Statistical Inference Course Project (Part 2)

MsGret

10 August 2020

The project consists of two parts: simulation (part 1) and basic inferential data analysis (part 2). In part 2 we do basic data analysis base on the ToothGrowth data in the R datasets package.

Load the ToothGrowth data

Load the data and perform some basic exploratory data analyses

```
data("ToothGrowth")
dim(ToothGrowth)
```

```
## [1] 60  3
```

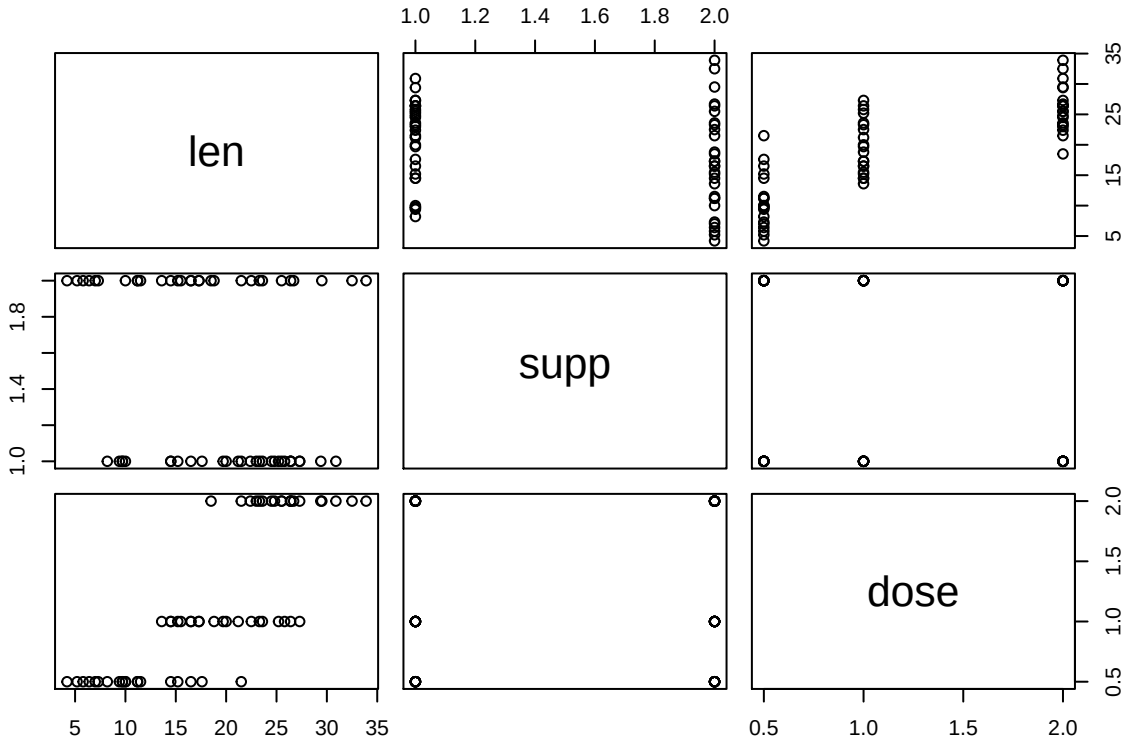
```
head(ToothGrowth)
```

```
##    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
```

```
str(ToothGrowth)
```

