Tooth Growth Analysis

About the data

Reviewing the help file "?ToothGrowth", we can see the breakdown as follows:

- A Study on the effect of two delivery methods Orange Juice or plain old Vitamin C for tooth growth.
- Sample size of 60 guinea pigs.
- Each animal recieving one of each three doses which are equally divided among 60 guinea pigs.
  - 20 receiving .05 mg OJ/VC, 20 receiving 1 mg OJ/VC, and 20 receiving 2 mg OJ/VC
- The reference link is a paywall but provides us with;
  - a title name "The Growth of the Odontoblasts of the **Incisor** Tooth as a Criterion of the Vitamin C Intake of the Guinea Pig: Five Figures"
  - Date Published: 01 May 1947.
  - First page of the article summarises "Canadian Government undertaking study to provide good source of vitamin c for soldiers during war time."

str(ToothGrowth) provided the technical information of the dataframe.

Environment

 view mycode

```
library(tidyverse)
library(ggpubr)
library(glue)
library(cowplot)
library(tinytex)
library(matrixStats)
downloadDate <- date()

glue("Simulation date is {downloadDate}")
```

```
## Simulation date is Thu May 21 12:29:30 2020
```

```
sessionInfo()
```

```
## R version 4.0.0 (2020-04-24)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04 LTS
##
## Matrix products: default
```

```
## BLAS:   /usr/lib/x86_64-linux-gnu/blas/libblas.so.3.9.0
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.9.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
##  [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
##  [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
##  [9] LC_ADDRESS=C               LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] matrixStats_0.56.0 tinytex_0.23       cowplot_1.0.0      glue_1.4.1
##  [5] ggpubr_0.3.0       forcats_0.5.0      stringr_1.4.0      dplyr_0.8.5
##  [9] purrr_0.3.4        readr_1.3.1        tidyr_1.1.0        tibble_3.0.1
## [13] ggplot2_3.3.0      tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.4.6      lubridate_1.7.8   lattice_0.20-41   assertthat_0.2.1
##  [5] digest_0.6.25     R6_2.4.1          cellranger_1.1.0  backports_1.1.7
##  [9] reprex_0.3.0      evaluate_0.14     httr_1.4.1        pillar_1.4.4
## [13] rlang_0.4.6       curl_4.3          readxl_1.3.1      rstudioapi_0.11
## [17] data.table_1.12.8 car_3.0-7         rmarkdown_2.1     foreign_0.8-79
## [21] munsell_0.5.0     broom_0.5.6       compiler_4.0.0    modelr_0.1.8
## [25] xfun_0.14         pkgconfig_2.0.3   htmltools_0.4.0   tidyselect_1.1.0
## [29] rio_0.5.16        fansi_0.4.1       crayon_1.3.4      dbplyr_1.4.3
## [33] withr_2.2.0       grid_4.0.0        nlme_3.1-147      jsonlite_1.6.1
## [37] gtable_0.3.0      lifecycle_0.2.0   DBI_1.1.0         magrittr_1.5
## [41] scales_1.1.1      zip_2.0.4         cli_2.0.2         stringi_1.4.6
## [45] carData_3.0-3     ggsignif_0.6.0    fs_1.4.1          xml2_1.3.2
## [49] ellipsis_0.3.1    generics_0.0.2    vctrs_0.3.0       openxlsx_4.1.5
## [53] tools_4.0.0       hms_0.5.3         abind_1.4-5       yaml_2.2.1
## [57] colorspace_1.4-1  rstatix_0.5.0     rvest_0.3.5       knitr_1.28
## [61] haven_2.2.0
```

Data Loading

```
attach(ToothGrowth)
```

Format

ToothGrowth shows 60 observations on 3 variables

- len - tooth length
- supp - Supplement type (VC or OJ)
- dose - in milligrams/day

Using view(toothGrowth) we can see the following:

- range 1:30 len have a supp valuing “VC”
- range 31:60 len have a supp valuing “OJ”

```
str(ToothGrowth)
```

