STATISTICAL PROGRAMMING FOR BUSINESS ANALYTICS

ASSIGNMENT NO.9

TANAY BHALERAO

U47707491

APRIL 17, 2015

UNIVERSITY OF SOUTH FLORIDA
Management Information System

Homework for Chapter 12

1. Refer to the USEDCARS dataset. Suppose the future record will always be in this format. Write a macro that will summarize the used cars for a selected manufacturer. We need the number of cars (sample size), the mean price, the median price and the standard deviation of the sample in the output data set. The arguments are input date set name, manufacturer and output data set name. Test your macro by the USEDCARS dataset for Ford cars. Print the information in the output dataset.

```
%MACRO MFG (AIN = , BIN = , CIN = );
PROC SORT DATA = &AIN;
BY MANUF;
RUN;

DATA TEMP_D;
SET &AIN;
BY MANUFACTURER;
IF MANUFACTURER = &BIN;
RUN;
PROC MEANS NOPRINT DATA = TEMP_D;
VAR PRICE;
BY MANUFACTURER;
OUTPUT OUT = &CIN MEAN = MEAN_C MEDIAN = MEDIAN_C STD = STD_DEV_C;
RUN;
%MEND;

DATA USEDCARS;
INFILE
"\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\usedcars.txt" DLM =
"09"X FIRSTOBS = 2 OBS = 50;
INPUT YEAR 1-2  @@ MANUFACTURER $ 9-18 MODEL $ 24-34 @38 MILES comma6. @50
PRICE comma6. DEALER $ 61-83;
RUN;

%MFG(AIN = USEDCARS, BIN = 'Ford', CIN = AGGR_MFG);
RUN;

PROC PRINT DATA = AGGR_MFG;
VAR MANUFACTURER _FREQ_ MEAN_C MEDIAN_C STD_DEV_C;
RUN;
```

SELECTIVE CAR STATS

Obs	MANUFACTURER	_FREQ_	MEDIAN_C	MEAN_C	STD_DEV_C
1	Ford	11	9994	9595.82	4010.86

2. Write a macro that can take a random sample size of n from an input data set. The arguments for the macro should be the dataset name, sample size, seed for random numbers, and the output file name. Test your macro by using the USEDCARS dataset. Use two seeds to take two

random samples of size 10 and print your sample in the original data format. (Hint: Attach each datum with a random number, delete the datum if its rank is larger than the sample size.)

```
%MACRO RAND_DAT(INPUT_D = ,DATA_COUNT = , SEED = , OUT_SAMPLE = );
DATA RAN;
DO INDEX = 1 TO 100;
RAN_NO = RANUNI(&SEED);
OUTPUT;
END;
DROP INDEX;
RUN;

DATA DATA_TEMP;
MERGE &INPUT_D (IN = LM) RAN;
IF LM;
RUN;

PROC SORT DATA=DATA_TEMP;
BY RAN_NO;
RUN;

DATA &OUT_SAMPLE;
SET DATA_TEMP;
IF _N_ <= &DATA_COUNT;
DROP RANDOM;
RUN;
%MEND;


DATA USEDCARS;
INFILE
"\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\usedcars.txt" DLM =
"09"X FIRSTOBS = 2 OBS = 50;
INPUT YEAR 1-2  @@ MANUF $ 9-18 MODEL $ 24-34 @38 MILES comma6. @50 PRICE
comma6. DEALER $ 61-83;
RUN;
%RAND_DAT(INPUT_D = USEDCARS, DATA_COUNT = 10, SEED = 246, OUT_SAMPLE =
RAN_CARS);
RUN;
PROC PRINT DATA = RAN_CARS;
TITLE "RANDOM CAR SELECTION";
RUN;
%RAND_DAT(INPUT_D = USEDCARS, DATA_COUNT = 10, SEED = 876, OUT_SAMPLE
=R_CARS);
RUN;
PROC PRINT DATA = R_CARS;
TITLE "RANDOM CAR SELECTION II";
RUN;
```

RANDOM CAR SELECTION

Obs	YEAR	MANUF	MODEL	MILES	PRICE	DEALER	RAN_NO
1	97	Honda	Civic	8000	14450	Gainesville Nissan	0.04417
2	90	Audi	100	.	5995	Taylor Volvo	0.05644
3	97	Plymouth	Neon	.	7998	Budget Car Sales	0.06250
4	92	Toyota	Corolla	.	5800	Hometown Motors	0.07854
5	97	Geo	Metro	.	6988	Budget Car Sales	0.09293
6	93	Mazda	Protege	.	5787	Hawes Chrysler Plymouth	0.10972
7	97	Toyota	Camry	.	15998	Budget Car Sales	0.11682
8	95	Volvo	850	.	19995	Taylor Volvo	0.16333
9	94	Oldsmobile	Cutlass	.	7995	Kraft Motorcar	0.17694
10	92	Mitsubishi	Mirage	.	2995	Magic Imports	0.18823

