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] 10.477509  8.975158 11.560227  8.905613  9.067715  9.289651 11.696547
##  [8] 10.099063  9.477168 10.367362
```

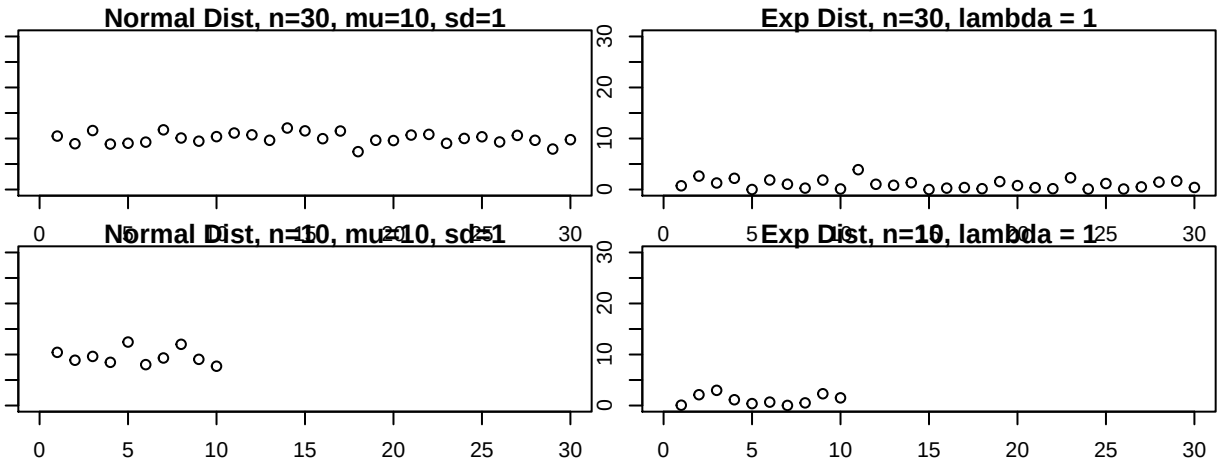
```
##  [1] 10.420229  8.873135  9.620543  8.468136 12.442250  8.014530  9.315546
##  [8] 11.999559  9.053778  7.702295
```

```
##  [1] 0.73143820 2.63086324 1.28065547 2.19954540 0.00479028 1.87168999
##  [7] 1.04191649 0.27118047 1.85264248 0.10722223
```

```
##  [1] 0.06753096 2.11728439 2.97678428 1.13116010 0.37469663 0.66730997
##  [7] 0.02094424 0.51798450 2.30835659 1.50322578
```

