

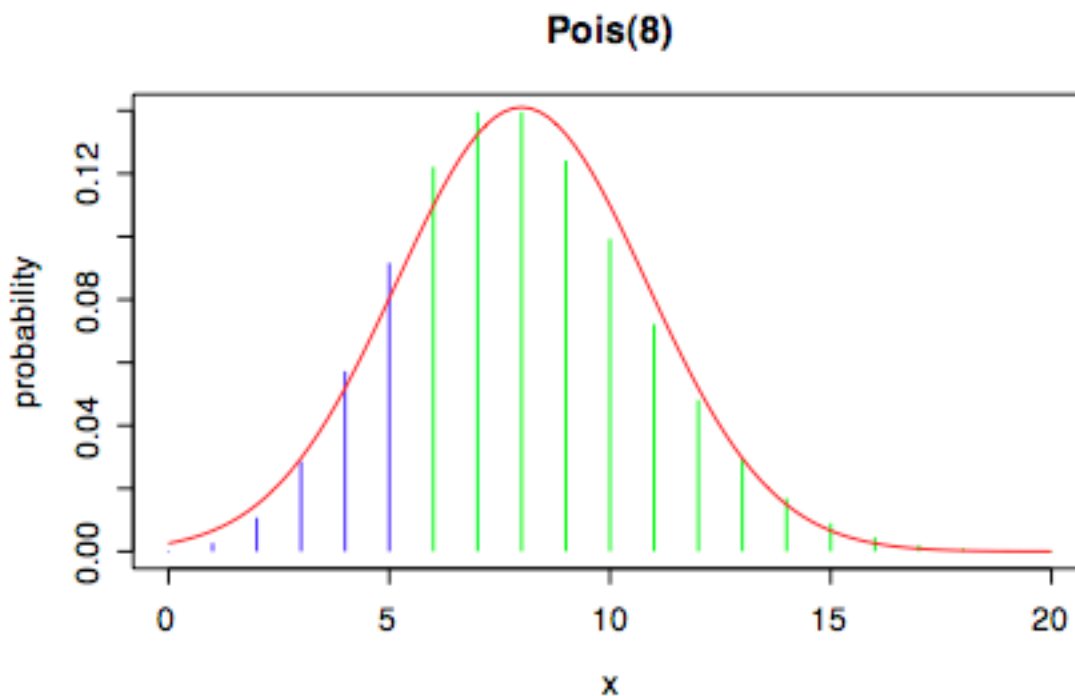
STATS 3N03/3J04
ASSIGNMENT #2 - SOLUTIONS

Part A

Question 1 [10]

The probability of getting 6 or more from a Poisson distribution with mean of 8 is 0.80876 by the exact Poisson calculation, 0.8116 by the normal approximation to the Poisson with continuity correction and 0.8556 by the normal approximation without continuity correction.

```
> plot(0:20,dpois(0:20,8),type="n", xlab="x", ylab="probability",
main="Pois(8)")
> lines(6:20,dpois(6:20,8),type="h", col="green")
> lines(0:5,dpois(0:5,8),type="h", col="blue")
> xgr<-seq(0,20,len=101)
> lines(xgr, dnorm(xgr,8,sqrt(8)), col="red")
> 1-ppois(5, 8)
[1] 0.808764
> 1-pnorm(5.5, 8, sqrt(8))
[1] 0.8116204
> 1-pnorm(5, 8, sqrt(8))
[1] 0.8555778
```



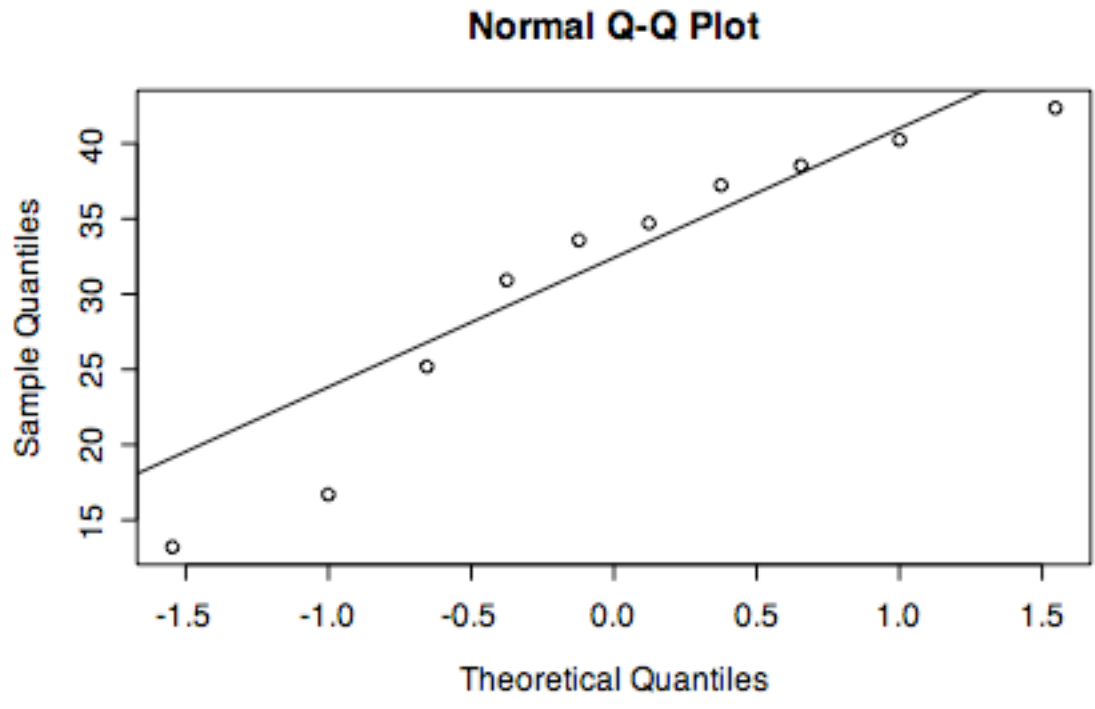
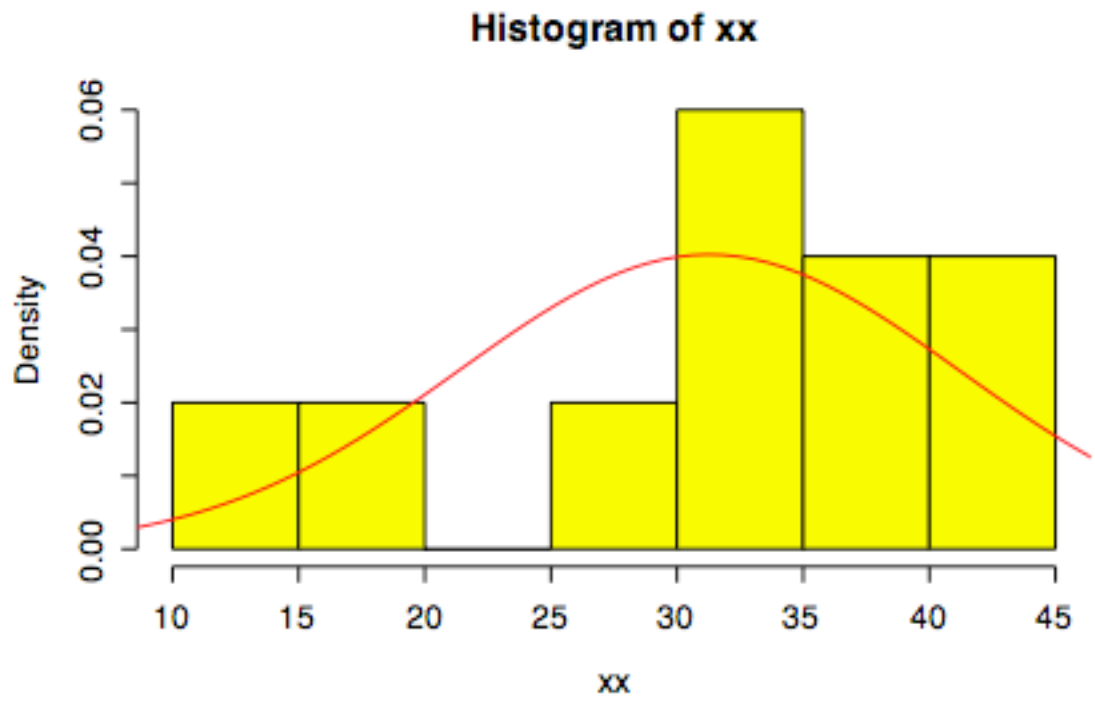
Question 2 [10]

