

Central Limit Theorem with Exponentials and Applied Inference with ToothGrowth Data

Thomas Ljungberg Kristensen

2025-10-02

Overview

In this project, we use simulations in R to study the distribution of averages of exponential random variables, illustrating the Central Limit Theorem (CLT). We then analyze the ToothGrowth dataset, performing exploratory analysis and statistical inference to investigate the effect of supplement type and dose on tooth growth in guinea pigs.

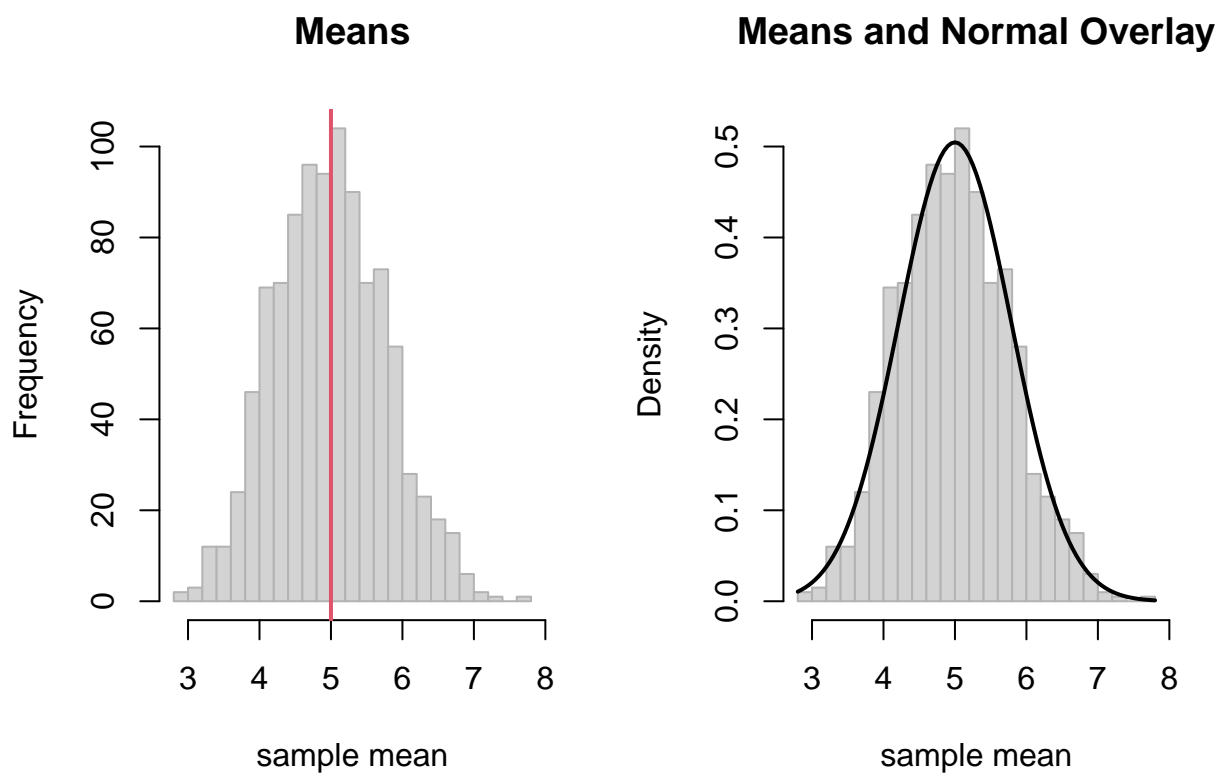
Part I — Simulation: Exponential and the CLT

The simulation is set up as the exponential distribution with rate 0.2 and 40 observations. See appendix for more details

This shows that the theoretical mean is 5 and simulated mean of sample means is 4.985. Theoretical variance of the mean is 0.625 (and sd 0.7906); simulated variance is 0.6 (and sd 0.7746).

