Statistical Inference - Course Project - Part 2 - ToothGrowth data

Jose Maria Vilaverde (jmvilaverde)

Saturday, May 23, 2015

Overview

Analyze the ToothGrowth data in the R datasets package.

In this investigation are presented some hypothesis about influence in tooth growth in guinea pigs. The hypothesis are based on three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

Simulation

Load the ToothGrowth data and perform some basic exploratory data analyses

```
data(ToothGrowth)
names(ToothGrowth)
```

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

```
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 ...
```

Provide a basic summary of the data.

```
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
```

Obtains all the groups from the data:

```
GroupOJ05 <- ToothGrowth[ToothGrowth$supp == "OJ"&ToothGrowth$dose == 0.5,]$len
GroupOJ1 <- ToothGrowth[ToothGrowth$supp == "OJ"&ToothGrowth$dose == 1,]$len
GroupOJ2 <- ToothGrowth[ToothGrowth$supp == "OJ"&ToothGrowth$dose == 2,]$len
GroupVC05 <- ToothGrowth[ToothGrowth$supp == "VC"&ToothGrowth$dose == 0.5,]$len
GroupVC1 <- ToothGrowth[ToothGrowth$supp == "VC"&ToothGrowth$dose == 0.5,]$len
GroupVC2 <- ToothGrowth[ToothGrowth$supp == "VC"&ToothGrowth$dose == 0.5,]$len
```

Initial analysis

Analysis by supp

```
orangeJuiceGroup <- ToothGrowth[ToothGrowth$supp == "OJ",]$len
vitaminCGroup <- ToothGrowth[ToothGrowth$supp == "VC",]$len
difference <- vitaminCGroup - orangeJuiceGroup
mn <- mean(difference)
s <- sd(difference)
```

The mean of growth for guinea pigs that received Orange Juice was: 20.6633333

The mean of growth for guinea pigs that received Vitamin C was: 16.9633333

The mean of difference Vitamin C Group - Orange Juice Group is: -3.7

The standard deviation of difference Vitamin C Group - Orange Juice Group is: 6.1363249

Analysis by dose

```
Group05 <- ToothGrowth[ToothGrowth$dose == 0.5,]$len
Group1 <- ToothGrowth[ToothGrowth$dose == 1,]$len
Group2 <- ToothGrowth[ToothGrowth$dose == 2,]$len
difference1to05 <- Group1 - Group05
mn1to05 <- mean(difference1to05)
s1to05 <- sd(difference1to05)

difference2to05 <- Group2 - Group05
mn2to05 <- mean(difference2to05)
s2to05 <- sd(difference2to05)

difference2to1 <- Group2 - Group1
mn2to1 <- mean(difference2to1)
s2to1 <- sd(difference2to1)
```

The mean of growth for guinea pigs that received a dose of 0.5 is: 10.605

The mean of growth for guinea pigs that received a dose of 1 is: 19.735

The mean of growth for guinea pigs that received a dose of 2 is: 26.1

The mean of difference dose of 1 - dose of 0.5 is: 9.13

The standard deviation of difference dose of 1 - dose of 0.5 is is: 5.8606718

Hyphotesis

Hyphotesis: $H_0 : \mu_{OrangeJuice} > \mu_{vitaminC}$

```
t.test(orangeJuiceGroup, vitaminCGroup, alternative = "less", paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  orangeJuiceGroup and vitaminCGroup
## t = 1.9153, df = 55.309, p-value = 0.9697
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 6.931731
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

The hypothesis the group that received Orange Juice growth was higher growth is accepted with a p = 0.9697.

Hyphotesis: $H_0 : \mu_{dose=1} > \mu_{dose=0.5}$

```
t.test(Group1,Group05, alternative = "less",paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  Group1 and Group05
## t = 6.4766, df = 37.986, p-value = 1
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 11.50668
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

The hypothesis the group that received dose 1 growth more than group of dose 0.5 is accepted with a p = 1.

