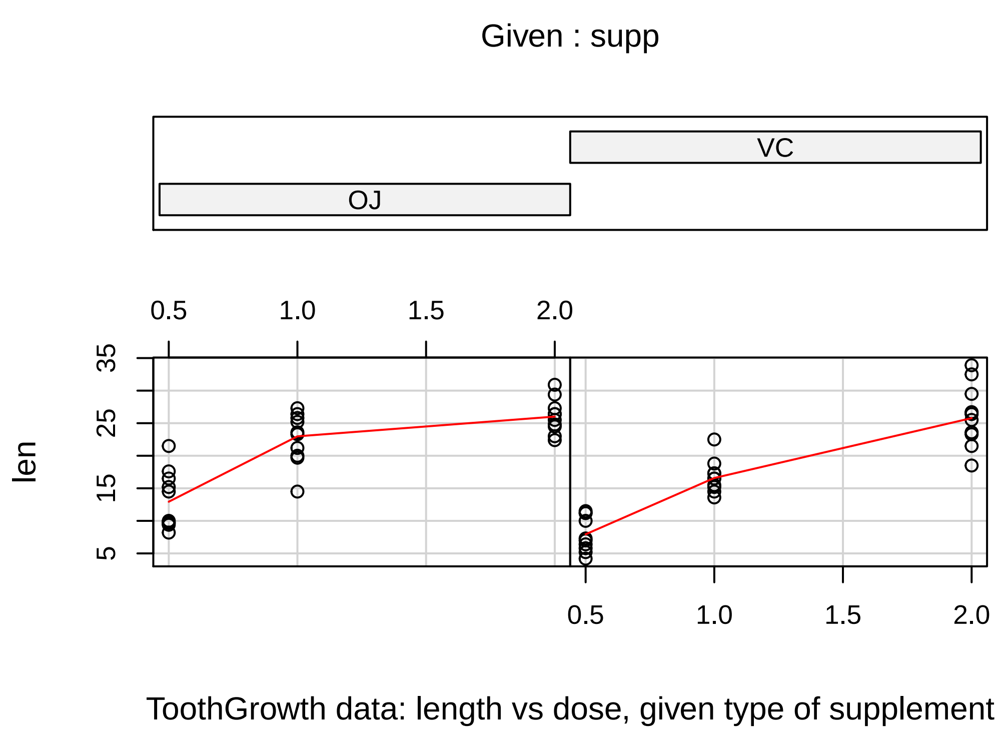
```r
plot(x=ToothGrowth$dose,y=ToothGrowth$len,xlab="Vitamin C Dose",
     ylab="Tooth Growth",
     main="Vitamin C Impact on Tooth Growth")
```




```r
gp <- ggplot(ToothGrowth, aes(x=interaction(supp, dose), y=len)) + geom_boxplot()
```




```r
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```



Sample t.test details

```r
a <- 0.05
m1 <- ToothGrowth[ToothGrowth$dose==0.5,"len"]
m2 <- ToothGrowth[ToothGrowth$dose==1.0,"len"]
t1 <- t.test(m2,m1,paired = FALSE,var.equal = FALSE)
t1
```

```
##
##  Welch Two Sample t-test
##
## data:  m2 and m1
## t = 6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

```r
m3 <- ToothGrowth[ToothGrowth$dose==2.0,"len"]
t1 <- t.test(m3,m2,paired = FALSE,var.equal = FALSE)
t1
```

```
##
##  Welch Two Sample t-test
##
## data:  m3 and m2
## t = 4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.733519 8.996481
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

```r
a  <- 0.05
m1 <- ToothGrowth[ToothGrowth$dose==0.5 & ToothGrowth$supp=="VC","len"]
m2 <- ToothGrowth[ToothGrowth$dose==0.5 & ToothGrowth$supp=="OJ","len"]
t1 <- t.test(m2,m1,paired = FALSE,var.equal = FALSE,conf.level=(1-a/2))
t1
```

```
##
##  Welch Two Sample t-test
##
## data:  m2 and m1
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 97.5 percent confidence interval:
##  1.125088 9.374912
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

assuming the significance level is 5%.