Statistical Inference Part 2

Simon Grasdal

2024-12-11

Overview

The project aim is to analyze the ToothGrowth data in the R datasets package.

Load the necessary packages

```
library(ggplot2)
library(tinytex)
library(datasets)
```

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
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, 4)
```

```
##    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
```

```
tail(ToothGrowth, 4)
```

```
##     len supp dose
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

Summary of the data

```r
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
```

2.Basic summary of the data

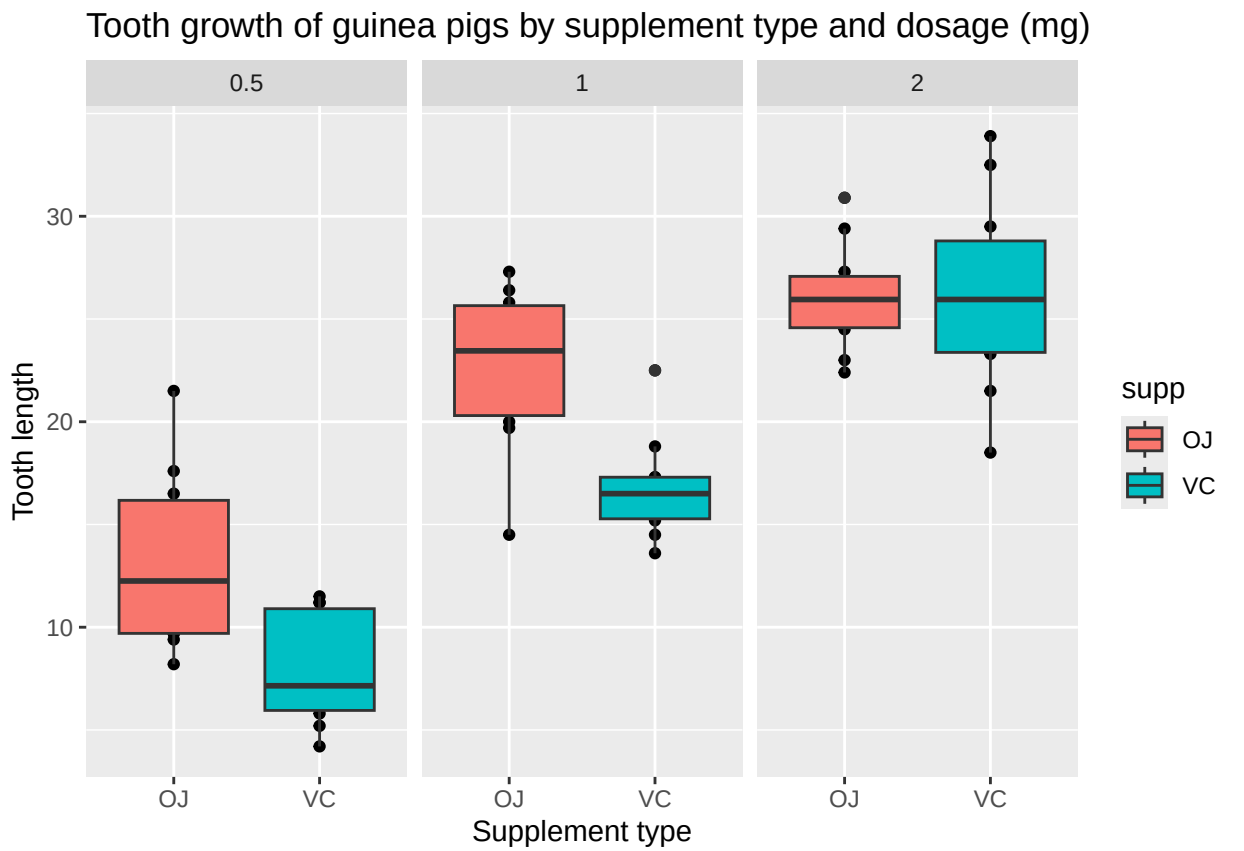
```r
# Calculatiing the mean of len based on the supplement methods
Supplement_mean = split(ToothGrowth$len, ToothGrowth$supp)
sapply(Supplement_mean, mean)
```

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

3. Using confidence intervals to compare growth of tooth by supplement dose

```r
qplot(supp,len,data=ToothGrowth,
      facets=~dose, main="Tooth growth of guinea pigs by supplement type and dosage (mg)",
      xlab="Supplement type",
      ylab="Tooth length") +
    geom_boxplot(aes(fill = supp))
```

```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



There appears to be a positive affect based on dosage as the dosage increases, the tooth grown also increases. Aside from the 2.0mg dosage, the OJ supplement generally records more tooth grown than VC.

4. Hypothesis Testing for Supplement

Assumptions: The variables must be independent and identically distributes (iid). - Variances of tooth growth are different when using different supplement and dosage. - Tooth growth follows a normal distribution.

Null Hypothesis: There is no difference in tooth growth when using supplement OJ or VC. Alternate Hypothesis: There will be more supplement growth for OJ than VC.

```
OJ <- ToothGrowth$len[ToothGrowth$supp == 'OJ']
VC <- ToothGrowth$len[ToothGrowth$supp == 'VC']
```

```
t.test(
    OJ,VC,
    alternative = "greater",
    paired = FALSE,
    var.equal = FALSE,
    conf.level = 0.95
)
```

Test

```
##
##  Welch Two Sample t-test
##
## data:  OJ and VC
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

Based on this One-tailed t-test we can see that the null is rejected (df = 55, p =.03). It can be concluded that the supplement OJ has a greater effect on tooth growth than supplement VC.

```
dose_half <- ToothGrowth$len[
    ToothGrowth$dose == 0.5
]
dose_one <- ToothGrowth$len[
    ToothGrowth$dose == 1
]
dose_two <- ToothGrowth$len[
    ToothGrowth$dose == 2
]
# One-tailed independant t-test with unequal variance
t.test(dose_half,
       dose_one,
       alternative = "less",
       paired = FALSE,
       var.equal = FALSE,
       conf.level = 0.95)
```

Hypothesis Test by Dosage

```
##
##  Welch Two Sample t-test
##
## data:  dose_half and dose_one
## t = -6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753323
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

```
t.test(
    dose_half,dose_two,
```

```
    alternative = "less",
    paired = FALSE,
    var.equal = FALSE,
    conf.level = 0.95
)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_half and dose_two
## t = -11.799, df = 36.883, p-value = 2.199e-14
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -13.27926
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

```
t.test(
    dose_one,dose_two,
    alternative = "less",
    paired = FALSE,
    var.equal = FALSE,
    conf.level = 0.95
)
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_one and dose_two
## t = -4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf -4.17387
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

Conclusion

When the dose is .5mg or .1mg there is a difference between the teeth growth supplements after taking OJ or VC. The assumption needed is we first assumed the whole population is normally distributed, then we assumed the population is normally distributed under each dose.