ISE 3293/5013 Laboratory 7
Sampling Distributions

In this lab we will investigate the idea of a *sampling distribution*. Most of the sampling will be done from a normal population. The procedure is as follows:

1. Sample from a Normal distribution using rnorm().
2. Create a statistic (i.e a function of the data).
3. Store the statistic.
4. Repeat the procedure for a designated number of iterations.
5. When finished create a histogram of the statistic.

The method for doing this will be to use a ready-made R script, adapt it and re-run it for the problems given below. This process will be very instructive and should help you to not only perform statistics but also give you the basis for much distributional theory.

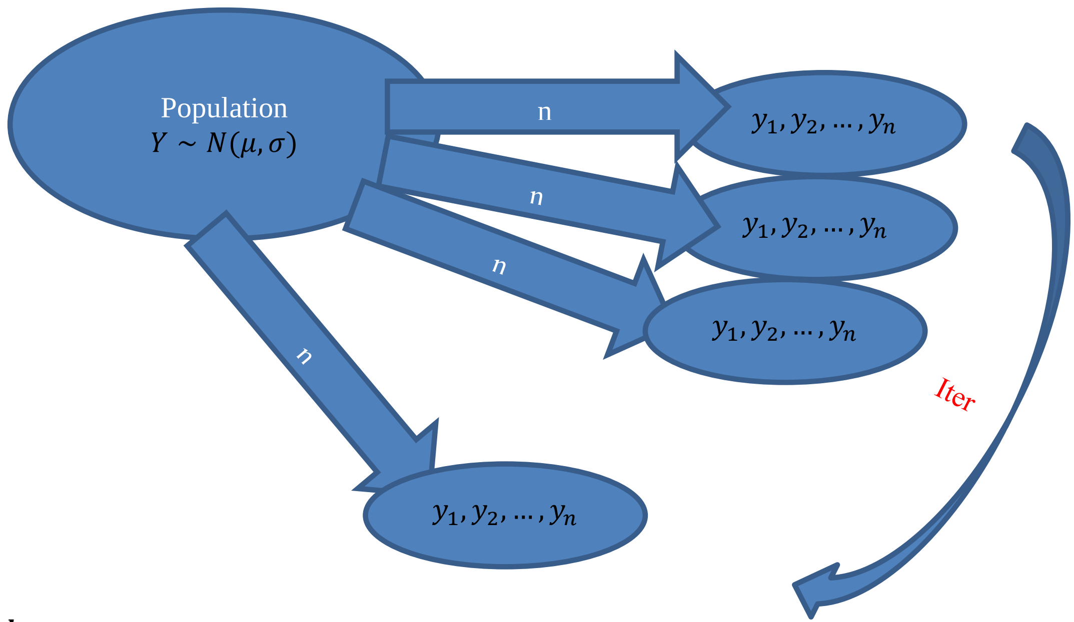
The lab is in two parts

1. One population sampling
2. Two population sampling

Objectives

In this lab you will learn how to:

1. Create a sample from one population.
2. Create statistics.
3. Create sampling distributions and appropriate graphs.
4. Sample from two different populations and create sampling distributions for statistics made from both samples.



Tasks

All output made please copy and paste into **this word file**. Save and place in the dropbox when completed. Anything you are asked to make should be recorded under the question in this document. There will be two files you need to upload:

- a pdf of this document (pdf) or the word file (docx)
- a text file of all the code you used to create answers (txt)

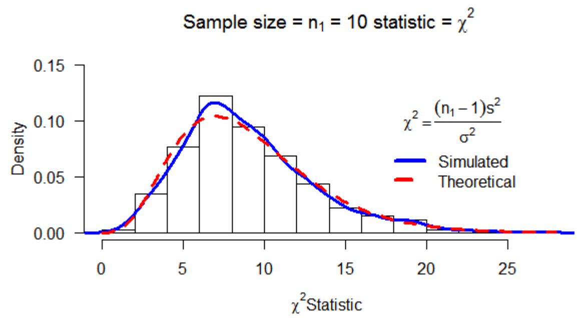
Note: All plots you are asked to make should be recorded in this document.

