The George Washington University
Department of Statistics

Stat 4197/6197 – Fall 2019

Week 13 – November 22, 2019

Lecture Topics: Miscellaneous

1) Simulating Data with SAS
2) Calling R in a SAS/IML Session
3) Review of Selected Topics of Macro Facility

Acknowledgements: Texts are primarily adapted/obtained from the SAS Blogs (The DO LOOP by Rick Wicklin) and other web sources. Portions of SAS' copyrighted SAS course content are reproduced here with permission of SAS Institute Inc., Cary, NC, USA. SAS Institute Inc. makes no warranties with respect to these materials and disclaims all liability therefor. Do not copy or circulate the handouts.

Simulations

Simulations provide a way to answer questions and explore properties of statistical estimators and procedures (Kleinman and Horton, 2014).

In statistics, data are commonly simulated to

- generate random samples from a statistical distribution (discrete or continuous) with known properties
- estimate the sampling distributions of basic measures such as means, median, Pearson correlations
- evaluate statistical techniques
- ascertain statistical power of hypothesis tests

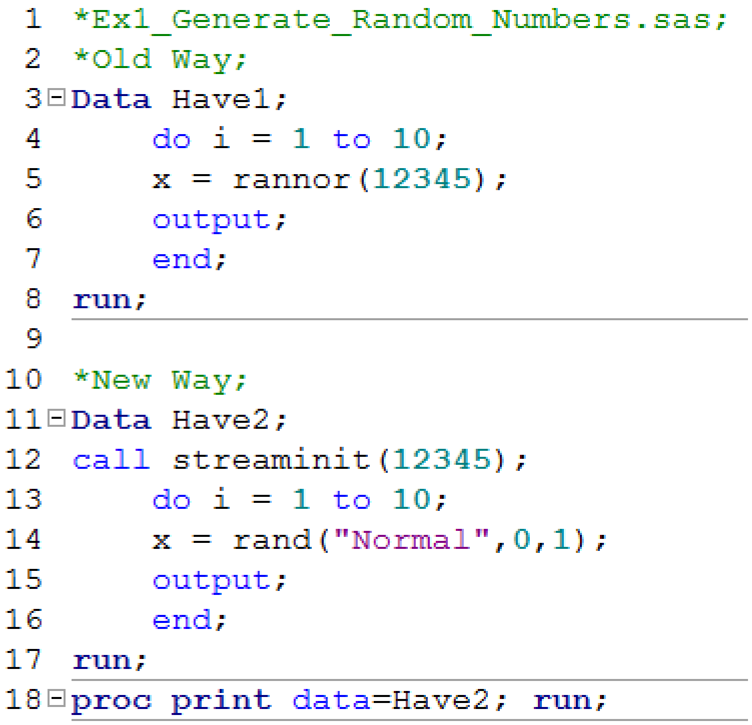
(Source: Simulating Data with SAS® by Rick Wicklin, 2013)

Some common functions used in simulation

- PDF Function
- CDF Function
- Quantile Function
- RAND Function

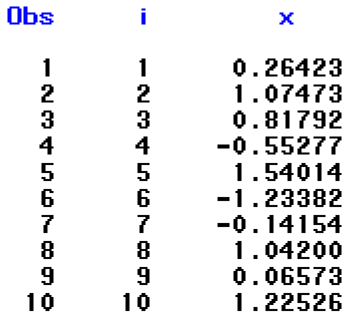
Acknowledgements: Texts are primarily adapted/obtained from the SAS Blogs (The DO LOOP by Rick Wicklin) and other web sources. Portions of SAS' copyrighted SAS course content are reproduced here with permission of SAS Institute Inc., Cary, NC, USA. SAS Institute Inc. makes no warranties with respect to these materials and disclaims all liability therefor. Do not copy or circulate the handouts.

