Statistical Inference Course Project Part2

Khalid

7/14/2020

Part Two

The project consists of two parts:

- A simulation exercise.
- **Basic inferential data analysis**.
  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.

Load the ToothGrowth data and perform some basic exploratory data analyses

Install & Load needed packages

```
data("ToothGrowth")

library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(datasets)
data(ToothGrowth)
library(ggplot2)
library(cowplot)
```

```
##
## ********************************************************
```

```
## Note: As of version 1.0.0, cowplot does not change the
```

```
##   default ggplot2 theme anymore. To recover the previous
```

```
##   behavior, execute:
##   theme_set(theme_cowplot())
```

```
## ********************************************************
```

Load the ToothGrowth data and perform some basic exploratory data analyses

```
ToothGrowth %>% group_by(supp, dose) %>%
                summarise(meanLength = mean(len), medianLength = median(len), sdLength = sd(le
n))
```

```
## `summarise()` regrouping output by 'supp' (override with `.groups` argument)
```

```
## # A tibble: 6 x 5
## # Groups:   supp [2]
##   supp   dose meanLength medianLength sdLength
##   <fct> <dbl>      <dbl>        <dbl>    <dbl>
## 1 OJ      0.5      13.2         12.2      4.46
## 2 OJ      1        22.7         23.5      3.91
## 3 OJ      2        26.1         26.0      2.66
## 4 VC      0.5       7.98         7.15     2.75
## 5 VC      1        16.8         16.5      2.52
## 6 VC      2        26.1         26.0      4.80
```

```
summary(ToothGrowth)
```

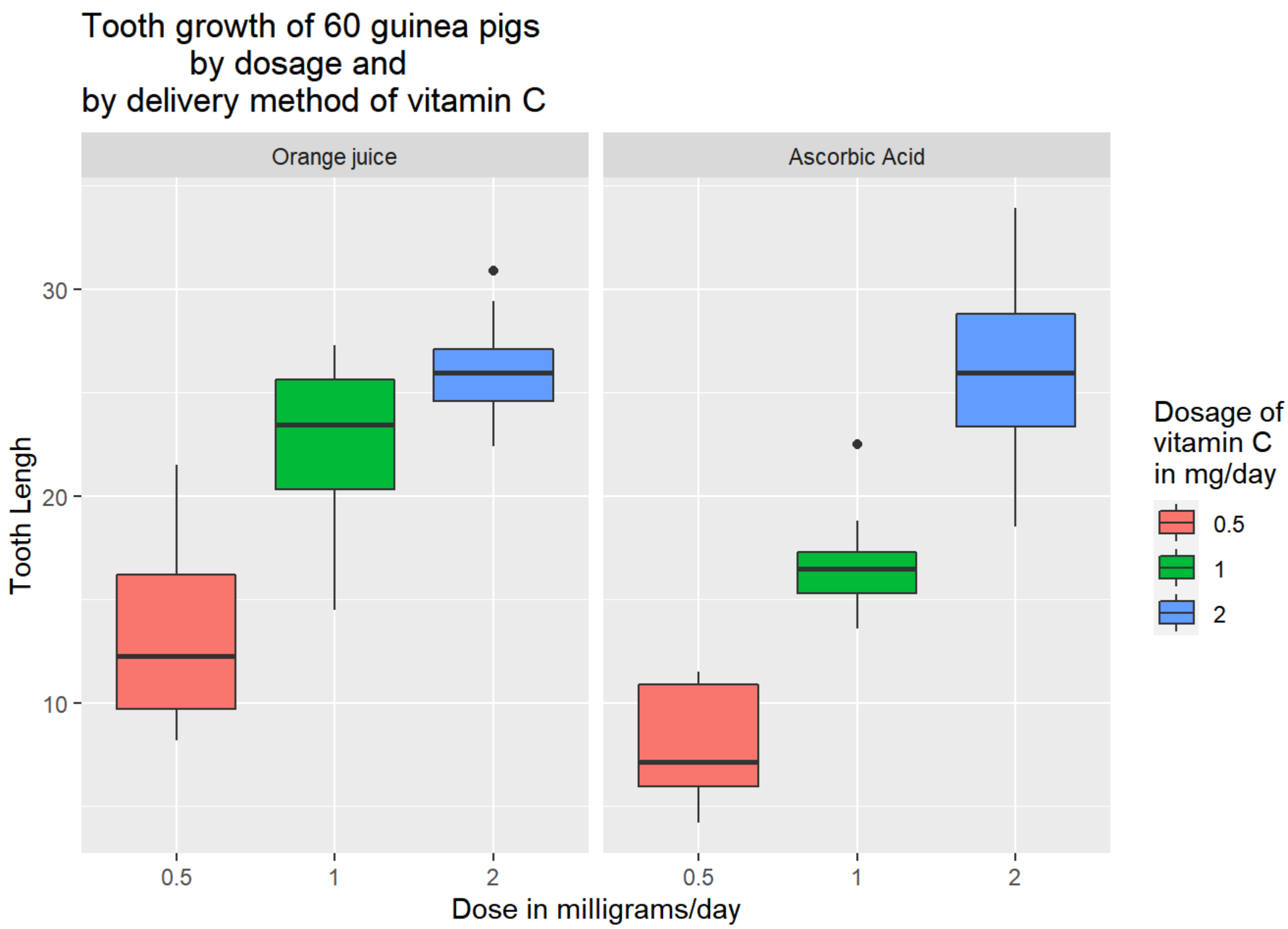
```
##       len        supp         dose      
##  Min.   : 4.20   OJ:30   Min.   :0.500  
##  1st Qu.:13.07   VC:30   1st Qu.:0.500  
##  Median :19.25           Median :1.000  
##  Mean   :18.81           Mean   :1.167  
##  3rd Qu.:25.27           3rd Qu.:2.000  
##  Max.   :33.90           Max.   :2.000  
```

```
head(ToothGrowth)
```