- Task 1
  - Make a folder LAB7
  - Download the file "lab7.r"
  - Place this file with the others in LAB7.
  - Start Rstudio
  - Open "lab7.r" from within Rstudio.
  - Go to the "session" menu within Rstudio and "set working directory" to where the source files are located.
  - Issue the function `getwd()` and copy the output here.

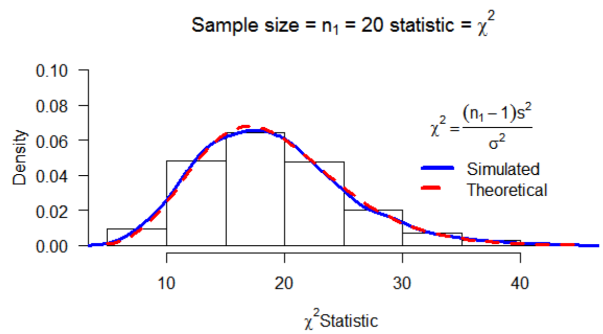
```
"F:/Google Drive - Saied/Courses/02 OU/11 Fundamentals of Engi
neering Statistical Analysis/02 Labs/07 Lab 7"
```

- Task 2
  - Make a new file for your code in RStudio editor, call it "mylab7.R" and place in it all the code you need to answer the tasks of this lab (copy and paste from lab7.R).
  - Use the hash # symbol and write your own comments in the code file explaining what the code does.
  - The first statistic we will make is the Chi-square statistic. This is created by the following formula $\chi^2 = \frac{(n-1)s^2}{\sigma^2}$, where $s^2$is the sample variance and $\sigma^2$is the population variance, where the population is Normal, $Y \sim N(\mu, \sigma^2)$, and n is the sample size.
  - The function you will use us called `mychisim()`

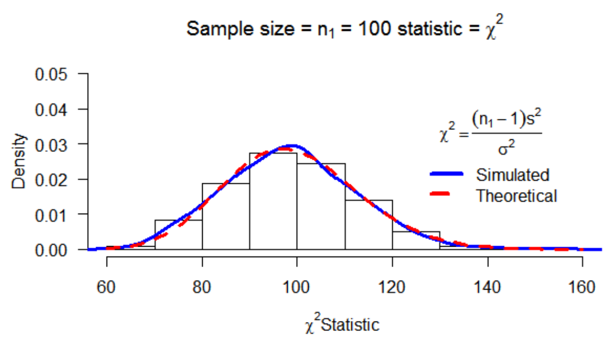
- Make four plots according to the following options (the function will require you to click into the graph to complete its operation) – you may need to adjust ymax.
  - $n_1 = 10, iter = 1000, \mu_1 = 10, \sigma_1 = 4$



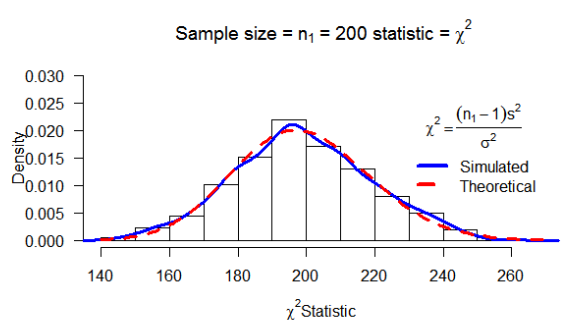
  - $n_1 = 20, iter = 1000, \mu_1 = 10, \sigma_1 = 4$



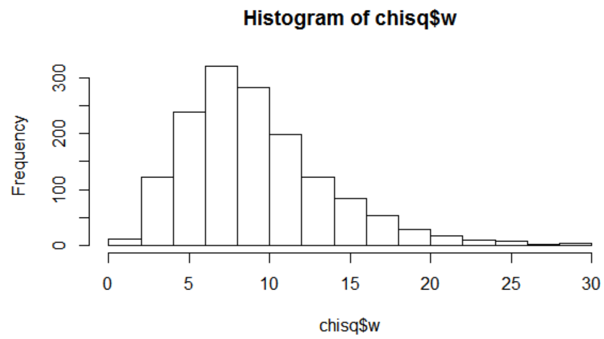
  - $n_1 = 100, iter = 1000, \mu_1 = 10, \sigma_1 = 4$