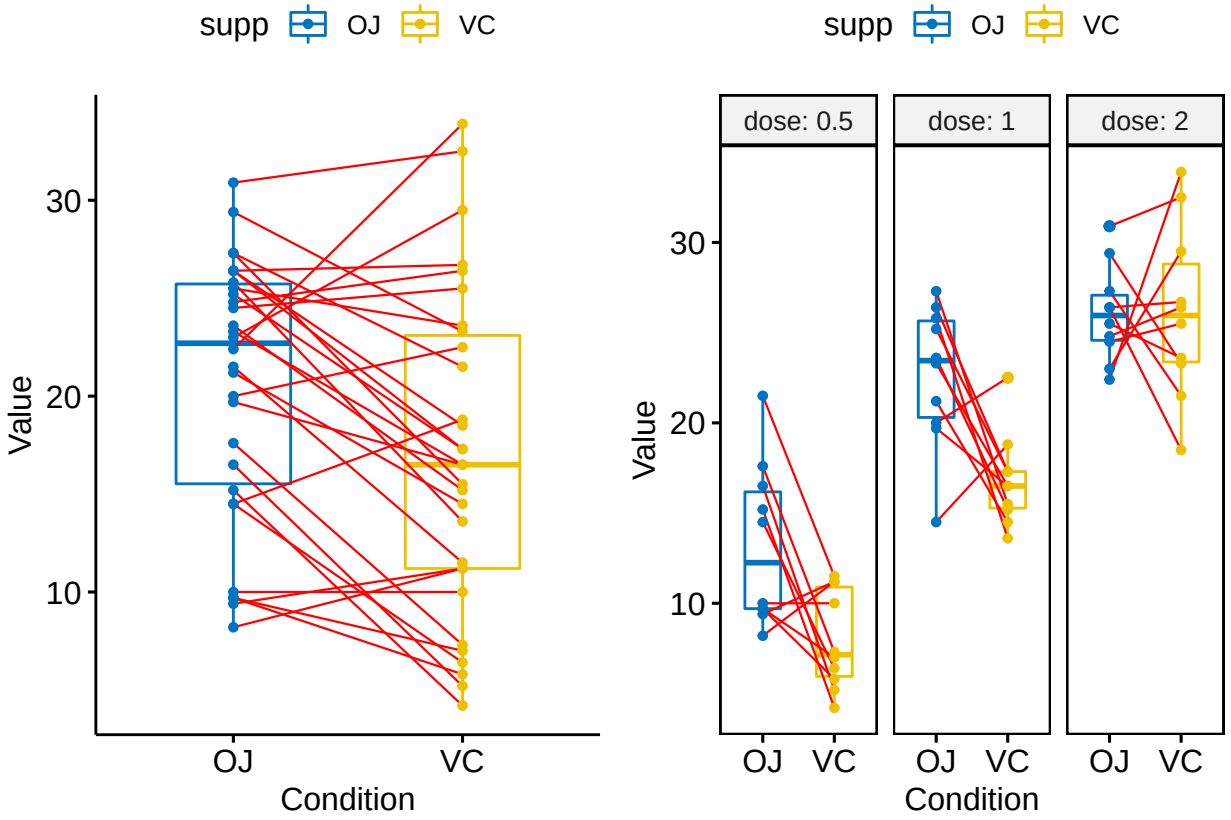
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
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
```

Exploring the data

```
e<-ggpaired(ToothGrowth, x = "supp", y = "len",
         color = "supp", line.color = "red", line.size = 0.4,
         palette = "jco")


e1 <- ggpaired(ToothGrowth, x = "supp", y = "len",
          color = "supp", palette = "jco",
          line.color = "red", line.size = 0.4,
          facet.by = "dose", short.panel.labs = FALSE)


plot_grid(e,e1)
```



The graph on the left tells us the following:

- The median is higher in OJ vs VC
- The range is more spread out in VC compared to OJ
- The OJ minimum point is higher than VC
- The VC max point is higher than OJ
- Look at the quantiles below to see the spread of data across the boxplot
- OJ quantiles between 50%-75% are in 22.70-25.725 vs. VC which are 16.5-23.1 the distance between them are also staggering 25.725-22.70 = 3.025 for OJ vs. 23.1-16.5 = 6.6. Mother nature is very effective at delivering vitamin C through OJ.

