RLisbonaMSDS6306_Week4_BootstrapSampling

Randy Lisbona

June 4, 2016

Create the normal and exponential sample datasets

Use rnorm and rexp to create some sample datasets

Print the datasets, inlcude the first 10 records from each dataset

```
##  [1] 10.467260 10.808825 10.205415  9.419556  8.633051 10.397145  9.996729
##  [8] 10.049278 10.869570 10.101158
```
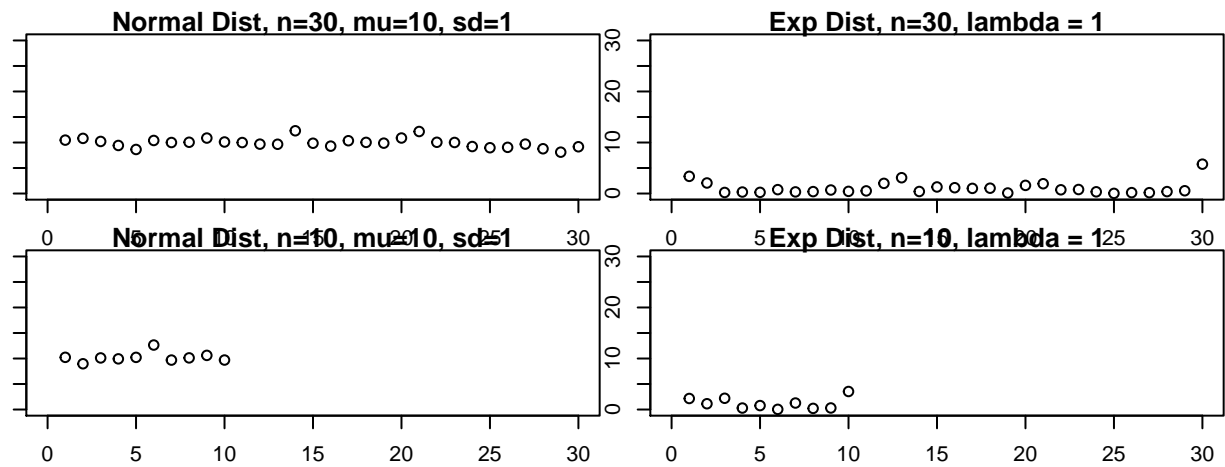
```
##  [1] 10.230742  8.969097 10.113847  9.929830 10.223464 12.640750  9.692279
##  [8] 10.121155 10.613238  9.699977
```

```
##  [1] 3.3686244 2.0671982 0.2060937 0.2934132 0.2291777 0.7628933 0.3188779
##  [8] 0.3770723 0.6897965 0.4241370
```

```
##  [1] 2.16463483 1.13422715 2.20406494 0.29421494 0.77443515 0.03714913
##  [7] 1.27610055 0.22577942 0.30506875 3.53379319
```
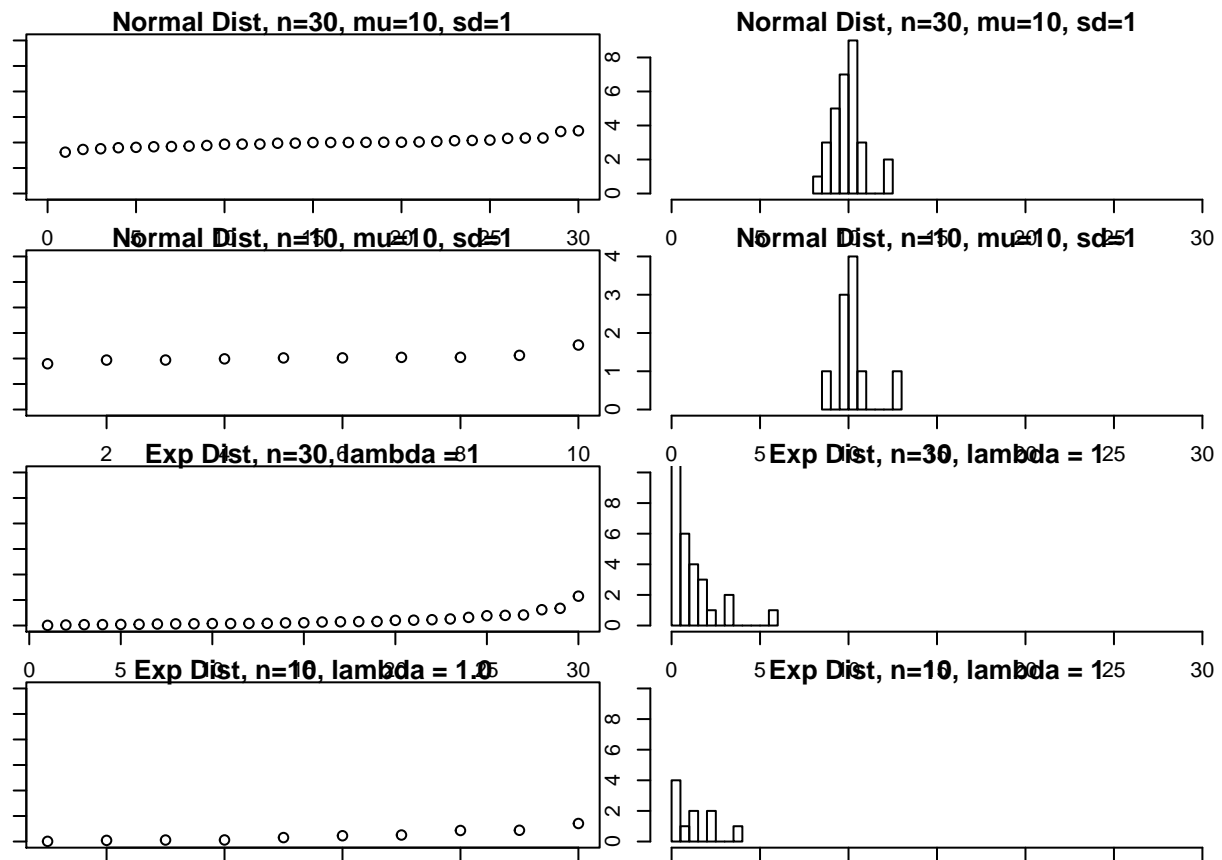
Explore the data with Plot of the normal and exponential sample datasets

