Statistical Inference Course Project

FarzanehBina

2024-03-01

Part1: Simulation Exercise

Overview: This part is going to execute simulations and data analysises to illustrate application of the central limit theorem. R programming will be the major tool to realize the mentioned goal.For this analysis, the lambda will be set to 0.2 for all of the simulations. This investigation will compare the distribution of averages of 40 exponentials over 1000 simulations.

Simulations

Set the simulation variables lambda, exponentials, and seed.

```
ECHO=TRUE
set.seed(1337)
lambda = 0.2
exponentials = 40
```

Run Simulations with variables

```
simMeans = NULL
for (i in 1 : 1000) simMeans = c(simMeans, mean(rexp(exponentials, lambda)))
```

Sample Mean versus Theoretical Mean

Sample Mean   Calculating the mean from the simulations with give the sample mean.

```
mean(simMeans)
```

```
## [1] 5.055995
```

Theoretical Mean   The theoretical mean of an exponential distribution is lambdaˆ-1.

```
lambdaˆ-1
```

```
## [1] 5
```

Comparison   There is only a slight difference between the simulations sample mean and the exponential distribution theoretical mean.

```r
abs(mean(simMeans)-lambda^-1)
```

```
## [1] 0.05599526
```

Sample Variance versus Theoretical Variance

Sample Variance Calculating the variance from the simulation means with give the sample variance.

```r
var(simMeans)
```

```
## [1] 0.6543703
```

Theoretical Variance The theoretical variance of an exponential distribution is (lambda * sqrt(n))^-2.

```r
(lambda * sqrt(exponentials))^-2
```

```
## [1] 0.625
```

Comparison There is only a slight difference between the simulations sample variance and the exponential distribution theoretical variance.

```r
abs(var(simMeans)-(lambda * sqrt(exponentials))^-2)
```

```
## [1] 0.0293703
```

Distribution

This is a density histogram of the 1000 simulations. There is an overlay with a normal distribution that has a mean of lambda^-1 and standard deviation of (lambda*sqrt(n))^-1, the theoretical normal distribution for the simulations.

```r
library(ggplot2)
ggplot(data.frame(y=simMeans), aes(x=y)) +
  geom_histogram(aes(y=..density..), binwidth=0.2, fill="#0072B2",
                 color="black") +
  stat_function(fun=dnorm, arg=list(mean=lambda^-1,
                                    sd=(lambda*sqrt(exponentials))^-1),
                size=2) +
  labs(title="Plot of the Simulations", x="Simulation Mean")
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```

```
## Warning in stat_function(fun = dnorm, arg = list(mean = lambda^-1, sd = (lambda
## * : Ignoring unknown parameters: 'arg'
```

```
## Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

Plot of the Simulations
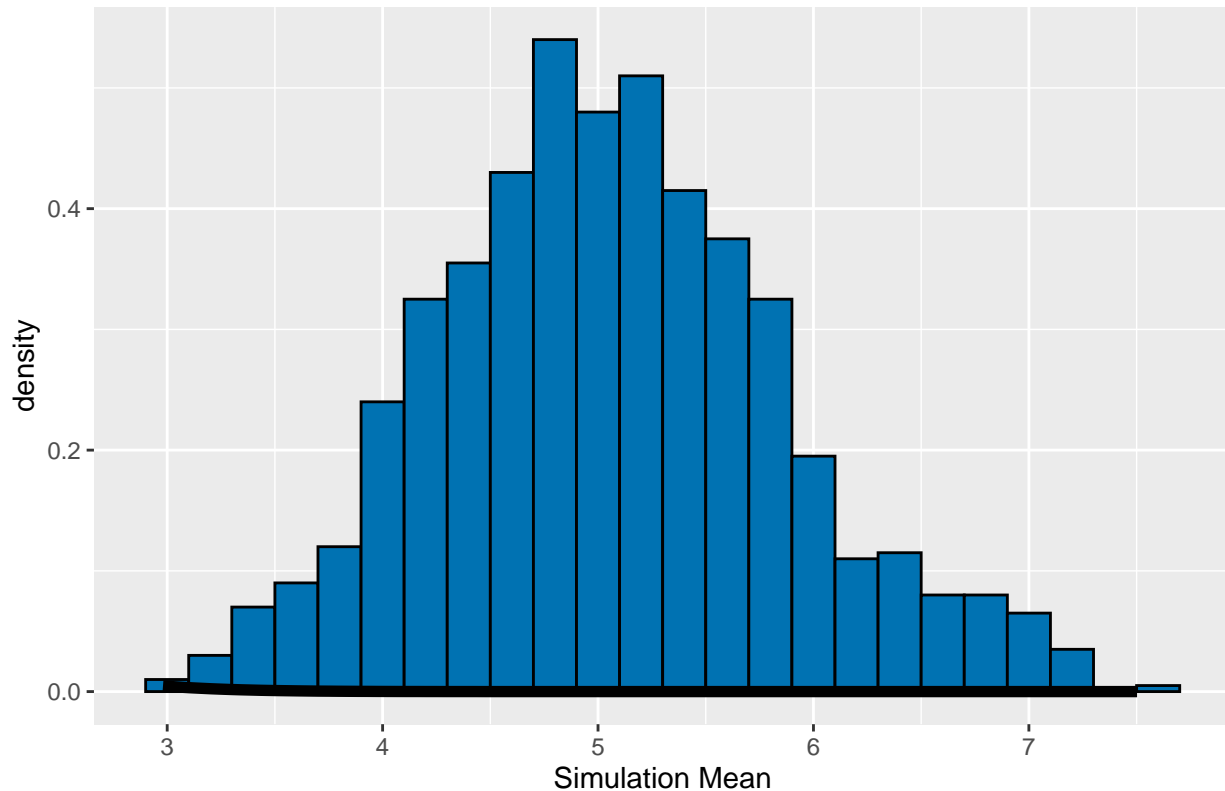
Part 2: Analyze the ToothGrowth data in the R datasets package