Generating Data



```
1  *Ex1_Generate_Random_Numbers.sas;
2  *Old Way;
3  Data Have1;
4        do i = 1 to 10;
5        x = rannor(12345);
6        output;
7        end;
8  run;
9
10 *New Way;
11 Data Have2;
12 call streaminit(12345);
13       do i = 1 to 10;
14       x = rand("Normal",0,1);
15       output;
16       end;
17 run;
18 proc print data=Have2; run;
```

Line 12: The routine is used to set the seed for the random number stream.

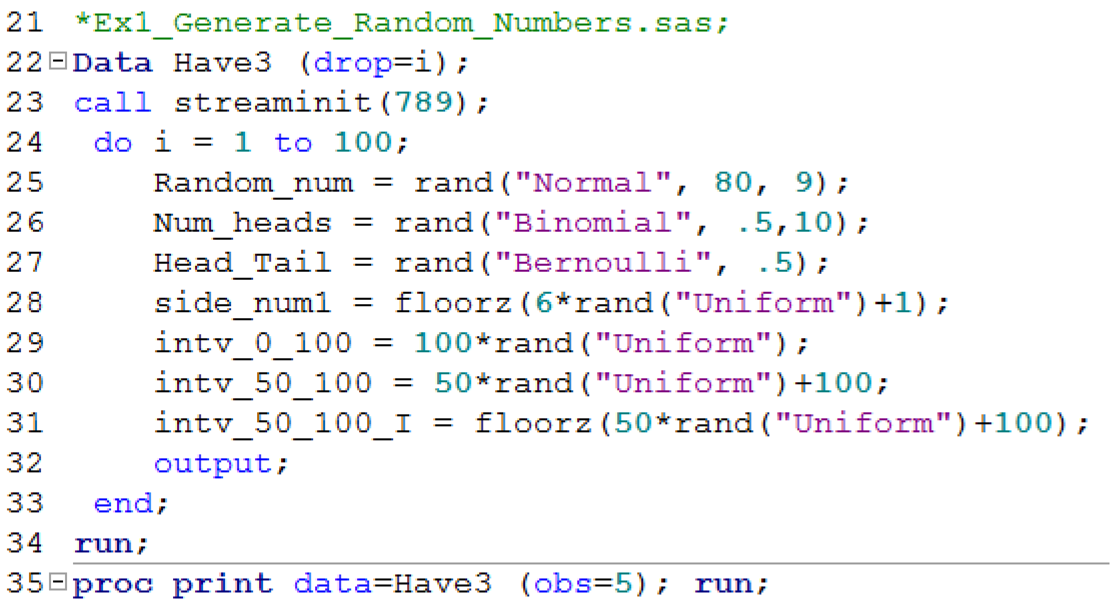


```
Obs     i          x

  1     1     0.26423
  2     2     1.07473
  3     3     0.81792
  4     4    -0.55277
  5     5     1.54014
  6     6    -1.23382
  7     7    -0.14154
  8     8     1.04200
  9     9     0.06573
 10    10     1.22526
```

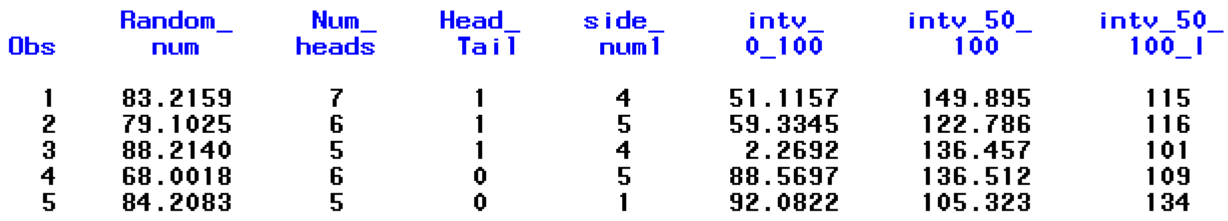
Acknowledgements: Texts are primarily adapted/obtained from the SAS Blogs (The DO LOOP by Rick Wicklin) and other web sources. Portions of SAS' copyrighted SAS course content are reproduced here with permission of SAS Institute Inc., Cary, NC, USA. SAS Institute Inc. makes no warranties with respect to these materials and disclaims all liability therefor. Do not copy or circulate the handouts.