```
## $OJ
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20   15.53   22.70   20.66   25.73   30.90
##
## $VC
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20   11.20   16.50   16.96   23.10   33.90
```

The graph on the right tell us the following:

- The median increase as the dosage increase
- VC delivery method is not an effective way of sourcing vitamin c unless it reaches the dose 2 mg per day.
  - During wartime, carrying an orange juice vs. a tablet is probably a different story.

- Orange juice is very effective delivery method for vitamin c.
- The OJ range of the box plot on 2 mg decreases - sign of diminishing returns?
- Their respective p-values
- Check-out the quantiles below

```
## : OJ
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    8.20    9.70   12.25   13.23   16.18   21.50
## ------------------------------------------------------------
## : VC
## : 0.5
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    4.20    5.95    7.15    7.98   10.90   11.50
## ------------------------------------------------------------
## : OJ
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   14.50   20.30   23.45   22.70   25.65   27.30
## ------------------------------------------------------------
## : VC
## : 1
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.60   15.28   16.50   16.77   17.30   22.50
## ------------------------------------------------------------
## : OJ
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   22.40   24.57   25.95   26.06   27.07   30.90
## ------------------------------------------------------------
## : VC
## : 2
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.50   23.38   25.95   26.14   28.80   33.90
```

Hypothesis Testing

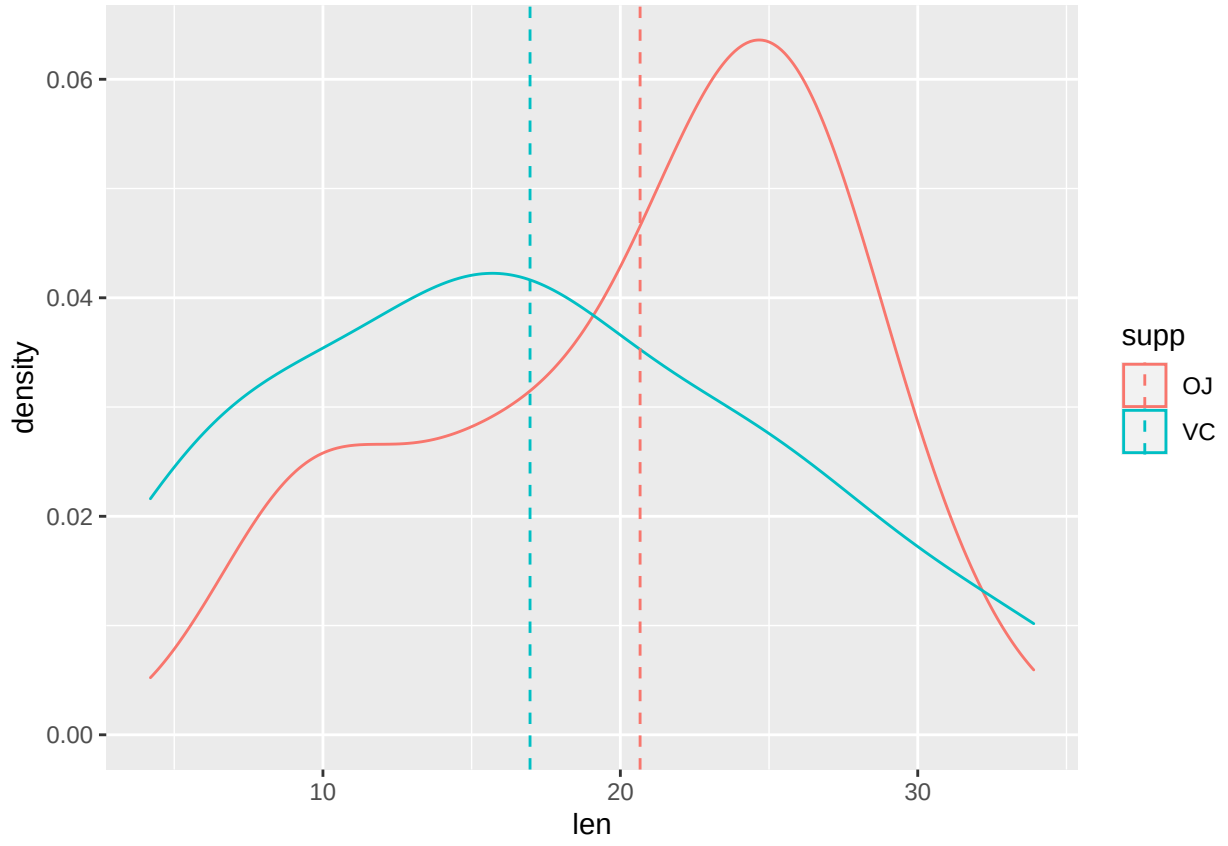
H0 = mean of both OJ and VC are equal and when subtracted should equal 0 Ha != H0 There is a difference in mean for VC and OJ

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

t.test(OJ$len,VC$len, paired = T, var.equal = T)
```

