Peer-graded Assignment: Statistical Inference Course Project Part II

Basic Inferential Data Analysis

Koen Vermeulen

10-8-2019

Overview

This report will analyze tooth growth.

Analysis

Loading and exploring data:

```
data("ToothGrowth")
dim(ToothGrowth)
```

[1] 60 3

```
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 …

Basic summary:

```
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
```

Subsetting vitamin C & orange juice:

```
VC <- dplyr::filter(ToothGrowth, supp == "VC")
OJ <- dplyr::filter(ToothGrowth, supp == "OJ")
t.test(VC$len, OJ$len, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -7.5710156  0.1710156
## attr(,"conf.level")
## [1] 0.95

t.test(VC$len, OJ$len, paired = FALSE, var.equal = FALSE)$p.value

## [1] 0.06063451
```

The interval contains zero, but the p-value is above 5% so it is not conclusive.

The next step is comparing the different doses. The hypothesis is that the dose has an effect on the tooth length.

```
dose05 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
dose1 <-  ToothGrowth$len[ToothGrowth$dose == 1.0]
dose2 <-  ToothGrowth$len[ToothGrowth$dose == 2.0]
t.test(dose05, dose1)$conf
```

[1] -11.983781 -6.276219 attr(,"conf.level") [1] 0.95

```
t.test(dose05, dose1)$p.value
```

[1] 1.268301e-07

```
t.test(dose05, dose2)$conf
```

[1] -18.15617 -12.83383 attr(,"conf.level") [1] 0.95

```
t.test(dose05, dose2)$p.value
```

[1] 4.397525e-14

```
t.test(dose1, dose2)$conf
```

[1] -8.996481 -3.733519 attr(,"conf.level") [1] 0.95

```
t.test(dose1, dose2)$p.value
```

[1] 1.90643e-05

As we can see from all 3 t.tests, the p-values are less than 0.05. Therefore we can accept the hypothesis that the dose has an effect on tooth length, in other words: the mean is not equal to zero. We also see that all of the 95% confidence intervals are below zero. This confirms that increasing the dose increases the tooth length.

In the third step we'll inspect the difference in mean lengths between the vitamin C (VC) group and orange juice (OJ) group for each dose level applied.

```
dose05VC <- VC$len[VC$dose == 0.5]
dose1VC <- VC$len[VC$dose == 1.0]
dose2VC <- VC$len[VC$dose == 2.0]
dose05OJ <- OJ$len[OJ$dose == 0.5]
dose1OJ <- OJ$len[OJ$dose == 1.0]
dose2OJ <- OJ$len[OJ$dose == 2.0]
```

As in previous cases, we'll assume that observations are not paired and that variances are not equal.

```
t.test(dose05VC, dose05OJ)$p.value
```

[1] 0.006358607

```
t.test(dose05VC, dose05OJ)$conf
```

[1] -8.780943 -1.719057 attr(,"conf.level") [1] 0.95

```
t.test(dose1VC, dose1OJ)$p.value
```

[1] 0.001038376

```
t.test(dose1VC, dose1OJ)$conf
```

[1] -9.057852 -2.802148 attr(,"conf.level") [1] 0.95

```
t.test(dose2VC, dose2OJ)$p.value
```

[1] 0.9638516

```
t.test(dose2VC, dose2OJ)$conf
```

[1] -3.63807 3.79807 attr(,"conf.level") [1] 0.95

We can see from the t-tests that the difference between group means for VC and OJ is statistically significant for dose levels 0.5 and 1 mg/day, because the p-values are less than 0.05. The dose of 2 mg/day has a large p-value and the confidence interval includes zero, therefore there is no significant difference.

Conclusions

The assumptions:

1. Samples used are random iid samples.

2. Each sample is indeendent of one another, in other words, they are not paired.

3. The population distribution of each samle must be approximately normal or mound shaped and roughly symetric.

Conclusions:

Supplement type alone has no effect on the growth of teeth, the dosage does (significantly). Increasing the dosage will increase the growth of teeth. These conclusions are based on the low p-values and confidence levels not containing zero.