Basic Inferential Data Analysis Instructions

Pribylov V.A.

02 09 2020

Overview

This is a course project needed to analyze the ToothGrowth data in the R datasets package.

Loading the data

```
TG <- ToothGrowth
```

Providing basic summary of the data

```
summary(TG)
```

```
##       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
```

```
OJ <- TG$len[TG$supp == 'OJ']
VC <- TG$len[TG$supp == 'VC']
```

Comparison of means for two factor-levels of supp column

```
mOJ <- mean(OJ)
vOJ <- var(OJ)
nOJ <- length(OJ)
OJ_confidence_interval <- mOJ + c(-1, 1)*qnorm(0.975)*sqrt(vOJ/nOJ)
paste0('OJ 95% confidence interval is (',round(OJ_confidence_interval[1], 2),
       ', ', round(OJ_confidence_interval[2], 2), ')')
```

```
## [1] "OJ 95% confidence interval is (18.3, 23.03)"
```

```
mVC <- mean(VC)
vVC <- var(VC)
nVC <- length(VC)
VC_confidence_interval <- mVC + c(-1, 1)*qnorm(0.975)*sqrt(vVC/nVC)
paste0('VC 95% confidence interval is (',round(VC_confidence_interval[1], 2),
       ', ', round(VC_confidence_interval[2], 2), ')')
```

```
## [1] "VC 95% confidence interval is (14.01, 19.92)"
```

As we can see, 95% confidence intervals for OJ and VC are intersecting each other. So let's try find 95% confidence interval for difference of their means.

```
pooled_variance <- sqrt(vOJ/nOJ + vVC/nVC)
difference_confidence_interval <- mOJ - mVC + c(-1, 1)*qnorm(0.975)*pooled_variance
paste0('95% confidence interval of difference between VC and OJ means is ('
       ,round(difference_confidence_interval[1], 2),', '
       , round(difference_confidence_interval[2], 2), ')')
```

```
## [1] "95% confidence interval of difference between VC and OJ means is (-0.09, 7.49)"
```

This confidence interval includes 0, so there can be no difference between their means, but for 90% confidence interval mean for OJ is bigger than mean for VC:

```
difference_confidence_interval_90 <- mOJ - mVC + c(-1, 1)*qnorm(0.95)*pooled_variance
paste0('90% confidence interval of difference between VC and OJ means is ('
       ,round(difference_confidence_interval_90[1], 2),', '
       , round(difference_confidence_interval_90[2], 2), ')')
```

```
## [1] "90% confidence interval of difference between VC and OJ means is (0.52, 6.88)"
```

Statistical characteristics for dose column

```
mns <- NULL
vars <- NULL
lens <- NULL
for (i in levels(as.factor(TG$dose))) {
    mns <- c(mns, mean(TG$len[TG$dose == i]))
    vars <- c(vars, var(TG$len[TG$dose == i]))
    lens <- c(lens, length(TG$len[TG$dose == i]))

}
```

There are some pairwise comparisons between levels of dose:

```
confidence_interval_05 <- mns[1] + c(-1, 1)*qnorm(0.975)*sqrt(vars[1]/lens[1])
confidence_interval_05
```

```
## [1]  8.632928 12.577072
```

```
confidence_interval_1 <- mns[2] + c(-1, 1)*qnorm(0.975)*sqrt(vars[2]/lens[2])
confidence_interval_1
```

```
## [1] 17.79989 21.67011
```

```
confidence_interval_2 <- mns[3] + c(-1, 1)*qnorm(0.975)*sqrt(vars[3]/lens[3])
confidence_interval_2
```

```
## [1] 24.44594 27.75406
```

As at the previous case let's find 95% confidence intervals of pairwise differences:

```
difference_confidence_interval_05_1 <- mns[2] - mns[1] +
    c(-1, 1)*qnorm(0.975)*sqrt(vars[1]/lens[1] + vars[2]/lens[2])
paste0('95% confidence interval of difference between 0.5 and 1 levels means is ('
       ,round(difference_confidence_interval_05_1[1], 2),', '
       , round(difference_confidence_interval_05_1[2], 2), ')')
```

```
## [1] "95% confidence interval of difference between 0.5 and 1 levels means is (6.37, 11.89)"
```

```
difference_confidence_interval_1_2 <- mns[3] - mns[2] +
    c(-1, 1)*qnorm(0.975)*sqrt(vars[2]/lens[2] + vars[3]/lens[3])
paste0('95% confidence interval of difference between 1 and 2 levels means is ('
       ,round(difference_confidence_interval_1_2[1], 2),', '
       , round(difference_confidence_interval_1_2[2], 2), ')')
```

```
## [1] "95% confidence interval of difference between 1 and 2 levels means is (3.82, 8.91)"
```

```
difference_confidence_interval_05_2 <- mns[3] - mns[1] +
    c(-1, 1)*qnorm(0.975)*sqrt(vars[1]/lens[1] + vars[3]/lens[3])
paste0('95% confidence interval of difference between 0.5 and 2 levels means is ('
       ,round(difference_confidence_interval_05_2[1], 2),', '
       , round(difference_confidence_interval_05_2[2], 2), ')')
```

```
## [1] "95% confidence interval of difference between 0.5 and 2 levels means is (12.92, 18.07)"
```

So with the increasing of dose level, we also can see increasing of mean length.