The following DATA step simulates 100 independent values from various distributions. The subsequent PROC PRINT step displays the first five observations.



```
21 *Ex1_Generate_Random_Numbers.sas;
22 Data Have3 (drop=i);
23 call streaminit(789);
24  do i = 1 to 100;
25     Random_num = rand("Normal", 80, 9);
26     Num_heads = rand("Binomial", .5,10);
27     Head_Tail = rand("Bernoulli", .5);
28     side_num1 = floorz(6*rand("Uniform")+1);
29     intv_0_100 = 100*rand("Uniform");
30     intv_50_100 = 50*rand("Uniform")+100;
31     intv_50_100_I = floorz(50*rand("Uniform")+100);
32     output;
33  end;
34 run;
35 proc print data=Have3 (obs=5); run;
```

Line 23: This routine sets the seed value.
Line 24: The DO loop iterates 100 times.
Line 25: Normal distribution with mean = 80 and standard deviation = 9.
Line 26: Binomial distribution with p (a numeric probability of success) = .5 and n (an integer parameter that counts the number of independent Bernoulli trials) = 10.
Line 27: Bernoulli distribution with p = .5.
Line 28: The Floorz converts the result to integers.



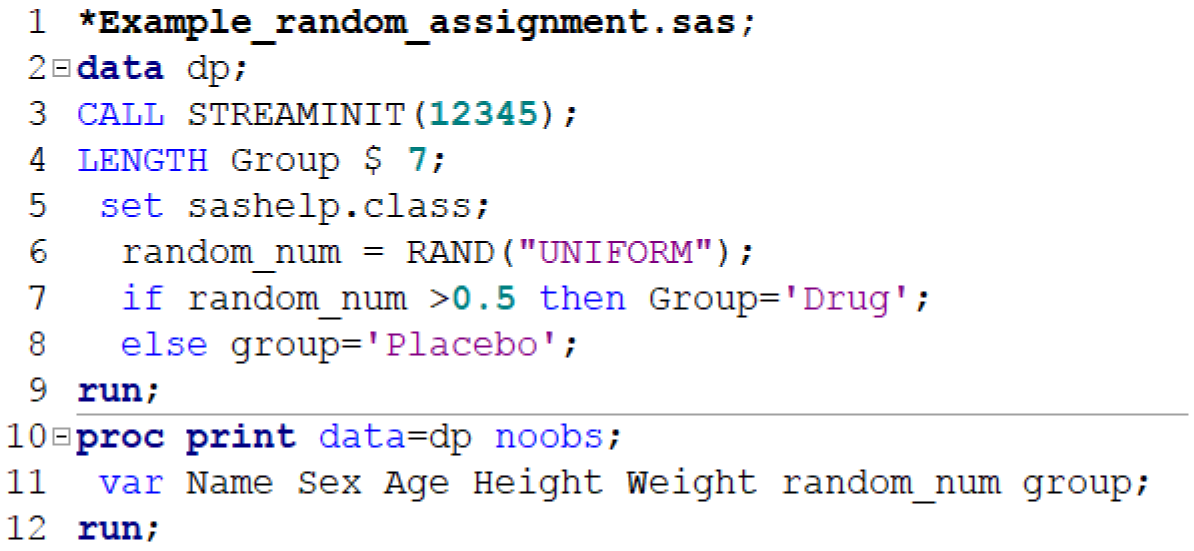
```
         Random_     Num_    Head_    side_     intv_      intv_50_    intv_50_
  Obs      num       heads    Tail     num1     0_100        100        100_I

    1    83.2159       7        1        4     51.1157     149.895       115
    2    79.1025       6        1        5     59.3345     122.786       116
    3    88.2140       5        1        4      2.2692     136.457       101
    4    68.0018       6        0        5     88.5697     136.512       109
    5    84.2083       5        0        1     92.0822     105.323       134
```

