

Assignment1

Matteo Gambera

25 marzo 2020

This is the project for the statistical inference class. In it, you will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

- A simulation exercise.
- Basic inferential data analysis.

You will create a report to answer the questions. Given the nature of the series, ideally you'll use knitr to create the reports and convert to a pdf. (I will post a very simple introduction to knitr). However, feel free to use whatever software that you would like to create your pdf.

Each pdf report should be no more than 3 pages with 3 pages of supporting appendix material if needed (code, figures, etcetera).

```
set.seed(100)
lambda <- 0.2
n <- 40
numsim <- 1000
```

Generate values

```
dt <- matrix(rexp(n*numsim,lambda),numsim)
```

Calculate theory statistics

```
TheoryMean <- 1/lambda
RowMeans <- apply(dt,1,mean); ActualMean <- mean(RowMeans)
TheorySD <- ((1/lambda) * (1/sqrt(n))); ActualSD <- sd(RowMeans)
TheoryVar <- TheorySD^2 ; ActualVar <- var(RowMeans)
```

1. Show the sample mean and compare it to the theoretical mean of the distribution.

```
round(ActualMean,5)
```

```
## [1] 4.9997
```

```
TheoryMean
```

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

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
round(ActualVar,3)
```

```
## [1] 0.634
```

```
TheoryVar
```

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

3. Show that the distribution is approximately normal.

```
library(ggplot2)
dfRowMeans <- data.frame(RowMeans)
mp <- ggplot(dfRowMeans, aes(x=RowMeans))
mp <- mp + geom_histogram(binwidth = lambda, fill="grey", color="black", aes(y = ..density..))
mp <- mp + labs(title="Distribution of 1,000 Instances of 40 Samples ", x="Mean ", y="Frequency")
mp <- mp + geom_vline(xintercept=ActualMean, size=1.0, color="green")
mp <- mp + stat_function(fun=dnorm, args=list(mean=ActualMean, sd=ActualSD), color = "green", size = 1.0)
mp <- mp + geom_vline(xintercept=TheoryMean, size=1.0, color="red")
mp <- mp + stat_function(fun=dnorm, args=list(mean=TheoryMean, sd=TheorySD), color = "blue", size = 1.0)
mp
```

Distribution of 1,000 Instances of 40 Samples

