

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, include the first 10 records from each dataset

```
## [1] 9.375238 12.042547 9.551198 10.263642 9.609880 9.475843 10.093066
## [8] 10.009806 11.559126 10.670639

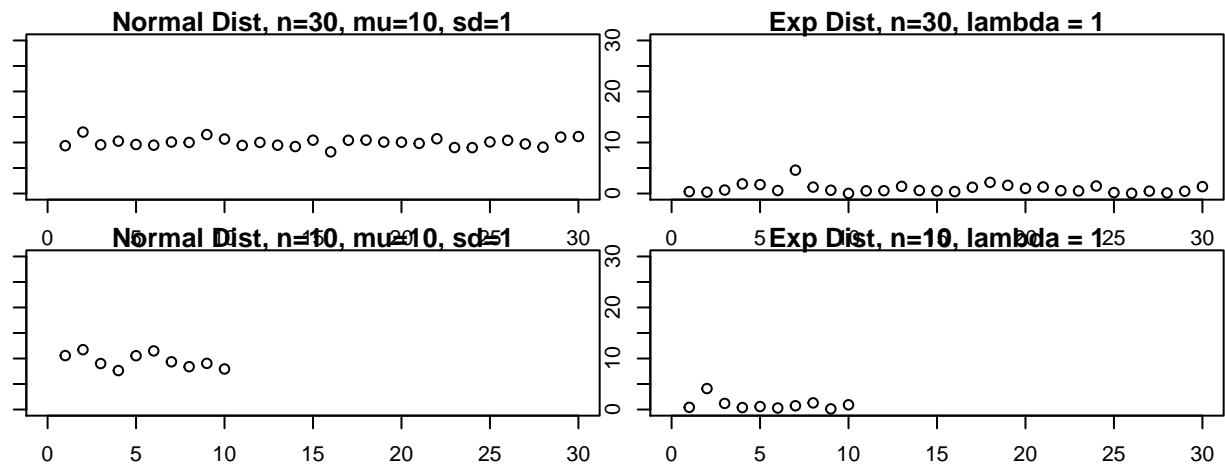
## [1] 10.571837 11.728115 9.008957 7.638036 10.541675 11.487561 9.351434
## [8] 8.392174 9.047143 7.947642

## [1] 0.36791059 0.26898470 0.68497053 1.89173806 1.75132854 0.58634140
## [7] 4.57834219 1.26032563 0.64594186 0.02722699

## [1] 0.4280592 4.1090919 1.1921795 0.3848168 0.5788840 0.2988509 0.7363150
## [8] 1.3036542 0.1252027 0.9225025
```

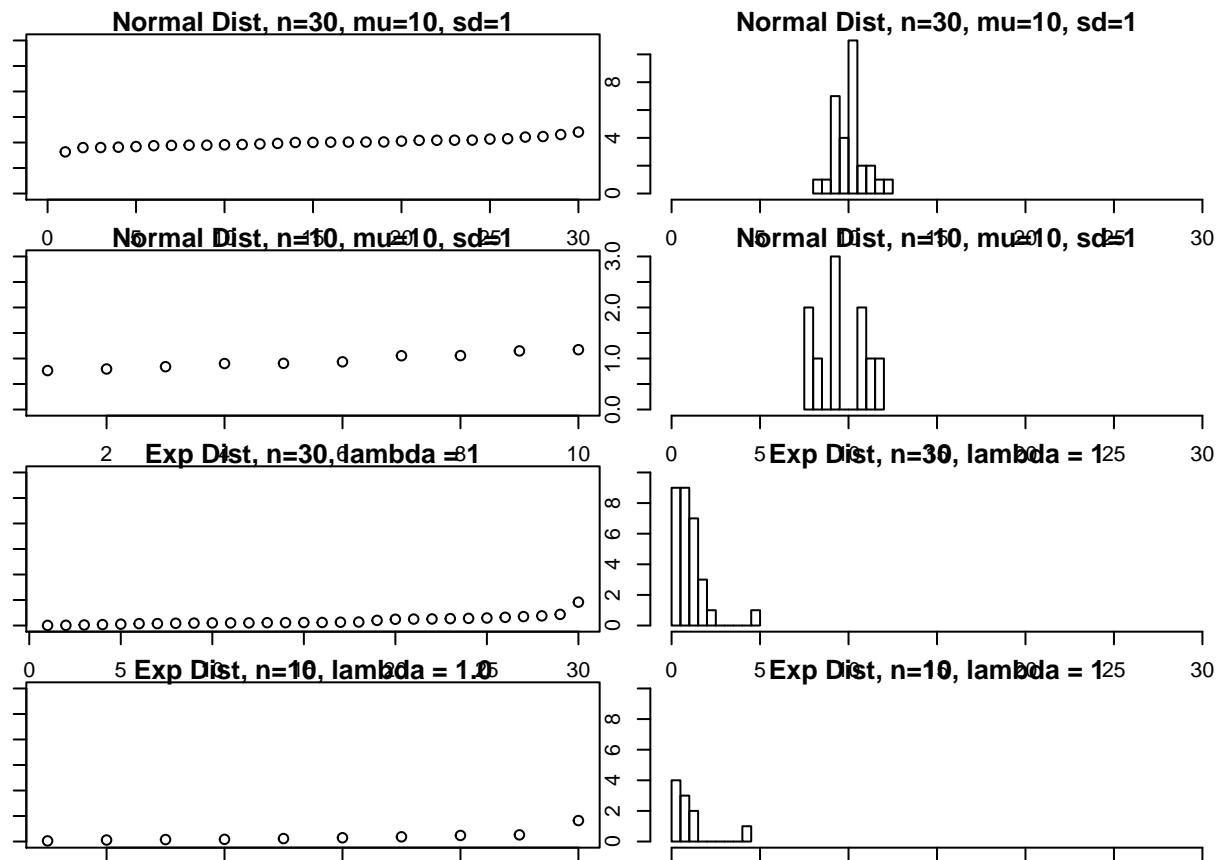
Explore the data with Plot of the normal and exponential sample datasets