Use plot and hist to compare the datasets set x and y limits to make it easier to compare plots

Normal Dist, n=30, mu=10, sd=1

Exp Dist, n=30, lambda = 1

Normal Dist, n=10, mu=10, sd=1

Exp Dist, n=10, lambda = 1

Sort the records ascending and plot again. Include histograms

set x and y limits to make it easier to compare plots

2

Normal Dist, n=30, mu=10, sd=1

Normal Dist, n=30, mu=10, sd=1

Normal Dist, n=10, mu=10, sd=1

Normal Dist, n=10, mu=10, sd=1

Exp Dist, n=30, lambda = 1

Exp Dist, n=30, lambda = 1

Exp Dist, n=10, lambda = 1.0

Exp Dist, n=10, lambda = 1

Resample the datasets, compare original histogram to resampled histogram # notice that the resampled histograms closely resemble a normal distribution, illustrating the central limit theorem
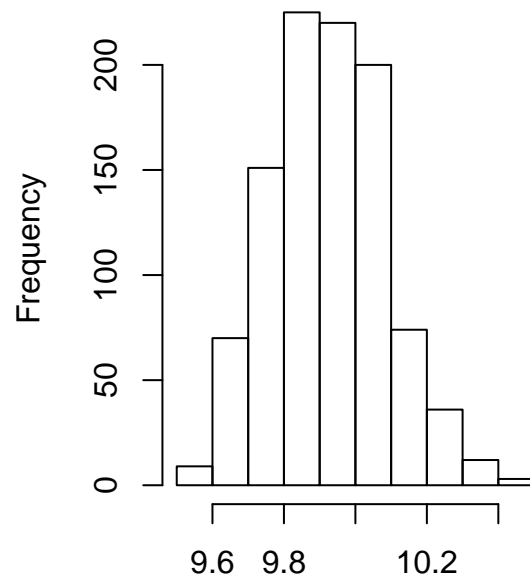
[1] 9.933821