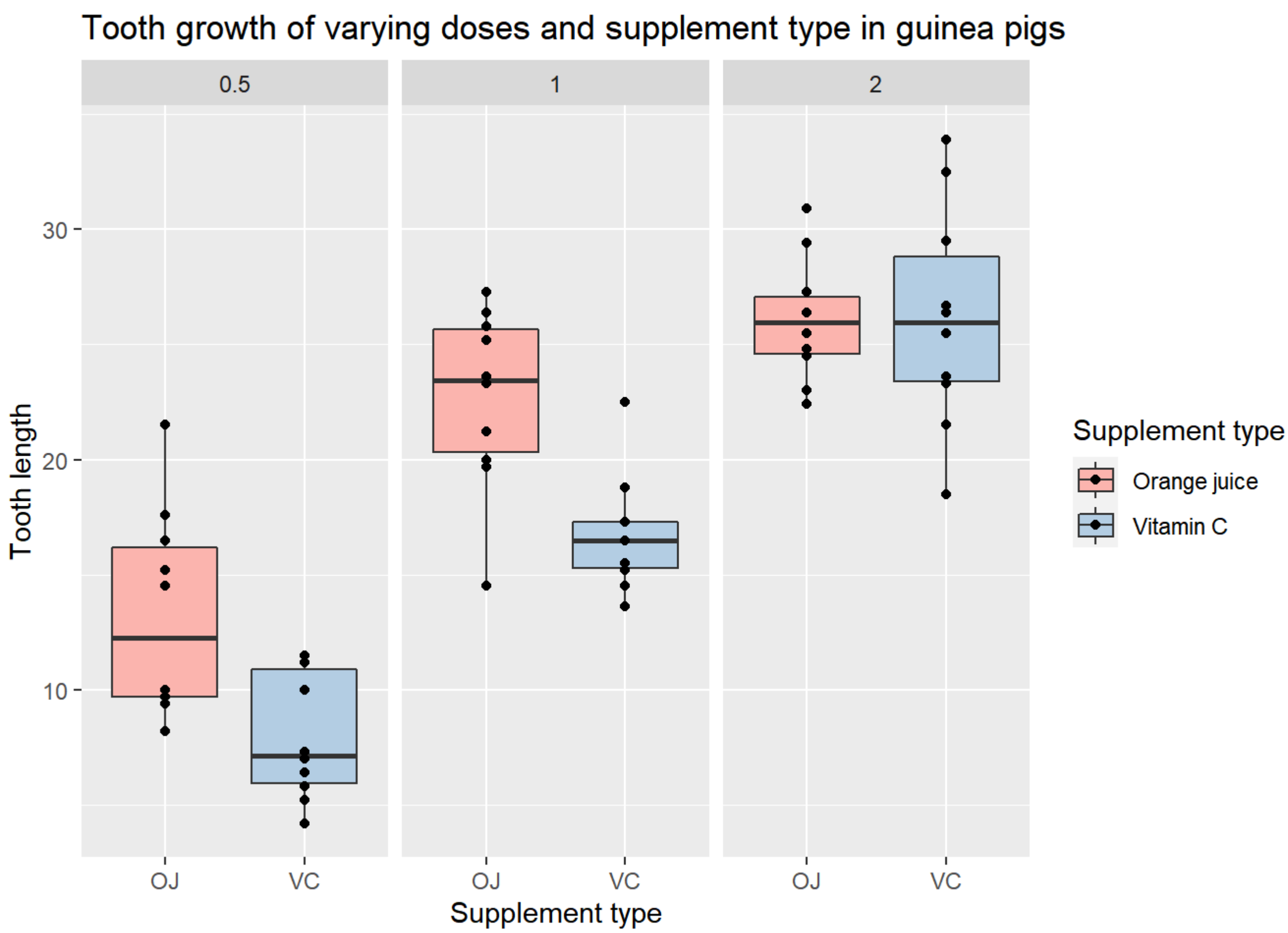
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