The following figures (see appendix for more information) shows

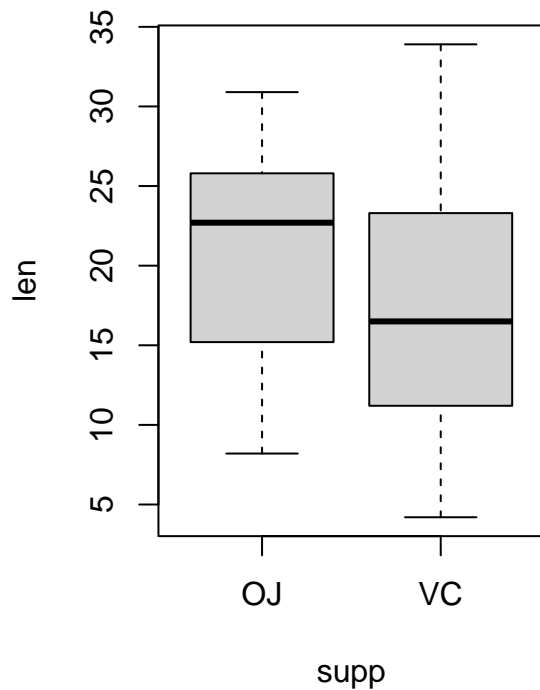
- 1) How variable the sample is (via variance) and compare it to the theoretical variance of the distribution,
- 2) Show that the distribution is approximately normal



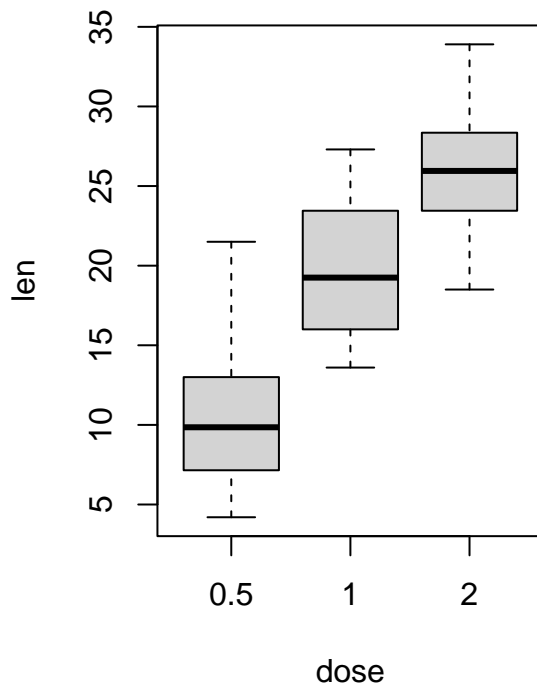
Part II — ToothGrowth: EDA and Inference

Plotting the ToothGrowth data, see figures or the appendix for more details, shows that length increases with dose but supplement only appears to do modest changes.

Tooth length by supplement



Tooth length by dose



Using the Welch t-test to determine p-value and confidence intervals for tooth growth by supplement (see appendix for details)

P-value = 0.0606. 95% Confidence interval for (OJ - VC) difference: [-0.17, 7.57].

And p-values and confidence intervals for pairwise comparison for tooth growth by dose, see appendix for details

```
## $'0.5 vs 1'
## [1] -11.983781 -6.276219
## attr("conf.level")
## [1] 0.95
##
## $'1 vs 2'
## [1] -8.996481 -3.733519
## attr("conf.level")
## [1] 0.95
##
## $'0.5 vs 2'
## [1] -18.15617 -12.83383
## attr("conf.level")
## [1] 0.95

## p(0.5 vs 1)    p(1 vs 2)  p(0.5 vs 2)
## 1.268301e-07  1.906430e-05  4.397525e-14
```

Assumptions

We are assuming that the observations are independent.

Conclusions

For the CLT simulation, we see that means of 40 exponentials center near 5 with variance approx 25/40, and their distribution is approximately normal.

For ToothGrowth, we see strong, statistically significant dose effect on tooth length but the supplement effect is smaller and context-dependent.