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



Sort the records ascending and plot again. Include histograms

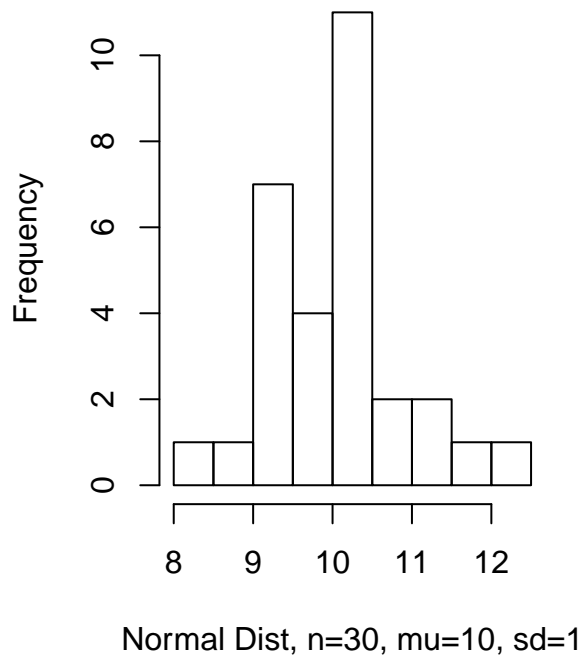
set x and y limits to make it easier to compare plots



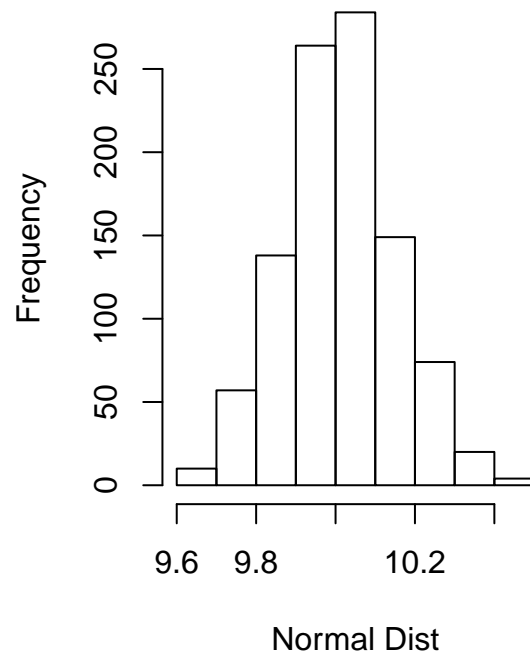
Resample the datasets, compare original histogram to resampled histogram # notice that the resampled histograms closely resemble a normal distribution, illustrating the central limit theorem