visualization

```
ggplot(ToothGrowth, aes(factor(dose), len, fill = factor(dose))) +
      geom_boxplot() +
      # facet_grid(.~supp)+
      facet_grid(.~supp, labeller = as_labeller(
            c("OJ" = "Orange juice",
              "VC" = "Ascorbic Acid"))) +
      labs(title = "Tooth growth of 60 guinea pigs
           by dosage and\nby delivery method of vitamin C",
           x = "Dose in milligrams/day",
           y = "Tooth Lengh") +
      scale_fill_discrete(name = "Dosage of\nvitamin C\nin mg/day")
```



```
ggplot(data = ToothGrowth, aes(x = supp, y = len, fill = supp)) +
        geom_boxplot() +
        geom_point() +
        facet_grid(. ~ dose) +
        xlab('Supplement type') +
        ylab('Tooth length') +
        ggtitle('Tooth growth of varying doses and supplement type in guinea pigs') +
        scale_fill_brewer(palette = 'Pastel1',
                          labels = c('Orange juice', 'Vitamin C'),
                          name = 'Supplement type')
```




```
low_dose <- ToothGrowth[ToothGrowth$dose == 0.5, ]
mid_dose <- ToothGrowth[ToothGrowth$dose == 1, ]
h_dose <- ToothGrowth[ToothGrowth$dose == 2, ]
```

hypothesis testing for low dose (0.5mg/day)

```
t.test(low_dose$len ~ low_dose$supp, alternative = 'two.sided', paired = F, var.equal = F)$p.val
ue
```

```
## [1] 0.006358607
```

hypothesis testing for medium dose (0.5mg/day)

```
t.test(mid_dose$len ~ mid_dose$supp,
       alternative = 'two.sided', paired = F, var.equal = F, conf.level = 0.95)$p.value
```

```
## [1] 0.001038376
```

Perform hypothesis testing for high dose (2mg/day)

```
t.test(h_dose$len ~ h_dose$supp,
       alternative = 'two.sided', paired = F, var.equal = F, conf.level = 0.95)$p.value
```

```
## [1] 0.9638516
```