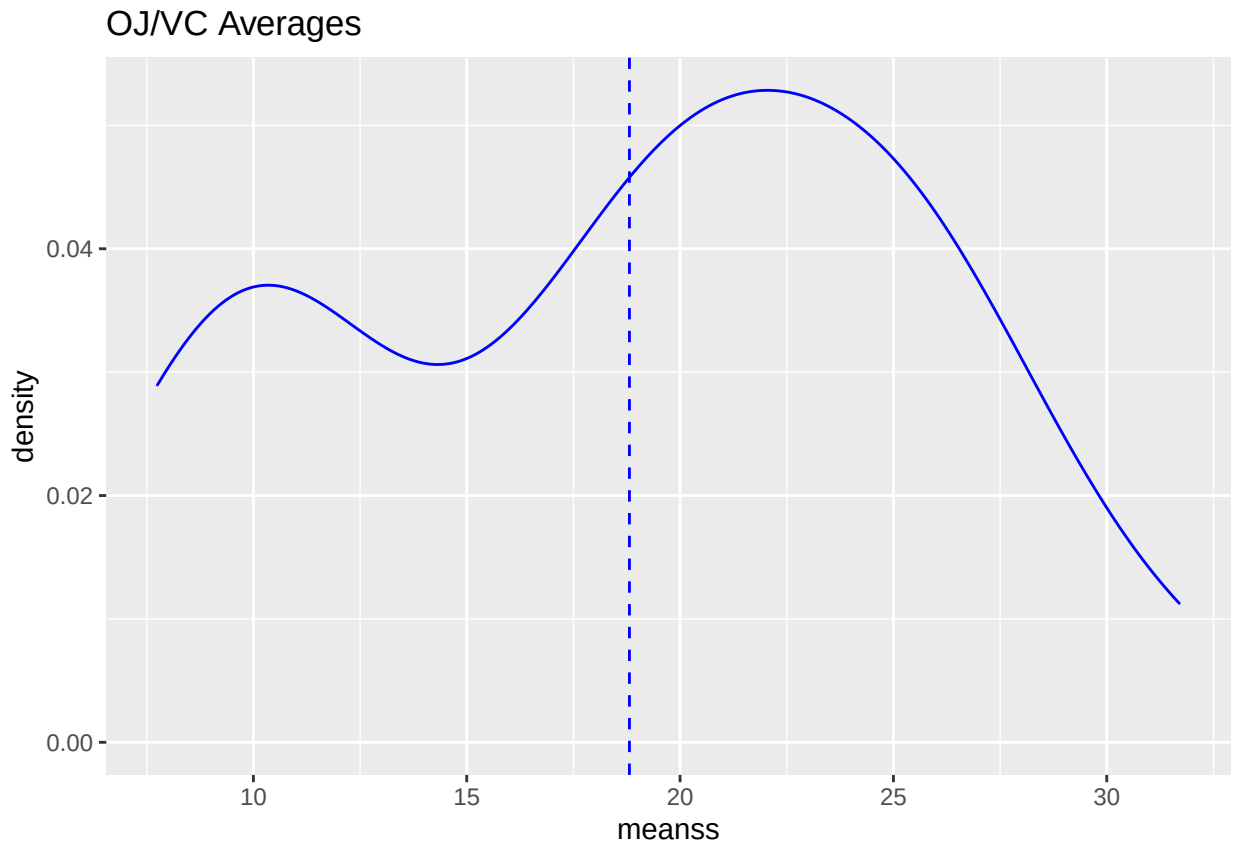
```
##
##  Paired t-test
##
## data:  OJ$len and VC$len
## t = 3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.408659 5.991341
```

```
## sample estimates:
## mean of the differences
##                     3.7
```

```
mu <- ToothGrowth %>% group_by(supp) %>% dplyr::summarise("means"=mean(len))

ggplot(data = ToothGrowth, aes(x=len, color = supp))+geom_density()+
  geom_vline(data=mu, aes(xintercept=means, color=supp),
             linetype="dashed")
```




```
x <- OJ$len
y <- VC$len

OJVC <-data.frame(x,y)
combOJVC <- OJVC %>% mutate("meanss" =rowMeans(OJVC))
mu1<-mean(combOJVC$meanss)

ggplot()+geom_density(data = combOJVC, aes(x=meanss), color="blue")+geom_vline(xintercept = mu1, linety
```