Original dataset   **Bootstrap resampled**
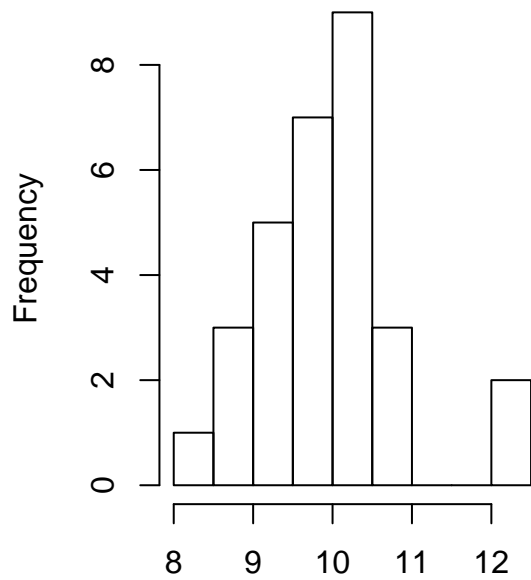
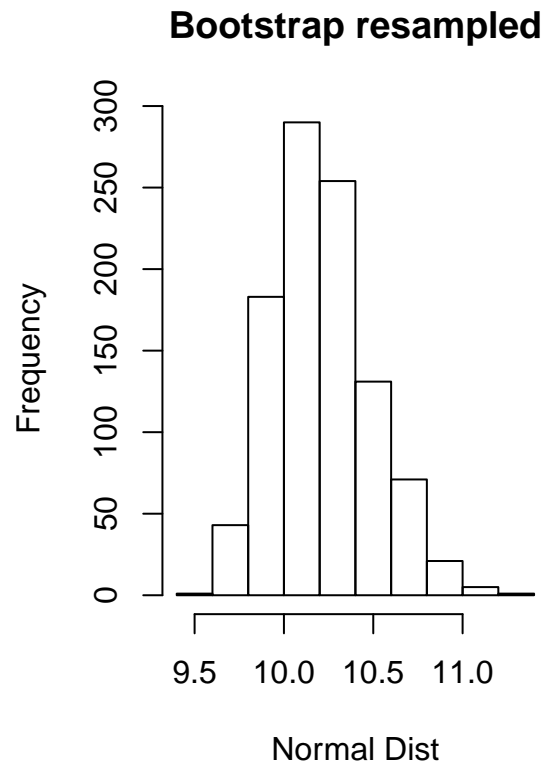Normal Dist, n=30, mu=10, sd=1   Normal Dist
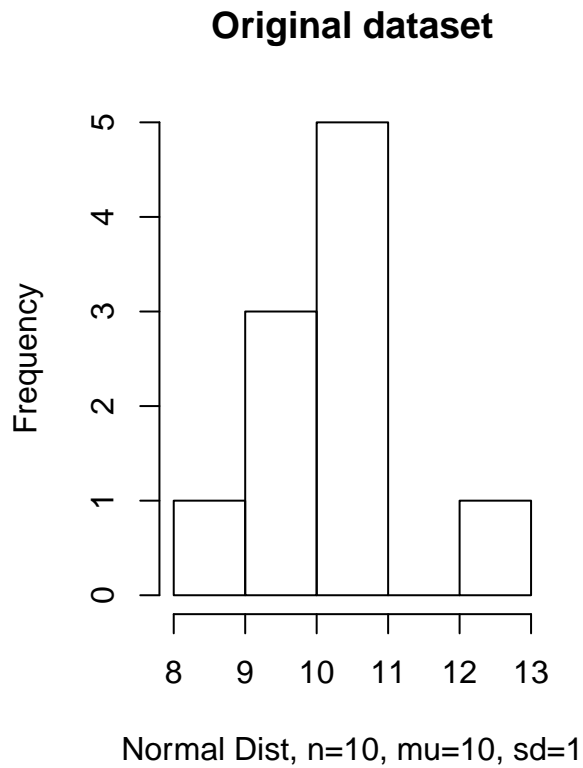
```
## [1] "Original dataset"

##    vars  n mean   sd median trimmed  mad min  max range skew kurtosis   se
## 1     1 30 9.93 0.91     10    9.88 0.65 8.1 12.3  4.19 0.63     0.71 0.17

## [1] "Resampled dataset"

##    vars    n mean   sd median trimmed  mad  min   max range skew kurtosis
## 1     1 1000 9.93 0.16   9.92    9.92 0.16 9.51 10.49  0.98 0.26    -0.06
##      se
## 1 0.01

## [1] 10.22344
```