With 100 observations, the histogram is roughly normal and the QQ Normal plot is quite straight; with 1000 observations the QQ plot is very straight and the histogram is quite smooth. This suggests that at least 100 observations are needed to establish normality. The QQ Normal plot appears to be more useful than the histogram for this purpose.

N(30, 10²) data, n = 10

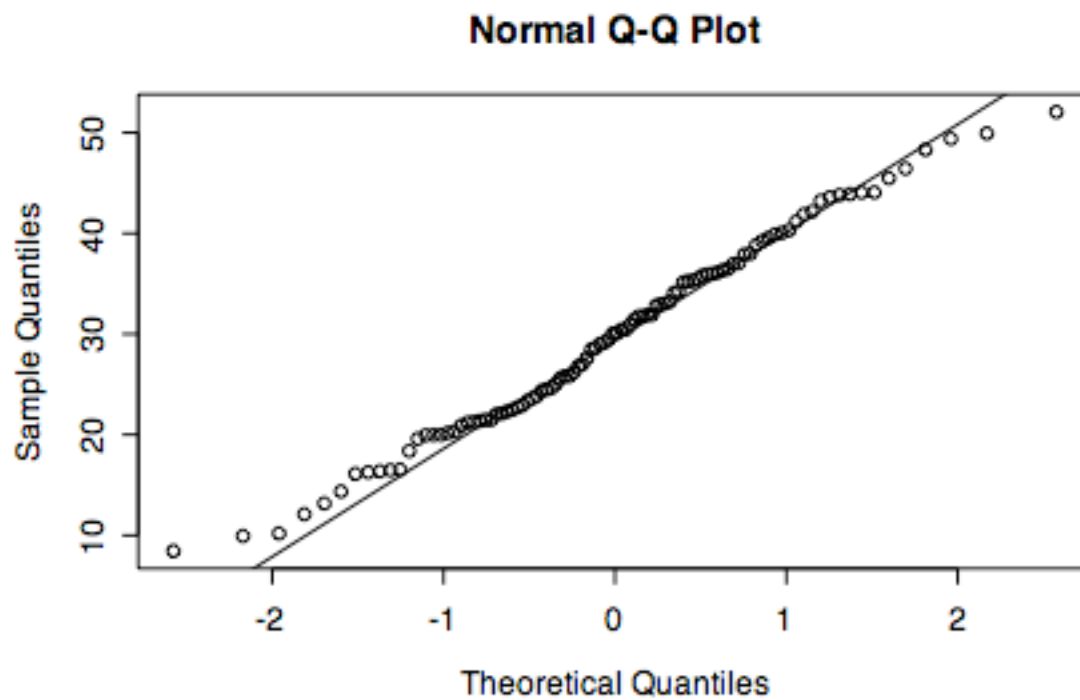
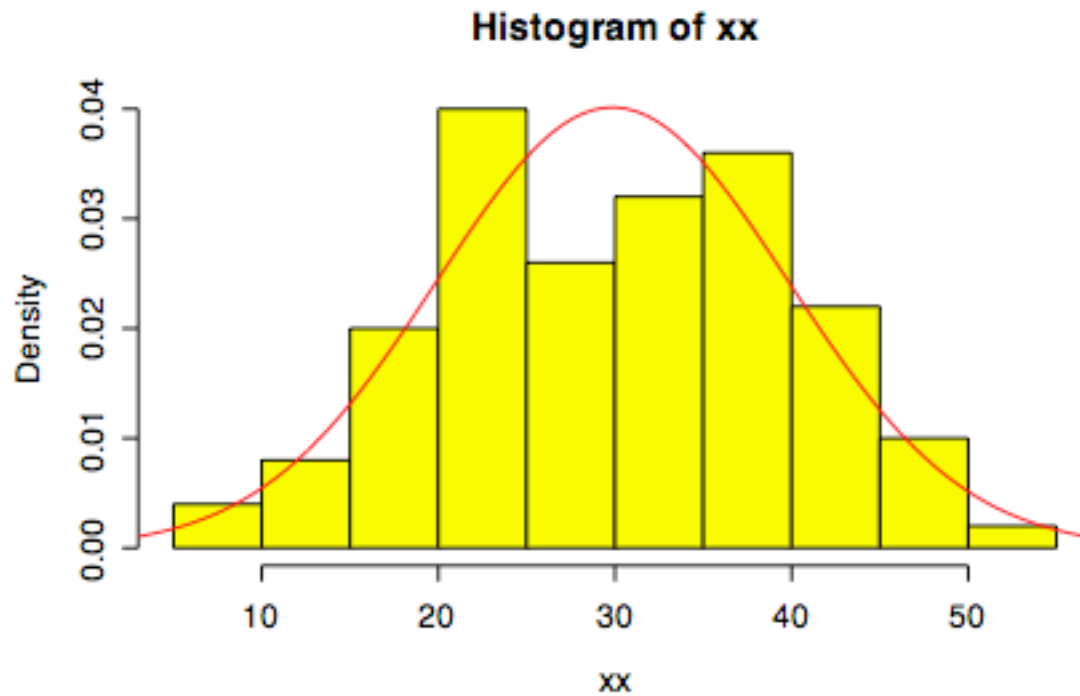
```
> xx <- rnorm(10, 30, 10)
> hist(xx, prob=T, col="yellow")
> xgr <- seq(0,60,len=101)
> lines(xgr, dnorm(xgr,mean(xx),sd(xx)), col="red")
```

```
> qqnorm(xx)  
> qqline(xx)
```



