

Simulation and inferential data analysis

Kirill Setdekov

13 may 2019

Part 1.A simulation exercise.

Overview

In a few (2-3) sentences explain what is going to be reported on.

Simulations:

Include English explanations of the simulations you ran, with the accompanying R code. Your explanations should make clear what the R code accomplishes.

Sample Mean versus Theoretical Mean:

Include figures with titles. In the figures, highlight the means you are comparing. Include text that explains the figures and what is shown on them, and provides appropriate numbers.

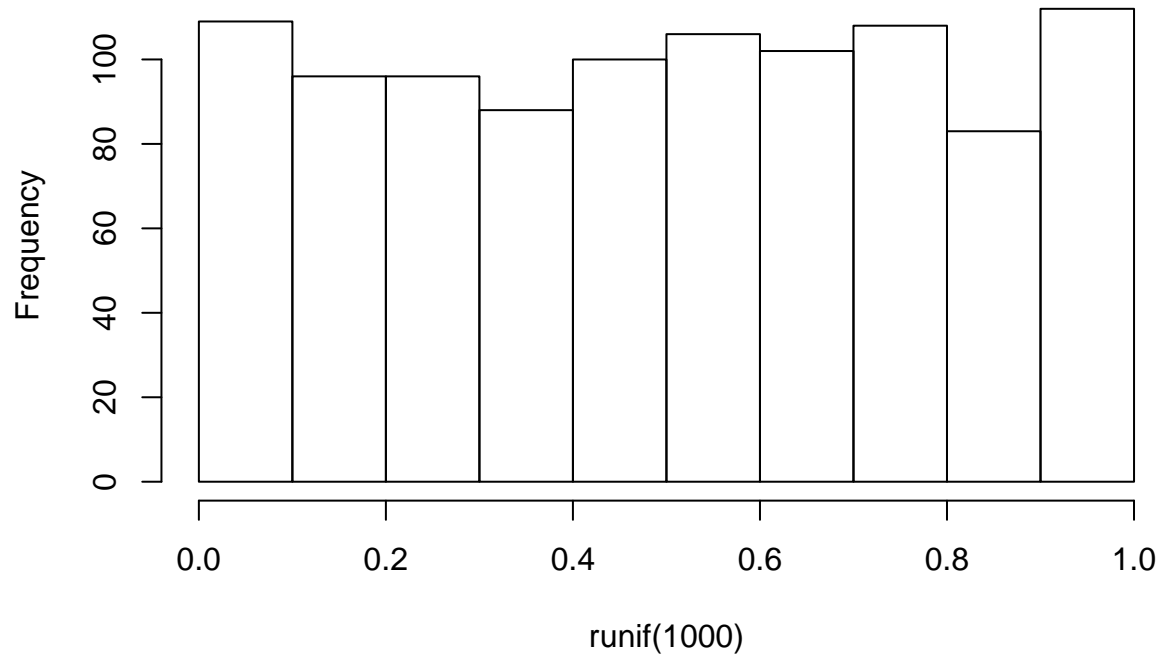
Sample Variance versus Theoretical Variance:

Include figures (output from R) with titles. Highlight the variances you are comparing. Include text that explains your understanding of the differences of the variances.

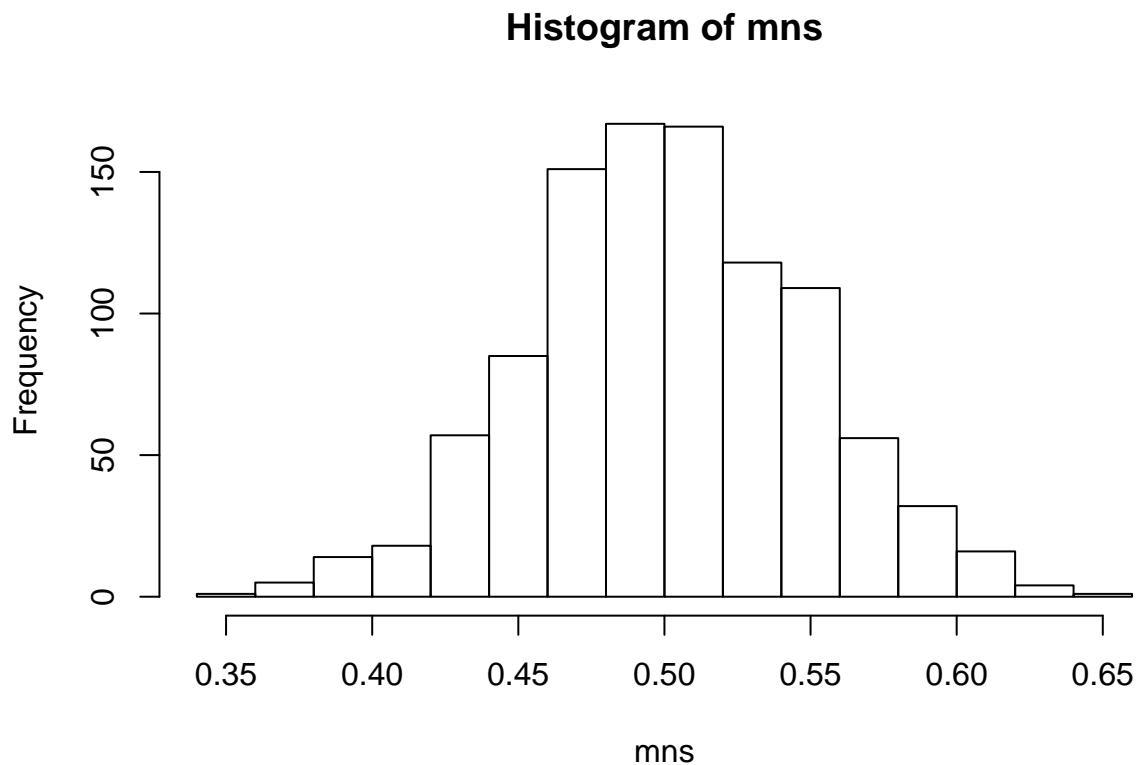
Distribution: Via figures and text, explain how one can tell the distribution is approximately normal.

```
hist(runif(1000))
```