Overview: In this part we will do some statistical data analysises about the Toothlength data.Load the ToothGrowth data and perform some basic exploratory data analyses.

Load the ToothGrowth data and perform exploratory data analyses

```
library(datasets)
data(ToothGrowth)
str(ToothGrowth)
head(ToothGrowth)
summary(ToothGrowth)
```

```
library(ggplot2)
t = ToothGrowth
levels(t$supp) <- c("Orange Juice", "Ascorbic Acid")
```

```
ggplot(t, aes(x=factor(dose), y=len)) +
  facet_grid(.~supp) +
  geom_boxplot(aes(fill = supp), show_guide = FALSE) +
  labs(title="Guinea pig tooth length by dosage for each type of supplement",
    x="Dose (mg/day)",
    y="Tooth Length")
```

```
## Warning: The 'show_guide' argument of 'layer()' is deprecated as of ggplot2 2.0.0.
## i Please use the 'show.legend' argument instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```


Guinea pig tooth length by dosage for each type of supplement
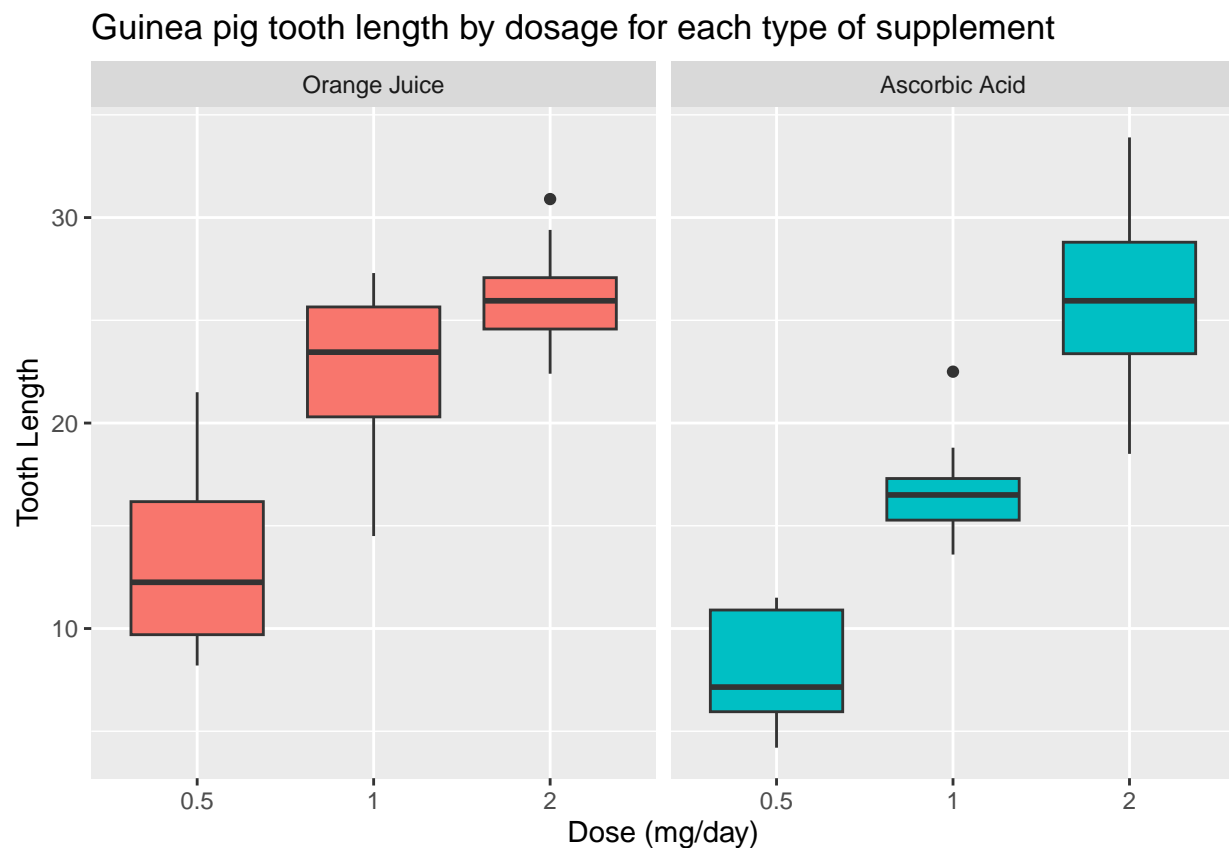
Basic summary of the data

The box plots seem to show, increasing the dosage increases the tooth growth. Orange juice is more effective than ascorbic acid for tooth growth when the dosage is .5 to 1.0 milligrams per day. Both types of supplements are equally as effective when the dosage is 2.0 milligrams per day.

Use confidence intervals & hypothesis tests to compare tooth growth by supplement and dose

Hypothesis #1   Orange juice & ascorbic acid deliver the same tooth growth across the data set.

