

point2

2023-12-24

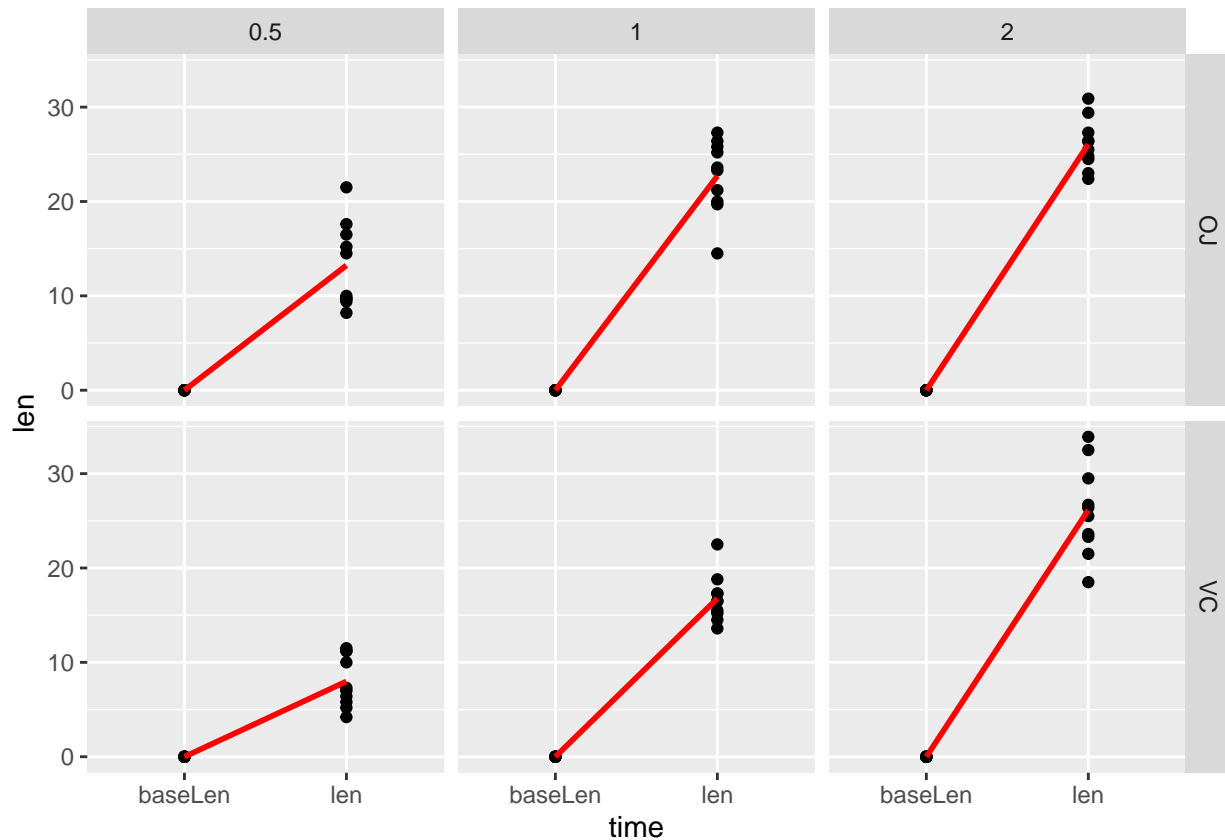
Exploratory analysis

A simple analysis is performed in which it is defined that all test subjects started with a base length = 0.

```
data(ToothGrowth)
library(ggplot2)
library(reshape2)
ToothGrowth$baseLen<-0
ToothGrowth$id<-1:60
Too<-melt(ToothGrowth,id=c("supp","dose","id"),measure.vars = c("baseLen","len"))
```

Plot

This graph divides the data into supplement and dosage. In addition, it has a line that represents the average of each length.



Testing on data

Carrying out tests to determine which supplement and at what dose is best. To carry out the tests, it will first be determined with which dose greater tooth growth is obtained with each supplement and then between supplements.

first get the best dosage for each supplement

#OJ

I begin hypothesis testing for the OJ supplement $\mu_1 = \text{mean the 0.5}$ and $\mu_2 = \text{mean the 1}$ ho: $\mu_1 = \mu_2$ ha: $\mu_1 \neq \mu_2$ alpha = 0.05

```
alpha<- 0.05
g1<- subset(TooSepForSupp$OJ, dose %in% 0.5 )$len
g2<- subset(TooSepForSupp$OJ, dose %in% 1)$len

u1 <-mean(g1)
n1<- length(g1)
u2<- mean(g2)
n2<- length(g2)
s1<- sd(g1)
s2<- sd(g2)
df<- (((s1^2)/n1 + (s2^2)/n2 )^2)/((((s1^2)/n1)^2/(n1-1) + ((s2^2)/n2)^2/(n2-1)))
mn<- u1-u2

est<-mn/sqrt(((s1^2)/n1)+((s2^2)/n2))
prob<-pt(est,df)
prob<alpha/2
```

```
## [1] TRUE
```