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

```
plot(ToothGrowth)
```



Basic Exploratory Data Analyses

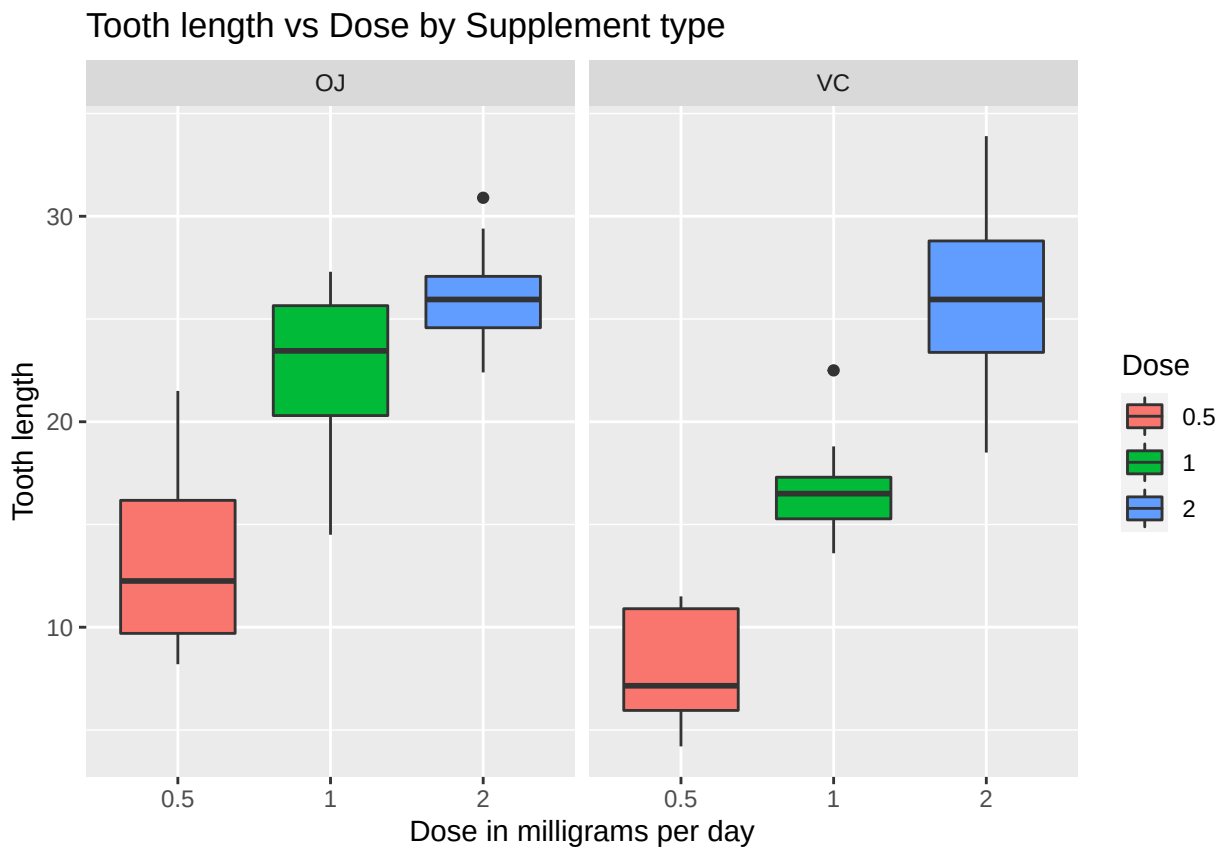
From the plot we can see that the variable dose takes three values. Use these variable as factor and plot results of basic exploratory data analyses.

```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)

require(ggplot2)
```