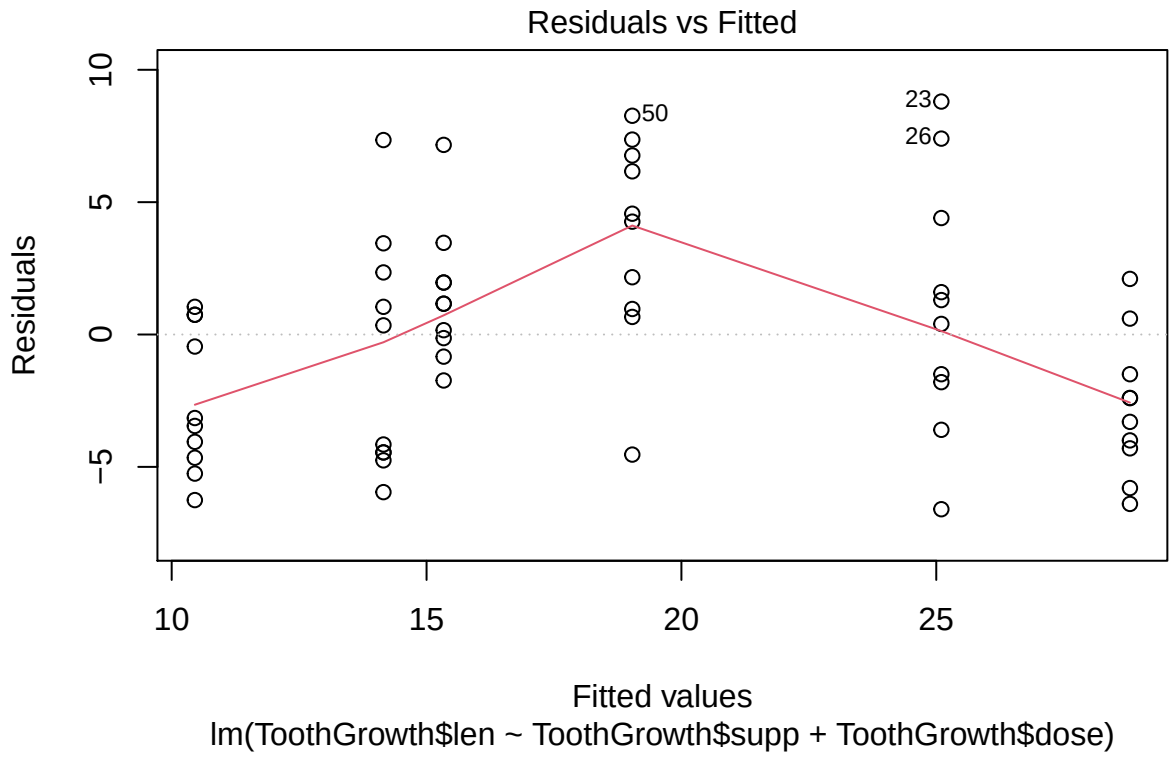


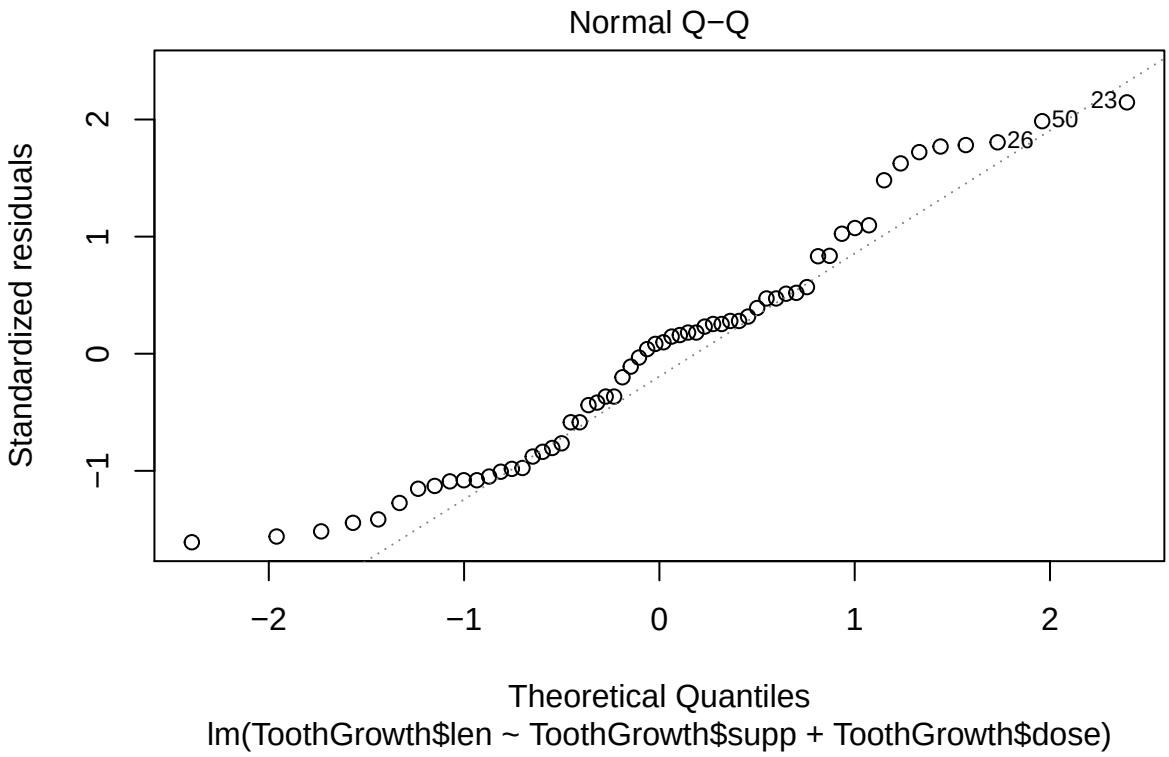

```
lmfit <- lm(ToothGrowth$len~ToothGrowth$supp+ToothGrowth$dose)
summary(lmfit)
```