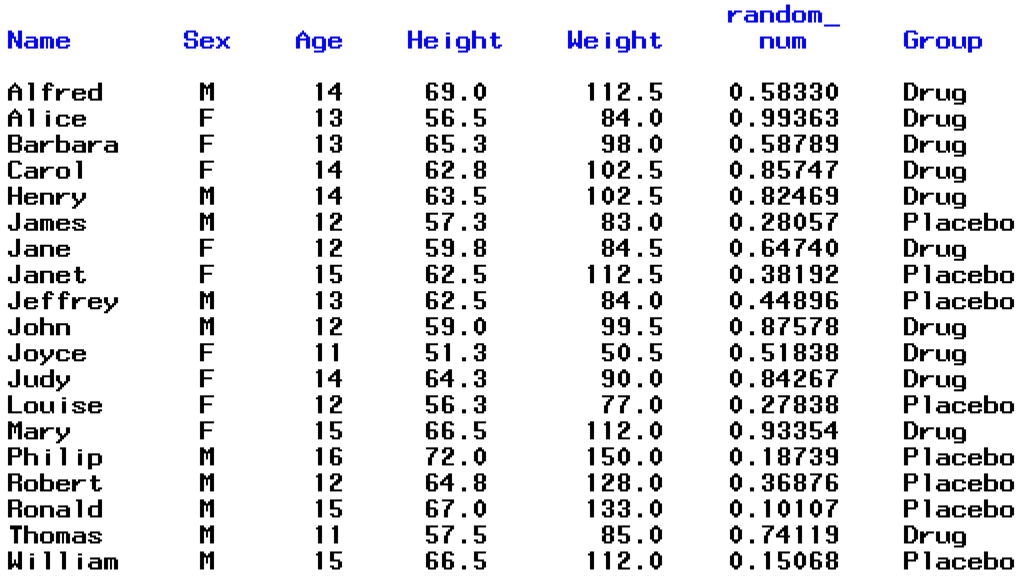
Some Other Distributions: Geometric, Poisson, and Exponential.

Acknowledgements: Texts are primarily adapted/obtained from the SAS Blogs (The DO LOOP by Rick Wicklin) and other web sources. Portions of SAS' copyrighted SAS course content are reproduced here with permission of SAS Institute Inc., Cary, NC, USA. SAS Institute Inc. makes no warranties with respect to these materials and disclaims all liability therefor.   Do not copy or circulate the handouts.

Assigning 19 Subjects in the data set SASHELP.CLASS to the Drug or Placebo Randomly



```
*Example_random_assignment.sas;
data dp;
CALL STREAMINIT(12345);
LENGTH Group $ 7;
 set sashelp.class;
  random_num = RAND("UNIFORM");
  if random_num >0.5 then Group='Drug';
  else group='Placebo';
run;
proc print data=dp noobs;
 var Name Sex Age Height Weight random_num group;
run;
```