Appendix

Data - Part I — Simulation: Exponential and the CLT

```
lambda <- 0.2
n      <- 40
nosim  <- 1000

# 1000 sample-means of 40 exponentials
means_exp <- replicate(nosim, mean(rexp(n, rate = lambda)))

# Theoretical center/variance for the sample mean
theo_mean <- 1/lambda
theo_var  <- (1/lambda)^2 / n
theo_sd   <- sqrt(theo_var)

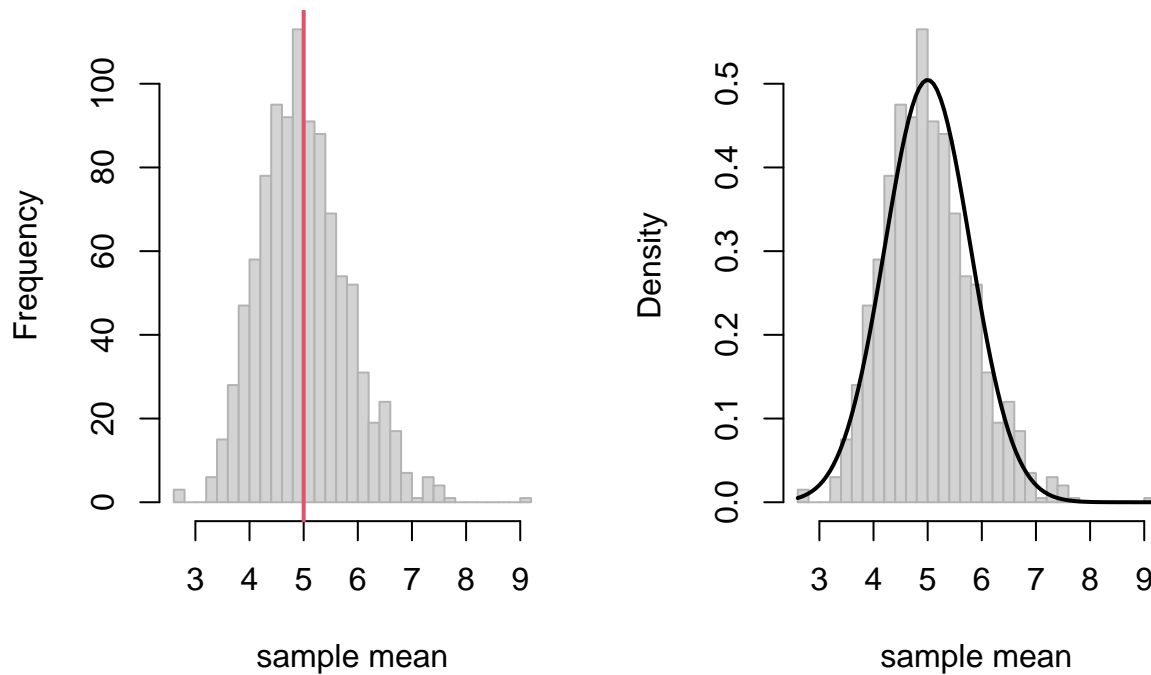
sim_mean  <- mean(means_exp)
sim_var   <- var(means_exp)
sim_sd    <- sd(means_exp)

par(mfrow=c(1,2))

hist(means_exp, breaks=30, main="Means of 40 Exponentials (n=1000)", xlab="sample mean", border="gray70",
     abline(v = theo_mean, col=2, lwd=2))

hist(means_exp, breaks=30, freq=FALSE, main="Means of 40 Exponentials with Normal Overlay", xlab="sample mean",
     curve(dnorm(x, mean=theo_mean, sd=theo_sd), add=TRUE, lwd=2))
```