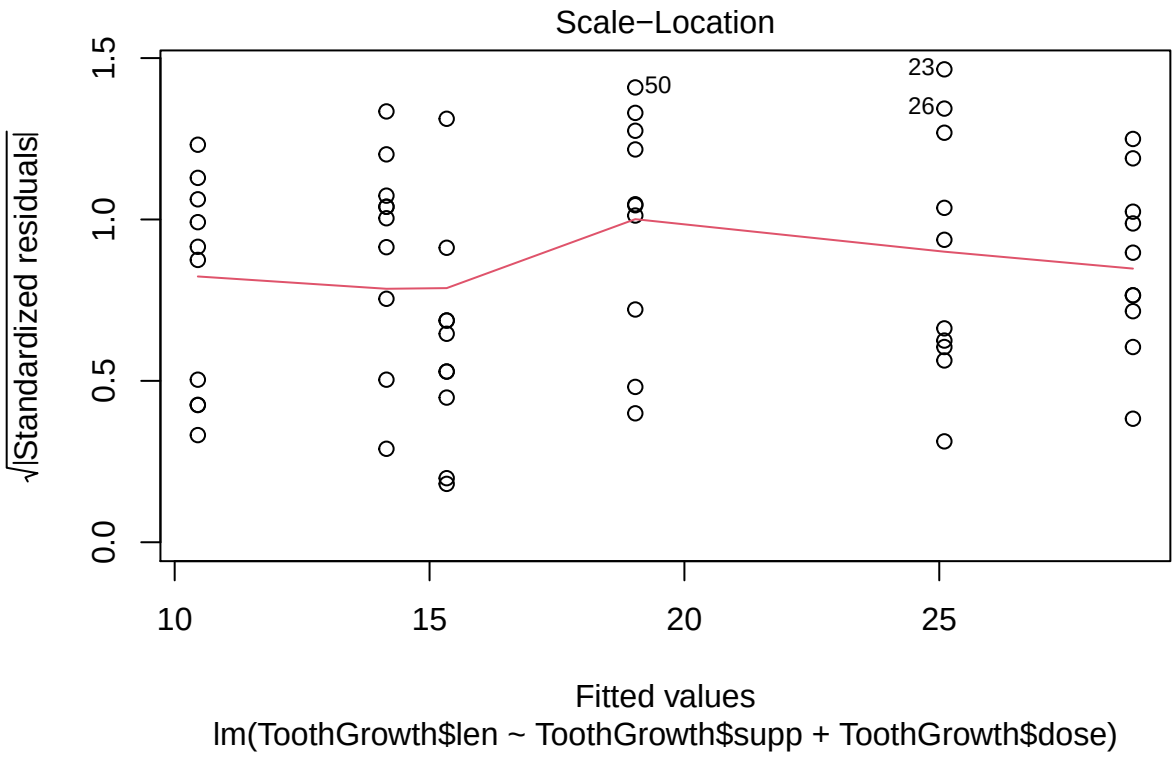
```
##
## Call:
## lm(formula = ToothGrowth$len ~ ToothGrowth$supp + ToothGrowth$dose)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.600 -3.700  0.373  2.116  8.800
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           9.2725     1.2824   7.231 1.31e-09 ***
## ToothGrowth$suppVC   -3.7000     1.0936  -3.383   0.0013 **