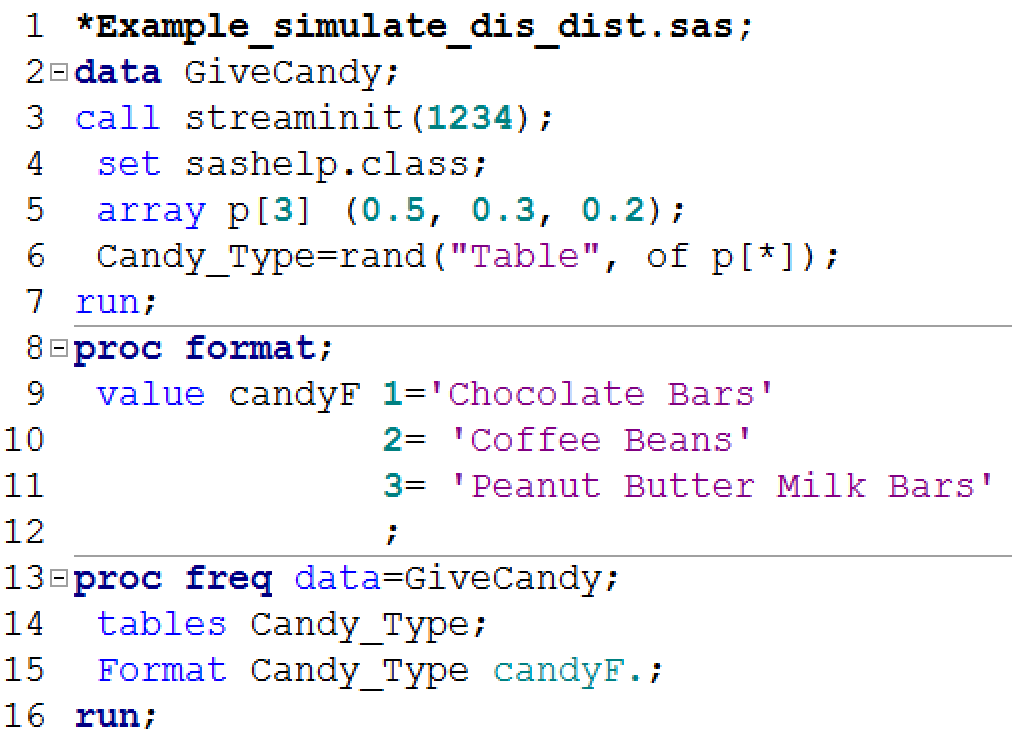


```
                                                       random_
Name         Sex     Age     Height      Weight        num       Group

Alfred        M       14      69.0        112.5      0.58330     Drug
Alice         F       13      56.5         84.0      0.99363     Drug
Barbara       F       13      65.3         98.0      0.58789     Drug
Carol         F       14      62.8        102.5      0.85747     Drug
Henry         M       14      63.5        102.5      0.82469     Drug
James         M       12      57.3         83.0      0.28057     Placebo
Jane          F       12      59.8         84.5      0.64740     Drug
Janet         F       15      62.5        112.5      0.38192     Placebo
Jeffrey       M       13      62.5         84.0      0.44896     Placebo
John          M       12      59.0         99.5      0.87578     Drug
Joyce         F       11      51.3         50.5      0.51838     Drug
Judy          F       14      64.3         90.0      0.84267     Drug
Louise        F       12      56.3         77.0      0.27838     Placebo
Mary          F       15      66.5        112.0      0.93354     Drug
Philip        M       16      72.0        150.0      0.18739     Placebo
Robert        M       12      64.8        128.0      0.36876     Placebo
Ronald        M       15      67.0        133.0      0.10107     Placebo
Thomas        M       11      57.5         85.0      0.74119     Drug
William       M       15      66.5        112.0      0.15068     Placebo
```

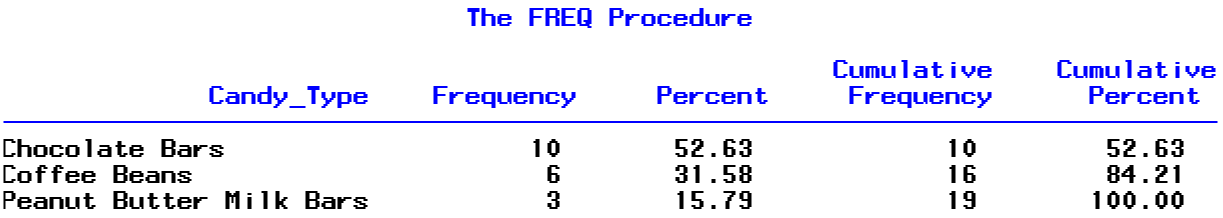
Acknowledgements: Texts are primarily adapted/obtained from the SAS Blogs (The DO LOOP by Rick Wicklin) and other web sources. Portions of SAS' copyrighted SAS course content are reproduced here with permission of SAS Institute Inc., Cary, NC, USA. SAS Institute Inc. makes no warranties with respect to these materials and disclaims all liability therefor. Do not copy or circulate the handouts.

