1. Nest the simulation of the two-group t-test from “week6” notes in a macro program (original code reproduced below). The parameters to this macro program should include: the sample sizes ($N_X, N_Y$), the population means ($\mu_X, \mu_Y$), the population sigmas ($\sigma_X, \sigma_Y$) and the number of simulation ($N_{sims}$).

Run your macro program to estimate the Type I error rates for combinations of $N_X$ and $N_Y = [5, 15, 25] \times [5, 15, 25]$ and $\sigma_Y/\sigma_X = 1, 2, 4$. Estimate Power for the same conditions with $\mu_X - \mu_Y = 1$. Note that you could use MACRO variables to build this code.

```r
/* Problem: Explore whether t-test really is robust to violations of the equal variance assumption */

Strategy: See if the t-test operates at the nominal Type I error rate when the unequal variance assumption is violated

Test case: n1=n2=10
Population 1: N(0,1)
Population 2: N(0,4)
*/

data twogroup;
array x{10} x1-x10;
array y{10} y1-y10;

do isim = 1 to 10000;
/* generate samples X~N(0,1) Y~N(0,4) - normal case */
do isample = 1 to 10;
x{isample} = ranor(0);
y{isample} = 2*ranor(0);
end;
/* calculate the t-statistic */
xbar = mean(of x1-x10);
ybar = mean(of y1-y10);
xvar = var(of x1-x10);
yvar = var(of y1-y10);
s2p = (9*xvar + 9*yvar)/18;
tstat = (xbar-ybar)/sqrt(s2p*(2/10));
```
Pvalue = 2*(1-probt(abs(tstat),18));
Reject05 = (Pvalue <= 0.05);

keep xbar ybar xvar yvar s2p tstat Pvalue Reject05;
output;
end;  * end of the simulation loop;

/*
proc print;
run;
*/
proc freq;
table Reject05;
run;

2. Write a macro program to fit a multiple regression model where the Y is identified as one macro variable and the Xs are identified as another macro variable. Your macro should
   i. fit the model
   ii. plot the residuals versus each of the X's
   iii. plot the residuals versus the Yhat.
   iv. plot the normal probability plot of the residuals.

The macro parameters provided to this program should include
   i. the dataset name
   ii. the RESPONSE variable
   iii. the PREDICTOR variable(s)
   iv. a title

Provide a check to make sure that an argument was passed for the dataset, response variable and the predictor variables. If an argument was not provided, generate an error message.

Test your macro program with data predicting the weights of squids as a function of various squid measurements.

/*
Squid data – from Myers
   y = squid weight
   x1 = rostral length (in)
   x2 = wing length (in)
   x3 = rostral to notch length
   x4 = notch to wing length
   x5 = width
Ref:  Myers
*/

data squid;
   input x1 x2 x3 x4 x5 y @@;
datalines;
1.31 1.07 0.44 0.75 0.35 1.95 1.55 1.49 0.53 0.90 0.47 2.90
0.99 0.84 0.34 0.57 0.32 0.72 0.99 0.83 0.34 0.54 0.27 0.81
1.05 0.90 0.36 0.64 0.30 1.09 1.09 0.93 0.42 0.61 0.31 1.22
1.08 0.90 0.40 0.51 0.31 1.02 1.27 1.08 0.44 0.77 0.34 1.93
0.99 0.85 0.36 0.56 0.29 0.64 1.34 1.13 0.45 0.77 0.37 2.08
1.30 1.10 0.45 0.76 0.38 1.98 1.33 1.10 0.48 0.77 0.38 1.90
1.86 1.47 0.60 1.01 0.65 8.56 1.58 1.34 0.52 0.95 0.50 4.49
1.97 1.59 0.67 1.20 0.59 8.49 1.80 1.56 0.66 1.02 0.59 6.17
1.75 1.58 0.63 1.09 0.59 7.54 1.72 1.43 0.64 1.02 0.63 6.36
1.68 1.57 0.72 0.96 0.68 7.63 1.75 1.59 0.68 1.08 0.62 7.78
2.19 1.86 0.75 1.24 0.72 10.15 1.73 1.67 0.64 1.14 0.55 6.88

[OPTIONAL] If you really want to go crazy with this assignment, then add the following options
* plot the residuals versus each X squared.
* plot the residuals versus $X_iX_j$ (pairs of the predictors).