Basic inferential data analysis of 'ToothGrowth' dataset

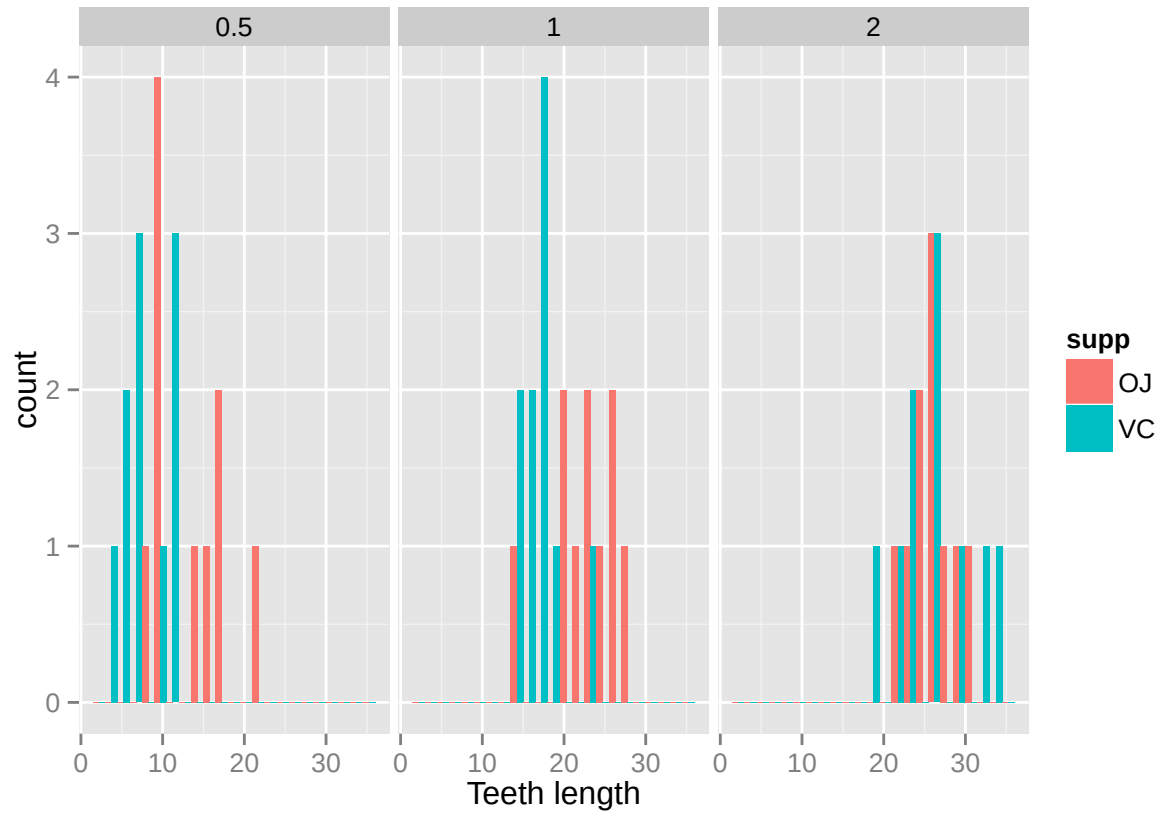
Elena Tikalenko

Data description

The data for investigation was extracted from *ToothGrowth* dataset. It's the information about the length of odontoblasts in 60 guinea pigs, each receiving one of three dose levels of Vitamin C (0.5, 1.0, and 2.0 mg) with one of two delivery methods (orange juice or an aqueous solution of ascorbic acid). So we have two independent groups.