Simulating Data from a Discrete Distribution (See Wicklin, 2015)



```
 1  *Example_simulate_dis_dist.sas;
 2 data GiveCandy;
 3  call streaminit(1234);
 4   set sashelp.class;
 5   array p[3] (0.5, 0.3, 0.2);
 6   Candy_Type=rand("Table", of p[*]);
 7  run;
 8 proc format;
 9   value candyF 1='Chocolate Bars'
10                2= 'Coffee Beans'
11                3= 'Peanut Butter Milk Bars'
12                ;
13 proc freq data=GiveCandy;
14   tables Candy_Type;
15   Format Candy_Type candyF.;
16  run;
```

Line 6: The RAND function uses a "Table" distribution to specify a table of probabilities for each outcome.



```
                          The FREQ Procedure

                                                      Cumulative    Cumulative
             Candy_Type    Frequency     Percent       Frequency       Percent
------------------------------------------------------------------------------
Chocolate Bars                    10       52.63              10         52.63
Coffee Beans                       6       31.58              16         84.21
Peanut Butter Milk Bars            3       15.79              19        100.00
```

Acknowledgements: Texts are primarily adapted/obtained from the SAS Blogs (The DO LOOP by Rick Wicklin) and other web sources. Portions of SAS' copyrighted SAS course content are reproduced here with permission of SAS Institute Inc., Cary, NC, USA. SAS Institute Inc. makes no warranties with respect to these materials and disclaims all liability therefor. Do not copy or circulate the handouts.

Calling R in a SAS Session with PROC IML

Requirements:

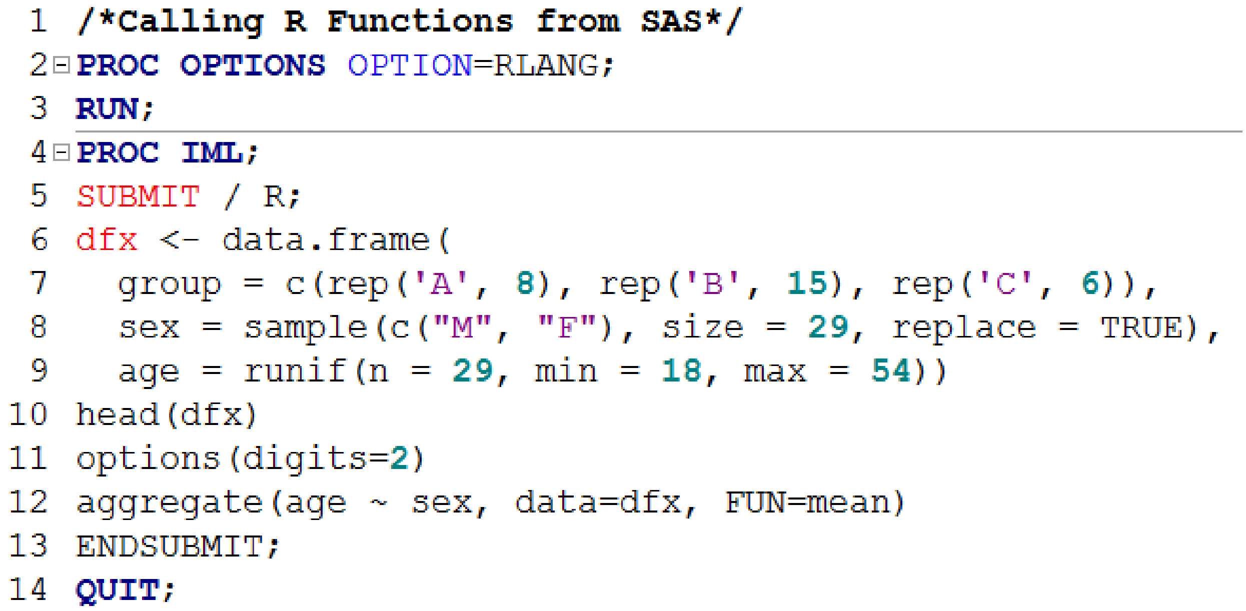
- Both R and SAS software packages must be installed
- The RLANG system option in SAS must also be enabled.

