StatInf_Assignment2_ToothGrowth

Willianto Asalim

17/05/2020

Statistical Inference Proj 2: Basic Inferential Data Analysis

---

The Effect Of Vitamin C On Tooth Growth In Guinea Pigs

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). Info on ToothGrowth dataset can be found at the following link: rdocumentation website <- click here

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

---

```
library(datasets) ##Load R Builtin Data sets
data("ToothGrowth") ##Load ToothGrowth Datasets
```

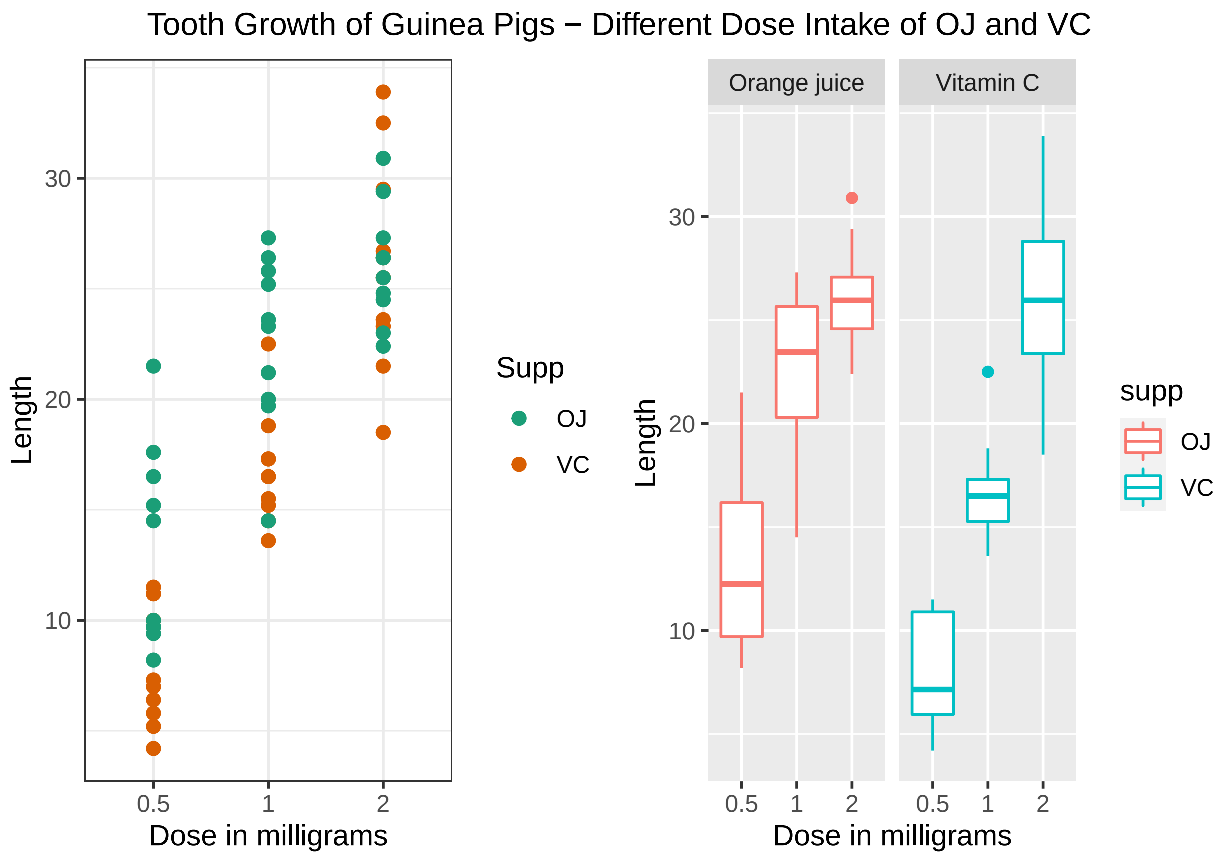
```
dim(ToothGrowth) ##No. observations and variables
names(ToothGrowth) ##Variables names
kable(head(ToothGrowth)) ##First 6 rows of data
```

```
## [1] 60  3
## [1] "len"  "supp" "dose"
```

