Large Numbers and the Central Limit Theorem

Martí Cuquet

Abstract

In this report, we analyze the ToothGrowth data from the R datasets package. We use confidence intervals to compare tooth growth by supplement type and dose.

Exploratory data analysis

```
data(ToothGrowth)
```

The `ToothGrowth` data set reports on the effect of vitamin C on tooth growth in guinea pigs. It consists of 60 observations of 3 variables:

- `len`: The length of odontoblasts (teeth), a numerical variable ranging between 4.2 and 33.9.
- `supp`: The supplement type ("VC"" for ascorbic acid or "OJ" for orange juice), encoded as a factor.
- `dose`: The dose level in mg, a numerical variable with values 0.5, 1.0 or 2.0.

To get a bit more insight, we plot in Figure 1 three density plots of the teeth length depending on the supplement type (OJ or VC) and the dose level (0.5, 1.0 or 2.0 mg).

```
g1 <- ggplot(ToothGrowth, aes(x = len, fill = supp))
g1 <- g1 + geom_density(alpha = 0.6) + facet_wrap(~ dose)
g1 <- g1 + labs(title = "Length of odontoblasts", x = "Length", y = "Density")
g1 <- g1 + theme(text = element_text(size=8))
```

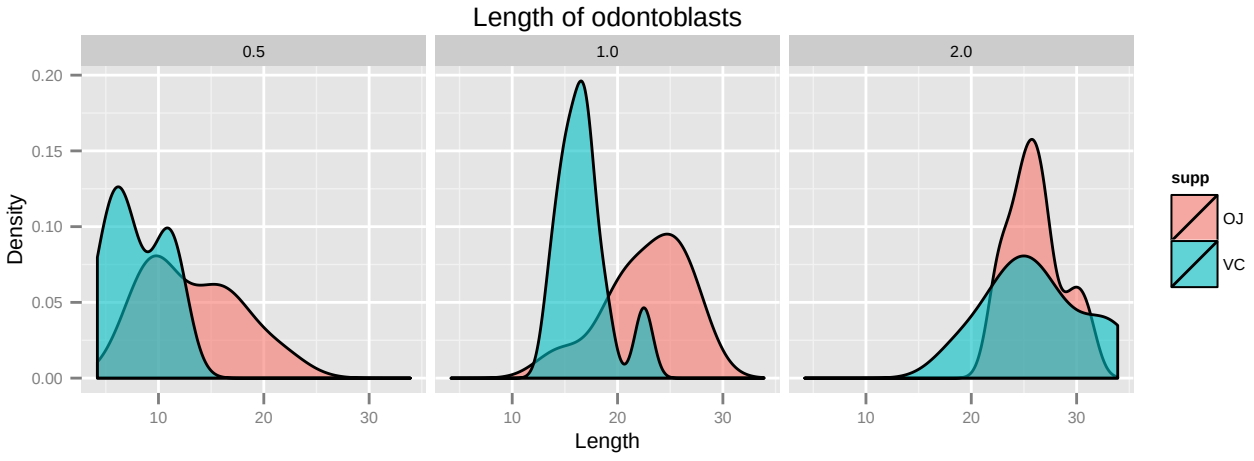

Figure 1: Density plots of the length of odontoblasts. Each panel corresponds to a dose level (0.5, 1.0 or 2.0 mg), red to orange juice supplement and turquoise to ascorbic acid supplement.

At first glance, it seems that orange juice triggers a higher teeth length for small dosages (0.5 and 1.0 mg), while for a higher dosage (2.0 mg) the results is basically the same.

We can have a closer look at the mean and variance with respect to dosage and supplement type:

```
ind <- list(ToothGrowth$supp, ToothGrowth$dose)
pantb <- tapply(ToothGrowth$len, ind,
                function(x) sprintf("%.2f(%.2f)", mean(x), sd(x)))
```

This is shown in the following table (standard deviations are shown in parenthesis):

	0.5	1	2
OJ	13.23(4.46)	22.70(3.91)	26.06(2.66)
VC	7.98(2.75)	16.77(2.52)	26.14(4.80)

Tooth growth depending on supplement type and dose

Here we compare tooth growth by supplement type and dosage. The exploratory data analysis suggests that the growth is higher for orange juice supplement if the dosage is of either 0.5 or 1 mg, and roughly the same if the dosage is 2 mg. As we have a small sample, we use Gosset's t confidence intervals to compare the average growth. We expect it to work well, as the distributions are roughly symmetric and mound shaped.

We calculate the three 95% confidence intervals of the difference of teeth length by supplement type corresponding to the three diferent doses.

```
tests <- data.frame(do.call(rbind,
                            lapply(unique(ToothGrowth$dose), function(x) {
                              t.test(len ~ factor(supp), data = ToothGrowth,
                                     paired = TRUE, subset = dose == x)
                              })))
tests <- data.frame(cbind(dose = unique(ToothGrowth$dose),
                          tests[, c("estimate", "conf.int")]))
names(tests) <- c("Dose", "Mean of the differences", "95% confidence interval")
```

The following table shows the result of the tests:

Dose	Mean of the differences	95% confidence interval
0.5	5.25	1.263458, 9.236542
1	5.93	1.951911, 9.908089
2	-0.08	-4.328976, 4.168976

These tests confirmed the observations in the exploratory data analysis. For a dose of 0.5 or 1 mg, the 95% confidence intervals do not contain 0 and hence imply that there is a statistically significant difference regarding the supplement type. More specifically, as the mean of the differences is positive, in these two cases (dose of 0.5 or 1 mg) orange juice leads to a higher teeth growth than ascorbic acid. (Note that in the paired test we substracted ascorbic acid from orange juice.) On the contrary, we can not observe any significant difference between the supplement types when a dose of 2.0 mg is given.

Conclusions

In conclusion, there is a statistically significant higher teeth growth for orange juice as opposed to ascorbic acid supplement, if a dose of 0.5 or 1 mg is given, while no significant difference exists for a dose of 2 mg. This result has been obtained assuming roughly symmetric and mound shaped distributions and using 95% confidence intervals via Gosset's t distribution.