# Statistical Inference - Course project Part 2

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

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/day

## Load and Explore the ToothGrowth data

```
data(ToothGrowth)
print(paste("Loaded ", nrow(ToothGrowth)," rows"))
```

```
## [1] "Loaded  60  rows"
```

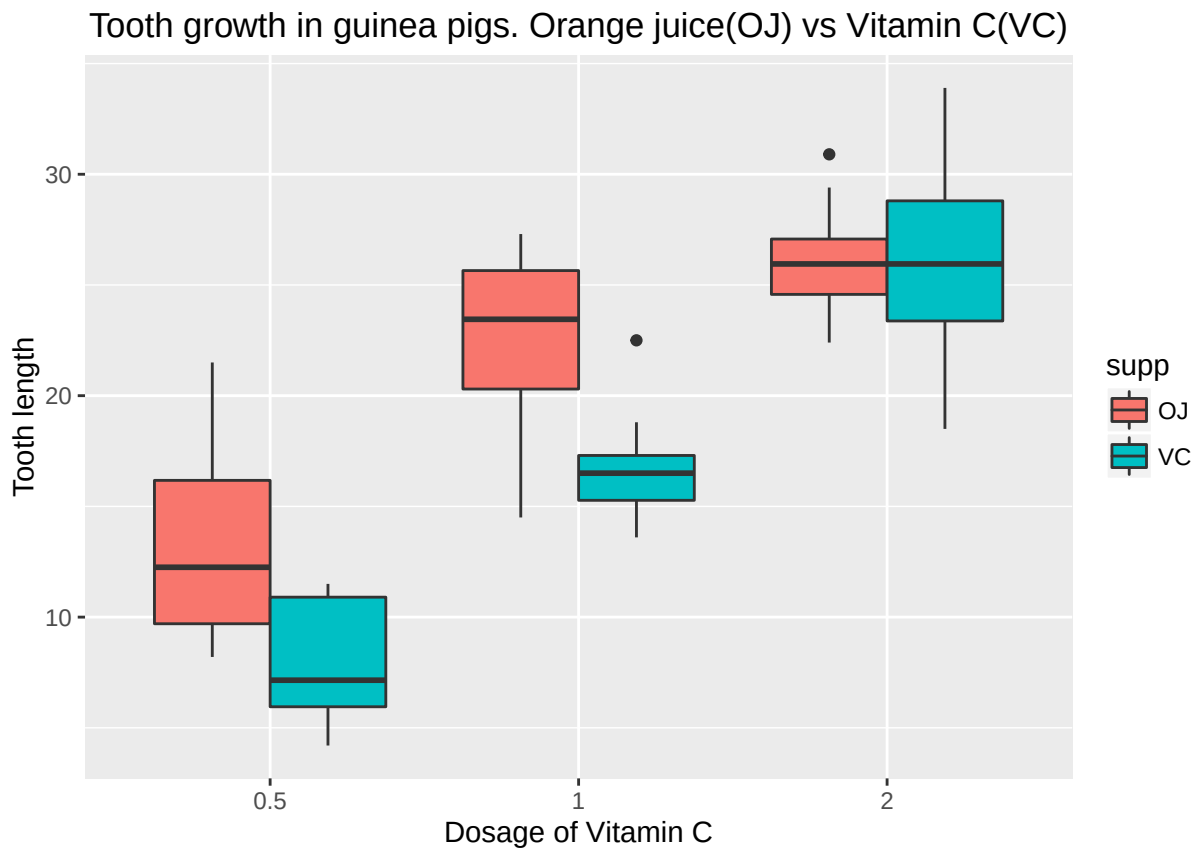
```
head(ToothGrowth)
```

```
##    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
```

```
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
```

```
g <- ggplot(ToothGrowth, aes(factor(dose), len)) +
  geom_boxplot(aes(fill = supp)) +
  xlab("Dosage of Vitamin C") +
  ylab("Tooth length") +
  ggtitle('Tooth growth in guinea pigs. Orange juice(OJ) vs Vitamin C(VC)')
g
```




Here we obviously see that dosage increases the tooth lentgh, but the Supplement type is probably not as significant if at all, especially at higher dosages.

### Provide a basic summary of the data

```
tapply(ToothGrowth$len,ToothGrowth$dose,mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

```
tapply(ToothGrowth$len,ToothGrowth$supp,mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

### Use confidence intervals and/or hypothesis tests to compare tooth growth

```
#split by Supplement type
vc <-subset(ToothGrowth,supp=="VC")$len
oj <-subset(ToothGrowth,supp=="OJ")$len

t.test(vc, oj, paired = FALSE, var.equal = FALSE)$p.value
```

```
## [1] 0.06063451
```

Looks like the p.value is pretty low, but still provides not enough confidence to say that means of datasets split by Supplement type is different.

Now let's look at dosage dependability

```
#split by dosage
d1 <-subset(ToothGrowth,dose==0.5)$len
d2 <-subset(ToothGrowth,dose==1)$len
d3 <-subset(ToothGrowth,dose==2)$len
t.test(d1, d2, paired = FALSE, var.equal = FALSE)$p.value
```

```
## [1] 1.268301e-07
```

```
t.test(d2, d3, paired = FALSE, var.equal = FALSE)$p.value
```

```
## [1] 1.90643e-05
```

P.values are pretty much absolute 0. There's near absolute confidence that new tooth growth lengths means are rising with increased dosage.

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

As above, there's to little confidence say that supplement type makes any difference. However, the increased dosage definitely makes for a larger tooth growth.