[1] 10.01863

Original dataset



Bootstrap resampled



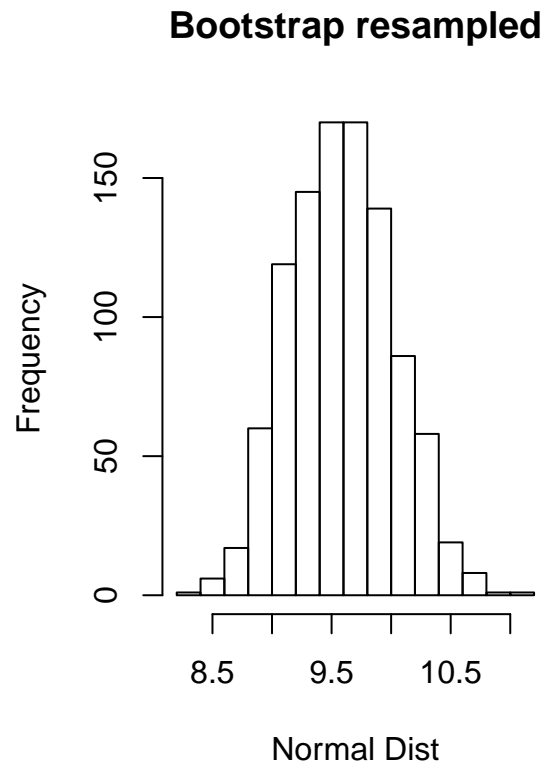
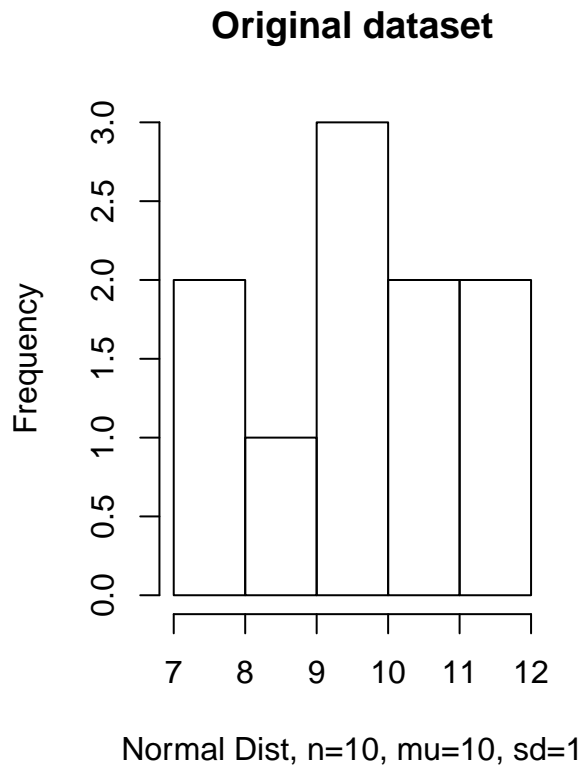
```
## [1] "Original dataset"

##   vars  n mean  sd median trimmed  mad min   max range skew kurtosis
## 1    1 30 10.02 0.82  10.04    9.98 0.77 8.16 12.04  3.88 0.29    0.06
##   se
## 1 0.15

## [1] "Resampled dataset"

##   vars    n mean  sd median trimmed  mad min   max range skew kurtosis
## 1    1 1000 10.01 0.14  10.01   10.01 0.13 9.64 10.47  0.83 0.12   -0.02
##   se
## 1    0

## [1] 9.571457
```



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

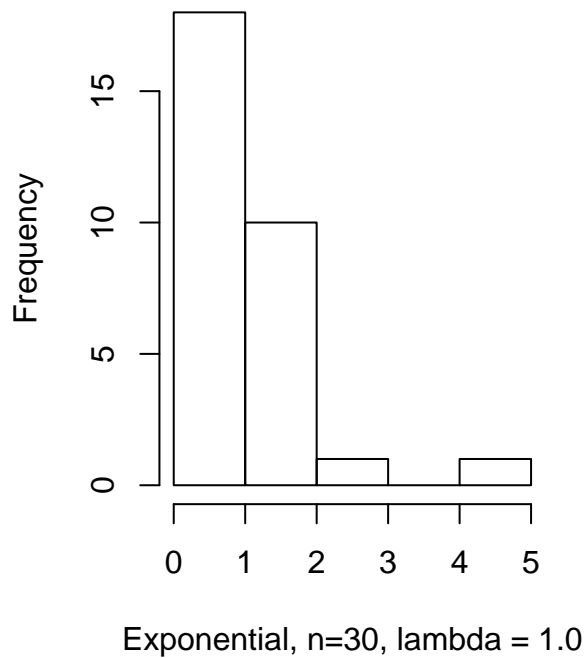
##   vars  n mean  sd median trimmed  mad min   max range skew kurtosis
## 1    1  10 9.57 1.44   9.2   9.54 1.92 7.64 11.73  4.09 0.18    -1.6
##    se
## 1 0.46

