INTRODUCTION TO R PROGRAMMING DAY 1 EXERCISE:

A time-tested strategy to tackle what may appear to be a complex programming task is to begin by "getting working" the very simplest prototype version of just *one element* of the function, test it, and then incrementally add the additional features that are required.

1) Write a function **T.test()** that performs a two sample t-test, similar to the R function **t.test()**. The formula for the statistic is:

$$ t = \frac{\bar{y}_1 - \bar{y}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} $$

Where **n1** and **n2** are the two sample sizes, **y-bar1** and **y-bar2** are the two sample means, and **s** is the square root of the pooled estimate of the variance (**s-squared**), given by:

$$ s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)} $$

We can use the following header for the function:

```
> T.test <- function(y1,y2) { ...
+ }
```

In the body of the function:

a) First find **n1** and **n2**
b) Then calculate the numerator of **t**

c) Then calculate the pooled variance and hence **s**

d) Finally, calculate the value of **t** and return that as the value of the function

Please give this a try…we will likely build on this in future days…..

```
######################################################
#####      HOMEMADE TWO SAMPLE T-TEST FUNCTION    #####
######################################################

# The Long Way:

T.test <- function(y1, y2){ # In body of function
  # first find n1 and n2
  n1  <- length(y1)
  n2  <- length(y2)
  y.bar1 <- mean(y1)
  y.bar2 <- mean(y2)
  # then calculate the numerator of t
  t.num <- (y.bar1 - y.bar2)
  s1  <- var(y1)
  s2  <- var(y2)
  # Then calculate pooled variance and hence s
  s.sq.num <- ((n1-1)*s1)+((n2-1)*s2)
  s.sq.den <- (n1-1)+(n2-1)
  s.sq <- s.sq.num/s.sq.den
  s <- sqrt(s.sq)
  # Calculate value of t and return it
  t.value <- t.num/(s*sqrt(1/n1 + 1/n2))
  return(t.value)
}

male <- rnorm(1000,0,100)
hist(male)
summary(male)

female <- rnorm(1000,10,100)
hist(female)
summary(female)

# This one is homemade:
tstat <- T.test(female,male); tstat

# Compare homemade results to the t.test() function in R
tst <- t.test(female,male); tst

# Homemade function can be written more compactly as:
```

```
T.test2 <- function(y1, y2) {
  n1  <- length(y1); n2  <- length(y2)
  yb1 <- mean(y1);    yb2 <- mean(y2)
  s1  <- var(y1);     s2  <- var(y2)
  s <- ((n1-1)*s1 + (n2-1)*s2)/(n1+n2-2)
  t.value <- (yb1 - yb2)/sqrt(s*(1/n1 + 1/n2))
  t.value
}

tstat <- T.test2(female,male); tstat
```