STATISTICAL PROGRAMMING FOR BUSINESS ANALYTICS

ASSIGNMENT NO. 11

Tanay Bhalerao

U47707491

MAY 1, 2015
UNIVERSITY OF SOUTH FLORIDA
Management Information Systems

Homework 11

1. Use PROC SQL to list all the observations from data set Inventory where Price is greater than 20.

```
libname tan '\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\';
PROC SQL;
select * from tan.inventory where price GT 20;
RUN;
```

The SAS System

Model	Price
M567	$23.50
L776	$159.98
X999	$29.95

2. Use PROC SQL to create a new, temporary, SAS data set (N_Sales) containing the observations from Sales where Region has a value of North. Include only the variables Name and TotalSales in the new data set.

```
PROC SQL;
create table N_sales as select Name,TotalSales from tan.sales where region
like 'North';
RUN;
```

The SAS System

Name	TotalSales
George Smith	449.5
George Smith	15597
George Smith	5129
Glenda Johnson	52.5
George Smith	15597

3. Write the necessary PROC SQL statements to accomplish the same goal as the program here:

```
data third;
 set learn.first learn.second;
run;
```

```
Proc sql;
create table third as
select * from learn.first
union
select * from learn.second
;
      RUN;
```

4. Chapter 15: 15.2, 15.4, 15.8

15.2

```
DATA TEMPERATURE;
INPUT TF1-TF10;
ARRAY TC[10];
ARRAY TF[10];

DO I=1 TO 10;
TC[I]=5/9*(TF[I]-32);
END;
DROP I;
DATALINES;
32 212 -40 10 20 30 40 50 60 70
-10 0 10 20 30 40 50 60 70 80
;
RUN;
PROC PRINT DATA=TEMPERATURE NOOBS;
   TITLE "Listing of Data Set TEMPERATURE";
      RUN;
```

Listing of Data Set TEMPERATURE

TF1	TF2	TF3	TF4	TF5	TF6	TF7	TF8	TF9	TF10	TC1	TC2	TC3	TC4	TC5	TC6	TC7	TC8	TC9	TC10	I
32	212	-40	10	20	30	40	50	60	70	0.0000	100.000	-40.0000	-12.2222	-6.66667	-1.11111	4.4444	10.0000	15.5556	21.1111	11
-10	0	10	20	30	40	50	60	70	80	-23.3333	-17.778	-12.2222	-6.6667	-1.11111	4.44444	10.0000	15.5556	21.1111	26.6667	11

15.4

```
DATA MIXED;
INFORMAT A1-A3 B C $5.;
INPUT X1-X3 Y Z A1-A3 B C;
ARRAY NUMV[*] X1-X3 Y Z;
ARRAY CHARV[*]  A1-A3 B C ;
ARRAY LOGNUM[*] LX1-LX3 LY LZ;

DO I=1 TO DIM(NUMV);
LOGNUM[I]=LOG(NUMV[I]);
END;

DO T=lbound(CHARV) TO hbound(CHARV);
IF CHARV[T] EQ '?' THEN CHARV[T] = ' ';
END;

DROP I T;
DATALINES;
10 20 30 40 50 ONE TWO THREE ? ?
```

```
11 22 33 44 55 ? LLL MMM ? VVV
;
PROC PRINT DATA=MIXED NOOBS;
    TITLE "Listing of Data Set MIXED";
RUN;
```

Listing of Data Set MIXED

A1	A2	A3	B	C	X1	X2	X3	Y	Z	LX1	LX2	LX3	LY	LZ
ONE	TWO	THREE			10	20	30	40	50	2.30259	2.99573	3.40120	3.68888	3.91202
	LLL	MMM		VVV	11	22	33	44	55	2.39790	3.09104	3.49651	3.78419	4.00733

15.8

```
DATA EXPER;
    INPUT TIME0-TIME4;
DATALINES;
100 200 300 400 500
55 110 130 150 170
;

DATA EXPERIMENT;
SET EXPER;
ARRAY TIMES[*] TIME0-TIME4;
ARRAY MIN[*] MIN0-MIN4;
DO I=1 TO DIM(MIN);
MIN[I]=ROUND((TIMES[I]/60),0.1);
END;
DROP I TIME0-TIME4;
RUN;

PROC PRINT DATA= EXPERIMENT;
TITLE "SECOND-MINUTE CONVERSION";
        RUN;
```

SECOND-MINUTE CONVERSION

Obs	MIN0	MIN1	MIN2	MIN3	MIN4
1	1.7	3.3	5.0	6.7	8.3
2	0.9	1.8	2.2	2.5	2.8

5. Chapter 16: 16.2 and 16.4

    16.2

```
DATA QUES;
    INPUT ID $ REASON1-REASON4;
DATALINES;
```

```
001 3 6 13 17
002 8 3 4 .
003 20 2 . .
004 8 4 20 19
;

DATA QUESTION_FREQ;
SET QUES;

ARRAY REASON(001:004) REASON1-REASON4;
DO I = 001 to 004 ;
    REASON_T = REASON(I);
    OUTPUT;
  END;

  DROP REASON1-REASON4 I;
RUN;

PROC PRINT DATA=QUESTION_FREQ;
TITLE "TRANSPOSE";
RUN;

PROC FREQ DATA=QUESTION_FREQ;
TITLE "FREQ";
TABLES REASON_T;
      RUN;
```