Original dataset
                                                      Bootstrap resampled

Normal Dist, n=10, mu=10, sd=1   Normal Dist

```
## [1] "Original dataset"

##   vars  n  mean   sd median trimmed  mad  min   max range skew kurtosis
## 1    1 10 10.22 0.96  10.12   10.08 0.45 8.97 12.64  3.67 1.36      1.4
##     se
## 1 0.3

## [1] "Resampled dataset"

##   vars    n  mean   sd median trimmed  mad min   max range skew kurtosis
## 1    1 1000 10.21 0.27  10.19    10.2 0.27 9.6 11.21  1.61 0.48     0.06
##     se
## 1 0.01

## [1] 1.058215
```
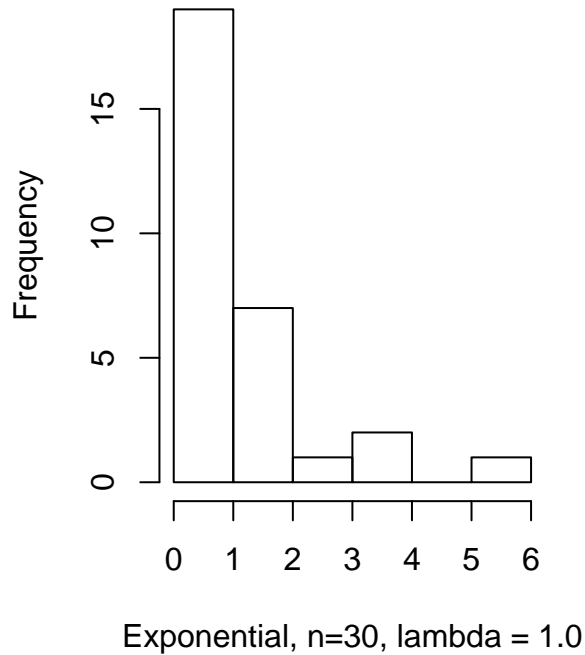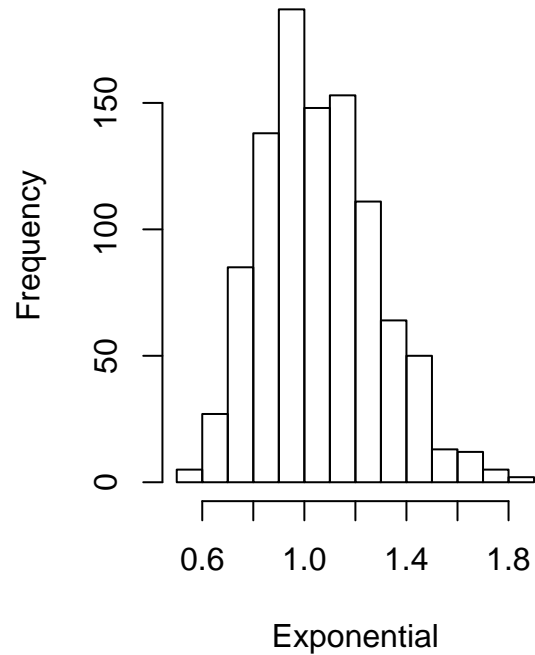
Original dataset