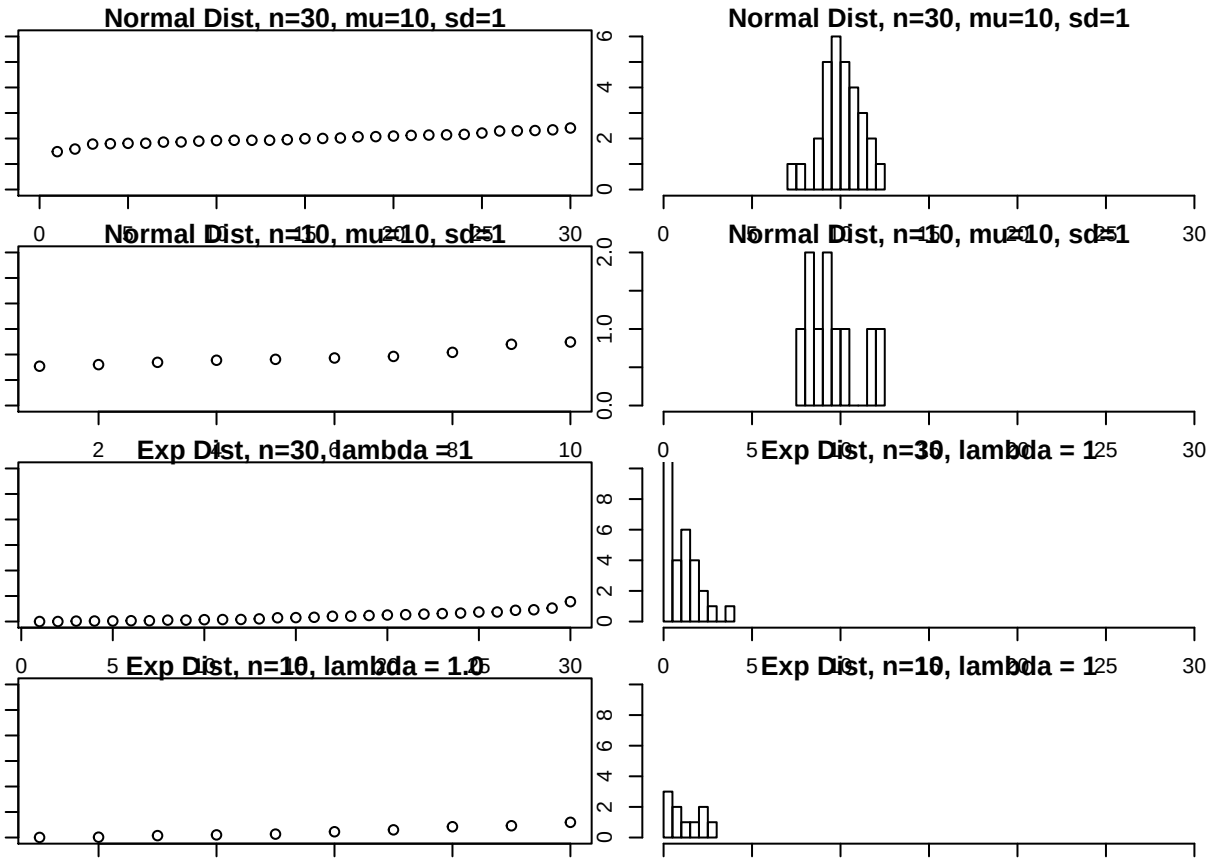
Explore the data with Plot of the normal and exponential sample datasets

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



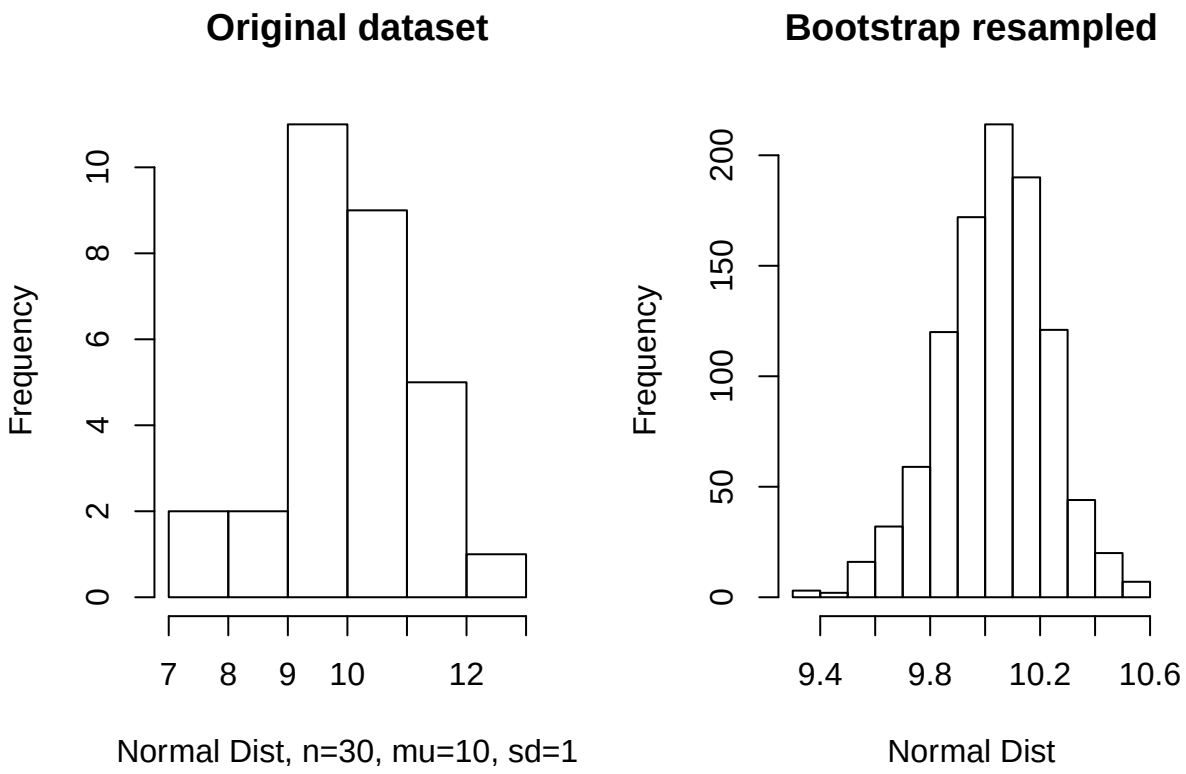
Sort the records ascending and plot again. Include histograms

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] 10.03915
```



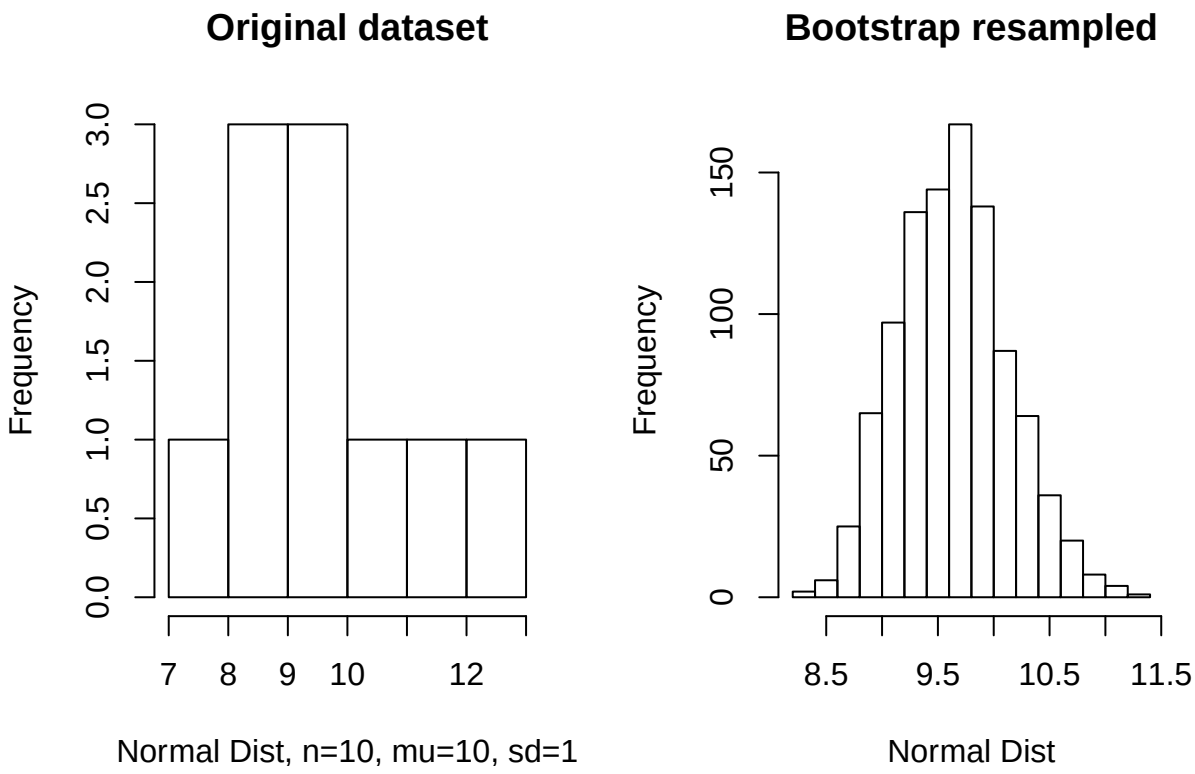

```
## [1] "Original dataset"
```

```
##   vars  n  mean   sd median trimmed  mad  min   max range  skew kurtosis
## 1    1 30 10.04 1.09   9.99   10.07 1.03 7.42 12.05  4.63 -0.22    -0.32
##    se
## 1 0.2
```

