Exercise part 2 (Toothgrowth)

GS

4/8/2021

```
library(tidyverse)
library(ggpubr)
library(tinytex)
theme_set(theme_bw())
```

The Effect of Vitamin C on Tooth Growth in Guinea Pigs

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).

I will compare the guinea tooth growth by supplement and dose. Firstly, I will do some basic EDA and then I will do some comparisons

Loading datset

```
data(ToothGrowth)
tg <- ToothGrowth
```

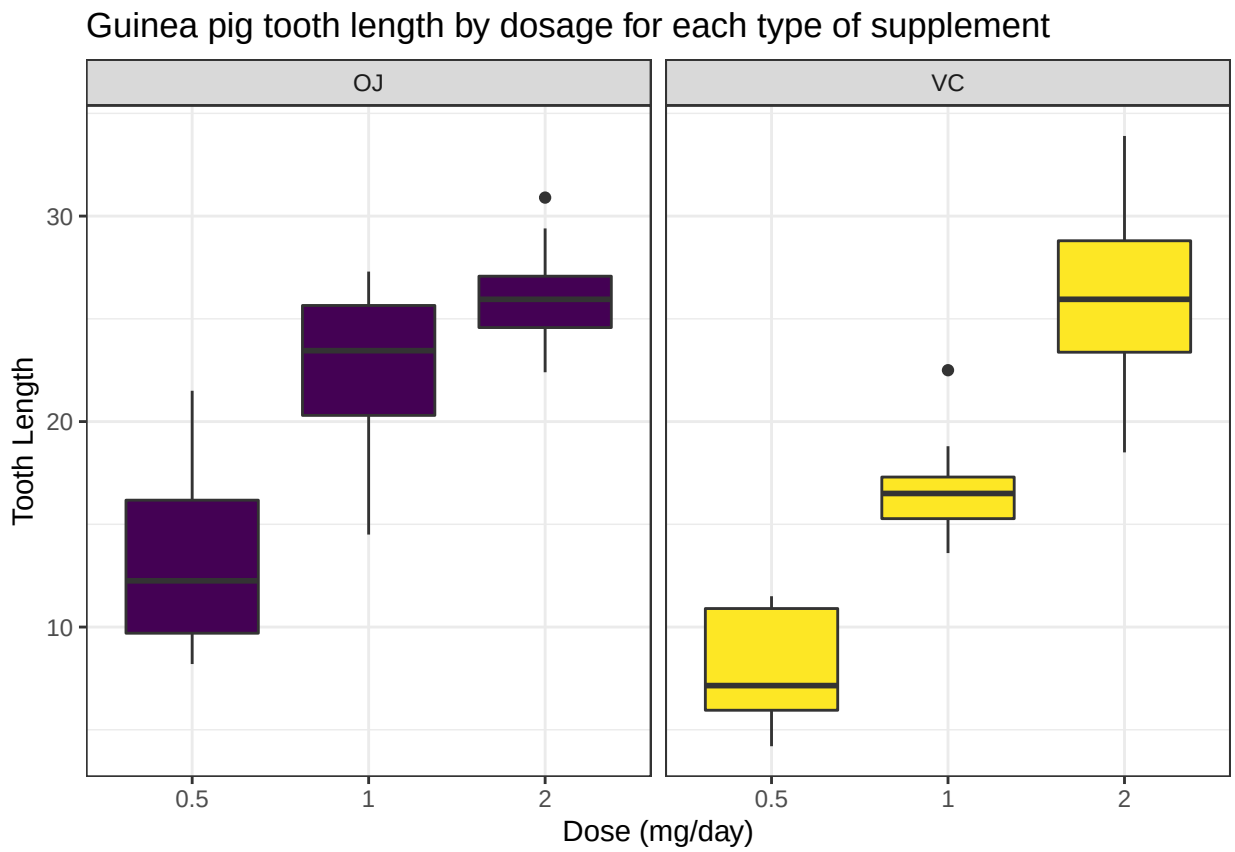
Exploratory data analysis

```
glimpse(tg)
```

```
Rows: 60
Columns: 3
$ len  <dbl> 4.2, 11.5, 7.3, 5.8, 6.4, 10.0, 11.2, 11.2, 5.2, 7.0, 16.5, 16.5,~
$ supp <fct> VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, V~
$ dose <dbl> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 1.0, 1.0, 1.0, ~
```

```
ggplot(tg, aes(x=factor(dose), y=len, fill = supp)) +
  geom_boxplot() +
  facet_grid(.~supp) +
```

```
  ggtitle('Guinea pig tooth length by dosage for each type of supplement') +
  xlab('Dose (mg/day)') +
  ylab('Tooth Length') +
  theme(legend.position = 'none') +
  scale_fill_viridis_d()
```



The box plots seem to show, increasing the dosage increases the tooth growth. Orange juice seem to do it better in lower dose but at high doses does not seem to be a diference.