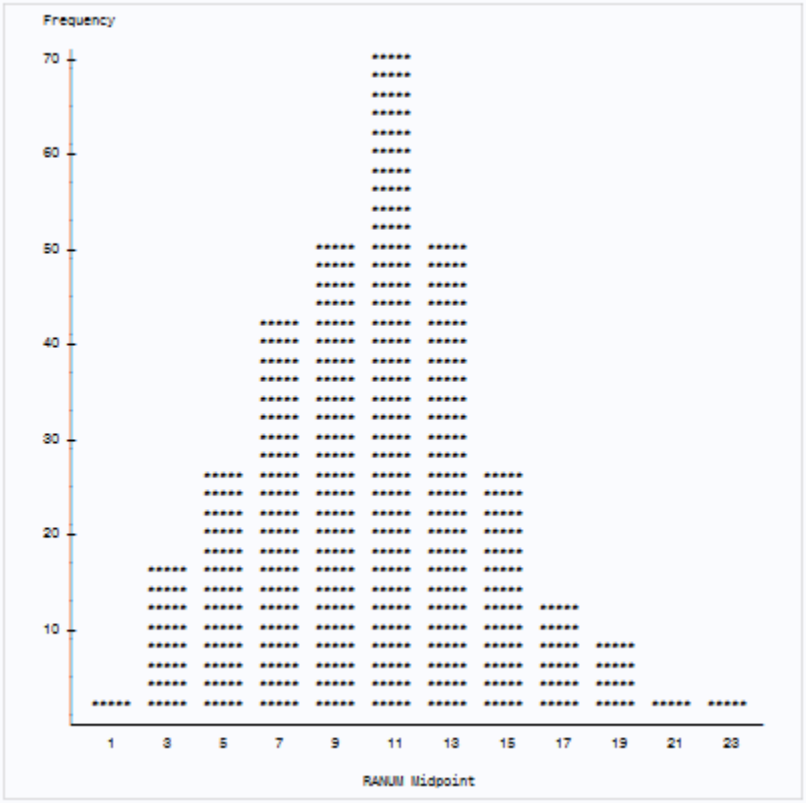
RANDOM CAR SELECTION II

Obs	YEAR	MANUF	MODEL	MILES	PRICE	DEALER	RAN_NO
1	83	Oldsmobile	Cutlass	.	1950	University Auto	0.02721
2	94	Oldsmobile	Cutlass	.	7995	Kraft Motorcar	0.04473
3	90	Mazda	626	.	4998	Gatorland Toyota	0.08696
4	98	GMC	Safari	.	19900	Tomlinson Motor Company	0.08788
5	94	Ford	Ranger	.	9994	White Ford	0.15945
6	94	Toyota	Corolla	.	7665	Saturn of Gainesville	0.20859
7	93	Pontiac	Grand Am	.	5485	Wade Raulerson	0.23321
8	91	Geo	Storm	.	2995	Magic Imports	0.24540
9	96	Buick	LeSabre	25000	16900	Tomlinson Motor Company	0.24609
10	96	Ford	Explorer	32000	17999	Santa Fe Ford	0.25432