Hyphotesis: $H_0 : \mu_{dose=2} > \mu_{dose=0.5}$

```
t.test(Group2,Group05, alternative = "less",paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  Group2 and Group05
## t = 11.799, df = 36.883, p-value = 1
```

```
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 17.71074
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

The hypothesis the group that received dose 2 growth more than group of dose 0.5 is accepted with a p = 1.

Hyphotesis: $H_0 : \mu_{dose=2} > \mu_{dose=1}$

```
t.test(Group2,Group1, alternative = "less",paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  Group2 and Group1
## t = 4.9005, df = 37.101, p-value = 1
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##     -Inf 8.55613
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

The hypothesis the group that received dose 2 growth more than group of dose 1 is accepted with a p = 1.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
arrange(summarize(group_by(ToothGrowth, dose, supp), meanGroup = mean(len), sdGroup = sd(len)), meanGrou
```

```
## Source: local data frame [6 x 4]
## Groups: dose
##
##   dose supp meanGroup  sdGroup
## 1  0.5   VC      7.98 2.746634
## 2  0.5   OJ     13.23 4.459709
## 3  1.0   VC     16.77 2.515309
## 4  1.0   OJ     22.70 3.910953
## 5  2.0   OJ     26.06 2.655058
## 6  2.0   VC     26.14 4.797731
```

By previous analysis, hypothesis and this summary seems that higher dose implies higher mean of growth, and at same dose Orange Juice have higher mean than Vitamin C, except for dose 2.0.

Let's test same dose and Vitamin C vs. Orange Juice:

Confidence Test for Oranje Juice 0.5 - Vitamin C 0.5

Hyphotesis: $H_0 : \mu_{OJ,0.5} > \mu_{VC,0.5}$

```
#Oranje Juice 0.5 - Vitamin C 0.5
t.test(GroupOJ05, GroupVC05, alternative = "less", paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  GroupOJ05 and GroupVC05
## t = 3.1697, df = 14.969, p-value = 0.9968
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##     -Inf 8.15396
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

The hypothesis that, with same dose 0.5, the group that received Orange Juice growth more than group Vitamin C is accepted with p-value = 0.9968.

Confidence Test for Orange Juice 1 - Vitamin C 1

Hyphotesis: $H_0 : \mu_{OJ,1} > \mu_{VC,1}$

```
#Oranje Juice 1 - Vitamin C 1
t.test(GroupOJ1, GroupVC1, alternative = "less", paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  GroupOJ1 and GroupVC1
## t = 9.7401, df = 16.141, p-value = 1
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##     -Inf 17.3571
## sample estimates:
## mean of x mean of y
##     22.70      7.98
```

The hypothesis that, with same dose 1, the group that received Orange Juice growth more than group Vitamin C is accepted with p-value = 1.

Confidence Test for Orange Juice 2 - Vitamin C 2

In this case the mean for Vitamin C is greater than Orange Juice, try two hypothesis:

Hyphotesis: $H_0 : \mu_{OJ,2} > \mu_{VC,2}$

```
#Oranje Juice 2 - Vitamin C 2
t.test(GroupOJ2, GroupVC2, alternative = "less", paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  GroupOJ2 and GroupVC2
## t = 14.967, df = 17.979, p-value = 1
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##      -Inf 20.17493
## sample estimates:
## mean of x mean of y
##     26.06      7.98
```

The hypothesis that, with same dose 2, the group that received Orange Juice growth more than group Vitamin C is accepted with p-value = 1.

Hyphotesis: $H_1 : \mu_{OJ,2} < \mu_{VC,2}$

```
#Oranje Juice 2 - Vitamin C 2
t.test(GroupOJ2, GroupVC2, alternative = "greater", paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  GroupOJ2 and GroupVC2
## t = 14.967, df = 17.979, p-value = 6.811e-12
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  15.98507      Inf
## sample estimates:
## mean of x mean of y
##     26.06      7.98
```

The hypothesis that, with same dose 1, the group that received Vitamin C growth more than group Orange Juice is rejected with p-value = 6.811e-12.