Run the following code in SAS to check whether the permission to use R within SAS is enabled.

```
PROC OPTIONS OPTION=RLANG; RUN;
```

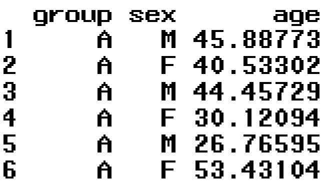
If the permission to use R within SAS is not enabled, you must modify the sasv9.cfg file, which is typically found in "C:\Program Files\SASHome\SASFoundation\9.3\nls\en\”, in order to reset this option by adding a –RLANG line to the config file.

Acknowledgements: Texts are primarily adapted/obtained from the SAS Blogs (The DO LOOP by Rick Wicklin) and other web sources. Portions of SAS’ copyrighted SAS course content are reproduced here with permission of SAS Institute Inc., Cary, NC, USA. SAS Institute Inc. makes no warranties with respect to these materials and disclaims all liability therefor. Do not copy or circulate the handouts.



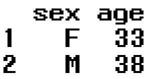
```
 1 /*Calling R Functions from SAS*/
 2 PROC OPTIONS OPTION=RLANG;
 3 RUN;
 4 PROC IML;
 5 SUBMIT / R;
 6 dfx <- data.frame(
 7   group = c(rep('A', 8), rep('B', 15), rep('C', 6)),
 8   sex = sample(c("M", "F"), size = 29, replace = TRUE),
 9   age = runif(n = 29, min = 18, max = 54))
10 head(dfx)
11 options(digits=2)
12 aggregate(age ~ sex, data=dfx, FUN=mean)
13 ENDSUBMIT;
14 QUIT;
```

Output from line 11 (Listing of 6 observations)



```
  group sex      age
1     A   M 45.88773
2     A   F 40.53302
3     A   M 44.45729
4     A   F 30.12094
5     A   M 26.76595
6     A   F 53.43104
```

Output from Lines 11-12



```
  sex age
1   F  33
2   M  38
```

Acknowledgements: Texts are primarily adapted/obtained from the SAS Blogs (The DO LOOP by Rick Wicklin) and other web sources. Portions of SAS' copyrighted SAS course content are reproduced here with permission of SAS Institute Inc., Cary, NC, USA. SAS Institute Inc. makes no warranties with respect to these materials and disclaims all liability therefor. Do not copy or circulate the handouts.

```
 [1] "./Ex1_Generate_Random_Numbers.sas"
 [2] "./Ex10_simulate.sas"
 [3] "./Ex11_simulate.sas"
 [4] "./Ex12_simulate.sas"
 [5] "./Ex13_simulate.sas"
 [6] "./Ex14_data_driven_simu.sas"
 [7] "./Ex15_calling_R_from_SAS.sas"
 [8] "./Ex16_calling_R_from_SAS.sas"
 [9] "./Ex17_calling_R_from_SAS.sas"
[10] "./Ex18_collapse_multiple_cols.sas"
[11] "./Ex19_count_repeated_rows.sas"
[12] "./Ex2_DoLoop_Mod_Func.sas"
[13] "./Ex20_macro_case_study.sas"
[14] "./Ex21_Load.sas"
[15] "./Ex22_R_List_of_Files.sas"
[16] "./Ex3_call_symputx_dates.sas"
[17] "./Ex4_generate_dates_sashelp_class.sas"
[18] "./Ex5_generate_SSN_dates_other_vars.sas"
[19] "./Ex6_random_assignment.sas"
[20] "./Ex7_simualte_randomID.sas"
[21] "./Ex8_simulate.sas"
[22] "./Ex9_simulate.sas"
```

Acknowledgements: Texts are primarily adapted/obtained from the SAS Blogs (The DO LOOP by Rick Wicklin) and other web sources. Portions of SAS' copyrighted SAS course content are reproduced here with permission of SAS Institute Inc., Cary, NC, USA. SAS Institute Inc. makes no warranties with respect to these materials and disclaims all liability therefor. Do not copy or circulate the handouts.