Discussion 6 Review

1. Estimation of a Population Proportion

   - If a sample can be considered a collection of iid RVs Y_i where the outcome of each is either zero or one, then we define the sample proportion:

     $$\text{Sample proportion: } \hat{\pi} = P = \frac{\sum_{i=1}^{n} Y_i}{n}.$$

   - $E(P) = \pi$, $VAR(P) = \frac{\pi(1-\pi)}{n}$, $SE(P) = \sqrt{\frac{\pi(1-\pi)}{n}}$.

   - So long as $n\pi > 5$ and $n(1-\pi) > 5$, the approximate distribution of P is:

     $$P \dot{\sim} N(\pi, \frac{\pi(1-\pi)}{n}).$$

   - So long as $n\pi > 5$ and $n(1-\pi) > 5$, an approximate $100(1-\alpha)\%$ CI for π would be of the form:

     $$P \pm z_{\alpha/2}\sqrt{\frac{P(1-P)}{n}}.$$

2.
```
#We can put multiple graphs in a single plot with the help of par() function
#the option mfrow=c(nrows, ncols) to create a matrix of nrows x ncols plots
#that are filled in by row
par(mfrow=c(2,3))

for (i in 1:6){
#set.seed function in R is used to reproduce results
#i.e. it produces the same sample again and again.
#When we generate randoms numbers without set.seed() function
#it will produce different samples at different time
set.seed(10*i)

data <- rnorm(10)
qqnorm(data, main = paste0("QQ-plot",i))
qqline(data)
}
par(mfrow=c(1,1))
```