Statistical Inference Project 1 - Part 2

Joao Clemencip

27 September 2015

Overview

In this project we will attempt to show that the averages of many exponential distributions follow the Central Limit Theorem, having an approximately Gaussian distibution.

```
data(ToothGrowth)
t = ToothGrowth %>% tbl_df
```

Exploratory Data Analysis

Data Documentation

Description

The response is the 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).

Format

A data frame with 60 observations on 3 variables.

len (numeric): Tooth length supp (factor): Supplement type (VC or OJ). dose (numeric): Dose in milligrams.

Basic Data Analysis

Since the dose is an explanatory variable, we will convert into a factor instead of numeric.

```
t$dose = as.factor(t$dose)
head(t)
```

```
## Source: local data frame [6 x 3]
##
##     len   supp   dose
##   (dbl) (fctr) (fctr)
## 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
```

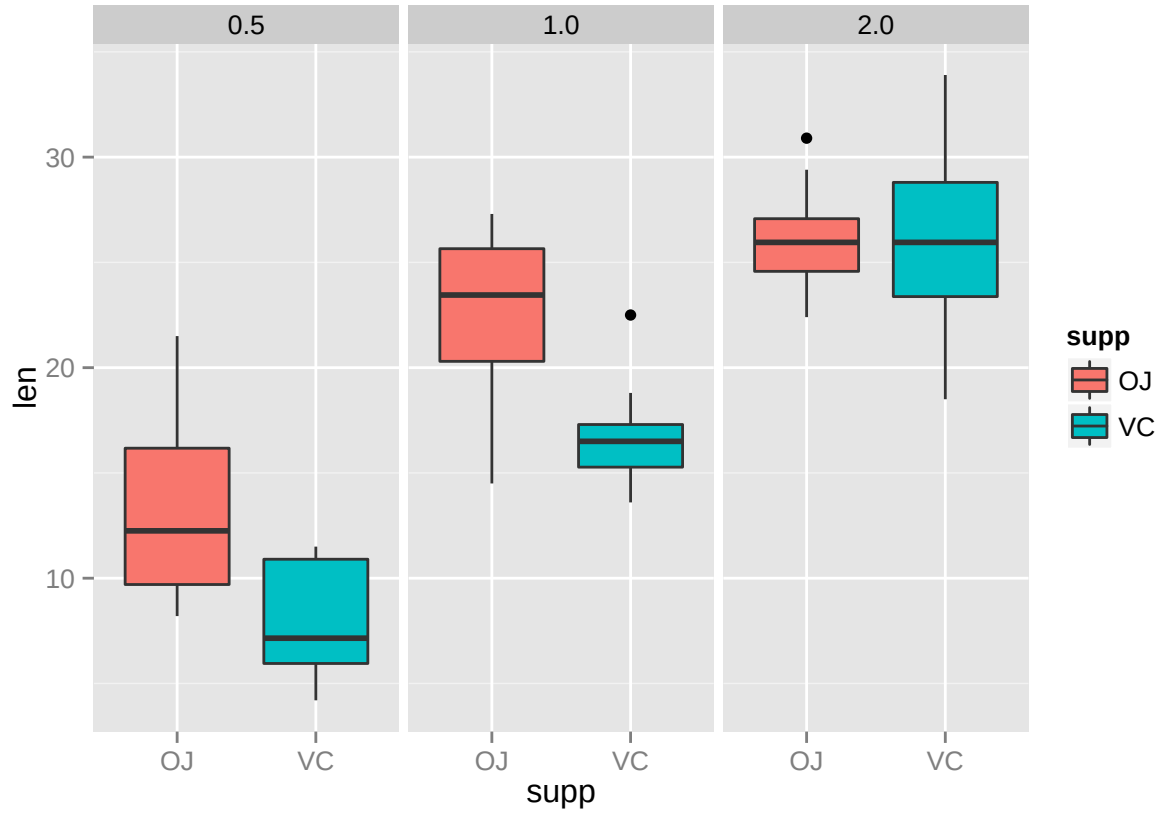
```
str(t)
```

```
## Classes 'tbl_df', 'tbl' and '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```
summary(t)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
ggplot(
  data = ToothGrowth,
  aes(x = supp, y = len)
  ) +
  geom_boxplot(
    aes(fill = supp)
  ) +
  facet_wrap(~ dose)
```



Based on this analysis, it appears that for both orange juice and ascorbic acid, there is a direct correlation with the dose supplied and the tooth growth. Orange juice appears to also be more effective at lower doses than for the equivalent dose of ascorbic acid. However, this difference is less prominent when higher doses were administered.

Confidence intervals and hypothesis testing

```
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  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 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
```

From these results, we cannot reject the null hypothesis that the supplement type has no effect on length.

We will now perorm a similar test, using the dose as a factor. For this, we will subset the data into 3 groups of dose combinations and perform tests between them.

```
t.test(len ~ dose, data = ToothGrowth %>% filter(dose != 0.5)) # 1.0 and 2.0
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means 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
```

```
t.test(len ~ dose, data = ToothGrowth %>% filter(dose != 1.0)) # 0.5 and 2.0
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means 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
```

```r
t.test(len ~ dose, data = ToothGrowth %>% filter(dose != 2.0)) # 0.5 and 1.0
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means 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
```

We can conclude that the dose has an effect on the growth.