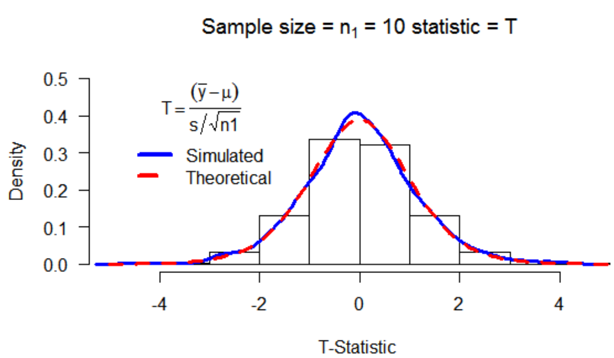
  - $n_1 = 200, iter = 1000, \mu_1 = 10, \sigma_1 = 4$



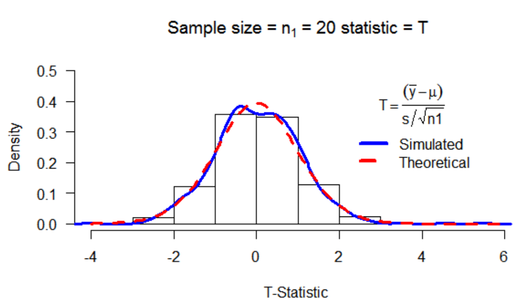
- The function returns a list of statistics, the statistic we are interested in is the χ^2value for each iteration. These values are in the vector called w. Invoke the function with $n_1 = 10, iter = 1500, \mu_1 = 20, \sigma_1 = 10$ and place the output into an object called chisq. Make a histogram of chisq$w.



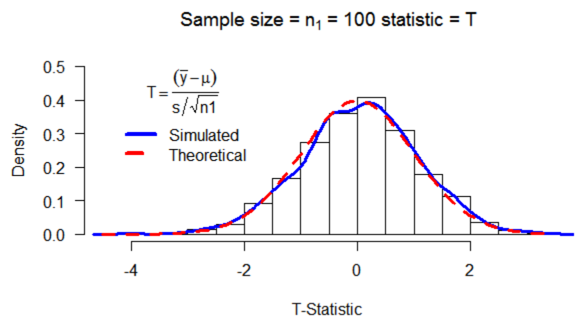
- Task 3
  - Now we will adjust the function mychisim() by copying it and replacing appropriate portions. Copy and paste the mychisim function into mylab7.R and rename the function, call it myTsim.
  - The statistic you will need to make is $T = \frac{\bar{y}-\mu}{\frac{s}{\sqrt{n}}}$, this can most easily be done by using the functions, mean() and sd().
  - **Once you have made the function make some simulations as before (make sure you have all the code ready to repeat at the end) – that is:**
    - **A) Make four plots according to the following options (the function will require you to click into the graph to complete its operation) – you may need to adjust ymax.**
      - $\boldsymbol{n_1 = 10, iter = 1000, \mu_1 = 10, \sigma_1 = 4}$



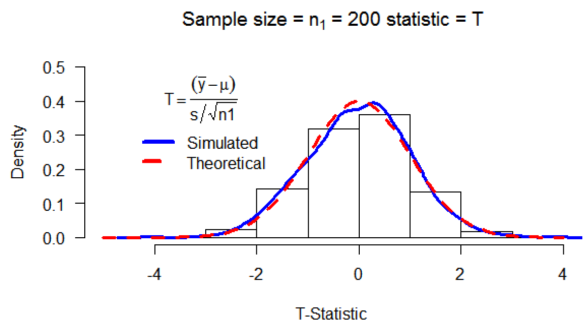
      - $\boldsymbol{n_1 = 20, iter = 1000, \mu_1 = 10, \sigma_1 = 4}$