Exponential, n=30, lambda = 1.0

Bootstrap resampled

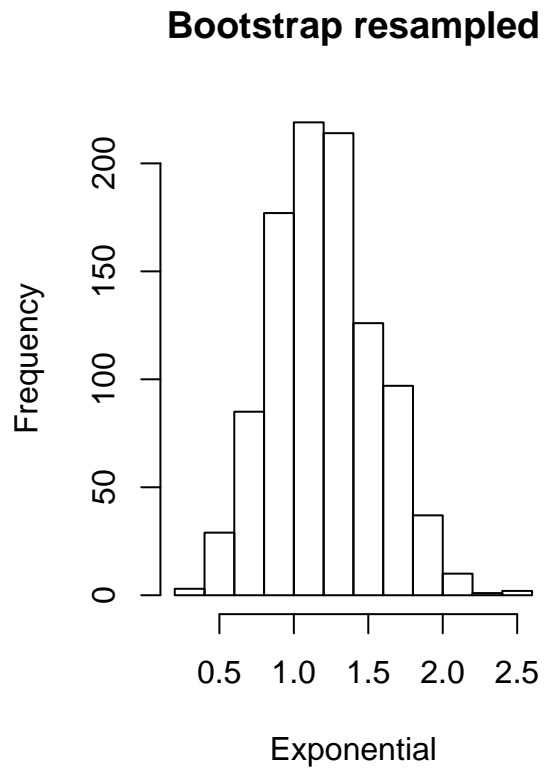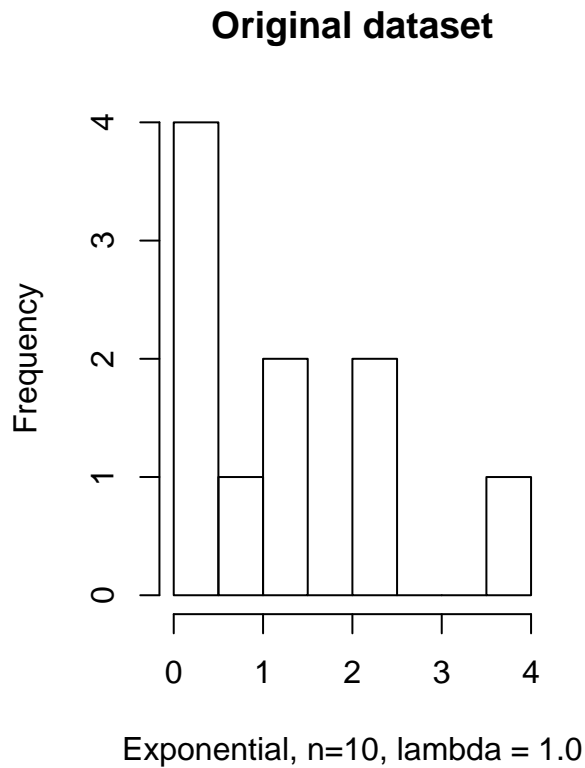
Exponential

```
## [1] "Original dataset"

##   vars  n mean   sd median trimmed mad  min  max range skew kurtosis   se
## 1    1 30 1.06 1.23   0.62     0.8 0.6 0.03 5.76  5.72 2.15     4.88 0.22

## [1] "Resampled dataset"

##   vars    n mean   sd median trimmed  mad  min  max range skew kurtosis
## 1    1 1000 1.06 0.22   1.04    1.05 0.23 0.52 1.82  1.31 0.44    -0.04
##      se
## 1 0.01

## [1] 1.194947
```

Original dataset

Exponential, n=10, lambda = 1.0

Bootstrap resampled

Exponential

```
## [1] "Original dataset"

##   vars  n mean   sd median trimmed  mad  min  max range skew kurtosis   se
## 1     1 10 1.19 1.13   0.95    1.05 1.03 0.04 3.53   3.5 0.75     -0.8 0.36

## [1] "Resampled dataset"

##   vars    n mean   sd median trimmed  mad  min  max range skew kurtosis
## 1     1 1000 1.21 0.35   1.19     1.2 0.35 0.26 2.51  2.26 0.29    -0.08
##      se
## 1 0.01
```

Conclusion

The bootstrap method can be used to create a sample distribution from small data sets that approximates a normal sample from the original population , demonstrating the central limit theorem.