## [1] "Resampled dataset"

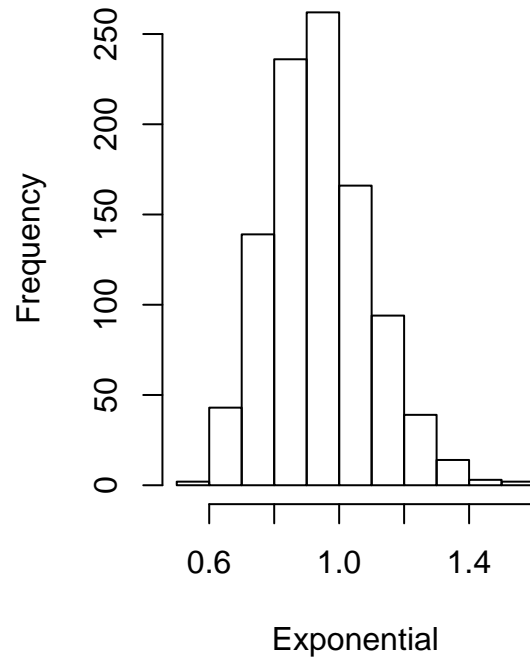
##   vars    n mean  sd median trimmed  mad min   max range skew kurtosis
## 1    1 1000 9.58 0.43   9.58   9.58 0.47 8.32 11.05  2.73 0.11    -0.3
##    se
## 1 0.01

## [1] 0.9492649
```

Original dataset



Bootstrap resampled



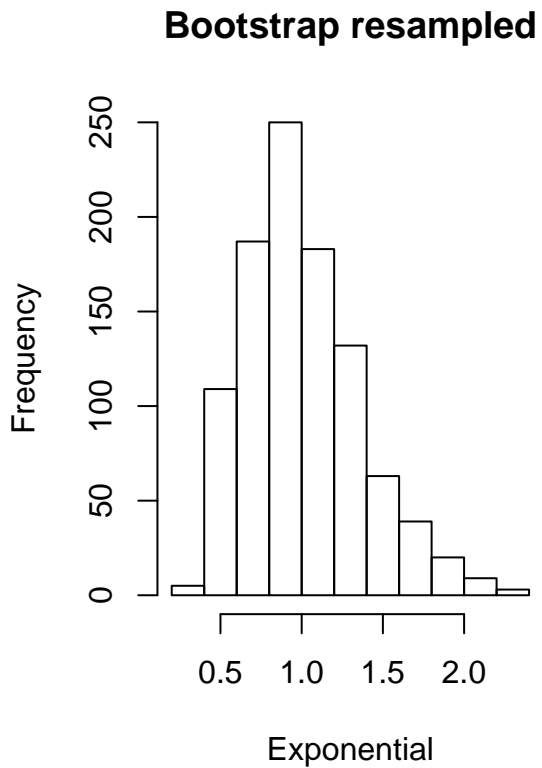
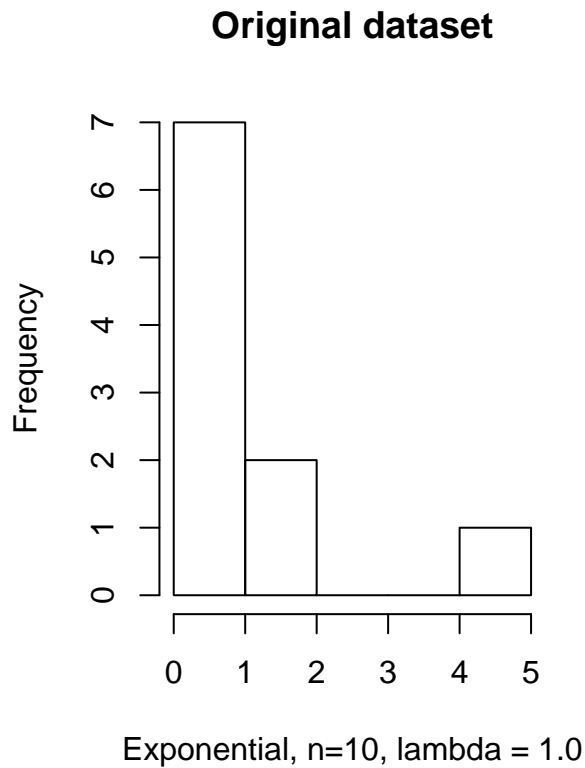
```
## [1] "Original dataset"

##   vars  n mean  sd median trimmed  mad min  max range skew kurtosis  se
## 1    1  30 0.95 0.9   0.59   0.82 0.61 0.03 4.58  4.55 2.19    6.19 0.16

## [1] "Resampled dataset"

##   vars    n mean   sd median trimmed  mad min  max range skew kurtosis se
## 1    1 1000 0.94 0.15   0.93   0.93 0.15 0.5 1.52  1.02 0.46    0.22 0

## [1] 1.007956
```



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

##   vars  n mean   sd median trimmed  mad min  max range skew kurtosis   se
## 1    1  10 1.01 1.16  0.66   0.73 0.47 0.13 4.11  3.98 1.83    2.26 0.37

## [1] "Resampled dataset"

##   vars    n mean   sd median trimmed  mad min  max range skew kurtosis
## 1    1 1000 1.01 0.35  0.97   0.98 0.36 0.35 2.38  2.04 0.74    0.37
##      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.