- $n_1 = 100, iter = 1000, \mu_1 = 10, \sigma_1 = 4$

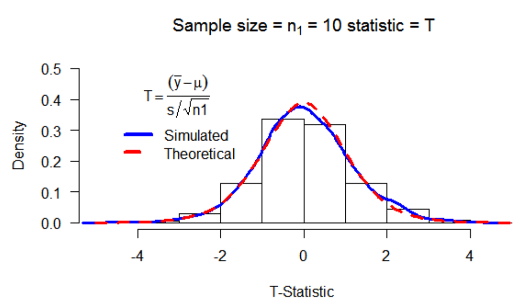


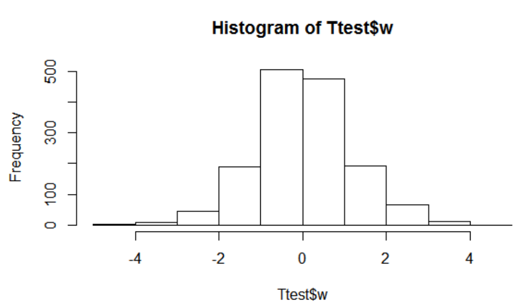
- $n_1 = 200, iter = 1000, \mu_1 = 10, \sigma_1 = 4$



  - **B) The function returns a list of statistics, the statistic we are interested in is the T value for each iteration. These values are in the vector called w. Invoke the function with $n_1 = 10, iter = 1500, \mu_1 = 20, \sigma_1 = 10$ and place the output into an object called T. Make a histogram of T$w.**

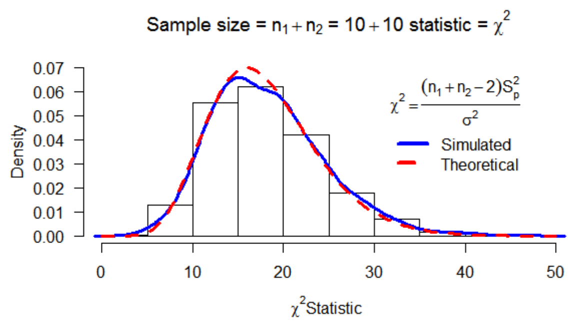
- **Record all plots here.**



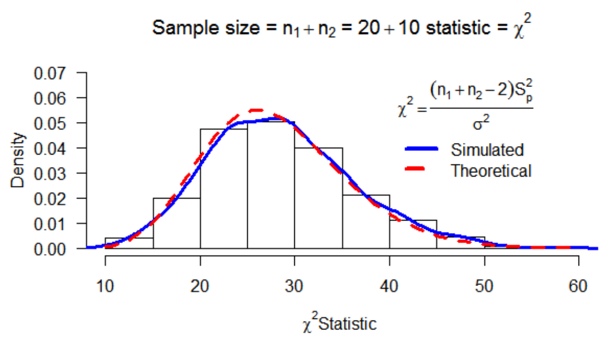


- Now start up BBFLASHBACK recorder and record the re-making of the plots made above in A) and B) by re-issuing the code you made, give a brief dialog as you record. Place the .fbr file into D2L Lab 7 dropbox.

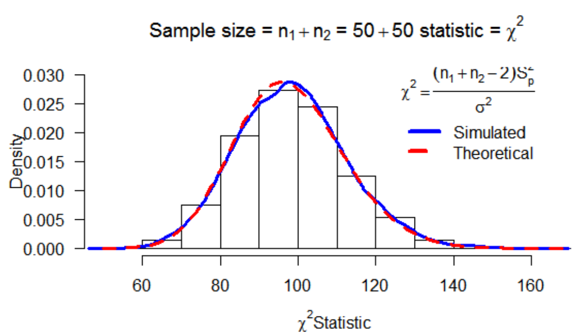
- Task 4
  - You will now make simulations from two populations and use the samples to make a statistc.
  - The first statistic is the two sample chisquare statistic. The function is called mychsim2().
  - The statistic is $\chi^2 = \frac{(n_1+n_2-2)S_p^2}{\sigma^2}$, where we assume that both populations have the same variance σ^2. $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$, where $S_i^2$is the sample variance from population i, n_i is the sample size and S_p^2 is the pooled sample variance.
  - Use mychisim2() to sample from two normal populations with the following parameters:
    - $n_1 = 10, n_2 = 10, \mu_1 = 5, \mu_2 = 10, \sigma_1 = \sigma_2 = 4, iter = 1000$