The value for my test is less than the established alpha, so the difference between the 2 means is large enough to consider that the null hypothesis should be rejected. I use the t statistic to know which group to choose to continue comparing.

```
est
```

```
## [1] -5.048635
```

Since the statistic is negative that means that g2 is larger than g1 so I choose g2

$\mu_1 = \text{mean the 1}$ and $\mu_2 = \text{mean the 2}$ ho: $\mu_1 = \mu_2$ ha: $\mu_1 \neq \mu_2$ alpha = 0.05

```
g1<- g2
g2<- subset(TooSepForSupp$OJ, dose %in% 2 )$len

testBina<- t.test(g1,g2,paired= FALSE,var.equal = FALSE)

testBina$p.value/2 < alpha/2
```

```
## [1] TRUE
```

Again my p.values are below my alpha value so again the null hypothesis is rejected.

```
testBina$statistic
```

```
##          t
## -2.247761
```

I choose group 2, so in conclusion, for the OJ supplement, the dose that made the teeth grow the most is 2 doses.

#VC

u1 = mean the 0.5 and u2 = mean the 1 ho: u1=u2 ha: u2!=u1 alpha = 0.05

```
alpha<- 0.05
g1<- subset(TooSepForSupp$VC, dose %in% 0.5 )$len
g2<- subset(TooSepForSupp$VC, dose %in% 1)$len
u1 <-mean(g1)
n1<- length(g1)
u2<- mean(g2)
n2<- length(g2)
s1<- sd(g1)
s2<- sd(g2)
df<- (((s1^2)/n1 + (s2^2)/n2 )^2)/((((s1^2)/n1)^2/(n1-1) + (((s2^2)/n2)^2/(n2-1)))
mn<- u1-u2
testBina<-t.test(g1,g2,paires=FALSE,var.equal = FALSE)
est<-mn/sqrt(((s1^2)/n1)+((s2^2)/n2))
testBina$p.value/2<alpha/2
```

```
## [1] TRUE
```

```
pt(est,df)<alpha/2
```

```
## [1] TRUE
```

I reject the null hypothesis so I use the t statistic to know which of the 2 groups I am going to compare with now.

```
testBina$statistic
```

```
##          t
## -7.46343
```

```
est
```

```
## [1] -7.46343
```

Being the negative statistic, by the order of the Y-X groups. So, the largest group or the largest mean in this case is X=g2

u1 = mean the 1 and u2 = mean the 2 ho: u1=u2 ha: u2!=u1 alpha = 0.05

```
g1<- g2
g2<- subset(TooSepForSupp$VC, dose %in% 2 )$len
testBina<- t.test(g1,g2,paired= FALSE,var.equal = FALSE)
testBina$p.value/2 < alpha/2
```

```
## [1] TRUE
```

Again my p.values are below my alpha value so again the null hypothesis is rejected.

```
testBina$statistic
```

```
##          t
## -5.469814
```

the statistic is negative so my choice of the largest group is $X=g2$. In conclusion, the teeth for the VC supplement grew more with 2 doses

first conclusion

Identify the dose that makes teeth grow the most in both supplements, the next thing is to obtain the supplement with the greatest growth

Get the best supplement using the previously chosen doses

$u1$ = average of the 2nd dose with VC supplement and $u2$ = average of the 2nd dose with OJ supplement
 $h_0: u1=u2$ $h_a: u1 \neq u2$ $\alpha = 0.05$

```
g1<-subset(TooSepForSupp$VC, dose %in% 2)$len
g2<-subset(TooSepForSupp$OJ, dose %in% 2)$len
```

I'm going to use resampling

```
n1<- length(g1)
n2<- length(g2)
nsim<- 10000

g1M<- matrix(sample(g1,nsim*n1,replace = TRUE),nsim,n1)
g2M<- matrix(sample(g2,nsim*n2,replace= TRUE),nsim,n2)
meanG1V<- apply(g1M,1,mean)
meanG2V<- apply(g2M,1,mean)
sdG1V<- apply(g1M,1,sd)
sdG2V<- apply(g2M,1,sd)
u1<- mean(meanG1V)
u2<- mean(meanG2V)
s1<- mean(sdG1V)
```

```

s2<- mean(sdG2V)
df<- (((s1^2)/n1 + (s2^2)/n2 )^2)/((((s1^2)/n1)^2)/(n1-1) + (((s2^2)/n2)^2/(n2-1)))

est<- (u1-u2)/sqrt((s1^2)/n1 + (s2^2)/n2)
prob<-pt(est,df,lower.tail = FALSE)
prob<alpha

```

```
## [1] FALSE
```

```
prob<alpha/2
```

```
## [1] FALSE
```

Conclusion

Due to the result obtained with the hypothesis test, the p-value is not less than $\alpha/2$ and not even less than α . It is concluded that between the VC and OJ supplement groups with 2 doses there is not a difference notable enough to obtain a supplement that more favors tooth growth. Then the null hypothesis is not rejected.