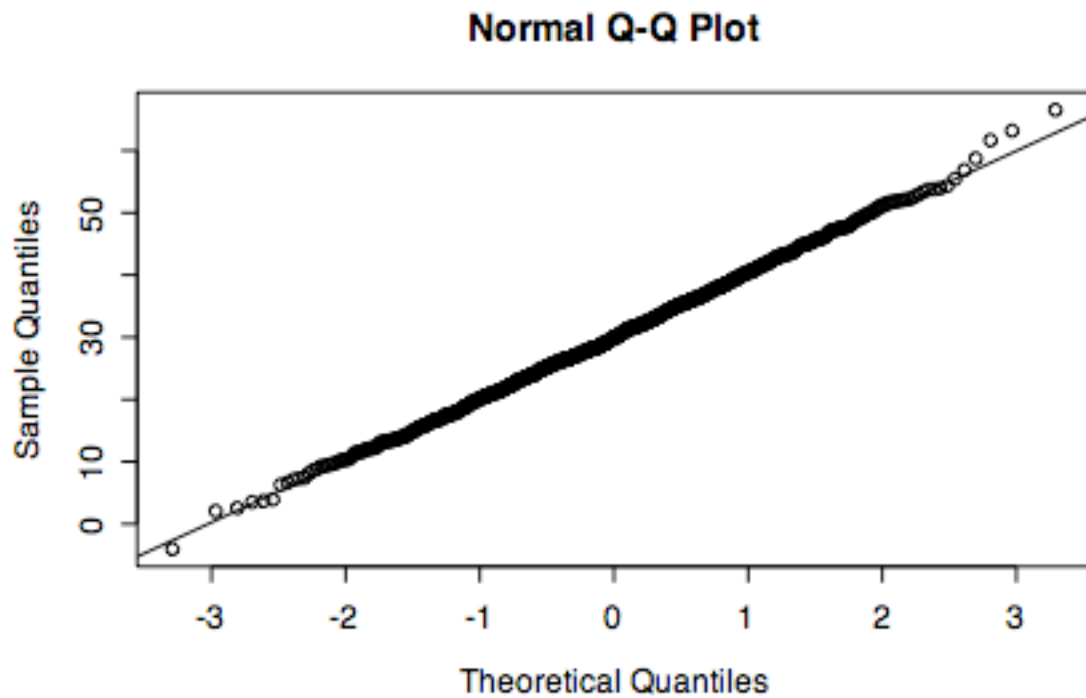
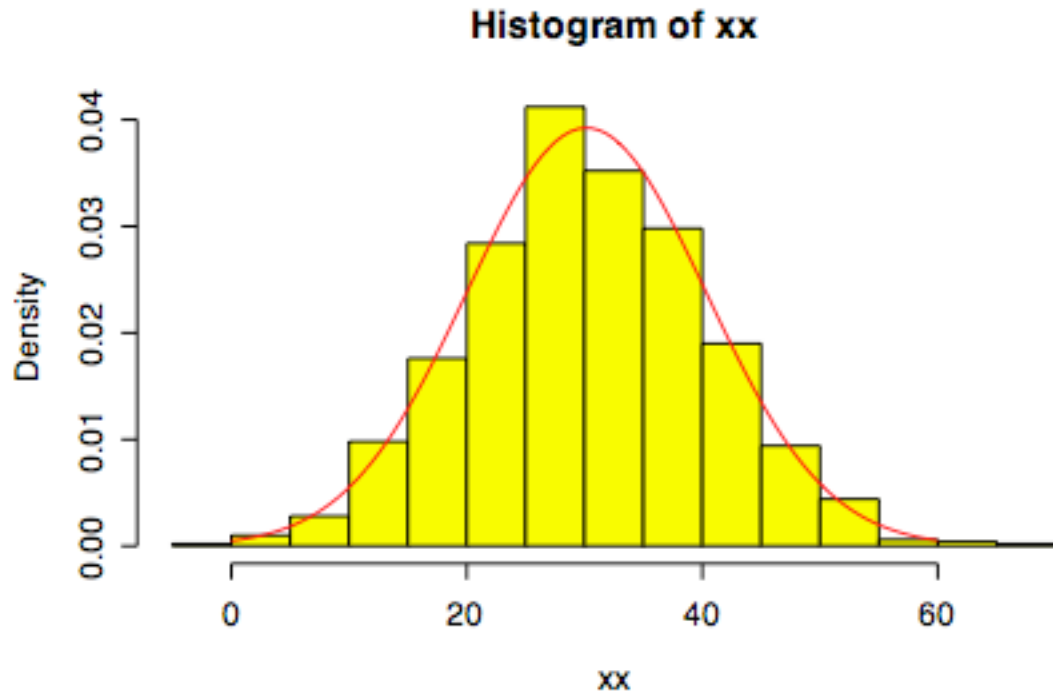
N(30, 10²) data, n = 100

```
> xx <- rnorm(100, 30, 10)  
> hist(xx, prob=T, col="yellow")  
> lines(xgr, dnorm(xgr, mean(xx), sd(xx)), col="red")  
> qqnorm(xx)  
> qqline(xx)
```



$N(30, 10^2)$ data, $n = 1000$

```
> xx <- rnorm(1000, 30, 10)
> hist(xx, prob=T, col="yellow")
> lines(xgr, dnorm(xgr, mean(xx), sd(xx)), col="red")
> qqnorm(xx)
> qqline(xx)
```