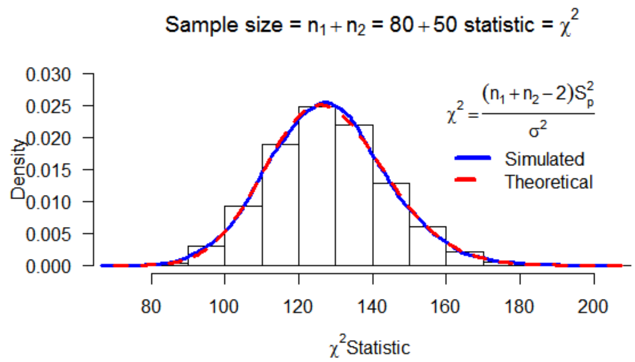
    - $n_1 = 20, n_2 = 10, \mu_1 = 3, \mu_2 = 5, \sigma_1 = \sigma_2 = 10, iter = 1000$



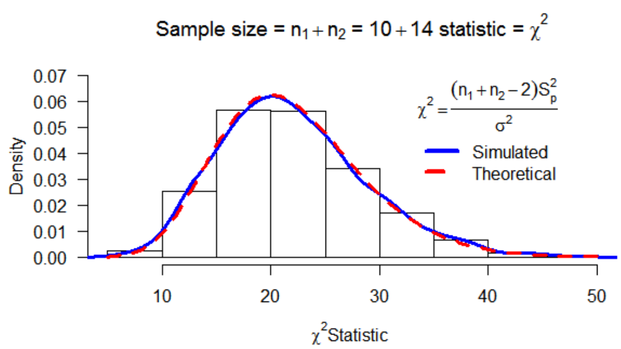
    - $n_1 = 50, n_2 = 50, \mu_1 = 5, \mu_2 = 10, \sigma_1 = \sigma_2 = 4, iter = 10000$

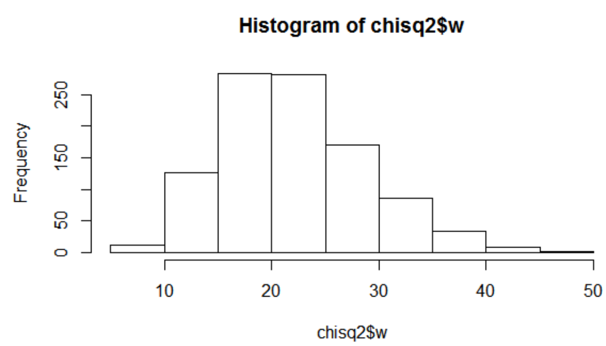


    - $n_1 = 80, n_2 = 50, \mu_1 = 3, \mu_2 = 5, \sigma_1 = \sigma_2 = 10, iter = 10000$



- Use default values in the function with $iter = 10000$ and use the output to make a histogram as before.





- Task 5
  - Alter the function myTsim2() to place the legend where you click with the mouse.
  - From the table taken from the book (MS page 252) and reproduced below write down the student's T statistic the function calculates, explain the notation.

I check the book, it seems the page number is wrong. I used the formula given in this document.

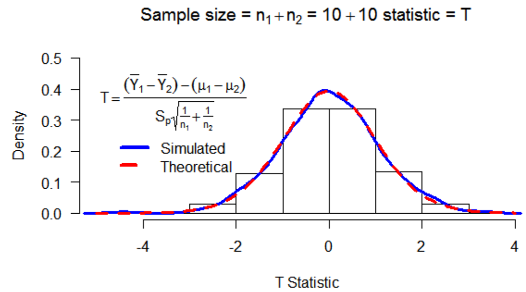
$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$\bar{Y}_1$: The mean of first population sample
$\bar{Y}_2$: The mean of second population sample
μ_1: The mean of first normal distribution we taking sample from
μ_2: The mean of second normal distribution we taking sample from