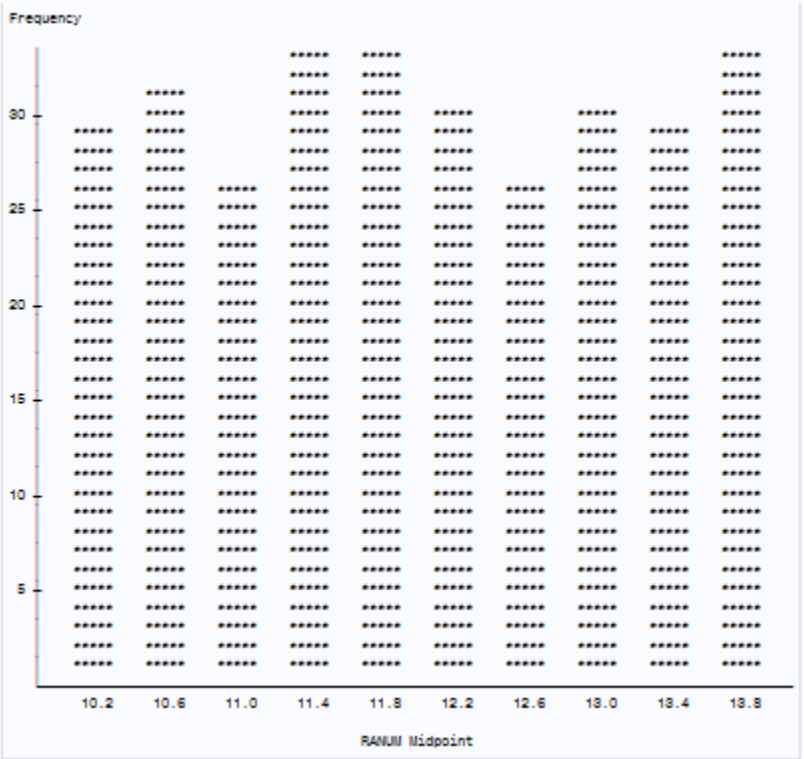
3. Write a macro that can generate a normal or uniform data with any required size. The arguments of the macro should contain the required mean, standard deviation (sd), distribution, seed, and output file name. Test your macro by calling it twice; one for n=300 normal, mean=10, sd=4, and one with 300 uniform with mean=10 and sd=4. Use PROC CHART to plot the histograms of the output file.

```
%MACRO RAND_DIS(COUNT = , DIST_TYPE = , SEED_D =  ,MEAN_D = , STD_DEV_D = ,
OUT = );
DATA &OUT;
IF &DIST_TYPE = "NORM" THEN
DO INDEX = 1 TO &COUNT;
RANUM = RANNOR(&SEED_D)* &STD_DEV_D + &MEAN_D;
OUTPUT;
END;
ELSE IF &DIST_TYPE = "UNI" THEN
DO INDEX = 1 TO &COUNT;
RANUM = RANUNI(&SEED_D)* &STD_DEV_D + &MEAN_D;
OUTPUT;
END;
DROP INDEX;
RUN;
%MEND;

%RAND_DIS(COUNT = 300, DIST_TYPE = "NORM", SEED_D = 999, MEAN_D = 10,
STD_DEV_D = 4,   OUT = NORM_DATA)
RUN;
%RAND_DIS(COUNT = 300, DIST_TYPE = "UNI", SEED_D = 999, MEAN_D = 10,
STD_DEV_D = 4,   OUT = UNI_DATA)
RUN;

PROC CHART DATA=NORM_DATA;
TITLE "NORMAL DISTRIBUTION";
VBAR RANUM;
RUN;

PROC CHART DATA=UNI_DATA;
TITLE "UNIFORM DISTRIBUTION";
VBAR RANUM;
RUN;
```

NORMAL DISTRIBUTION



UNIFORM DISTRIBUTION



4. Chapter 12: 12.14, 12.16, 12.18

12.14

```
DATA LISTER;
INPUT @1 (THREE1-THREE4)(3.)
      @10(TWO1-TWO4)(2.)
      @18(FOUR1-FOUR4)(4.);
DATALINES;
123121234217874444123872345873235432
192837465748392919283747372818182838
;

PROC PRINT DATA=LISTER;
TITLE "LIST";
RUN;
```

LIST

Obs	THREE1	THREE2	THREE3	THREE4	TWO1	TWO2	TWO3	TWO4	FOUR1	FOUR2	FOUR3	FOUR4
1	123	121	234	217	21	78	74	44	4123	8723	4587	3235
2	192	837	465	748	74	83	92	91	9283	7473	7281	8182

12.16

```
DATA SUBJ;
INPUT #1 ID 1-3 GENDER $ 5 @7 DO_B MMDDYY8. #2 HEIGHT 1-2 WEIGHT 4-6;
FORMAT do_b MMDDYY8.;
DATALINES;
001,M,06/14/1944
68,155
002,F,12/25/1967
52,99
003,M,07/04/1983
72,128
;


PROC PRINT DATA=SUBJ NOOBS;

TITLE "SUBJECT";
RUN;
```

SUBJECT

ID	GENDER	DO_B	HEIGHT	WEIGHT
1	M	06/14/19	68	155
2	F	12/25/19	52	99
3	M	07/04/19	72	128

12.18

```
DATA PROBLEM;
INPUT @11 GENDER $ 1.@;
IF GENDER = 'F';
INPUT @1 DATE MMDDYY8. @12 AGE 2. @14 SCORE 3. @;
OUTPUT;
FORMAT DATE MMDDYY8.;

DATALINES;
04/04/2004M15 90
05/12/2004F16 95
07/23/2004M18 88
01/20/2004F17100
;
RUN;
PROC PRINT DATA=PROBLEM;
RUN;
```

The SAS System

Obs	GENDER	DATE	AGE	SCORE
1	F	05/12/20	16	95
2	F	01/20/20	17	100