TRANSPOSE

Obs	ID	REASON_T
1	001	3
2	001	6
3	001	13
4	001	17
5	002	8
6	002	3
7	002	4
8	002	.
9	003	20
10	003	2
11	003	.
12	003	.
13	004	8
14	004	4
15	004	20
16	004	19

FREQ

The FREQ Procedure

REASON_T	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2	1	7.69	1	7.69
3	2	15.38	3	23.08
4	2	15.38	5	38.46
6	1	7.69	6	46.15
8	2	15.38	8	61.54
13	1	7.69	9	69.23
17	1	7.69	10	76.92
19	1	7.69	11	84.62
20	2	15.38	13	100.00

Frequency Missing = 3

16.4

```sas
DATA THIN;
    INPUT ID $ TIME X @@;
DATALINES;
001 1 10   001 2 12   001 3 15
004 1 17
003 1 14   003 2 18   003 3 22   003 4 28
002 1 18   004 2 28
;

PROC PRINT DATA=THIN;
RUN;

PROC SORT DATA=THIN OUT=THINSORT ;
  BY ID ;
RUN ;

DATA WIDE1 ;
  SET THINSORT ;
  BY ID ;

  KEEP ID X1-X4;
  RETAIN X1-X4;

  ARRAY AX[4] X1 - X4 ;

  IF first.ID THEN
  DO;
    DO i = 1 to 4 ;
      AX[i] = . ;
    END;
  END;
  AX[TIME] = X ;

  IF last.ID THEN OUTPUT ;

RUN;

PROC PRINT DATA=WIDE1 NOOBS;
TITLE "WIDE DATA";
    RUN;
```

WIDE DATA

ID	X1	X2	X3	X4
001	10	12	15	.
002	18	.	.	.
003	14	18	22	28
004	17	28	.	.

6. Chapter 17: 17.10

```sas
DATA NUM_CHAR;
      INPUT X $ Y $ Z $ DATE : $10. NUMERAL DOB : DATE9.;
      FORMAT DOB MMDDYY10.;
DATALINES;
10 20 30 10/21/1946 123 09SEP2004
1 2 3 11/11/2004 999 01JAN1960
;
Data Correct;
Set Num_Char;
X1 = Input(X, 8.);
Y1 = Input (Y, 8.);
Z1 = Input (Z, 8.);
Date_n= Input(Date, MMDDYY10.);
Numeral_n = Put(Numeral, 8.);
DOB_char = Put(DOB, MMDDYY10.);
Drop X Y Z Date Numeral DOB;
Rename X1=X Y1=Y Z1=Z Date_n=Date Numeral_n=Numeral;
Format Date_n Date9.;
run;
Proc Print Data=Correct noobs;
Title "Formatting Correct";
run;
```

Formatting Correct

X	Y	Z	Date	Numeral	DOB_char
10	20	30	21OCT1946	123	09/09/2004
1	2	3	11NOV2004	999	01/01/1960

7. Chapter 18: 18.10, 18.12 (a) only

18.10

```sas
DATA WEIGHT;
INPUT SUBJ $ WEIGHT $ @@;
WEIGHT = UPCASE(WEIGHT);
IF INDEX(WEIGHT,'K') GE 1 THEN WEIGHT = INT(COMPRESS(WEIGHT, 'KGLBS.'))*2.2;
ELSE WEIGHT = INT(COMPRESS(WEIGHT, 'KGLBS.'));
DATALINES;
1 50KG 2 120 3 121LBS. 4 88KG. 5 200
6 80KG 7 250LB
;

PROC PRINT DATA = WEIGHT;
TITLE "WEIGHT";
      RUN;
```

WEIGHT

Obs	subj	weight
1	1	110
2	2	120
3	3	121
4	4	193.6
5	5	200
6	6	176
7	7	250

18.12

```
data INDEXES;
Length INPUTL $32;
INFILE DATALINES DSD;
input INPUTL $;
NUMBER = INT(SUBSTR(INPUTL,(INDEX(INPUTL,'XYZ')+ 3), 2));
datalines;
ABC123XYZ7823
NONE HERE
XYZ12345
12345XYZ9876
;
proc print data = INDEXES;
TITLE "INDEXES";
    run;
```

INDEXES

Obs	INPUTL	NUMBER
1	ABC123XYZ7823	78
2	NONE HERE	.
3	XYZ12345	12
4	12345XYZ9876	98