Part2 ToothGrowth

Shayaan Ahmed Farooqi

5/20/2020

```
library(tidyverse)
data("ToothGrowth")
ToothGrowth <- tbl_df(ToothGrowth)
```

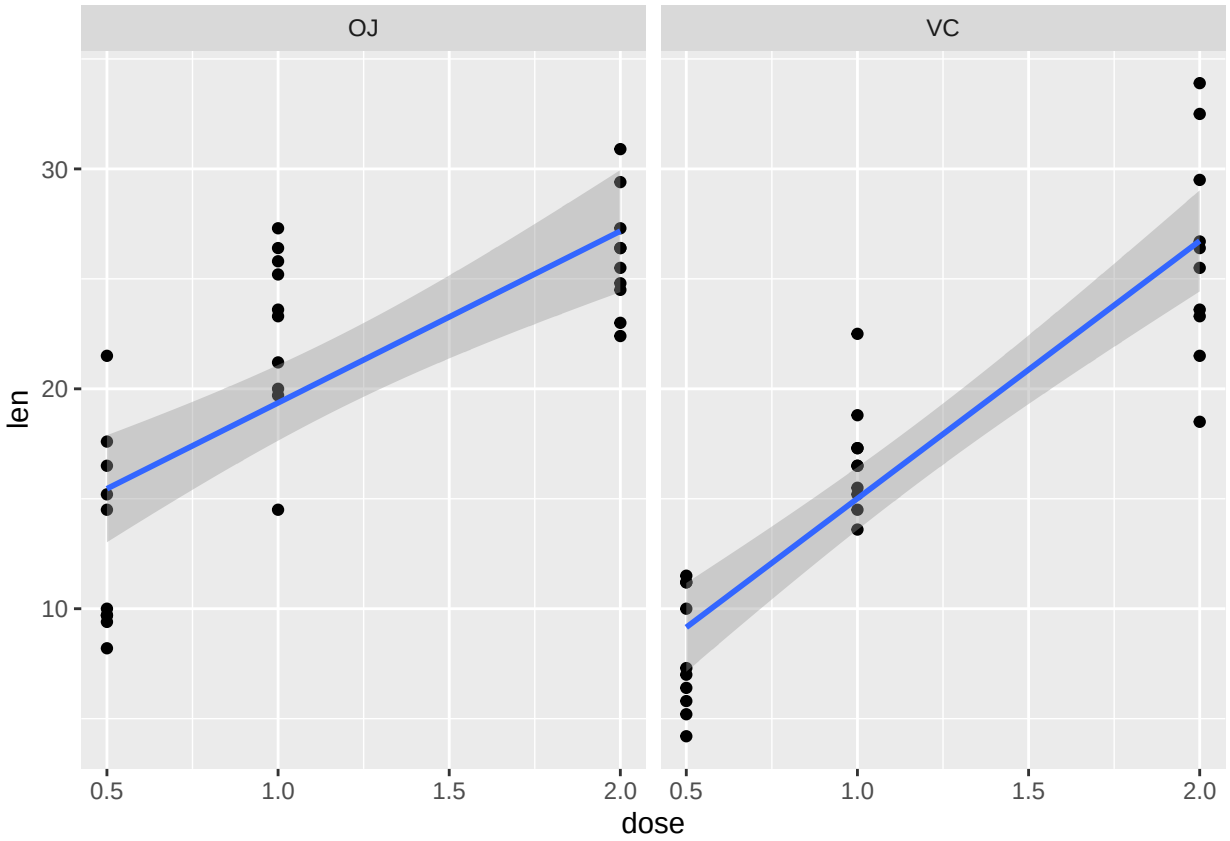
Exploratory Data Analysis:

```
str(ToothGrowth)
```

```
## tibble [60 x 3] (S3: tbl_df/tbl/data.frame)
##  $ len : num [1:60] 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 [1:60] 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
ggplot(data=ToothGrowth) +
    aes(x = dose, y = len) +
    geom_point() +
    facet_grid(cols=vars(supp)) +
    geom_smooth(method='lm')
```

```
## `geom_smooth()` using formula 'y ~ x'
```



Data Summary:

The ToothGrowth dataset studies how vitamin C supplements affect the tooth growth of Guinea Pigs. It has 3 variables: **len**, **supp** and **dose**. len represents the length of teeth of Guinea Pigs, It is numeric. supp represents the type of supplement given, it is either Orange Juice represented as "OJ" or it is ascorbic acid coded as "VC". The dose represents the amount of supplement given in milligrams/day.

Based on our initial exploratory graphs we can see a clear positive relation in dosage and the length of teeth. More dosage results in longer teeth and ascorbic acid seems to have a bigger impact on the length of teeth.

Tests:

Test 1

Alternate Hypothesis 1: Supplement VC has higher mean length than supplement OJ.
Null Hypothesis : Both supplements have the same mean lengths.

Assuming that both groups have the same variance and the population is normally distributed.

```
vc <- ToothGrowth %>%
    filter(supp=="VC")

oj <- ToothGrowth %>%
    filter(supp=="OJ")
```

```r
t.test(x=vc$len, y=oj$len, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  vc$len and oj$len
## t = -1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.5670064  0.1670064
## sample estimates:
## mean of x mean of y
##  16.96333  20.66333
```

Based on this result we fail to reject the Null Hypothesis since our confidence interval includes 0 and we have a P-value of 6%.

Test 2

Alternate Hypothesis 2: dose 2 has higher mean length than dose 0.5.
Null Hypothesis : Both dosages have the same mean lengths.

Assuming that both groups have the same variance and the population is normally distributed.

```r
d2 <- ToothGrowth %>%
    filter(dose==2)

d05 <- ToothGrowth %>%
    filter(dose==0.5)

t.test(x=d2$len, y=d05$len, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  d2$len and d05$len
## t = 11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  12.83648 18.15352
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

Based on these results we can comfortably reject the Null hypothesis since our P-Value is extremely small and our confidence interval also does not include 0.

Both conclusions are based on the assumptions that all Guinea pigs were chosen independently and randomly from a normally distributed population and we do not have a bad sampling situation where for example naturally fast growing guinea pigs were given high dosages which would confound our results. There is also a small chance that this particular sample was a rare coincidence and just happened to result in this conclusion as shown by the P-value.