Simulation Exercise

We want to look at samples of 40 variables drawn from the exponential distribution. Let's go ahead and do that. The following code will draw random variables from an exponential distribution with $\lambda = 0.2$, 40000 times, for 10000 samples of 40 variables:

```r
num_exps = 40;
ntrials = 10000;
set.seed(4023); exps = matrix(rexp(num_exps * ntrials, 0.2), nrow = num_exps, ncol = ntrials);
exp_means = colMeans(exps);
```

Both the mean and standard deviation of an exponential are $1/\lambda$, so 5 in this case. We expect the sample mean to also be equal to 5, and the sample variance to be $5^2/40 = 0.7906$. Here is what we measure for the sample mean and variance for our simulation, along with the percent difference from the expected values:

```r
exp_mean = mean(exp_means)
print(paste("Mean:", exp_mean)); print(paste("Percent Difference:",(exp_mean - 5)/5 * 100))
```

```
## [1] "Mean: 5.00017624385373"
```

```
## [1] "Percent Difference: 0.00352487707466764"
```

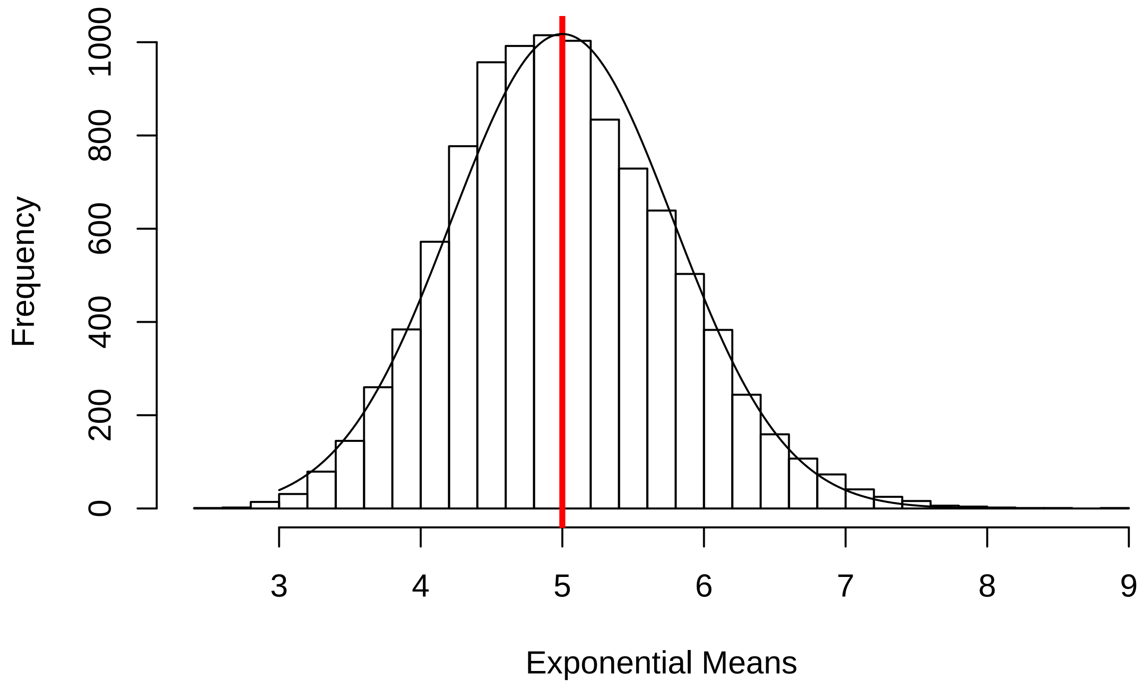
```r
exp_var = var(exp_means)
print(paste("Variance:", exp_var)); print(paste("Percent Difference:",(exp_var - 25/40)/(25/40) * 100))
```

```
## [1] "Variance: 0.614821034868332"
```

```
## [1] "Percent Difference: -1.62863442106692"
```

The mean is less than 0.01% difference than the theoretical value, while the variance is less than 2% different. So, they're pretty close. Now, we want to see how close the distribution of these means are to a normal distribution. So, let's plot them up, along with a normal distribution with the same mean and variance:

```r
hist(exp_means, 30, xlab  = "Exponential Means", main = '')
abline(v = 5, col = 'red', lwd=3)
xdummy = seq(3,8,0.01)
lines(xdummy,exp(-0.5 * (xdummy - 5)^2/exp_var)/sqrt(2 * pi * exp_var) * 10000 * 0.2)
```



Just looking at the plot, the distributions look very similar. The distribution of exponential means does seem skewed slightly towards the left. We want to determine if our distribution is significantly different form normal, though. A good way to do that is with the Kolmogorov-Smirnov Test (K-S Test). The K-S Test compares one distribution to another one, or one distribution to a known cumulative distribution (e.g., normal, Poisson). The null hypothesis is that the first distribution is drawn from the second one. In our case, we have the null hypothesis that our distribution of exponential means is drawn from a normal distribution. Let's do a K-S Test using our distribution and *pnorm*, which is the cumulative normal distribution.

```
ks.test(exp_means, "pnorm", mean = exp_mean, sd = sqrt(exp_var))
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  exp_means
## D = 0.025135, p-value = 6.51e-06
## alternative hypothesis: two-sided
```

With a p-value of less than 1e-05, we can reject the null hypothesis at the 99.999% level. In conclusion, our simulation suggests the distribution of exponential means is significantly different from a normal distribution.