```
## [1] "Resampled dataset"
```

```
##   vars    n  mean   sd median trimmed mad min   max range  skew kurtosis
## 1    1 1000 10.04 0.19  10.05   10.04 0.2 9.3 10.56  1.25 -0.28     0.21
##     se
## 1 0.01
```

```
## [1] 9.591
```



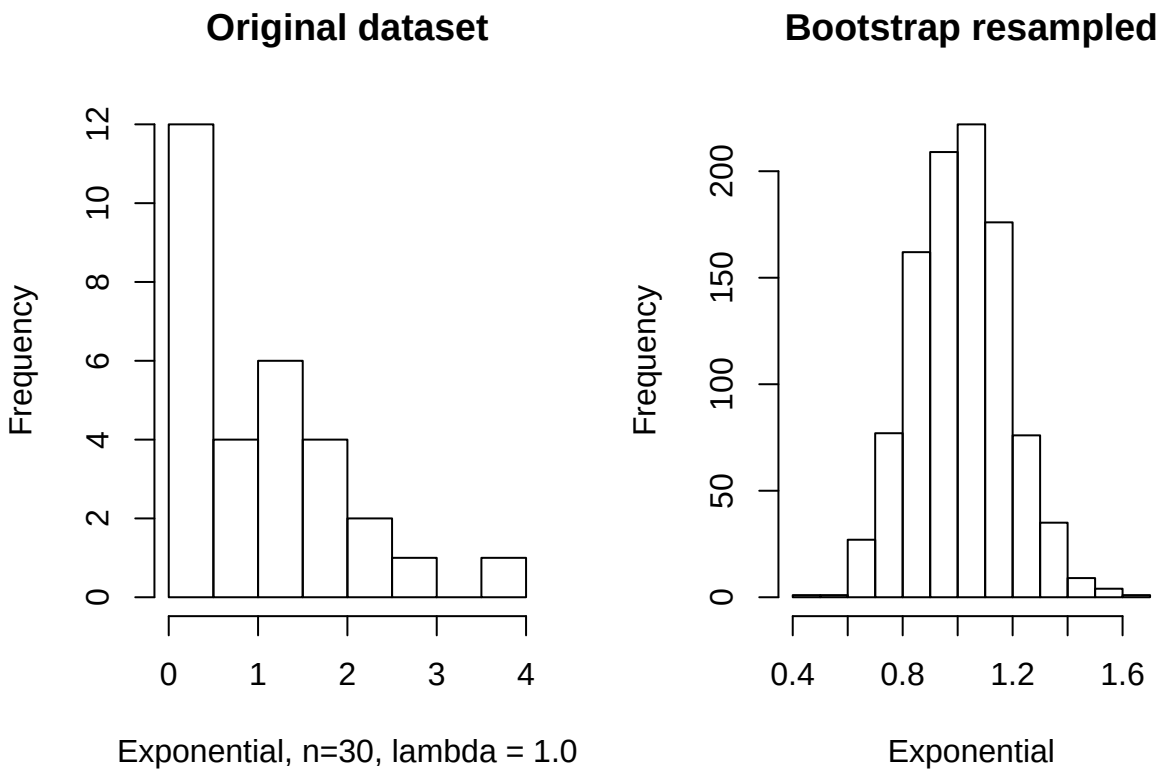

```
## [1] "Original dataset"
```

```
##   vars  n mean   sd median trimmed mad min   max range skew kurtosis  se
## 1    1 10 9.59 1.59   9.18    9.47 1.4 7.7 12.44  4.74 0.62    -1.12 0.5
```

```
## [1] "Resampled dataset"
```

```
