2.15 Simulations to Study the ANOVA HOV Assumption

- The file *simanova.r* posted on the course webpage contains R code that will do the following assessing the impact of equal ($\sigma_1 = \sigma_2 = \sigma_3$) and unequal (not all σ_i are equal) on the probability of rejecting $H_0 : \mu_1 = \mu_2 = \mu_3$.
- Various cases can be studied by varying any or all of the following:
  - The values of σ_1, σ_2, and σ_3.
  - The values of μ_1, μ_2, and μ_3.
  - The values of n_1, n_2, and n_3.
- The program will output the estimated probability of rejecting $H_0 : \mu_1 = \mu_2 = \mu_3$ assuming α levels of .01. .05, and .10 from a oneway ANOVA. That is, the proportion of samples that lead to a rejection of H_0 using $\alpha = .01..05, .10$.
- If $H_0 : \mu_1 = \mu_2 = \mu_3$ is **true**, these estimated probabilities represent estimates of the **Type I error** (i.e. the probability of incorrectly rejecting H_0 when H_0 it is true). Thus, the values should be close to the nominal (stated) α levels of .01, .05, and .10.
- If $H_0 : \mu_1 = \mu_2 = \mu_3$ is **false**, these estimated probabilities represent estimates of the **power of the F-test**.

  The **power** of a test equals the probability of correctly rejecting H_0 when H_0 is false

  $= 1 -$ **the probability of not rejecting H_0 when H_0 is false**

  $= 1 -$ **Type II error**.

R Code to perform the simulations

```
# ANOVA Simulation with 3 treatments

#################################################################
#  ASSESSING THE IMPACT OF EQUAL vs UNEQUAL STANADARD DEVIATIONS
#################################################################

  a    <- 3    # Enter number of treatments

  sd.1 <- 1    # Enter sigma_1  1 1 1
  sd.2 <- 1    # Enter sigma_2  1 2 2
  sd.3 <- 1    # Enter sigma_3  1 3 3

  n.1 <- 9     # Enter n_1      9 9 9
  n.2 <- 9     # Enter n_2      9 9 6
  n.3 <- 9     # Enter n_3      9 9 3

  mu.1 <- 10   # Enter mu_1     10 10 10
  mu.2 <- 10   # Enter mu_2     10 10 10
  mu.3 <- 10   # Enter mu_3     10 10 10

  iter <- 10000   # Enter the number of t-statistics to simulate

## Simulate F-statistics
  N <- n.1 + n.2 + n.3
  df.MSE   <- N - a
  df.MSE
```

```
  df.MStrt <- a -1
  df.MStrt

  F.stat <- numeric(iter)

  for (i in 1:iter) {
     sample1 <- rnorm(n.1, mean=mu.1, sd=sd.1)
     sample2 <- rnorm(n.2, mean=mu.2, sd=sd.2)
     sample3 <- rnorm(n.3, mean=mu.3, sd=sd.3)

     var.1 <- var(sample1)
     var.2 <- var(sample2)
     var.3 <- var(sample3)

     SS.E  <- (n.1-1)*var.1 + (n.2-1)*var.2 + (n.3-1)*var.3
     MS.E  <- SS.E/df.MSE

     all.dat  <- c(sample1,sample2,sample3)
     all.dat
     SS.TOTAL <- (N-1)*var(all.dat)

     SS.TRT   <- SS.TOTAL - SS.E
     MS.TRT   <- SS.TRT/df.MStrt

     F.stat[i] <- MS.TRT/MS.E
     }

windows()
Fmax = max(F.stat)

hist(F.stat, freq=FALSE, nclass=50, xlim=c(-.01,Fmax), ylim=c(0,.8),
     main="Histogram of F-statistics with superimposed F pdf")
curve(df(x,df.MStrt,df.MSE), add=TRUE, col=2, lwd=2)

  F.01 <- qf(.99,df.MStrt,df.MSE)
  F.05 <- qf(.95,df.MStrt,df.MSE)
  F.10 <- qf(.90,df.MStrt,df.MSE)
  F.01
  F.05
  F.10

# Simulated rejection probabilities for alpha = .01, .05, .10.
# If Ho: mu.1 = mu.2 = mu.3 is true, this is the estimated Type I error.
# If Ho is not true, this is the estimated Power = 1 - Type II error.

  reject.01  <- ifelse(F.stat >= F.01,1,0)
  pvalue.01  <- sum(reject.01)/iter
  pvalue.01

  reject.05  <- ifelse(F.stat >= F.05,1,0)
  pvalue.05  <- sum(reject.05)/iter
  pvalue.05

  reject.10  <- ifelse(F.stat >= F.10,1,0)
  pvalue.10  <- sum(reject.10)/iter
  pvalue.10
```