```
hypoth1<-t.test(len ~ supp, data = t)
hypoth1$conf.int
```

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

```
hypoth1$p.value
```

```
## [1] 0.06063451
```

The confidence intervals includes 0 and the p-value is greater than the threshold of 0.05. The null hypothesis cannot be rejected.

Hypothesis #2   For the dosage of 0.5 mg/day, the two supplements deliver the same tooth growth.

```
hypoth2<-t.test(len ~ supp, data = subset(t, dose == 0.5))
hypoth2$conf.int
```

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

```
hypoth2$p.value
```

```
## [1] 0.006358607
```

The confidence interval does not include 0 and the p-value is below the 0.05 threshold. The null hypothesis can be rejected. The alternative hypothesis that 0.5 mg/day dosage of orange juice delivers more tooth growth than ascorbic acid is accepted.

Hypothesis #3   For the dosage of 1 mg/day, the two supplements deliver the same tooth growth

```
hypoth3<-t.test(len ~ supp, data = subset(t, dose == 1))
hypoth3$conf.int
```

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

```
hypoth3$p.value
```

```
## [1] 0.001038376
```

The confidence interval does not include 0 and the p-value is smaller than the 0.05 threshold. The null hypothesis can be rejected. The alternative hypothesis that 1 mg/day dosage of orange juice delivers more tooth growth than ascorbic acid is accepted.

Hypothesis #4   For the dosage of 2 mg/day, the two supplements deliver the same tooth growth

```
hypoth4<-t.test(len ~ supp, data = subset(t, dose == 2))
hypoth4$conf.int
```

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

```
hypoth4$p.value
```

```
## [1] 0.9638516
```

The confidence interval does include 0 and the p-value is larger than the 0.05 threshold. The null hypothesis cannot be rejected.

Conclusions & assumptions

Orange juice delivers more tooth growth than ascorbic acid for dosages 0.5 & 1.0. Orange juice and ascorbic acid deliver the same amount of tooth growth for dose amount 2.0 mg/day. For the entire data set we cannot conclude orange juice is more effective that ascorbic acid.

Assumptions

- Normal distribution of the tooth lengths
- No other unmeasured factors are affecting tooth length