# Project report - Statistical Inference

*Kirty Vedula*

*04/24/2015*

## Libraries

```
library(datasets)
library(ggplot2)
library(knitr)
library(markdown)
```

## Simulating exponential distribution with rexp commmand

```
set.seed(3)
lambda <- 0.2
num_sim <- 1000
sample_size <- 40
sim <- matrix(rexp(num_sim*sample_size, rate=lambda), num_sim, sample_size)
row_means <- rowMeans(sim)
```

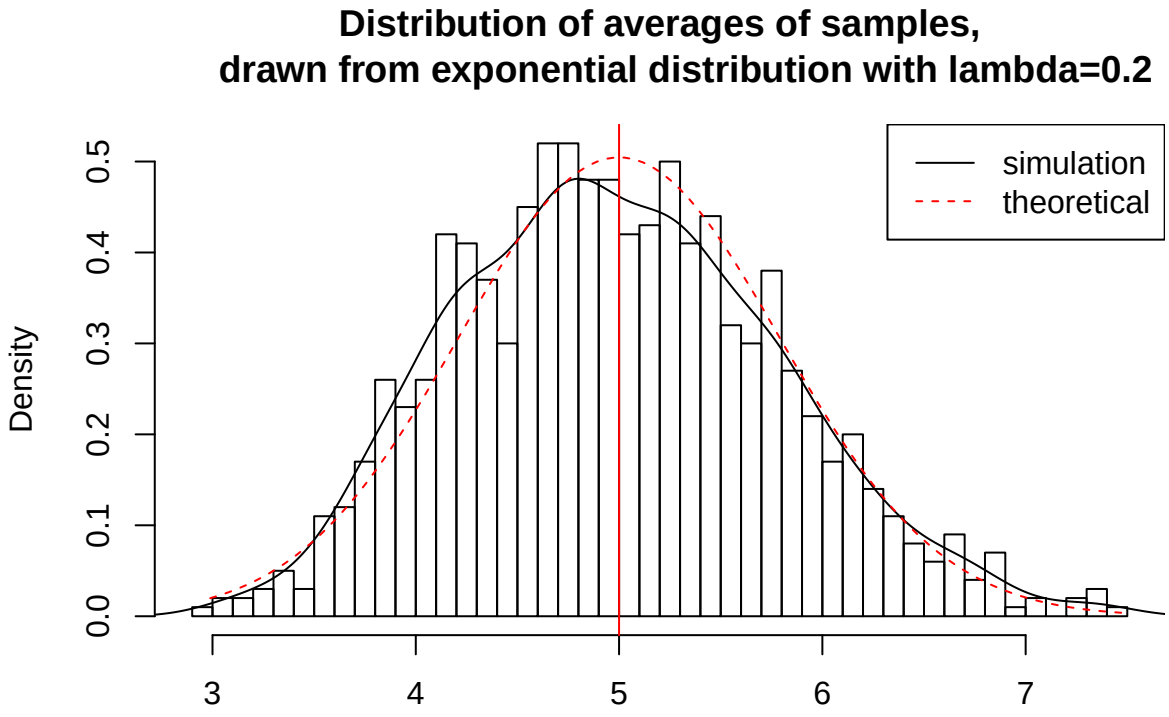
---

## Plot the histogram of averages

### Density of the averages of samples

### Theoretical center of distribution

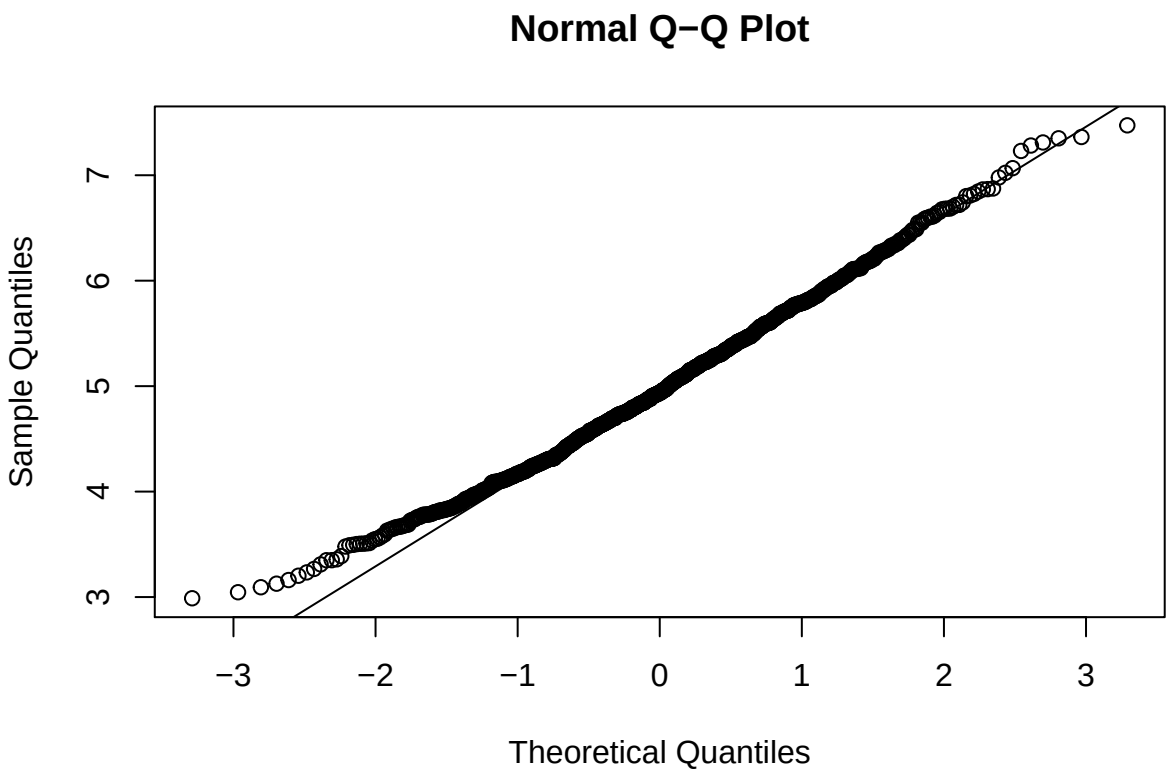
```
hist(row_means, breaks=50, prob=TRUE,
     main="Distribution of averages of samples,
     drawn from exponential distribution with lambda=0.2",
     xlab="")
lines(density(row_means))
abline(v=1/lambda, col="red")
xfit <- seq(min(row_means), max(row_means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sample_size)))
lines(xfit, yfit, pch=22, col="red", lty=2)
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("black", "red"))
```