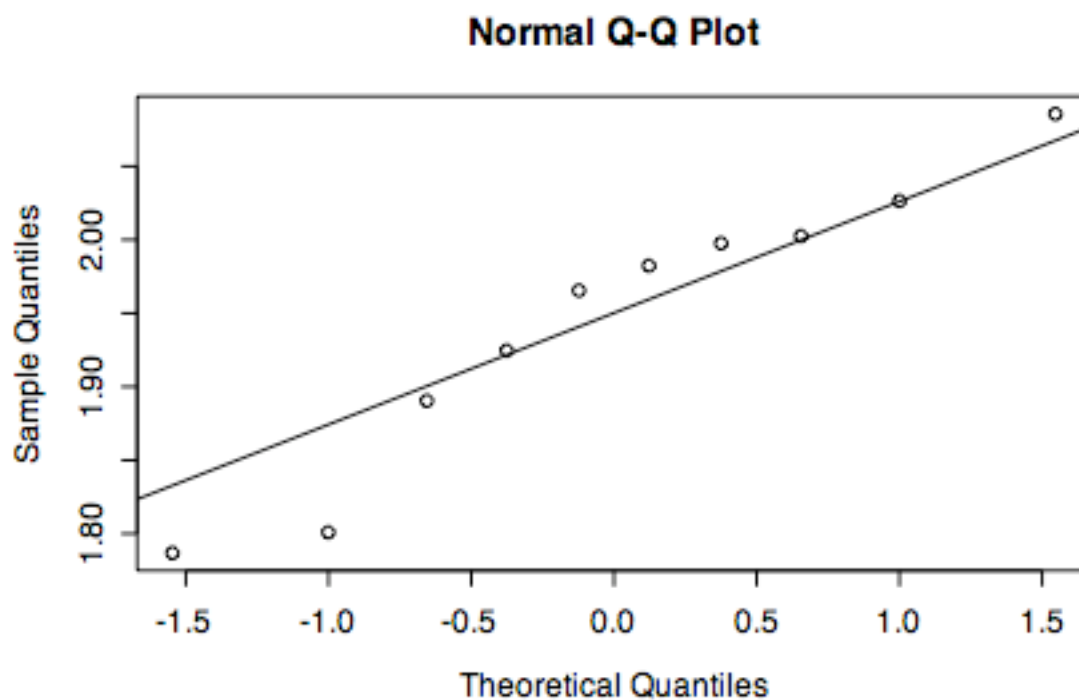
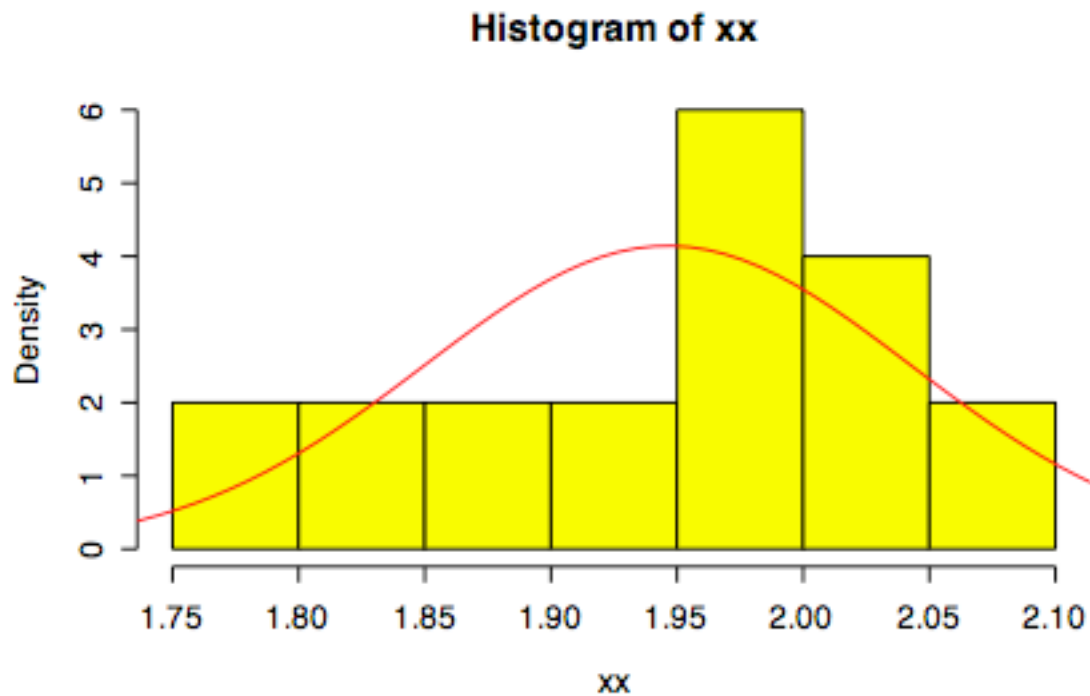


Question 3 [10]

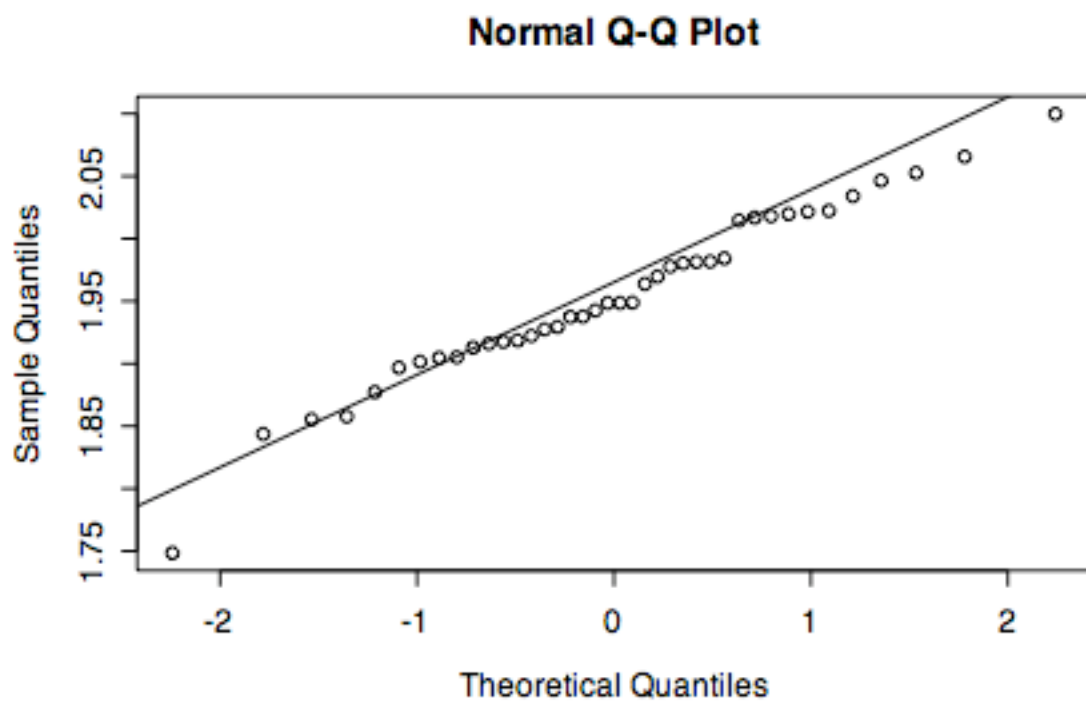
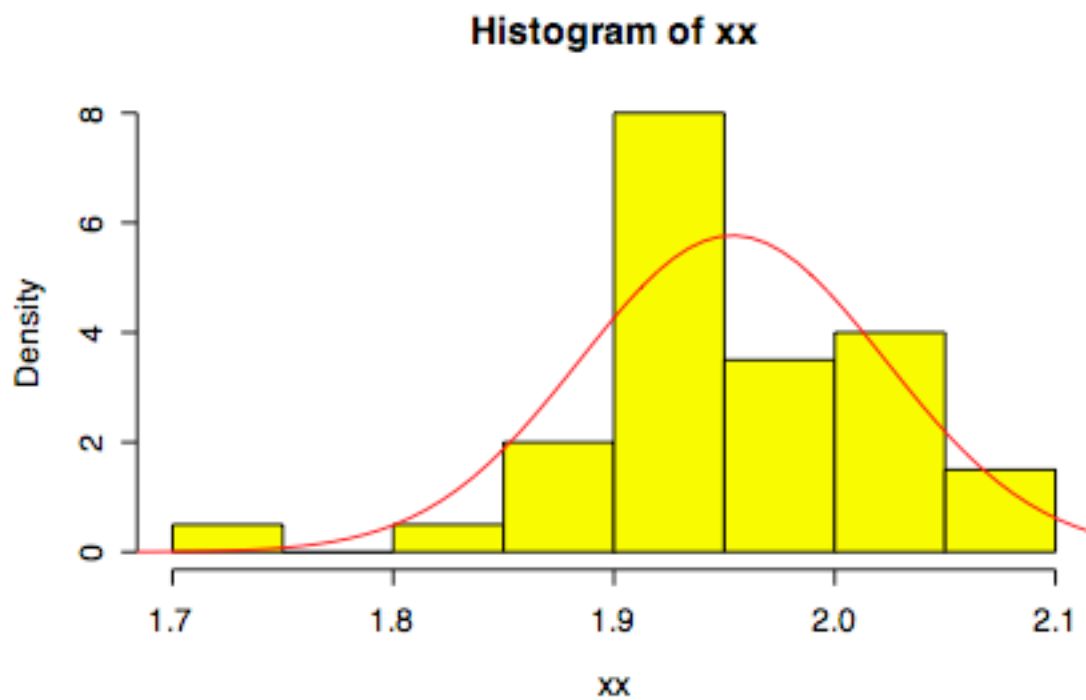
With 40 observations the QQ Normal plot looks reasonably straight, while the histogram does not reveal the shape of the density very well. With 100 or 1000 observations the negative skewness in the histogram is more apparent and the curvature in the QQ Normal line is well defined. We can conclude that at least 100 observations are needed to distinguish this Weibull distribution from a normal distribution. The QQ Normal plot appears to be more useful than the histogram for this purpose.

