Statistical Inference Project Part 2

bzg

12-1-2022

Synopsis

This is a project for the Coursera Statistical Inference Class. The project consists of two parts: - Simulation Exercise to explore inference - Basic inferential analysis using the ToothGrowth data in the R datasets package

Part 2 - Inferential Analysis using ToothGrowth data

Overview

Analyse the ToothGrowth data in the R datasets package.

Instructions

- Load the ToothGrowth data and perform some basic exploratory data analysis
- Provide a basic summary of the data
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose
- State your conclusions and the assumptions needed for your conclusions

Preparing the environment

Load all necessary libraries.

```
# Load packages
library(dplyr, warn.conflicts = F)
library(ggplot2)
library(knitr)
library(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

Load the Tooth Growth dataset

```
# Load data
toothGrowth <- data.table(ToothGrowth)
```

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).

Format

A data frame with 60 observations on 3 variables.

Exploratory data analysis

```
# Explore dataset
head(toothGrowth); tail(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
```

```
##     len supp dose
## 1: 24.8   OJ    2
## 2: 30.9   OJ    2
## 3: 26.4   OJ    2
## 4: 27.3   OJ    2
## 5: 29.4   OJ    2
## 6: 23.0   OJ    2
```

The three variables are **length**, **supplement**, and **dose**

```
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
```

```
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 ...
```

There are two levels (delivery types) for supp (supplement): **OJ** (Orange Juice) and **VC** (Vitamin C).

Length and dose are both numeric values but we cannot determine how many values exist for dose. View the unique values of **dose**.

```
unique(ToothGrowth$dose)
```

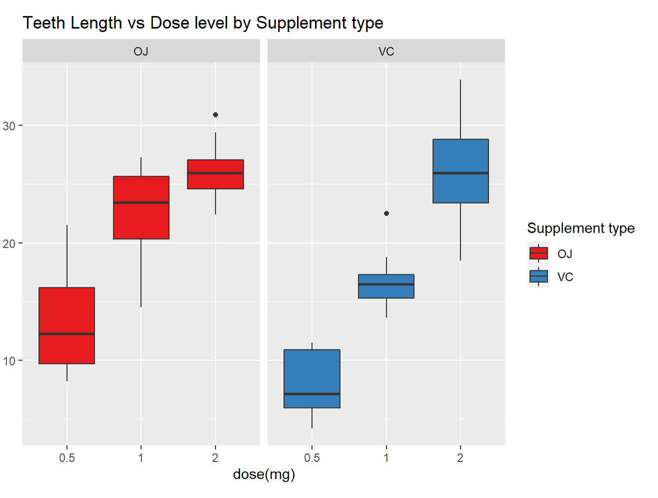
```
## [1] 0.5 1.0 2.0
```

There are three discrete levels for dose: 0.5, 1.0, and 2. We can conveniently convert it to a factor variable with three levels

```
# Convert to factor
toothGrowth <- toothGrowth %>% mutate(dose = as.factor(dose))
```

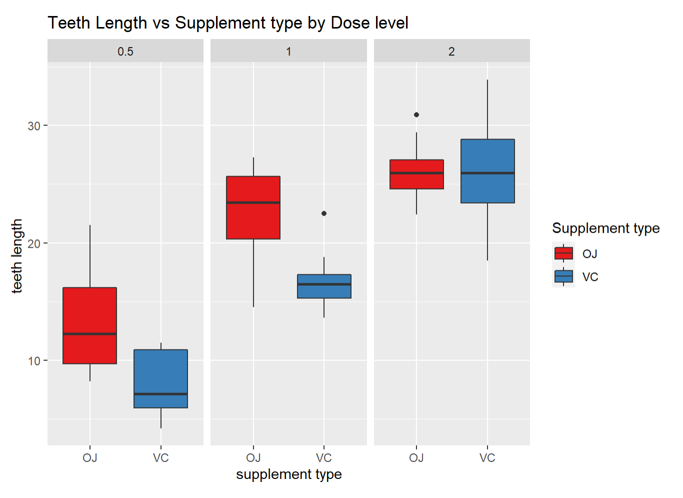
Visually examine the data by looking at the tooth length compared to dose by supplement.

```
# Plot the relationship between teeth length and dose level for each supplement type
ggplot(toothGrowth, aes(x=dose, y=len, fill = supp)) +
  geom_boxplot() +
  facet_grid(. ~ supp) +
  scale_fill_brewer(palette = "Set1") +
  ggtitle("Teeth Length vs Dose level by Supplement type") +
  labs(x="dose(mg)", y= "teeth length ") +
  guides(fill=guide_legend(title="Supplement type"))
```



This multipanel plot emphasizes the relationship between teeth length and dose level for each supplement type. It appears to be a positive relationship for both supplement types. In other words, as the amount of supplement increases, so does teeth length.

```
# Plot the relationship between supplement type and teeth length emphasizing
ggplot(toothGrowth, aes(x = supp, y = len)) +
  geom_boxplot(aes(fill = supp)) +
  facet_wrap(~ dose) +
  scale_fill_brewer(palette = "Set1") +
  ggtitle("Teeth Length vs Supplement type by Dose level ") +
  labs(x="supplement type", y= "teeth length ") +
  guides(fill=guide_legend(title="Supplement type"))
```



This second plot shows the relationship between supplement type and teeth length emphasizing direct comparison between supplement types. Here the relationship is much less clear. Orange juice OJ appears to be more effective at dosage levels 0.5 and 1. On the other hand, at dosage level 2 there doesn't appear to be any significant difference.

