statinf_cp1b_joannanw

joannanw

Sunday, March 22, 2015

The goal of this assignment is to explore and analyze the ToothGrowth data according to confidence intervals and hypothesis testing. The assignment will cover basic analyses to justification on the conclusion of the experiment.

Analyses

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

```
data(ToothGrowth)
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 ...
```

```
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
```

```
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
```

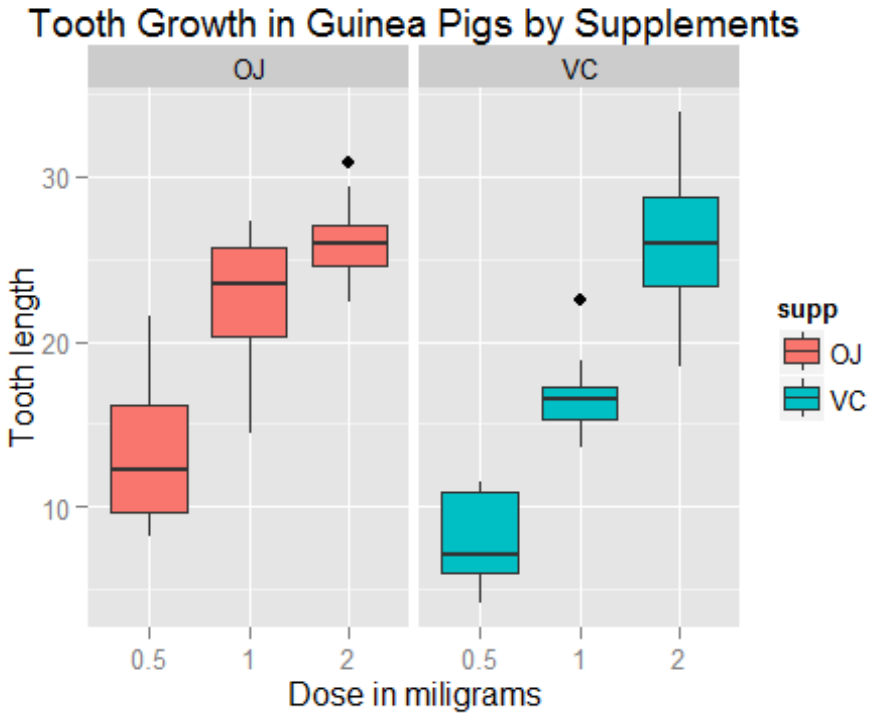
```
data <- ToothGrowth %>%
        group_by(supp,dose) %>%
        summarise(Mean=mean(len),Min=min(len),Max=max(len),"SD"=sd(len))

library(reshape2)
wideData <-
dcast(melt(data,id.vars=c("supp","dose")),dose~supp+variable,fun.aggregate=su
m)
wideData
```

```
##   dose OJ_Mean OJ_Min OJ_Max    OJ_SD VC_Mean VC_Min VC_Max    VC_SD
## 1  0.5   13.23    8.2   21.5 4.459709    7.98    4.2   11.5 2.746634
## 2  1.0   22.70   14.5   27.3 3.910953   16.77   13.6   22.5 2.515309
## 3  2.0   26.06   22.4   30.9 2.655058   26.14   18.5   33.9 4.797731
```

The ToothGrowth data in the R datasets package is a set of 60 observations, length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1 and 2 mg) with each of two delivery methods (orange juice or ascorbic acid)

2. Provide a basic summary of the data.

```
library(ggplot2)
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
geom_boxplot(notch = F, show_guide = TRUE) +
facet_grid(. ~ supp) +
xlab("Dose in miligrams") +
ylab("Tooth length") +
ggtitle("Tooth Growth in Guinea Pigs by Supplements")
```



The graph shows that as the dose of Orange Juice and Ascorbic Acid increases, so does the length of the tooth. The overall data seems to show that doses of Orange Juice generally yields longer tooth length than doses of Ascorbic Acid. However, as the dose of each supplement reaches 2, the average length of tooth reaches around the same value, indicating possible lack of effectiveness of the supplements.

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

- Which supplement yields greater effect on tooth growth?

To test this, construct a table of hypothetical output using the Student's t-Test on the length of tooth by supplement type. Assume that Orange Juice has more effect on tooth growth.

H0: Mu_OJ - Mu_VC = 0, Ha: Mu_OJ - Mu_VC > 0

```
VC <- ToothGrowth[ToothGrowth$supp == "VC", 1]
OJ <- ToothGrowth[ToothGrowth$supp == "OJ", 1]
var(VC); var(OJ)
```

```
## [1] 68.32723
```

```
## [1] 43.63344
```

```
test0 <- t.test(OJ, VC, alternative = "greater", var.equal = FALSE, paired = TRUE)
# alternative = "less" indicates VC less than OJ. Assume different standard deviation and paired samples.
```

```
# Or use t.test(len ~ supp, data = ToothGrowth, alternative = "greater",
var.equal = FALSE, paired = TRUE)
test0

##
##  Paired t-test
##
## data:  OJ and VC
## t = 3.3026, df = 29, p-value = 0.001275
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  1.796409      Inf
## sample estimates:
## mean of the differences
##                     3.7
```

The t-test gives p = 0.0012749 < 0.05, this means the null hypothesis (H0) can be rejected. There is strong evidence of a mean increase in tooth length between using Orange Juice versus using Ascorbic Acid.

- Does more dosage lead to longer tooth length?

Assume that higher dosage causes longer tooth length.

```
test1 <- t.test(len ~ dose, data=subset(ToothGrowth, dose %in% c(0.5,1.0)),
var.equal=T, alternative="l")
test1

##
##  Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 38, p-value = 6.331e-08
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -6.753344
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735

test2 <- t.test(len ~ dose, data=subset(ToothGrowth, dose %in% c(1.0, 2.0)),
var.equal=T, alternative="l")
test2

##
##  Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 38, p-value = 9.054e-06
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -4.175196
```

```
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

```
test3 <- t.test(len ~ dose, data=subset(ToothGrowth, dose %in% c(0.5, 2.0)),
var.equal=T, alternative="l")
test3
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 38, p-value = 1.419e-14
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -13.28093
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

The t-test gives $p < 0.05$, which means that the null hypothesis can be rejected. The bigger the dosage of the supplement, the longer the tooth length.

4. State your conclusions and the assumptions needed for your conclusions.

- Each supplement has different effect on the test subjects.
- Dosage level has different effect on the test subjects.
- The same test subjects are used to test the supplements and dosage levels.