Weibull(30, 2) data, n = 10

```
> xx <- rweibull(10, 30, 2)
> hist(xx, prob=T, col="yellow")
> xgr <- seq(1.5, 2.2, len=101)
> lines(xgr, dnorm(xgr, mean(xx), sd(xx)), col="red")
> qqnorm(xx)
> qqline(xx)
```

**Weibull(30, 2) data, n = 40**

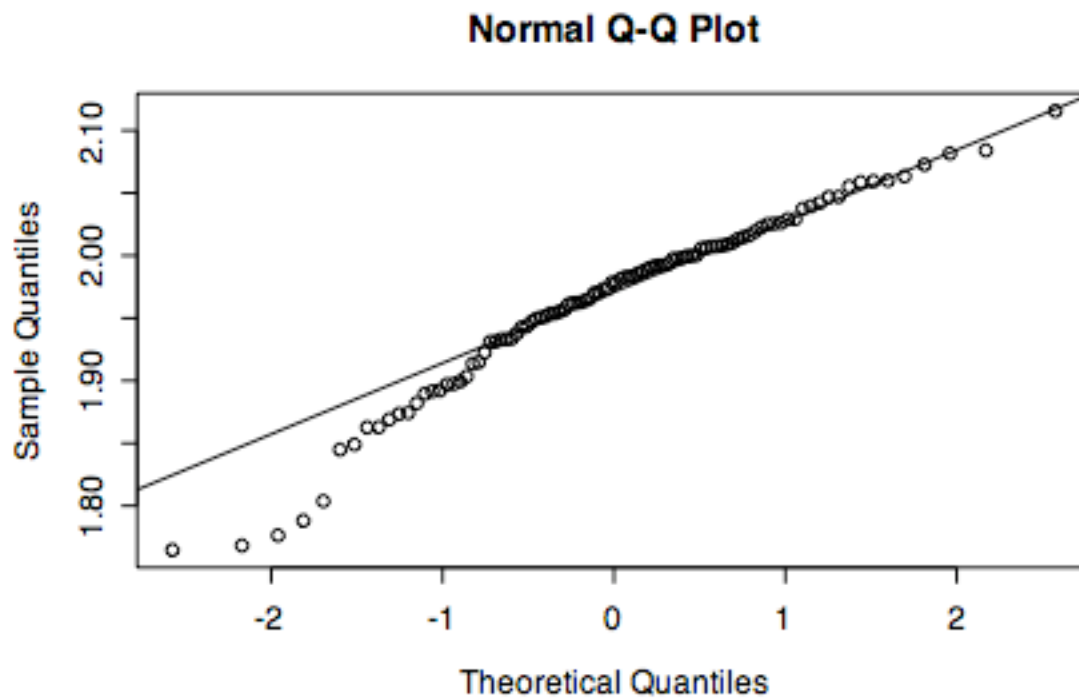
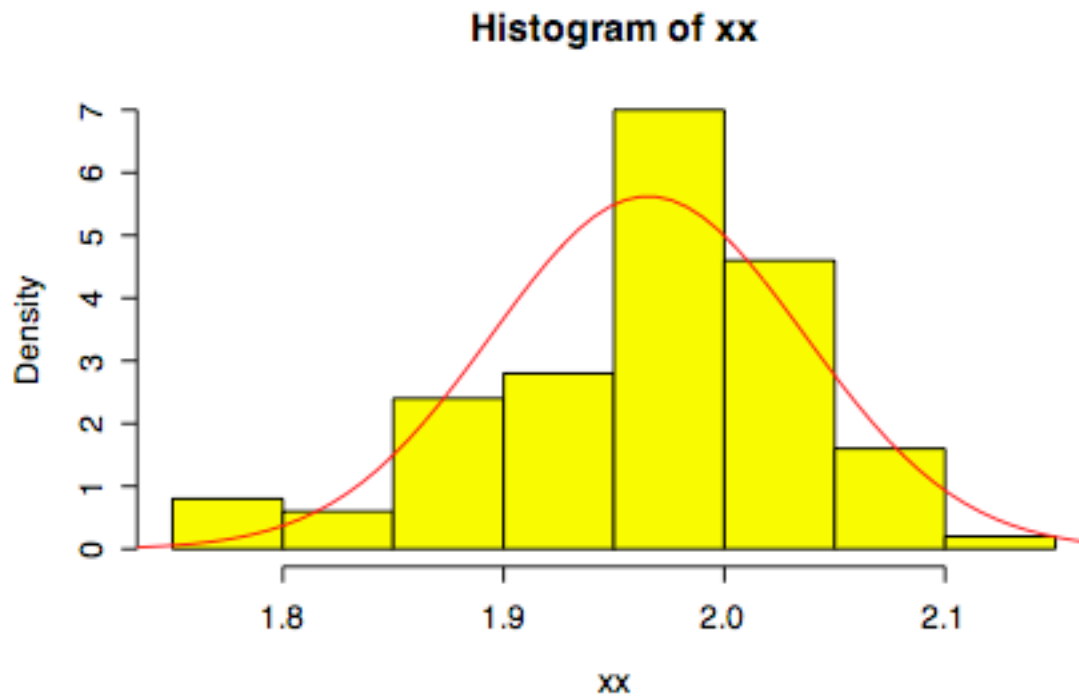
```
> xx <- rweibull(40, 30, 2)
> hist(xx, prob=T, col="yellow")
> lines(xgr, dnorm(xgr, mean(xx), sd(xx)), col="red")
> qqnorm(xx)
> qqline(xx)
```