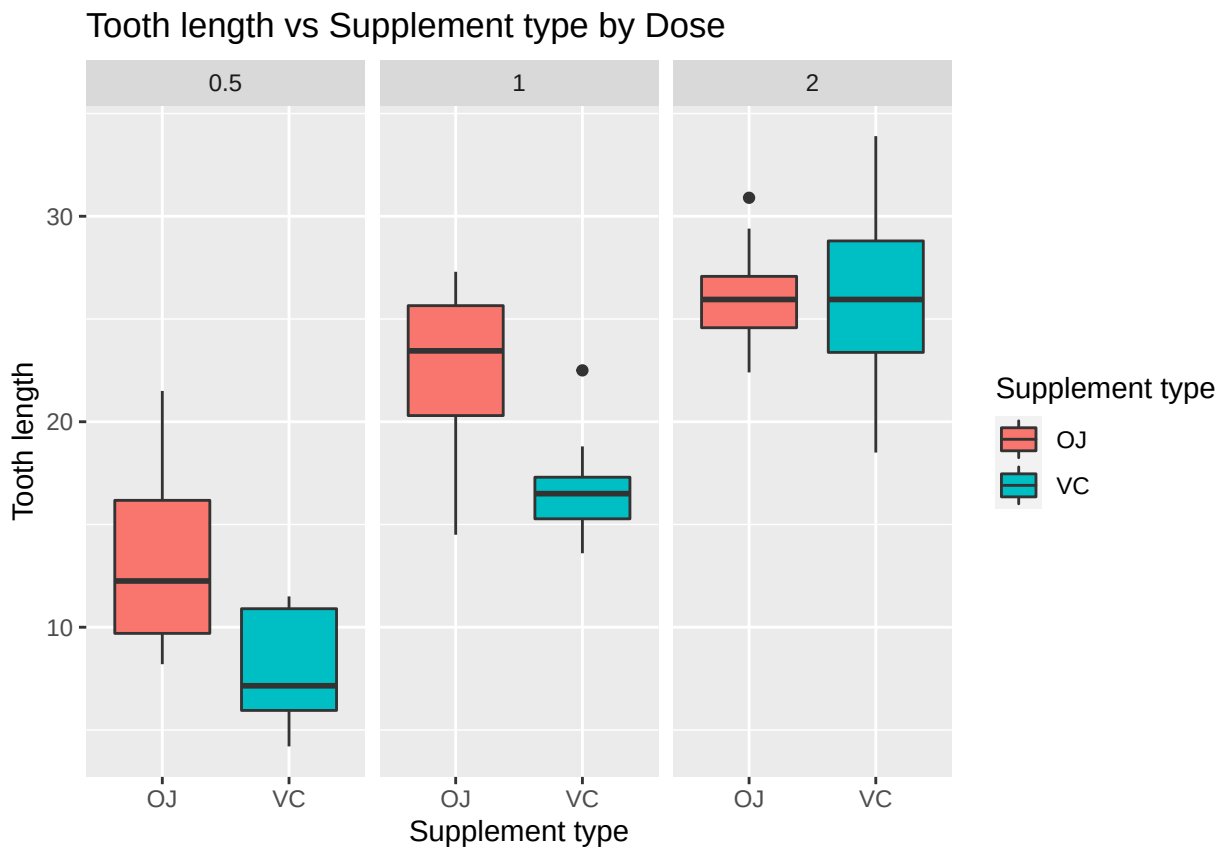
```
## Loading required package: ggplot2
```

```
ggplot(data = ToothGrowth, aes(x = dose, y = len, fill = dose)) +
    geom_boxplot() +
    facet_grid(. ~ supp) +
    labs(x = "Dose in milligrams per day", y = "Tooth length",
         title = "Tooth length vs Dose by Supplement type") +
    scale_fill_discrete(name = "Dose")
```



Plots show the increase in tooth length with increasing doses.

```
ggplot(data = ToothGrowth, aes(x = supp, y = len, fill = supp)) +
    geom_boxplot() +
    facet_grid(. ~ dose) +
    labs(x = "Supplement type", y = "Tooth length",
         title = "Tooth length vs Supplement type by Dose") +
    scale_fill_discrete(name = "Supplement type")
```



Orange juice (coded as OJ) is more effective than ascorbic acid (a form of vitamin C and coded as VC) for tooth growth at doses of 0.5 and 1 milligrams per day.

Compare tooth growth by supp

Use hypothesis test to compare tooth growth by supp.

Delivery methods have no impact on tooth growth. H_0: Difference of means (OJ and VC) is 0. H_A: Difference of means (OJ and VC) is not 0.

```
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The p-value of this test was 0.06 (greater than the significance level alpha = 0.05) and confidence interval of the test contains zero. We can say that there is not enough evidence to reject the null hypothesis that supplement types have no impact on Tooth growth.

Compare tooth growth by dose

Use hypothesis test to compare tooth growth by dose.

Doses have no impact on tooth growth. H_0: Difference of means (.5 and 1, or .5 and 2, or 1 and 2) is 0. H_A: Difference of means (.5 and 1, or .5 and 2, or 1 and 2) is not 0.

```
t.test(len ~ dose, data = ToothGrowth, subset = ToothGrowth$dose == 0.5 | ToothGrowth$dose == 1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
t.test(len ~ dose, data = ToothGrowth, subset = ToothGrowth$dose == 0.5 | ToothGrowth$dose == 2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```
t.test(len ~ dose, data = ToothGrowth, subset = ToothGrowth$dose == 1 | ToothGrowth$dose == 2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

In all three cases the p-value of this test less than the significance level alpha = 0.05 and confidence intervals of the test don't contain zero. We can say that there is enough evidence to reject the null hypothesis that doses have no impact on Tooth growth.

Conclusion

According to the analysis performed, we can conclude that

- supplement types have no impact on Tooth growth;

- doses have impact on Tooth growth.

The conclusions are based on the following assumptions:

- the sample is representative of the pig population;
- observations are independent.