Statistical inference part 2

Sergei Keidzh

May 13 2018

```
setwd("/statistical inference")
library(datasets)
library(ggplot2)
data(ToothGrowth)
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 ...
```

```
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
```

to make it reproducible it's better to convert variable to factor

in this case variable dose

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

let's review again

```
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

provide a basic summary

```
summary(ToothGrowth)
```

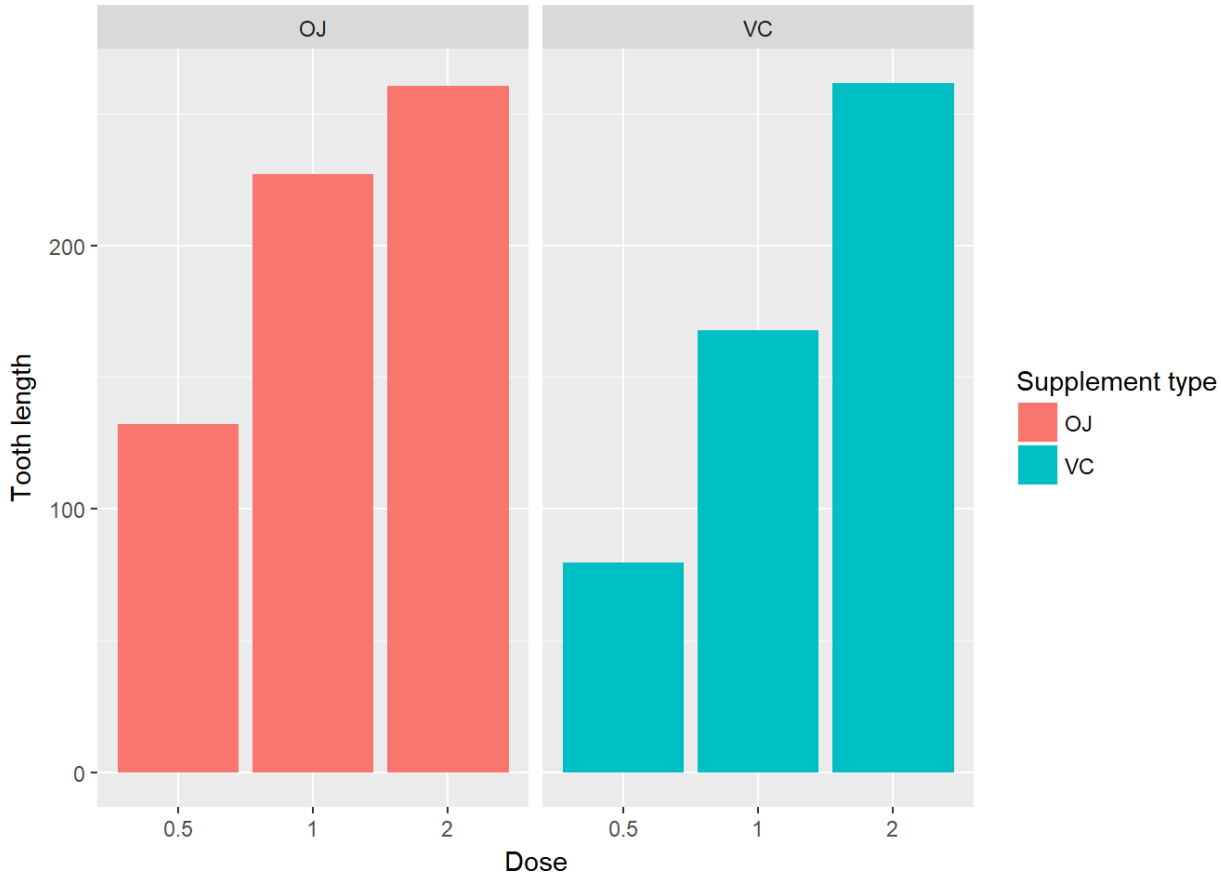
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

split of cases between different dose levels and delivery methods

```
table(ToothGrowth$dose, ToothGrowth$supp)
```

```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp))+
  geom_bar(stat="identity",)+
  facet_grid(. ~ supp)+
  xlab("Dose")+
  ylab("Tooth length")+
  guides(fill=guide_legend(title="Supplement type"))
```



#on the figure we can see how vitamin C influenced in positive way on #tooth length which has been dependend on dose of vitamin.

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

(Only use the techniques from class, even if there's other approaches worth considering)

```
data2 <- lm(len ~ dose + supp, data=ToothGrowth)


summary(data2)
```

```
##
## Call:
## lm(formula = len ~ dose + supp, data = ToothGrowth)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -7.085 -2.751 -0.800  2.446  9.650
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.4550     0.9883  12.603  < 2e-16 ***
## dose1         9.1300     1.2104   7.543 4.38e-10 ***
## dose2        15.4950     1.2104  12.802  < 2e-16 ***
## suppVC       -3.7000     0.9883  -3.744 0.000429 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.828 on 56 degrees of freedom
## Multiple R-squared:  0.7623, Adjusted R-squared:  0.7496
## F-statistic: 59.88 on 3 and 56 DF,  p-value: < 2.2e-16
```

As we can see the Vitamin C provides more tooth growth than ascorbic acid at lower dosages (0.5 Mg - 1 Mg).

The tooth growth rate is not statistically different between supplement methods at the higher dosage

Dose is the key factor in tooth growth. If the dose is more higher, the more grow for the teeth.

The variances between the sample popluations are not equal

The sample data is not paired

The sample population distribution is mound shaped and not skewed