Checking if data is normally distributed

```
norm.test <- shapiro.test(tg$len)

norm.test.group <- tg %>%
  group_by(supp) %>%
  summarize(n = n(),
            shapiro = shapiro.test(len)$p.value)
```

```
[1] "Tha variable len is normally distributed, p-value:  0.1091"
```

```
[1] "Tha variable len subset QJ is not normally distributed, p-value:  0.0236"
```

```
[1] "Tha variable len subset VC is normally distributed, p-value:  0.4284"
```

Results

Hypothesis testing comparing tooth growth by type of supplement

H0 = No difference between groups (mean difference equal to 0)

H1 = Groups are different (means difference not equal to 0)

```
hypo_1<-t.test(len ~ supp, data = tg)
```

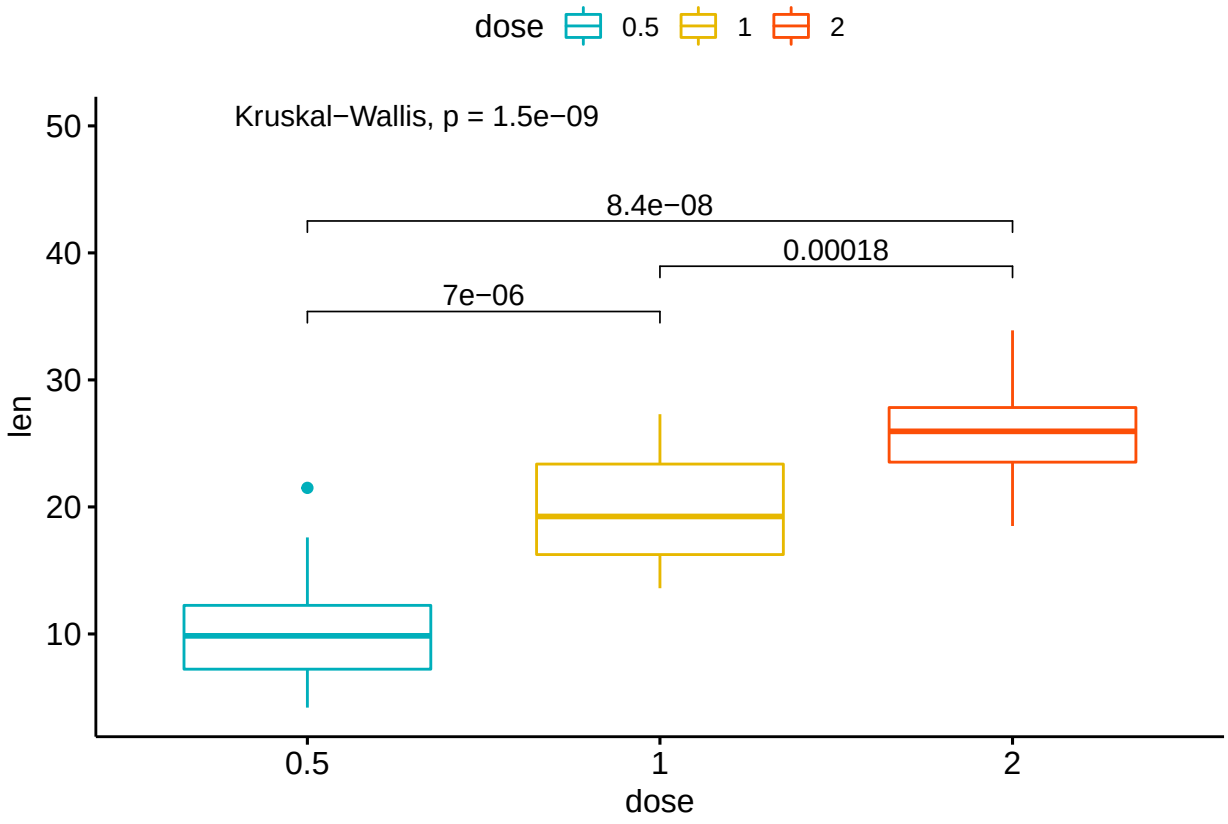
Interpretation:

- p.value: 0.06063 (greater than 0.05)
- 95%CI: includes 0 [-0.1710 - 7.5710]
- We cannot reject the null hypothesis (H0)

Graphical comparisons of the 3 doses in term of len (Kruskall - Wallis test)

```
g <- ggboxplot(tg, x = "dose", y = "len",
          color = "dose", palette =c("#00AFBB", "#E7B800", "#FC4E07"),
          shape = "dose")

my_comparisons <- list( c("0.5", "1"), c("1", "2"), c("0.5", "2") )
g + stat_compare_means(comparisons = my_comparisons) +
  stat_compare_means(label.y = 50)
```



Interpretation:

- H0: all groups are equal
- H1: at least one group is diferrent

Comparison of different doses

Dose: 0.5mg/day

```
hypo_2<-t.test(len ~ supp, data = subset(tg, dose == 0.5))
```

Interpretation:

- p.value: 0.006359 (lower than 0.05)
- 95%CI: does not include 0 [1.7190 - 8.7809]
- We reject the null hypothesis (H0) of equality

Dose: 1mg/day

```
hypo_3<-t.test(len ~ supp, data = subset(tg, dose == 1))
```

Interpretation:

- p.value: 0.001038 (lower than 0.05)
- 95%CI: does not include 0 [2.8021 - 9.0578]
- We reject the null hypothesis (H0) of equality

Dose: 2mg/day

```
hypo_4<-t.test(len ~ supp, data = subset(tg, dose == 2))
```

Interpretation:

- p.value: 0.9639 (greater than 0.05)
- 95%CI: does not include 0 [-3.7980 - 3.6380]
- We cannot reject the null hypothesis (H0) of equality

CONCLUSIONS

Overall, orange juice (OJ) produce the same amount of toth grwoth than ascorbic acid (AC).

By dosages, OJ produces more tooth growth than ascorbic acid AC for dosages of 0.5mg/day & 1.0mg/day and OJ and AC produces the same amount of tooth growth for dose og 2.0 mg/day.

Assumption: no other variable are influencing the tooth growth (eg.feeding)

```
sessionInfo()
```

```
R version 4.1.0 (2021-05-18)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 19043)

Matrix products: default

locale:
[1] LC_COLLATE=Spanish_Mexico.1252  LC_CTYPE=Spanish_Mexico.1252
[3] LC_MONETARY=Spanish_Mexico.1252 LC_NUMERIC=C
[5] LC_TIME=Spanish_Mexico.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] tinytex_0.32    ggpubr_0.4.0    forcats_0.5.1   stringr_1.4.0
 [5] dplyr_1.0.7     purrr_0.3.4     readr_1.4.0     tidyr_1.1.3
 [9] tibble_3.1.2    ggplot2_3.3.5   tidyverse_1.3.1

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.7         lubridate_1.7.10   assertthat_0.2.1   digest_0.6.27
 [5] utf8_1.2.1         R6_2.5.0           cellranger_1.1.0   backports_1.2.1
 [9] reprex_2.0.0       evaluate_0.14      highr_0.9          httr_1.4.2
[13] pillar_1.6.1       rlang_0.4.11       curl_4.3.2         readxl_1.3.1
[17] rstudioapi_0.13    data.table_1.14.0  car_3.0-11         rmarkdown_2.9
[21] labeling_0.4.2     foreign_0.8-81     munsell_0.5.0      broom_0.7.8
[25] compiler_4.1.0     modelr_0.1.8       xfun_0.24          pkgconfig_2.0.3
[29] htmltools_0.5.1.1  tidyselect_1.1.1   rio_0.5.27         viridisLite_0.4.0
[33] fansi_0.5.0        crayon_1.4.1       dbplyr_2.1.1       withr_2.4.2
[37] grid_4.1.0         jsonlite_1.7.2     gtable_0.3.0       lifecycle_1.0.0
[41] DBI_1.1.1          magrittr_2.0.1     scales_1.1.1       zip_2.2.0
[45] cli_3.0.0          stringi_1.6.2      carData_3.0-4      farver_2.1.0
[49] ggsignif_0.6.2     fs_1.5.0           xml2_1.3.2         ellipsis_0.3.2
[53] generics_0.1.0     vctrs_0.3.8        openxlsx_4.2.4     tools_4.1.0
[57] glue_1.4.2         hms_1.1.0          abind_1.4-5        yaml_2.2.1
[61] colorspace_2.0-2   rstatix_0.7.0      rvest_1.0.0        knitr_1.33
[65] haven_2.4.1
```