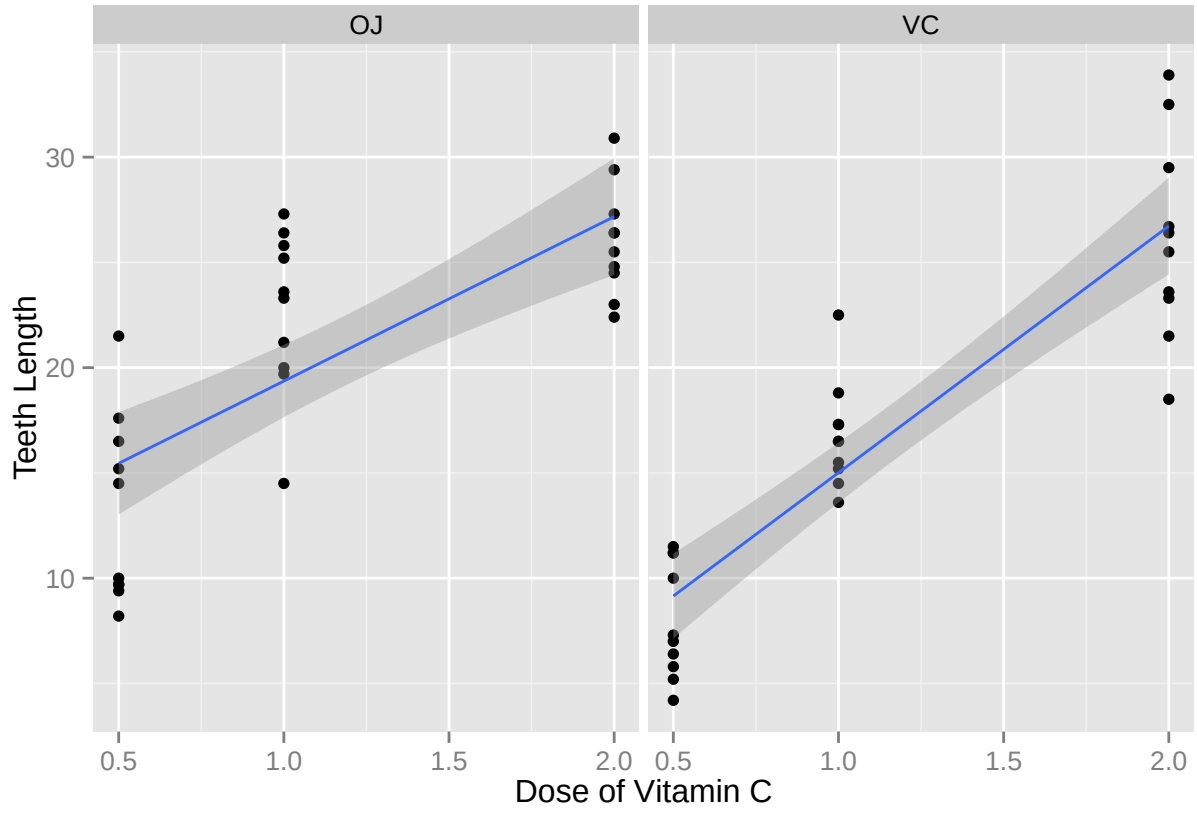
Load the dataset and let's see the histogram of tooth length according to the dose for both supplement types. It looks like that orange juice is better way for receiving of Vitamin C for dose 0.5 and 1. But there is no difference in a way of receiving for dose 2 mg.

```
data(ToothGrowth)
ggplot(ToothGrowth, aes(x = len, fill = supp)) +
        geom_histogram(position = "dodge", binwidth = 1.5) +
        facet_grid(.~ dose) + labs(x = "Teeth length")
```



From the next plot it's obviously that teeth length is increasing with dose increasing.

```
ggplot(ToothGrowth, aes(x = dose, y = len)) + geom_point() + facet_grid(.~ supp) +
        stat_smooth(method = "lm") + labs(x = "Dose of Vitamin C", y = "Teeth Length")
```



Analysis

Let's take three subsets of dataset *ToothGrowth* by dose and do t-test for each subset.

```
TG0.5 <- ToothGrowth[ToothGrowth$dose == 0.5,]
TG1 <- ToothGrowth[ToothGrowth$dose == 1,]
TG2 <- ToothGrowth[ToothGrowth$dose == 2,]
```

We used in t-tests paired = FALSE because we have six totally separate groups of pigs.

```
t.test(len ~ supp, data = TG0.5, paired = FALSE, var.equal = TRUE)[c("conf.int","estimate")]
```

```
## $conf.int
## [1] 1.770262 8.729738
## attr(,"conf.level")
## [1] 0.95
##
## $estimate
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len ~ supp, data = TG1, paired = FALSE, var.equal = TRUE)[c("conf.int","estimate")]
```

```
## $conf.int
## [1] 2.840692 9.019308
## attr(,"conf.level")
## [1] 0.95
##
## $estimate
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len ~ supp, data = TG2, paired = FALSE, var.equal = TRUE)[c("conf.int","estimate")]
```

```
## $conf.int
## [1] -3.722999  3.562999
## attr(,"conf.level")
## [1] 0.95
##
## $estimate
## mean in group OJ mean in group VC
##            26.06            26.14
```

Conclusion

In first two tests confidence intervals are entirely above zero. It means that supplement Orange Juice is better than Vitamin C for dose 0.5 mg and 1 mg. But in third test confidence interval contains zero, means in both groups are approximately equal. It means that it's impossible to define which way is the best for dose 2 mg.