Statistical_Inference_Assignment 2

DC

29/05/2019

```r
library(ggplot2)
library(tidyverse)
```

Overview

This report will be the second of 2 assessments in the statistical inference class.

The given criteria is as follows:

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.

1) *Load the ToothGrowth data and perform some basic exploratory data analyses*

2) *Provide a basic summary of the data.*

3) *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)*

4) *State your conclusions and the assumptions needed for your conclusions.*

Tooth Growth Data

As the guidance for this assignment is fairly less directional then the first assignment, a good place to start, if not the only place, will be to take a look at the data we're to examine.
Taken from r's help function using `?ToothGrowth` we have this descriptive summary of the dataset:

ToothGrowth {datasets} R Documentation
The Effect of Vitamin C on Tooth Growth in Guinea Pigs/jump

Description
The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs.
Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day)
by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C
and coded as VC).

Usage: ToothGrowth
Format: A data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length
[,2] supp factor Supplement type (VC or OJ).
[,3] dose numeric Dose in milligrams/day

Source *C. I. Bliss (1952). The Statistics of Bioassay. Academic Press.*

References
McNeil, D. R. (1977). Interactive Data Analysis. New York: Wiley.

Crampton, E. W. (1947). The growth of the odontoblast of the incisor teeth as a criterion of vitamin C intake of the guinea pig. The Journal of Nutrition, 33(5), 491–504. doi: 10.1093/jn/33.5.491.

Exploratory Analysis of the Supplement Variable

Now we have some intuition on what our dataset is about, and the variables we have to work with, let's perform some basic exploratory analysis.

```
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 ...
```

```
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
```

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

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

Before we delve into any further, let's set our assumptions for the sup variable.
Since I am personally clueless when it comes to tooth growth or the differences in vitamin C administrations on rodents, and I've seen no evidence to sway my beliefs - I am going to set the ***null hypothesis*** to be as follows:

> H0: Mean tooth growth from OJ administration == mean tooth growth from VC administration across all doses

and the alternate hypothesis, H1, as:

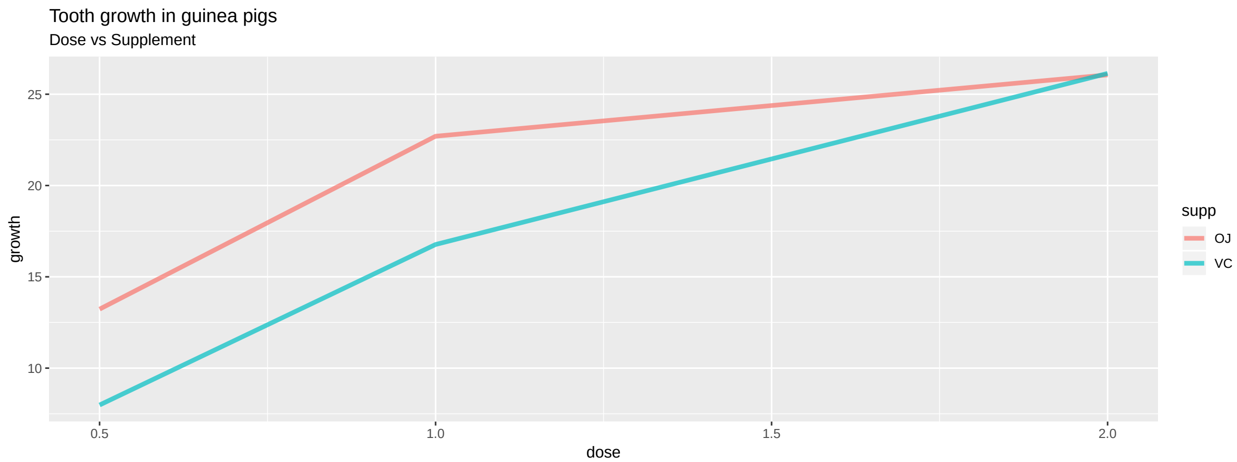
> H1: Mean tooth growth from OJ administration != mean tooth growth from VS administration across all doses

For this hypotheses, I will set our reject rate, or alpha, at a generic standard of **0.05**.
We will need a P-value of **0.05** or less from our following t.test to reject H0.