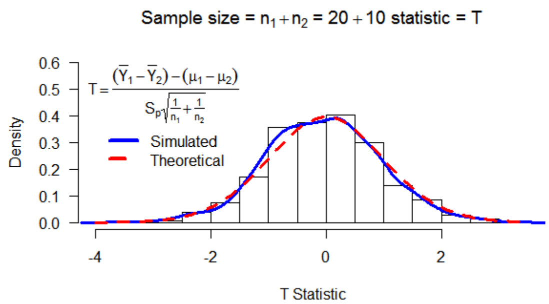
- Copy and paste from the code the part that calculates the statistic.

```
w=((ybar1-ybar2)-(mean1-mean2))/sqrt(spsq*(1/n1+1/n2))
```

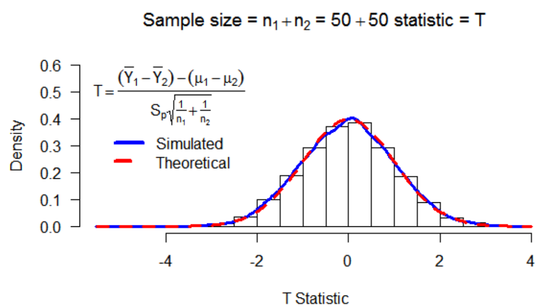
- Use myTsim2() to sample from two normal populations with the following parameters:
  - $n_1 = 10, n_2 = 10, \mu_1 = 5, \mu_2 = 10, \sigma_1 = \sigma_2 = 4, iter = 1000$



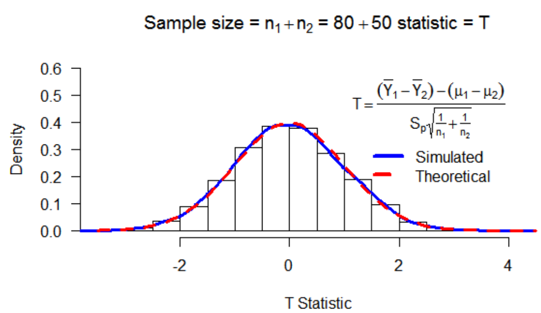
  - $n_1 = 20, n_2 = 10, \mu_1 = 3, \mu_2 = 5, \sigma_1 = \sigma_2 = 10, iter = 1000$



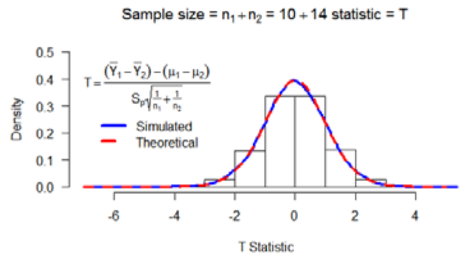
  - $n_1 = 50, n_2 = 50, \mu_1 = 5, \mu_2 = 10, \sigma_1 = \sigma_2 = 4, iter = 10000$

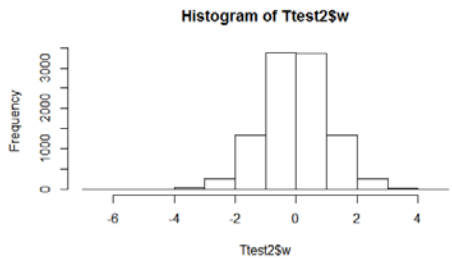


  - $n_1 = 80, n_2 = 50, \mu_1 = 3, \mu_2 = 5, \sigma_1 = \sigma_2 = 10, iter = 10000$