##   vars    n mean   sd median trimmed  mad  min   max range skew kurtosis
## 1    1 1000 9.64 0.49   9.63    9.63 0.48 8.33 11.27  2.94 0.25     -0.1
##     se
## 1 0.02
```

```
## [1] 1.014516
```



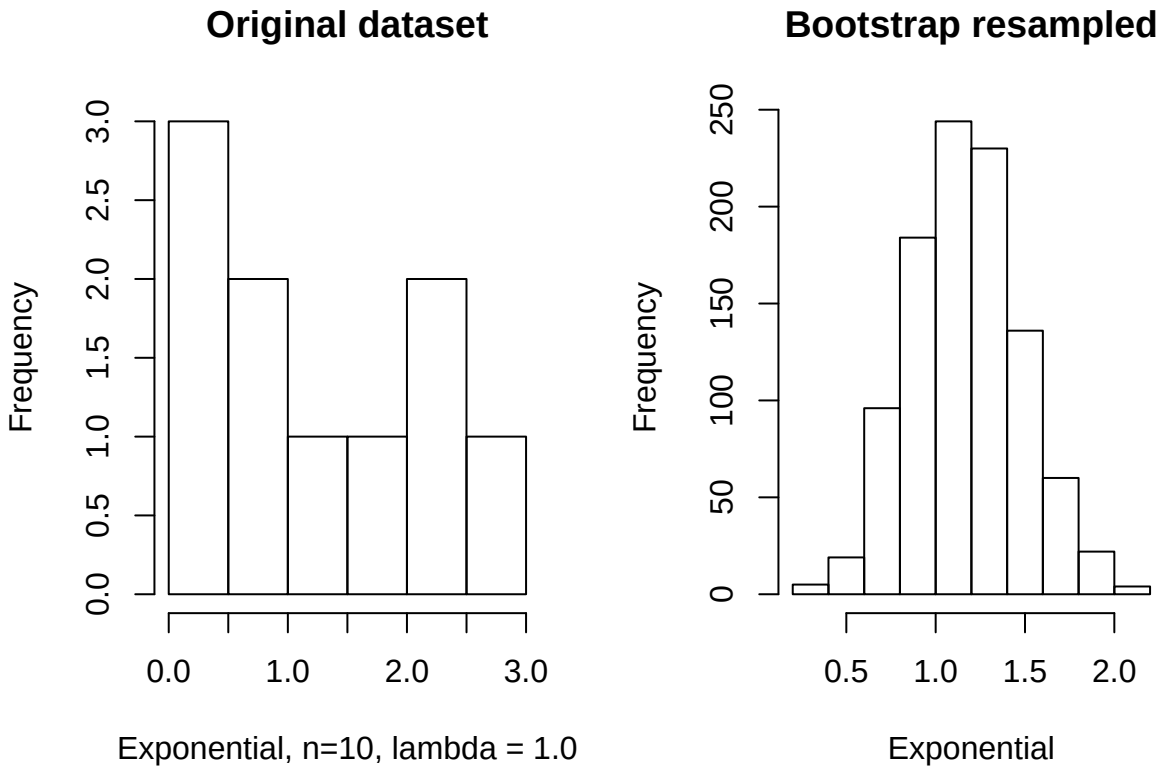

```
## [1] "Original dataset"
```

```
##   vars  n mean   sd median trimmed  mad min  max range skew kurtosis   se
## 1    1 30 1.01 0.93    0.8     0.9 0.95   0 3.89  3.89 1.08     0.87 0.17
```

```
## [1] "Resampled dataset"
```

```
##   vars    n mean   sd median trimmed  mad  min  max range skew kurtosis
## 1    1 1000 1.01 0.17   1.01    1.01 0.17 0.48 1.68   1.2 0.18     0.14
##     se
## 1 0.01
```

```
## [1] 1.168528
```




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

```
##   vars  n mean   sd median trimmed  mad  min  max range skew kurtosis   se
## 1    1 10 1.17 1.02    0.9    1.09 1.06 0.02 2.98  2.96 0.42    -1.44 0.32
```

```
## [1] "Resampled dataset"
```

```
##   vars    n mean   sd median trimmed  mad  min  max range skew kurtosis
## 1    1 1000 1.17 0.31   1.16    1.16 0.32 0.26 2.17   1.9 0.14    -0.14
##     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.