len	supp	dose
4.2	VC	0.5
11.5	VC	0.5
7.3	VC	0.5
5.8	VC	0.5
6.4	VC	0.5
10.0	VC	0.5

```
library(gridExtra) ##Provide side by side plotting
edaPlot <- ggplot(ToothGrowth, aes(factor(dose), len, color = supp))  ##ggplot basic info
p1 <- edaPlot + geom_point(size = 2) + theme_bw() +
        labs(x = "Dose in milligrams", y = "Length", color = "Supp") +
        scale_color_brewer(palette = "Dark2")
p2 <- edaPlot + geom_boxplot() +
      facet_grid(.~supp, labeller = as_labeller(
            c("OJ" = "Orange juice", "VC" = "Vitamin C"))) +
      labs(x = "Dose in milligrams", y = "Length") +
```

```
      scale_fill_discrete(name = "Dosage of\nvitamin C\nin mg/day")
## Side by side plotting
grid.arrange(p1, p2, ncol=2,
             top = "Tooth Growth of Guinea Pigs - Different Dose Intake of OJ and VC")
```



The ToothGrowth Dataset contains 60 observations and 3 variables: len, supp, dose. From observing the plots it seems that the 0.5 and 1.0 dose level of OJ intake has a longer tooth length than the VC intake. The 2.0 dose level of VC intake has longer tooth length for the guinea pigs.

2. Provide a basic summary of the data.

```
kable(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	NA	Median :1.000
Mean :18.81	NA	Mean :1.167
3rd Qu.:25.27	NA	3rd Qu.:2.000
Max. :33.90	NA	Max. :2.000

3. Use conf. intervals or hypothesis tests to compare tooth growth by supp & dose

---

Based on the exploratory data analysis and the summary of the ToothGrowth data performed above we know that there are a total of 60 guinea pigs: 30 are fed with OJ with 3 different dose levels and another 30 are fed with VC with 3 different dose levels. The hypothesis tests will be performed using th P value approach, if there is a significance difference with different dosage and supplement type. There are different tests we can carry out in the ToothGrowth dataset and in this instance we will be conducting 4 tests.

- Test based on 2 different supplements ie OJ vs VC

- Test based on 3 different dose levels in the same supplement. ie half dose vs one dose vs two dose in OJ supplement vice versa.

3 different dose levels data created (Result refer to the appendix)

```
## Dataset based on 0.5 dose
lowDose <- subset(ToothGrowth, dose %in% c("0.5"))
## Dataset based on 1.0 dose
midDose <- subset(ToothGrowth, dose %in% c("1"))
## Dataset based on 2.0 dose
highDose <- subset(ToothGrowth, dose %in% c("2"))
```

T Test based on supplement type (Result refer to Appendix)

```
## T Test on different supplements OJ vs VC
test1 <- t.test(len~supp, data = ToothGrowth, paired = FALSE, var.equal = TRUE)
```

T Test based on different dose of OJ vs VC (Result refer to Appendix)