Given the tasks at hand, to compare tooth growth to supplement and dose, it makes sense to again build our intuition for any immediate correlations via some visualizations.

```
ToothGrowth2 <- ToothGrowth %>%
    group_by(supp, dose) %>%
    summarise(growth = mean(len))

ggplot(data = ToothGrowth2, aes(x = dose, y = growth, colour = supp)) +
    geom_line(lwd = 1.5, alpha = 0.7) +
    ggtitle(label = "Tooth growth in guinea pigs", subtitle = "Dose vs Supplement")
```



From the above plot we can determine that there does indeed appear to be an advantage of providing orange juice over absorbic acid. However, both supplements fall inline with each other as the dose exceeds 1.0mg, matching up at 2.0mg.

Supplement vs Tooth Growth Testing

To back these preliminary findings up we will run the ToothData dataset through R's `t.test` function. But first, let's take a quick look whether we need to set the variance argument to TRUE. We can do this by using R's `var` function on the len variable of each of the supplement groups.

```
OJ_var <- var(ToothGrowth[ToothGrowth$supp == "OJ",]$len)
VC_var <- var(ToothGrowth[ToothGrowth$supp == "VC",]$len)
```

```
## OJ Var: 43.6334367816092
```

```
## VC Var: 68.3272298850575
```

Ok so we don't need to be math Gods to figure out that we need to set var.equal to FALSE. Additionally, since we're looking for the probability that H1 mean isn't equal to H0 mean then we need to use a two sided test.
Let's move on to splitting the data into OJ and VC variables and then performing the t.test.

```
OJ <- ToothGrowth[ToothGrowth$supp == "OJ", ]
VC <- ToothGrowth[ToothGrowth$supp == "VC", ]

t.test(OJ$len, VC$len, var.equal = F, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ$len and VC$len
## 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 of x mean of y
##  20.66333  16.96333
```

Outcome of Supplement Analysis

From the resulting t.test we can see that we have a confidence interval of *-0.1710156 to 7.5710156*. This interval, contains **0**, so we can't rule out the possibility that there is no difference between the means.
We can also see that this test returns a p-value of **0.06063**. This is above or pre-determined rejection rate so we **fail to reject H0: Mean tooth growth from OJ administration == mean tooth growth from VC administration across all doses**.

Exploratory Analysis of the Dose Variable

Again, before we delve deep into the analysis, let's set up our hypothesis's.
Following on from the first confession of ignorance, I'm coerced to admit that I again am clueless in the subject of rodent dentistry, therefore I will set the hypotheses for this inference as follows:

> H0: Mean tooth growth vs Supplement and Dose == 0
>
> H1: Mean tooth growth vs Supplement and Dose != 0

From the initial plot we can see that there does appear to be a positive linear correlation between dose and toothgrowth. However, this advantage appears to experience a reduced effect at doses over 1.0mg. Regardless of any reduced effects, the higher the dose also seems to provide the largest mean toothgrowth.
As is expected, let's back this up with some statistics!

Supplement vs Tooth Growth & Dose Testing

To being, we will need to subset our data into both supplements, and all dose levels.

```r
OJ_0.5 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5,]
OJ_1.0 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1,]
OJ_2.0 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2,]

VC_0.5 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5,]
VC_1.0 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 1,]
VC_2.0 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == 2,]
```

Now we have our data in a usable format, let's run through the t.tests for each dose and then move on to analyse the results.

0.5mg test

```r
t.test(OJ_0.5$len, VC_0.5$len, var.equal = FALSE, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ_0.5$len and VC_0.5$len
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

1.0mg test

```
t.test(OJ_1.0$len, VC_1.0$len, var.equal = FALSE, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ_1.0$len and VC_1.0$len
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

2.0mg test

```
t.test(OJ_2.0$len, VC_2.0$len, var.equal = FALSE, alternative = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ_2.0$len and VC_2.0$len
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

Outcome of Supplement vs Tooth Growth Testing

Now that we have our test results, let's create a table below to summarise what we've discovered.
I will highlight all statistically significant results using a **bold** font. These will be results which either do not contain 0 in the 95% confidence interval, or have a p-value <= the alpha we set at 0.05.

Statistic	0.5mg	1.0mg	2.0mg
Confidence Interval	**1.719057 - 8.780943**	**2.802148 - 9.057852**	-3.79807 - 3.63807
P-value	**0.006359**	**0.001038**	0.9639

Conclusion

Whilst we were unable to reject the null hypotheses for analyzing supplement vs tooth growth across all doses, we were able to find statistical significant results for supplement vs tooth growth for doses under 2.0mg. From this, we can conclude that for doses of **under 2.0mg**, providing **OJ** as a supplement will result in increased tooth growth.