Hypothesis testing: difference in tooth growth treated with different doses of supplements

```
- there is no difference in tooth growth in guinea pigs treated with different doses of orange j
uice or vitamin C
- there is a difference in tooth growth in guinea pigs treated with different doses of orange ju
ice or vitamin C
- Type 1 Error rate is less than 0.05 (p<0.05)
```

Perform hypothesis testing for orange juice at different doses Three different doses are compared to each other. The family-wise error rate will be controlled using the Bonferroni correction to adjust the p-value

```
orange_j <- ToothGrowth[ToothGrowth$supp == 'OJ', ]
orange_j_lm <- orange_j[orange_j$dose <= 1, ]
orange_j_lh <- orange_j[orange_j$dose != 1, ]
orange_j_mh <- orange_j[orange_j$dose >= 1, ]


vitamin_c <- ToothGrowth[ToothGrowth$supp == 'VC', ]
vitamin_c_lm <- vitamin_c[vitamin_c$dose <= 1, ]
vitamin_c_lh <- vitamin_c[vitamin_c$dose != 1, ]
vitamin_c_mh <- vitamin_c[vitamin_c$dose >= 1, ]
```

Perform hypothesis testing for orange juice at different doses

```
Three different doses are compared to each other. The family-wise error rate will be controlled
using the Bonferroni correction to adjust the p-value
```

```
orange_j_lm_p <- t.test(orange_j_lm$len ~ orange_j_lm$dose, alternative = 'two.sided', paired =
F, var.equal = F)$p.value

orange_j_lh_p <- t.test(orange_j_lh$len ~ orange_j_lh$dose, alternative = 'two.sided', paired =
F, var.equal = F)$p.value

orange_j_mh_p <- t.test(orange_j_mh$len ~ orange_j_mh$dose, alternative = 'two.sided', paired =
F, var.equal = F)$p.value

data.frame(comparisons = c('0.5 vs 1', '0.5 vs 2', '1 vs 2'),
           pvalue = c(orange_j_lm_p, orange_j_lh_p, orange_j_mh_p),
           bonferroni = p.adjust(c(orange_j_lm_p, orange_j_lh_p, orange_j_mh_p), method = 'bonfe
rroni'))
```

```
##   comparisons       pvalue   bonferroni
## 1    0.5 vs 1 8.784919e-05 2.635476e-04
## 2    0.5 vs 2 1.323784e-06 3.971352e-06
## 3      1 vs 2 3.919514e-02 1.175854e-01
```

#low vs medium dose

```
lowvc_lm_p <- t.test(vitamin_c_lm$len ~ vitamin_c_lm$dose, alternative = 'two.sided',
                  paired = F, var.equal = F, conf.level = 0.95)$p.value
```

#low vs high dose

```
lowvc_lh_p <- t.test(vitamin_c_lh$len ~ vitamin_c_lh$dose, alternative = 'two.sided',
                  paired = F, var.equal = F, conf.level = 0.95)$p.value
```

#medium vs high dose

```
lowvc_mh_p <- t.test(vitamin_c_mh$len ~ vitamin_c_mh$dose, alternative = 'two.sided',
                  paired = F, var.equal = F, conf.level = 0.95)$p.value
```

```
data.frame(comparisons = c('0.5 vs 1', '0.5 vs 2', '1 vs 2'),
           pvalue = c(lowvc_lm_p, lowvc_lh_p, lowvc_mh_p),
           bonferroni = p.adjust(c(lowvc_lm_p, lowvc_lh_p, lowvc_mh_p), method = 'bonferroni'))
```

```
##   comparisons       pvalue   bonferroni
## 1    0.5 vs 1 6.811018e-07 2.043305e-06
## 2    0.5 vs 2 4.681577e-08 1.404473e-07
## 3      1 vs 2 9.155603e-05 2.746681e-04
```

```
the Tooth Growth in guinea pigs treated, there is a does-dependent effect in that the higher the
dose is, the longer the tooth length is. Whether this relationship is linear requires further ex
ploration.
```