Weibull(30, 2) data, n = 100

```
> xx <- rweibull(100, 30, 2)
> hist(xx, prob=T, col="yellow")
```

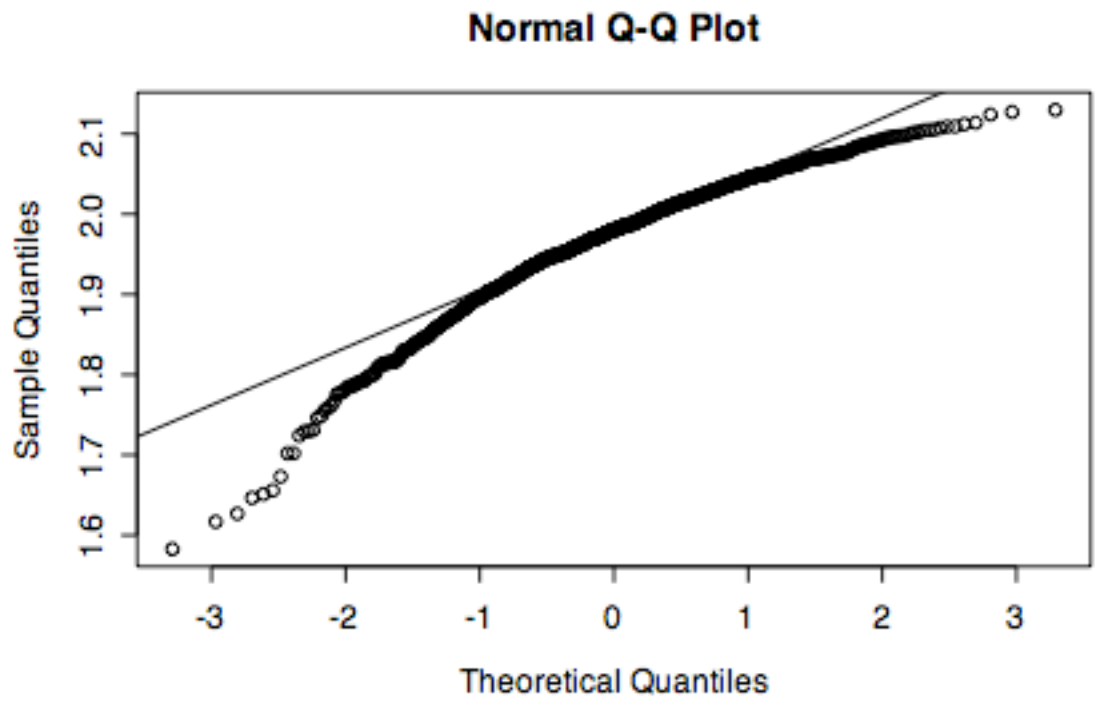
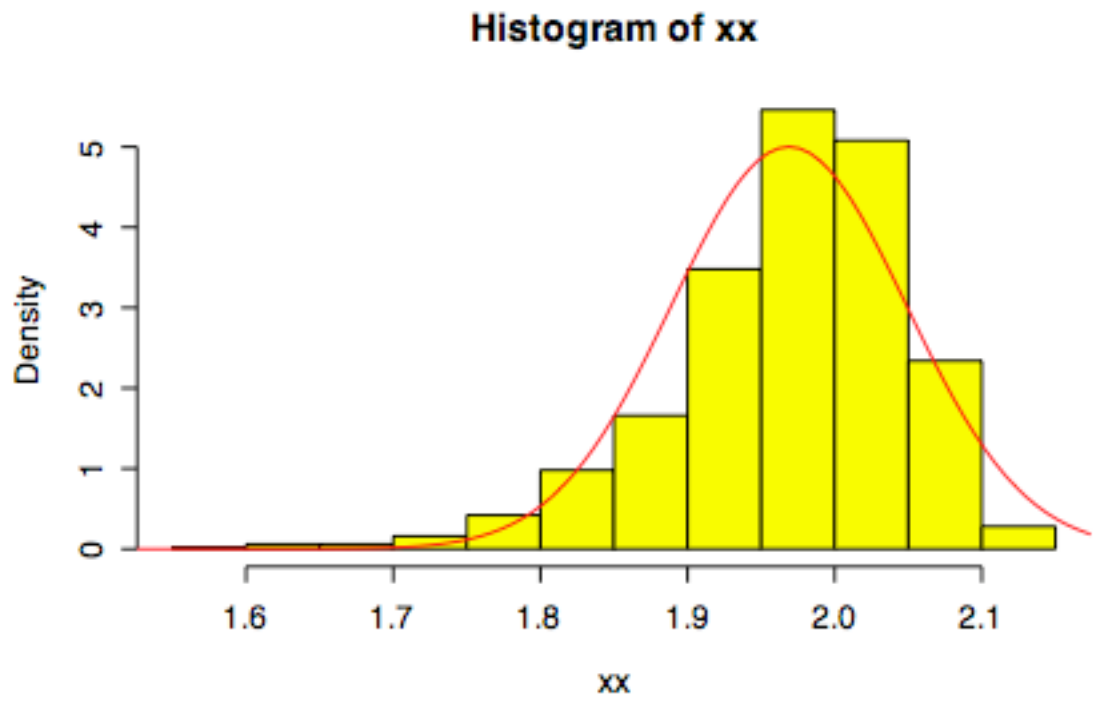
```
> lines(xgr, dnorm(xgr,mean(xx),sd(xx)), col="red")
> qqnorm(xx)
> qqline(xx)
```



Weibull(30, 2) data, n = 1000

```
> xx <- rweibull(1000, 30, 2)
> hist(xx, prob=T, col="yellow")
> lines(xgr, dnorm(xgr,mean(xx),sd(xx)), col="red")
> qqnorm(xx)
```

```
> qqline(xx)
```



Question 4 [10]

This Weibull distribution is negatively skewed. It has mean 1.96365, variance 0.0067271 and standard deviation 0.082019.