## ToothGrowth$dose      9.7636     0.8768  11.135 6.31e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.236 on 57 degrees of freedom
## Multiple R-squared:  0.7038, Adjusted R-squared:  0.6934
## F-statistic: 67.72 on 2 and 57 DF,  p-value: 8.716e-16
```

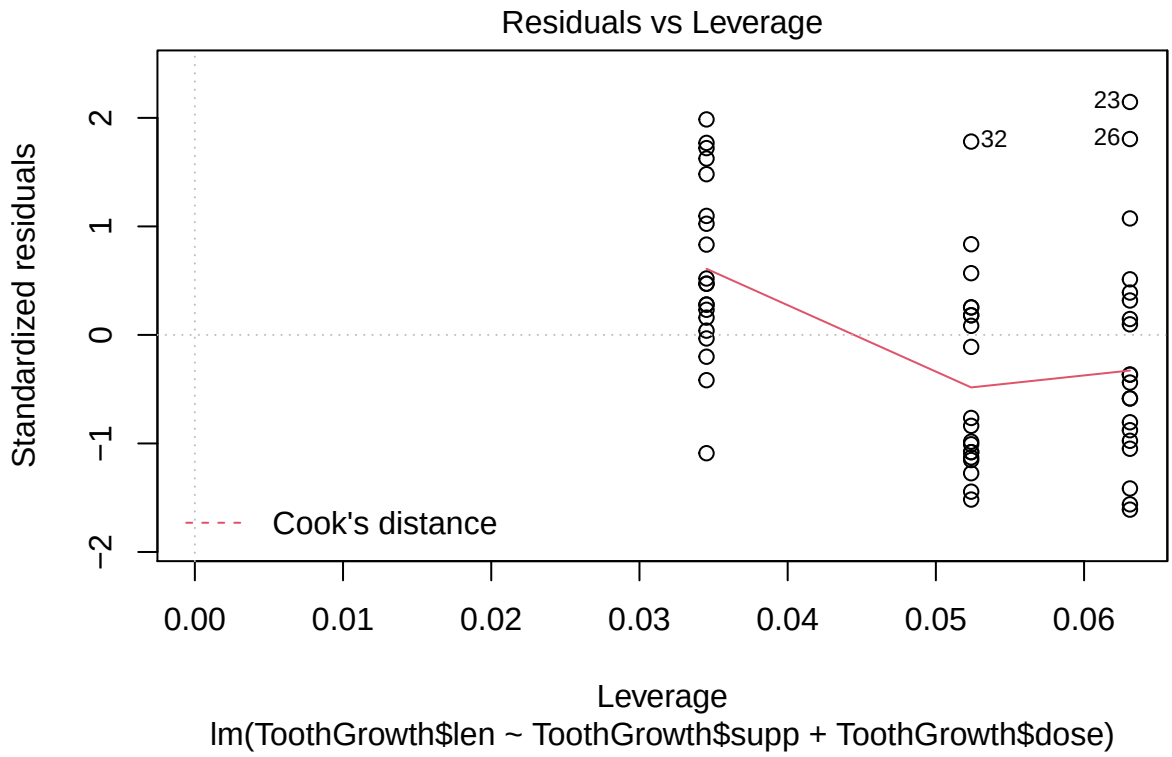
There is a statistically significant difference in means.

```r
plot(lmfit)
```