Histogram of runif(1000)



```
mns = NULL
for (i in 1 : 1000) mns = c(mns, mean(runif(40)))
hist(mns)
```



Basic inferential data analysis.

```
tooth <- datasets::ToothGrowth
```

```
summary(tooth)
```

```
##      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
```

```
library(ggplot2)
```

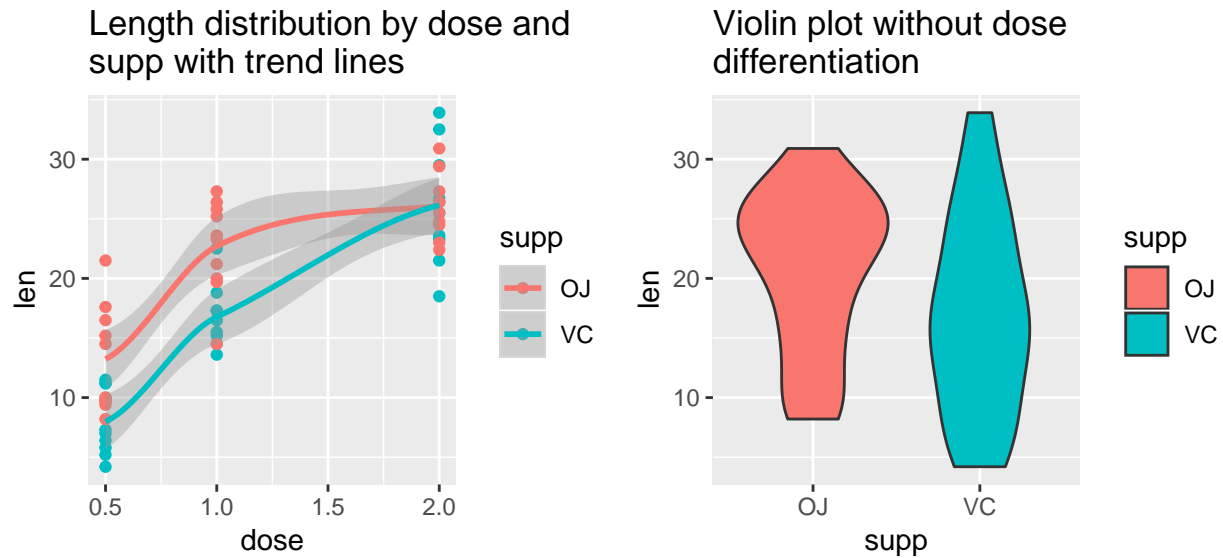
```
library(gridExtra)
```

```
p1 <-
```

```
  ggplot(tooth, aes(x = dose, y = len, color = supp)) + geom_point() + geom_smooth() +
  labs(title = "Length distribution by dose and \nsupp with trend lines")
```

```
p2 <-
```

```
  ggplot(tooth, aes(y = len, x = supp, fill = supp)) + geom_violin() + labs(title = "Violin plot with")
  grid.arrange(p1, p2, nrow = 1)
```



testing

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

conclusions

State your conclusions and the assumptions needed for your conclusions.

review criteria

- Did you show where the distribution is centered at and compare it to the theoretical center of the distribution?
- Did you show how variable it is and compare it to the theoretical variance of the distribution?
- Did you perform an exploratory data analysis of at least a single plot or table highlighting basic features of the data?
- Did the student perform some relevant confidence intervals and/or tests?
- Were the results of the tests and/or intervals interpreted in the context of the problem correctly?
- Did the student describe the assumptions needed for their conclusions?