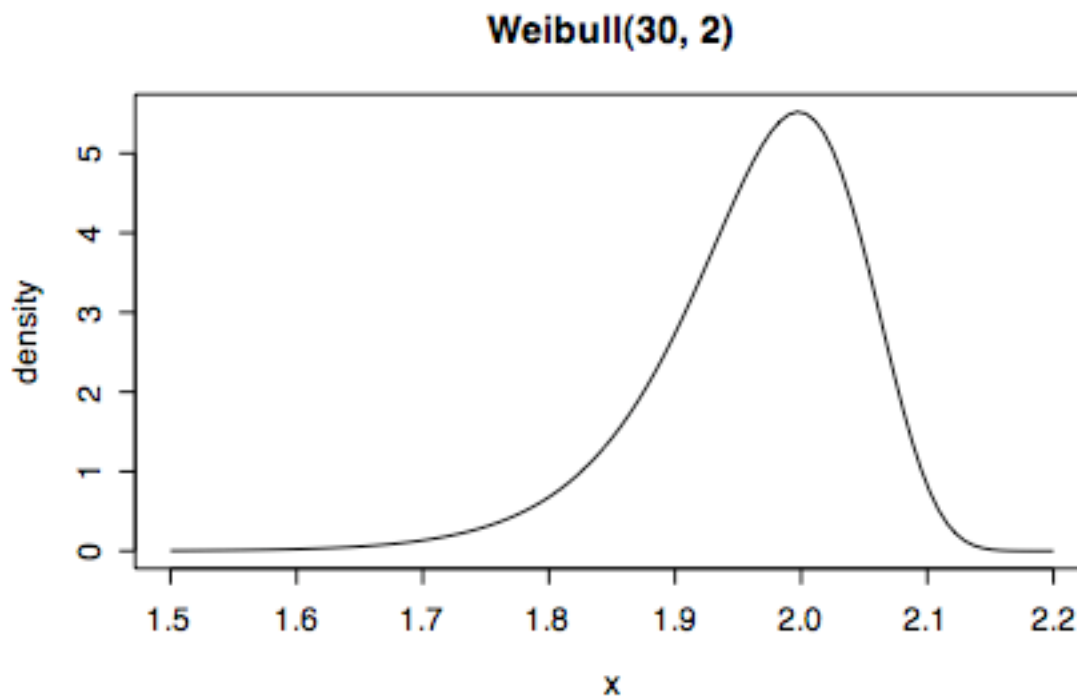
```
> xgr<-seq(1.5,2.2,len=101)
```



```

> plot(xgr,dweibull(xgr,30,2),type="l",
xlab="x",ylab="density",main="Weibull(30, 2)")
> wmean <- 2*gamma(1+1/30)
> wmean
[1] 1.963652
> wvar <- 4*gamma(1+2/30)-wmean^2
> wvar
[1] 0.006727122
> sqrt(wvar)
[1] 0.08201903

```



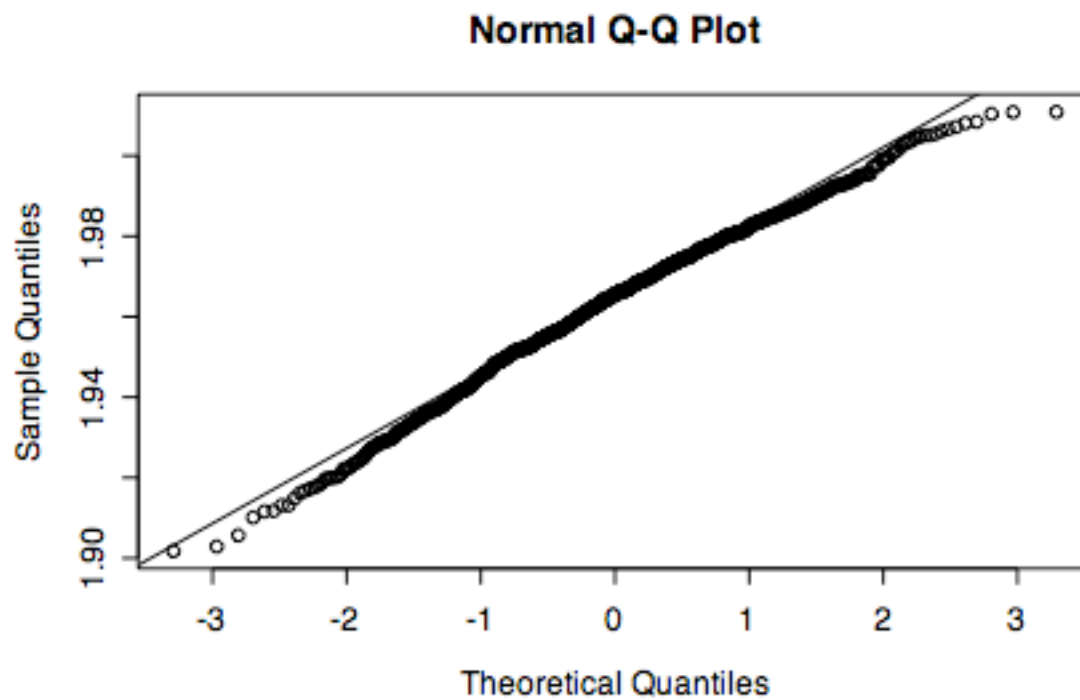
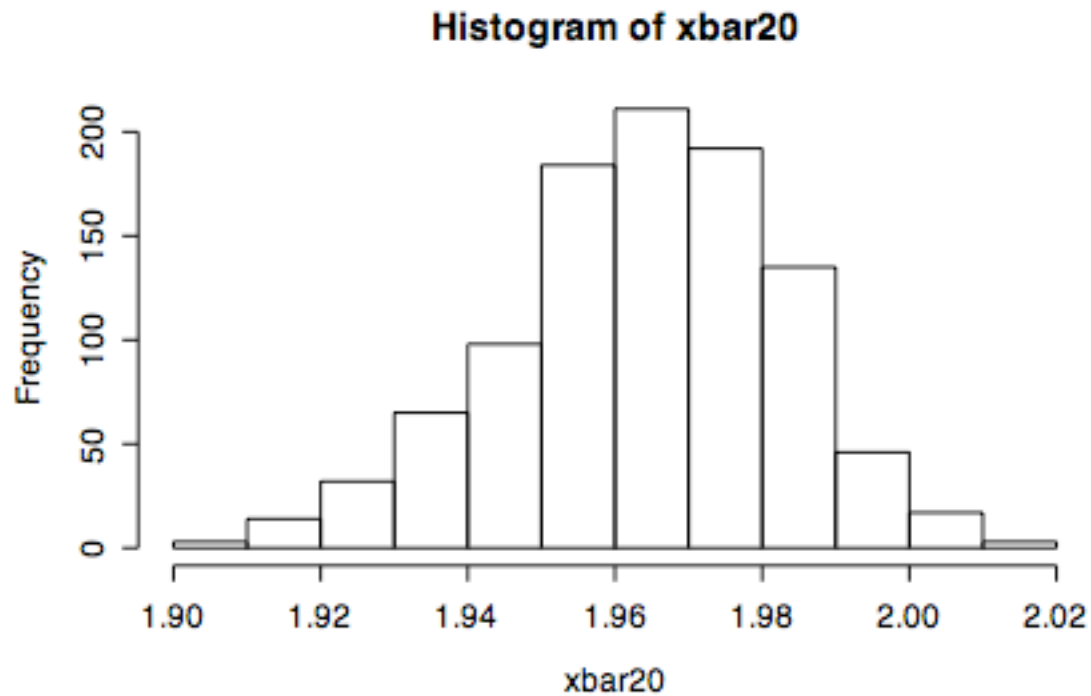
The Central Limit Theorem says that the sampling distribution of the sample mean will have mean 1.96365, standard deviation $0.082019/\sqrt{n}$, and become more normal in shape as n gets large. The examples below for $n = 20$ and $n = 200$ confirm these results.