```
##Comparing 0.5 dose level tooth growth based supplements
lowDoseTest <- t.test(len~supp, data=subset(lowDose, supp%in%c("OJ", "VC")),
                      paired=F, var.equal=T)
##Comparing 1.0 dose level tooth growth based supplements
midDoseTest <- t.test(len~supp, data=subset(midDose, supp%in%c("OJ", "VC")),
                      paired=F, var.equal=T)
##Comparing 2.0 dose level tooth growth based supplements
highDoseTest <- t.test(len~supp, data=subset(highDose, supp%in%c("OJ", "VC")),
                       paired=F, var.equal=T)
```

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

To measure the tooth growth impact of the guinea pigs due to the different dose levels and supplements, more than 95% confidence interval is required because anything less than that there is no significance difference for scientific studies. Therefore the P value must be less than 5% or 0.05 for significance difference to occur. If the P value is less than 0.05, it is likely that the supplement or dose level has impact on the tooth growth. If the P value is more than 0.05, it is unlikely that the supplement or dose level has impact on the tooth growth.

Based on the first T test conducted, the P value of guinea pigs consuming OJ or VC is 0.0603934. The P value is more than 0.05 which means there is no significance difference whether the guinea pigs consume OJ or VC to enhance the tooth growth. However for scientific research purpose we would like to check whether the different dose level of OJ or VC intake has any impact on the tooth growth of the guinea pigs. Observing the T tests conducted on different dose level 0.5 and 1.0, the P Values are 0.0053037 and 0.0007807262 respectively. These P values are less than 0.05 for both dose levels of supplements which means there is a significance difference. Guinea pigs that consumed 0.5 and 1.0 dose level of OJ will have higher tooth growth than the guinea pigs that consumed 0.5 and 1.0 dose level of VC. Whereas the dose level of 2.0, the P value is 0.9637098. The P value is more than 0.05 which means there is no significance difference in tooth growth whether the guinea pigs consumed 2.0 dose level of OJ or VC.

5. Appendix

T Test Results: T Test based on supplements - OJ vs VC

Table 3: Two Sample t-test: `len` by `supp` (continued below)

Test statistic	df	P value	Alternative hypothesis	mean in group OJ
1.915	58	0.06039	two.sided	20.66

mean in group VC
16.96

T Test based on 0.5 dose level of OJ vs VC

Table 5: Two Sample t-test: `len` by `supp` (continued below)

Test statistic	df	P value	Alternative hypothesis	mean in group OJ
3.17	18	0.005304 * *	two.sided	13.23

mean in group VC
7.98

T Test based on 1.0 dose level of OJ vs VC

Table 7: Two Sample t-test: `len` by `supp` (continued below)

Test statistic	df	P value	Alternative hypothesis
4.033	18	0.0007807 * * *	two.sided

mean in group OJ	mean in group VC
22.7	16.77

T Test based on 2.0 dose level of OJ vs VC

Table 9: Two Sample t-test: `len` by `supp` (continued below)

Test statistic	df	P value	Alternative hypothesis	mean in group OJ
-0.04614	18	0.9637	two.sided	26.06

mean in group VC
26.14

3 different dose levels data of both OJ and VC: 0.5 Low Dose Level Data

	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
7	11.2	VC	0.5
8	11.2	VC	0.5
9	5.2	VC	0.5
10	7.0	VC	0.5
31	15.2	OJ	0.5
32	21.5	OJ	0.5
33	17.6	OJ	0.5
34	9.7	OJ	0.5
35	14.5	OJ	0.5
36	10.0	OJ	0.5
37	8.2	OJ	0.5
38	9.4	OJ	0.5
39	16.5	OJ	0.5
40	9.7	OJ	0.5

1.0 Mid Dose Level Data

	len	supp	dose
11	16.5	VC	1
12	16.5	VC	1
13	15.2	VC	1
14	17.3	VC	1
15	22.5	VC	1
16	17.3	VC	1
17	13.6	VC	1
18	14.5	VC	1
19	18.8	VC	1
20	15.5	VC	1
41	19.7	OJ	1
42	23.3	OJ	1
43	23.6	OJ	1
44	26.4	OJ	1
45	20.0	OJ	1
46	25.2	OJ	1
47	25.8	OJ	1
48	21.2	OJ	1
49	14.5	OJ	1

	len	supp	dose
50	27.3	OJ	1

2.0 High Dose Level Data

	len	supp	dose
21	23.6	VC	2
22	18.5	VC	2
23	33.9	VC	2
24	25.5	VC	2
25	26.4	VC	2
26	32.5	VC	2
27	26.7	VC	2
28	21.5	VC	2
29	23.3	VC	2
30	29.5	VC	2
51	25.5	OJ	2
52	26.4	OJ	2
53	22.4	OJ	2
54	24.5	OJ	2
55	24.8	OJ	2
56	30.9	OJ	2
57	26.4	OJ	2
58	27.3	OJ	2
59	29.4	OJ	2
60	23.0	OJ	2

The platform specification used:

Spec	Description
OS	Windows 10 Pro - 64 bit
CPU	AMD Ryzen 5 - 3400G
RAM	16GB DDR4 3000MHz
Storage	500GB SSD - M.2 NVMe (PCIe)
Tool	RStudio