---

**Quantile plot for the row means calculated**

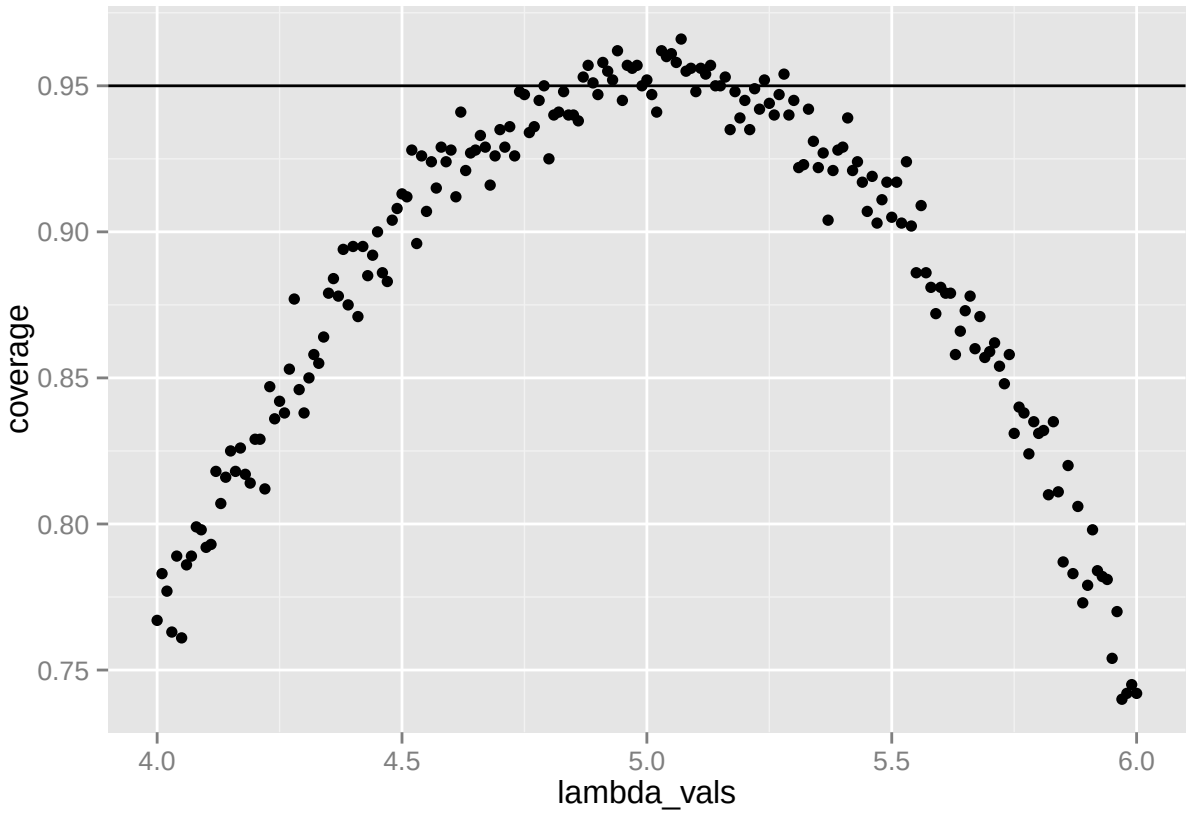
```
qqnorm(row_means)
qqline(row_means)
```