Weibull(30, 2) data, $n = 20$

```

> xbar20 <- apply(matrix(rweibull(1000*20,30,2),nrow=1000),1,mean)
> hist(xbar20)
> mean(xbar20)
[1] 1.963908
> sd(xbar20)
[1] 0.01877298
> sqrt(wvar/20)
[1] 0.01834001
> qqnorm(xbar20)
> qqline(xbar20)

```



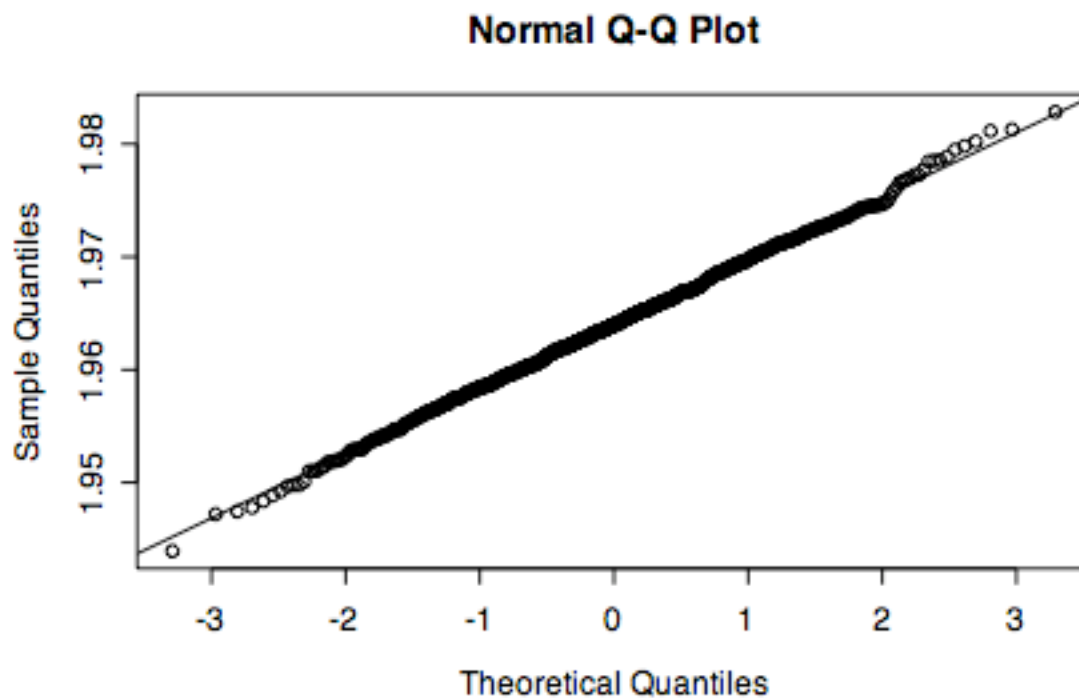
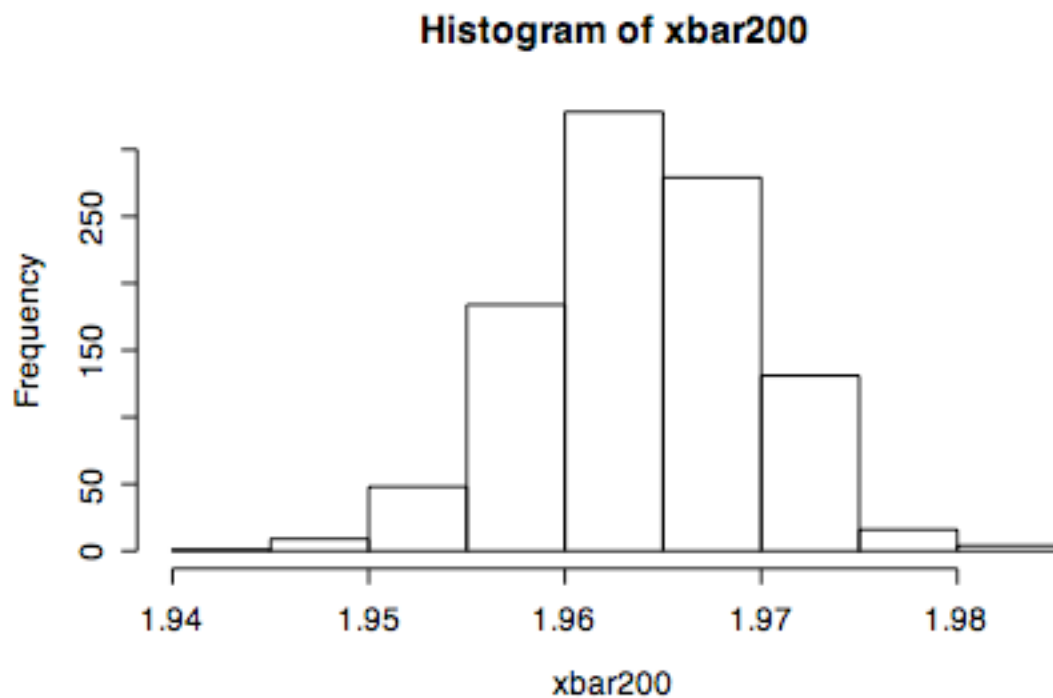
Weibull(30, 2) data, n = 200

```

> xbar200 <- apply(matrix(rweibull(1000*200,30,2),nrow=1000),1,mean)
> hist(xbar200)
> qqnorm(xbar200)
> qqline(xbar200)
> mean(xbar200)
[1] 1.963989
> sd(xbar200)
[1] 0.005719556
> sqrt(wvar/200)

```

[1] 0.005799622



Question 5 [10]

If σ is the standard deviation of the weight of one apple, the total weight of a bag is $N(1000, \sqrt{10} \cdot \sigma)$ so pp calculated below is the probability that a bag is underweight (less than 950 g). This means that the number of underweight bags out of 1000 is $\text{Binom}(1000, pp)$ so the probability of getting more than 60 underweight bags is computed from the binomial or by the

normal approximation to the binomial with continuity correction. The exact binomial calculation gives 0.3073 when $\sigma = 10$, 0.9652 when $\sigma = 11$ and 0.0003194 when $\sigma = 9$.

```
> sigma <- 10
> pp <- pnorm(950, 1000, sqrt(10)*sigma)
> pp
[1] 0.05692315
> 1-pbinom(60, 1000, pp)
[1] 0.3073176
> 1-pnorm(60.5, 1000*pp, sqrt(1000*pp*(1-pp)))
[1] 0.3127100
> sigma <- 9
> pp <- pnorm(950, 1000, sqrt(10)*sigma)
> pp
[1] 0.03947417
> 1-pbinom(60, 1000, pp)
[1] 0.0006926993
> 1-pnorm(60.5, 1000*pp, sqrt(1000*pp*(1-pp)))
[1] 0.0003193574
> sigma <- 11
> pp <- pnorm(950, 1000, sqrt(10)*sigma)
> pp
[1] 0.07530234
> 1-pbinom(60, 1000, pp)
[1] 0.9652065
> 1-pnorm(60.5, 1000*pp, sqrt(1000*pp*(1-pp)))
[1] 0.9619591
```

Full Marks:**Part A: 50****Part B: 40****Total: 90**