StatInfProject

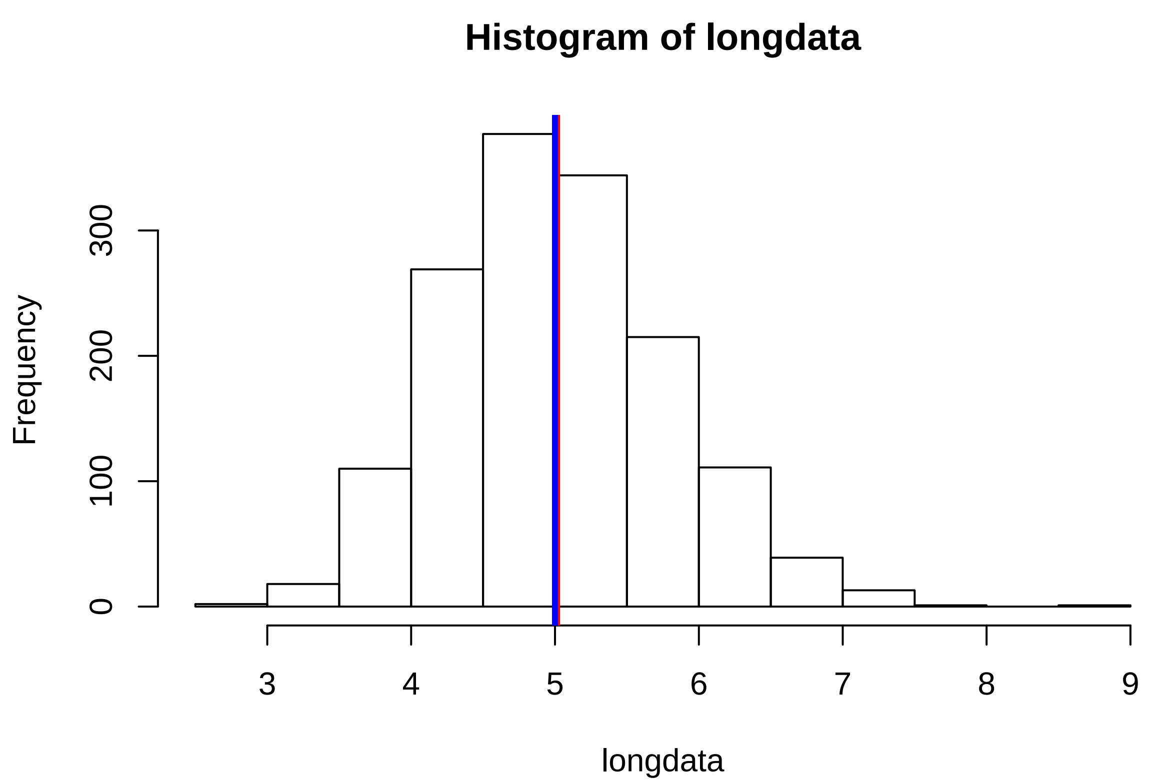
Jeff Wolbach

Sunday, August 17, 2014

The exponential distribution was investigated by simulating 1500, 40-member samples from an exponential distribution with a mean of 5 and a variance of 5. The mean of each 40-member sample was calculated and the behavior of the mean across all 1500 samples was compared to a normal distribution as a proof of the central limit theorem.

```
longdata=numeric(length=1500)
for (i in 1:1500) { longdata[i]=mean(rexp(40,0.2))}
mld<-mean(longdata)
sdld<-sd(longdata)
```

The distribution of the means of each sample is described in the following histogram.



The average of the calculated sample means (5.0143) is represented in the histogram as a vertical red line, and the true population mean (5.000) is represented in the histogram as a vertical blue line.

The calculated variance of the sample mean was 0.7739. This can be compared to a theoretical variance of (1/lambda)/sqrt(number of samples=40) = 5/sqrt(40)= 0.791.

As proof of the normality of the distribution of the sample means, we can calculate the fraction of the sample means that differ from the average of the sample means by specific fractions of the variance of the sample means.

```
norm_comp=matrix(nrow=5,ncol=2)
for (j in 1:5) {
  norm_comp[j,1]<- round(mean(longdata < (mld - (j*0.5)*sdld)),3)
  norm_comp[j,2]<- round(mean((mld + (j*0.5)*sdld) < longdata),3)
}
```

Where the first column of norm_comp represents the fraction of values below the average of the sample means (to the left), and the second column represents the fraction of values above the average of the sample means (to the right). For a normal distribution the two elements of each row should be equal, or the distribution should be symmetric about the center.

```
norm_comp
```

```
##           Left of Mean Right of Mean
## 0.5 sigma        0.329         0.289
## 1.0 sigma        0.152         0.164
## 1.5 sigma        0.056         0.073
## 2.0 sigma        0.011         0.032
## 2.5 sigma        0.003         0.011
```

The coverage of the confidence interval for (1/lambda) is calculated

```
ul<-mean(longdata)+1.96*sd(longdata)
ll<-mean(longdata)-1.96*sd(longdata)
cover<-mean(ll<longdata & ul>longdata)
cover
```

```
## [1] 0.9527
```

This result is very close to the theoretical coverage of 95% (or, 0.95), assuming a normal distribution.