Residuals vs Fitted
Residuals
10
5
0
−5
50
23
26
10
15
20
25
Fitted values
lm(ToothGrowth$len ~ ToothGrowth$supp + ToothGrowth$dose)


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(ToothGrowth$len ~ ToothGrowth$supp + ToothGrowth$dose)
23
50
26


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
50
23
26
Fitted values
lm(ToothGrowth$len ~ ToothGrowth$supp + ToothGrowth$dose)



Reject the null hypothesis because there are differences between the mean of OJ and VC.

```
OJdose_.5 <- ToothGrowth %>% filter(dose==0.5) %>%  filter(supp=="OJ")
VCdose_.5 <- ToothGrowth %>% filter(dose==0.5) %>%  filter(supp=="VC")

t.test(OJdose_.5$len, VCdose_.5$len, paired = T, var.equal = T)
```

```
##
##  Paired t-test
##
## data:  OJdose_.5$len and VCdose_.5$len
## t = 2.9791, df = 9, p-value = 0.01547
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.263458 9.236542
## sample estimates:
## mean of the differences
##                    5.25
```

```
OJdose_1 <- ToothGrowth %>% filter(dose==1.0) %>%  filter(supp=="OJ")
VCdose_1 <- ToothGrowth %>% filter(dose==1.0) %>%  filter(supp=="VC")

t.test(OJdose_1$len, VCdose_1$len, paired = T, var.equal = T)
```

```
##
```

```
##  Paired t-test
##
## data:  OJdose_1$len and VCdose_1$len
## t = 3.3721, df = 9, p-value = 0.008229
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.951911 9.908089
## sample estimates:
## mean of the differences
##                    5.93
```

```
OJdose_2 <- ToothGrowth %>% filter(dose==2.0) %>%  filter(supp=="OJ")
VCdose_2 <- ToothGrowth %>% filter(dose==2.0) %>%  filter(supp=="VC")

t.test(OJdose_2$len, VCdose_2$len, paired = T, var.equal = T)
```

```
##
##  Paired t-test
##
## data:  OJdose_2$len and VCdose_2$len
## t = -0.042592, df = 9, p-value = 0.967
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -4.328976  4.168976
## sample estimates:
## mean of the differences
##                   -0.08
```

P-value is .967 not enough to reject the null hypothesis. The difference between OJ mean and VC Mean is -.08 is very close to 0 but still not equal to zero.

Conclusion and Assumptions

There is a significant difference between tooth length and dose levels across different delivery methods. A higher dosage led to longer teeth. Based on our evidence, the orange juice delivery method is more effective than the alternative method. At dosage 2 mg, the OJ still has a comparable impact vs. VC but the effectiveness is not the same as the lower dosage. There is a .255% chance that H0 = Ha. This would be a very rare instance.

Our assumptions on the data - H0 = Ha The mean of OJ$lenandVC$len are equal - They are paired samples T-test and their variance are equal.