```
# Checking effectiveness dose levels
toothGrowth %>%
  filter(dose == 2) %>%
  group_by(supp) %>%
  summarise(avg.length = mean(len))
```

```
## # A tibble: 2 x 2
##   supp  avg.length
##   <fct>      <dbl>
## 1 OJ          26.1
## 2 VC          26.1
```

Summary of initial data analysis

There appears to be an impact on tooth growth by increasing the dosage of Orange Juice, until a dosage of two. Both Supplement types score an average length of **26.1** at that stage.

Use confidence intervals and/or hypothesis tests to compare tooth growth by supplement and dose

Use t.test to determine if there is a difference in the performance of the treatments. First, we will run the test based on supplement. Looking to see if the p-value is smaller than 0.05 and if the confidence interval crosses 0.

Testing by dose levels

Test A, dose = 0.5 and dose = 1 (p-value = 1.268e-07)

```
# Extract the len and dose vectors from toothGrowth
len_a <- toothGrowth %>%
  filter(dose %in% c(0.5,1)) %>%
  select(len) %>%
  unlist()

dose_a <- toothGrowth %>%
  filter(dose %in% c(0.5,1)) %>%
  select(dose) %>%
  unlist()

# Test A
(Test.a <- t.test(len_a~dose_a, paired = FALSE))
```

```
##
##  Welch Two Sample t-test
##
## data:  len_a by dose_a
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means between group 0.5 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

Test B, dose = 0.5 and dose = 2 (p-value = 4.398e-14)

```
# Extract the len and dose vectors from toothGrowth
len_b <- toothGrowth %>%
  filter(dose %in% c(0.5,2)) %>%
  select(len) %>%
  unlist()

dose_b <- toothGrowth %>%
  filter(dose %in% c(0.5, 2)) %>%
  select(dose) %>%
  unlist()

# Test B
(Test.b <- t.test(len_b~dose_b, paired = FALSE))
```

```
##
##  Welch Two Sample t-test
##
## data:  len_b by dose_b
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means between group 0.5 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

Test C, dose = 1 and dose = 2 (p-value = 1.906e-05)

```
#Extract the len and dose vectors from toothGrowth
len_c <- toothGrowth %>%
  filter(dose %in% c(1,2)) %>%
  select(len) %>%
  unlist()

dose_c <- toothGrowth %>%
  filter(dose %in% c(1,2)) %>%
  select(dose) %>%
  unlist()

#Test C
(Test.c <- t.test(len_c~dose_c, paired = FALSE))
```

```
##
##  Welch Two Sample t-test
##
## data:  len_c by dose_c
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

We went through all possible combinations of levels from the factor variable dose and in all cases the p-value is lower than the default significance level 0.05. We can safely reject the null. In other words there appears to be a positive relationship between dose level and teeth length.

Testing by Supplement

```
#Extract the len and supp vectors from toothGrowth
len <- toothGrowth %>%
  select(len) %>%
  unlist()

supp <- toothGrowth %>%
  select(supp) %>%
  unlist()

#Test T
t.test(len~supp, paired=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means between group OJ and group VC is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

We can see that the p-value of the test is *0.06*. Since the p-value is greater than 0.05 and the confidence interval of the test contains zero, there is not enough evidence to reject the null hypothesis.We cannot assume the delivery type has a significant effect on tooth growth.

Test the tooth length comparing the dosage of 1mg to 2mg to determine the effects of an increased dosage.

```
# Test T 1
t.test(toothGrowth$len[toothGrowth$dose==2],
       toothGrowth$len[toothGrowth$dose==1],
       paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  toothGrowth$len[toothGrowth$dose == 2] and toothGrowth$len[toothGrowth$dose == 1]
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

We see the p-value (**1.811e-05**) is very small, and is significant. Therefore, we can reject the null hypothesis and assume the dosage increase from 1mg to 2mg creates an positive effect on tooth growth.

Next, perform the test comparing the dosage of 0.5mg to 1mg.

```
t.test(toothGrowth$len[toothGrowth$dose==1],
       toothGrowth$len[toothGrowth$dose==0.5],
       paired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  toothGrowth$len[toothGrowth$dose == 1] and toothGrowth$len[toothGrowth$dose == 0.5]
## t = 6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276252 11.983748
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

Again, we see the p-value is still small although slighty larger than the previous test, therefore, it is significant. We can again reject the null hypothesis and assume the dosage increase from .5mg to 1mg creates an positive effect on tooth growth.

There is no need for more testing of dosages given the previous tests.

State conclusions and the assumptions needed for conclusions

In this experiment, we assume there is a common variance in the population and that the guinea pigs were chosen at random. The delivery type does not show a significant increase in tooth growth even though it does have a confidence level that crosses 0 at the 95% confidence.

However, there does appear to be a difference with an increase in tooth growth when the dosage is increased. The tests comparing the dosage show confidence intervals of differences never crossing zero.

Conclusion

Increasing the dosage leads to an increase in tooth growth in guinea pigs.