Exercise solution:

Question no. 1

Discussion question: In Chi-square case, you noticed that we calculated the upper tail and lower tail values. Why did we not do that in the normal dist and student's t distribution

- Because Normal and T dist. are symmetric about the mean where as chi-square is not. In other words the lower tail and upper tail for normal and t dist. will return the same value but opposite signs, e.g.,

```
qnorm(0.025);       qnorm(1-0.025)

## [1] -1.96
## [1] 1.96
```

```
qt(0.025, df=30);       qt(1-0.025, df=30)

## [1] -2.042
## [1] 2.042
```

Now check chi-square:

```
qchisq(0.025, df=30);       qchisq(1-0.025, df=30)

## [1] 16.79
## [1] 46.98
```

Exercise solution:

Question no. 2

Assume a sample of size (n=29), normally distributed with mean 20. The variance of the population (not sample) is 500? Calculate the 95% confidence interval of mean.

- We can directly use the normal distribution case since population variance is known.
- Confidence intervals can be calculated as follows:

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where standard error $(se) = \frac{\sigma}{\sqrt{n}}$ and for a normal dist. $z_{\alpha/2}$ at 95% confidence level is equal to 1.96

```
z=qnorm(1-0.025) # value of z_a/2 at 95% confidence level
n=29 # sample size
sigma2= 500;          sigma=sqrt(500) # population variance and standard deviation
m= 20 # mean of the variable
```

Now we can simply put in the values in the formula:

```
m + z*sigma/sqrt(n);                  m - z*sigma/sqrt(n)

## [1] 28.14
## [1] 11.86
```

Exercise solution:

Question no. 3

Assume we do not have the population mean, and the sample variance is 450, and sample mean is 20. Calculate the 95% confidence interval of mean.

- We cannot directly use the normal distribution case since population variance is unknown. But we are given sample variance, so we can proceed to use student's t distribution in this case.
- Confidence intervals can be calculated as follows:

$$\bar{X} \pm t_{v,\alpha/2}\frac{s}{\sqrt{n}}$$

where standard error $(se) = \frac{\sigma}{\sqrt{n}}$, v is the degree of freedom.

```
t=qt(1-0.025, 28) # value of t_a/2 at 95% confidence level with 28 degrees of freedom
n=29 # sample size
s2= 450;          s=sqrt(450) # population variance and standard deviation
m= 20 # mean of the variable
```

Now we can simply put in the values in the formula:

```
m + t*s/sqrt(n);                   m - t*s/sqrt(n)


## [1] 28.07
## [1] 11.93
```

Exercise solution:

Question no. 4

Given the sample variance (450), calculate 99% confidence interval of the variance in the above example

- The formula for calculating confidence interval for variance is:

  LCL is: $\frac{(n-1)}{\chi^2_{n-1,\alpha/2}} s^2$ and UCL is: $\frac{(n-1)}{\chi^2_{n-1,1-\alpha/2}} s^2$

```
chi_lower=qchisq(0.005, df=28) # value for lower tail dist. at 99%
chi_upper=qchisq(1-0.005, df=28) # value for upper tail dist. at 99%
```

Upper and lower confidence level:

```
(n-1)*s2/chi_lower;        (n-1)*s2/chi_upper


## [1] 1011
## [1] 247.1
```

Exercise solution:

Question no. 5

Calculate the 90% confidence interval of the variance of height in the example

```
data = read.csv("~/Dropbox/Teaching/Statistics/Lecture 8/data/ryder.csv", header = T)
x = as.numeric(data$hoejde) # we tell R this is a numerical data
x = na.omit(x) # We leave out individuals for whom data is missing
```

```
n <- length(x)          # calculates the no. of observations (n)
m <- mean(x)             # calculates the sample mean
s2 <- var(x);       s2   # calculates the sample variance


## [1] 92.84
```

Calculate the corresponding upper and lower tail values at 90% confidence level

```
chi_lower=qchisq(0.05, df=n-1) # value for lower tail dist. at 90%
chi_upper=qchisq(1-0.05, df=n-1) # value for upper tail dist. at 90%
```

Upper and lower confidence level:

```
(n-1)*s2/chi_lower;        (n-1)*s2/chi_upper


## [1] 97.21
## [1] 88.78
```

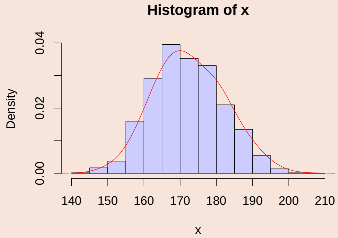
Exercise solution:

Question no. 6

Plot the distribution of heights. What kind of distribution does it look like?

```
data = read.csv("~/Dropbox/Teaching/Statistics/Lecture 8/data/ryder.csv", header = T)
x = as.numeric(data$hoejde) # we tell R this is a numerical data
x = na.omit(x) # We leave out individuals for whom data is missing
```

```
hist(x, prob=T, col = rgb(0.8, 0.8, 1));       lines(density(x, adjust=2), lwd = 2, col="red")
```



Yes, it resembles a normal distribution