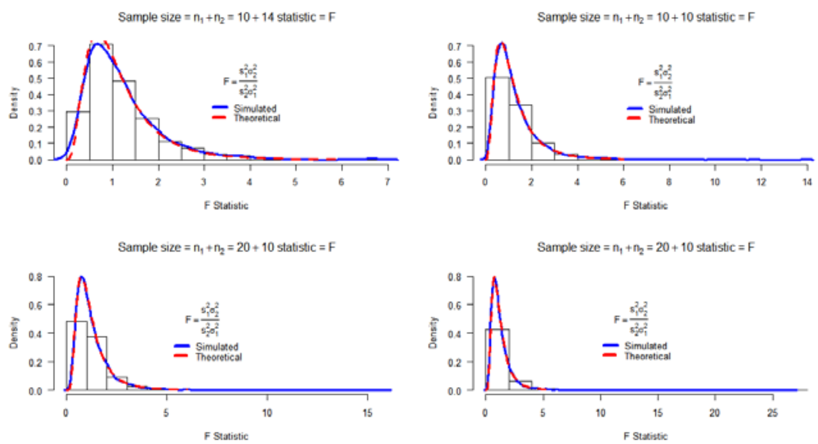


- Use default values in the function with $iter = 10000$ and use the output to make a histogram as before.

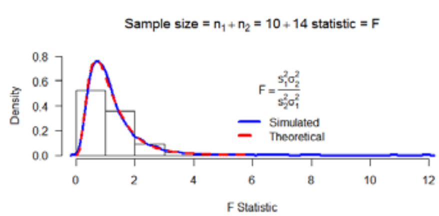




- Task 6
  - Now use myFsim2() to create F statistics from two normal populations.
  - Use the table below to write down the statistic that the function will calculate.
  - What assumptions are made?
  - Make four plots with different parameters.



  - Make a histogram from the function using default values.



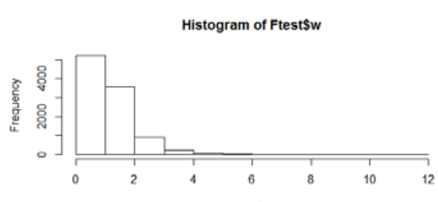


TABLE 6.3a Sampling Distributions of Statistics Based on Independent Random Samples Observations, Respectively, from Normally Distributed Populations with Parameters (μ_1

Statistic	Sampling Distribution	Additional Assumptions	
$\chi^2 = \frac{(n_1 + n_2 - 2)S_p^2}{\sigma^2}$ where $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$	Chi-square with $\nu = (n_1 + n_2 - 2)$ degrees of freedom	$\sigma_1^2 = \sigma_2^2 = \sigma^2$	
$T = \frac{(\overline{Y}_1 - \overline{Y}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ where $S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$	Student's T with $\nu = (n_1 + n_2 - 2)$ degrees of freedom	$\sigma_1^2 = \sigma_2^2 = \sigma^2$	Th an
$F = \left(\frac{S_1^2}{S_2^2}\right)\left(\frac{\sigma_2^2}{\sigma_1^2}\right)$	F distribution with $\nu_1 = (n_1 - 1)$ numerator degrees of freedom and $\nu_2 = (n_2 - 1)$ denominator degrees of freedom	None	The Def

TABLE 6.3b Sampling Distributions of Statistics Based on a Random Sample from a Single Normally Distributed Population with Mean μ and Variance σ^2

Statistic	Sampling Distribution	Additional Assumptions	Basis of Derivation of Sampling Distribution
$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$	Chi-square with $\nu = (n - 1)$ degrees of freedom	None	Methods of Section 6.7
$t = \frac{\overline{y} - \mu}{S/\sqrt{n}}$	Student's T with $\nu = (n - 1)$ degrees of freedom	None	Theorems 6.10–6.11 and Definition 6.15

################### LAB FINISHES HERE ###############################

- Task 7 – Extra for experts
  - Make a function that uses w to create confidence intervals – hint: you will need quantile()

```
myclevel<-function(w, cprob = 0.95, ...){

  pl = (1 - cprob) / 2
  ph = 1 - pl
  a = quantile(w, probs = pl)
  b = quantile(w, probs = ph)
  clevel = c(round(a,4), round(b,4))
  names(clevel)<- c("Low Tail", "High Tail")
  clevel

  return(list(summary=summary(w),clevel=clevel, a=a, b=b))
}

k = myclevel(chisq$w, cprob = 0.95)
myclevel(chisq$w, cprob = 0.95)
```