---

```
lambda_vals <- seq(4, 6, by=0.01)
coverage <- sapply(lambda_vals, function(lamb) {
    mu_hats <- rowMeans(matrix(rexp(sample_size*num_sim, rate=0.2),
                               num_sim, sample_size))
    ll <- mu_hats - qnorm(0.975) * sqrt(1/lambda**2/sample_size)
    ul <- mu_hats + qnorm(0.975) * sqrt(1/lambda**2/sample_size)
    mean(ll < lamb & ul > lamb)
})
```

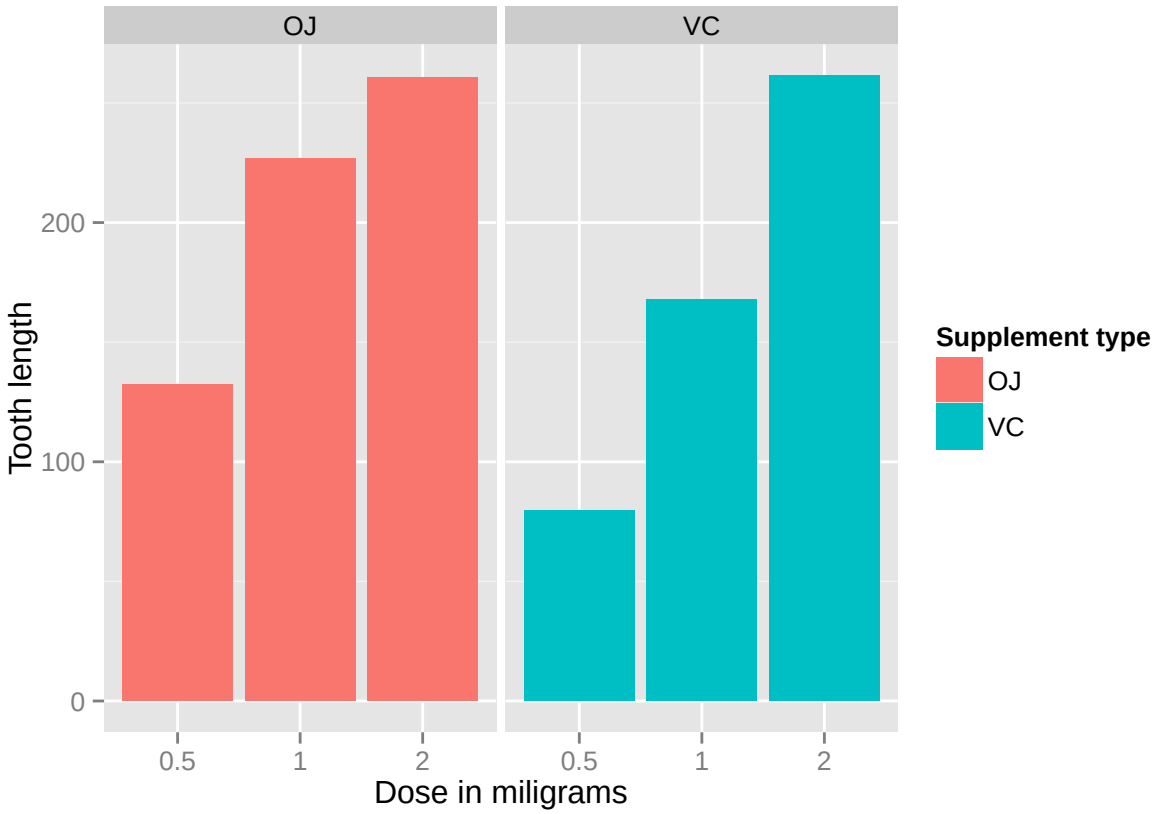
```
qplot(lambda_vals, coverage) + geom_hline(yintercept=0.95)
```




## Part 2

This part performs an exploratory data analysis of at least a single plot or table highlighting basic features of the data.

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
  geom_bar(stat="identity",) +
  facet_grid(. ~ supp) +
  xlab("Dose in miligrams") +
  ylab("Tooth length") +
  guides(fill=guide_legend(title="Supplement type"))
```




### Fitting a model using dose and supplements as features

```
fit <- lm(len ~ dose + supp, data=ToothGrowth)
summary(fit)
```

```
##
## Call:
## lm(formula = len ~ dose + supp, data = ToothGrowth)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.600 -3.700  0.373  2.116  8.800
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.2725     1.2824   7.231 1.31e-09 ***
## dose          9.7636     0.8768  11.135 6.31e-16 ***
## suppVC       -3.7000     1.0936  -3.383   0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.236 on 57 degrees of freedom
## Multiple R-squared:  0.7038, Adjusted R-squared:  0.6934
## F-statistic: 67.72 on 2 and 57 DF,  p-value: 8.716e-16
```

---

### Confidence interval for the fit

```
confint(fit)
```

```
##                 2.5 %    97.5 %
## (Intercept)  6.704608 11.840392
## dose         8.007741 11.519402
## suppVC      -5.889905 -1.510095
```