Means of 40 Exponentials (n=100)ns of 40 Exponentials with Normal



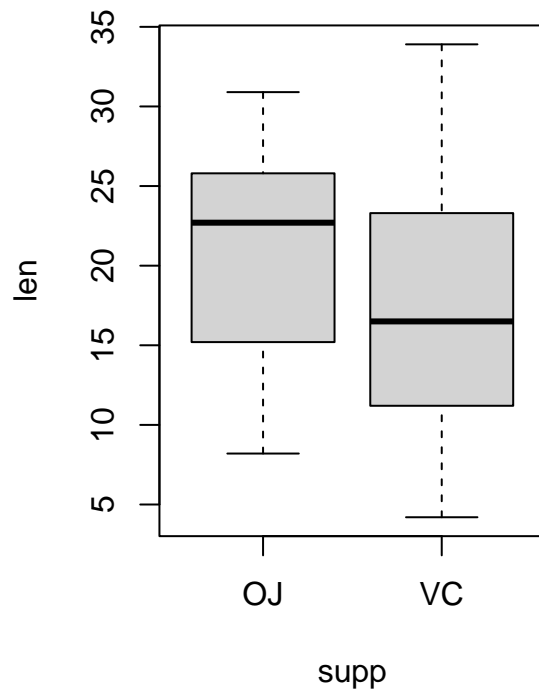
Data Part II — ToothGrowth: EDA and Inference

```
data(ToothGrowth)
TG <- ToothGrowth
```

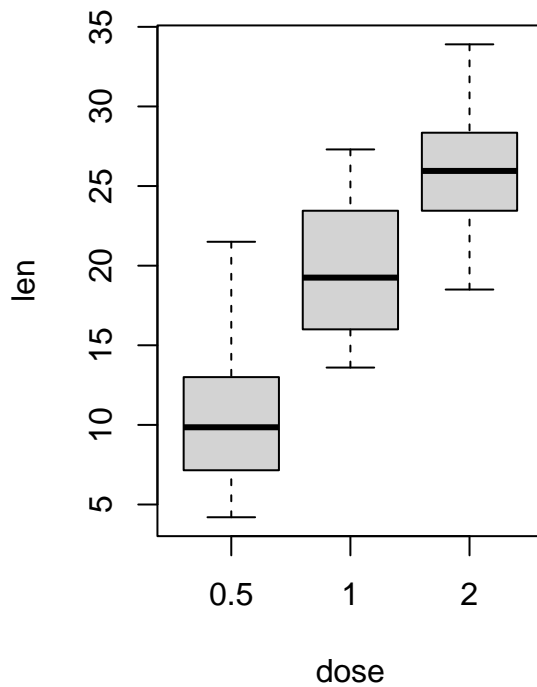
```
par(mfrow=c(1,2))

boxplot(len ~ supp, data=TG, main="Tooth length by supplement", ylab="len")
boxplot(len ~ dose, data=TG, main="Tooth length by dose", ylab="len")
```

Tooth length by supplement



Tooth length by dose



```
TG$dose <- factor(TG$dose) # treat as factor
t_05_1 <- t.test(len ~ dose, data=subset(TG, dose %in% c(0.5,1)))
t_1_2 <- t.test(len ~ dose, data=subset(TG, dose %in% c(1,2)))
t_05_2 <- t.test(len ~ dose, data=subset(TG, dose %in% c(0.5,2)))
```

```
list(`0.5 vs 1`=t_05_1$conf.int,
     `1 vs 2`=t_1_2$conf.int,
     `0.5 vs 2`=t_05_2$conf.int)
```

```
## $'0.5 vs 1'
## [1] -11.983781 -6.276219
## attr("conf.level")
## [1] 0.95
##
## $'1 vs 2'
## [1] -8.996481 -3.733519
## attr("conf.level")
## [1] 0.95
##
## $'0.5 vs 2'
## [1] -18.15617 -12.83383
## attr("conf.level")
## [1] 0.95
```

```
c(`p(0.5 vs 1)`=t_05_1$p.value,  
  `p(1 vs 2)` =t_1_2$p.value,  
  `p(0.5 vs 2)`=t_05_2$p.value)
```

```
## p(0.5 vs 1)    p(1 vs 2)  p(0.5 vs 2)  
## 1.268301e-07  1.906430e-05  4.397525e-14
```