

STATISTICAL PROGRAMMING FOR BUSINESS ANALYTICS

Assignment 10



Tanay Bhalerao

U47707491

APRIL 24, 2015

UNIVERSITY OF SOUTH FLORIDA
Management Information System

Homework 10

1. Refer to the AIRPORTS dataset. Write a SAS program to read the data file from your diskette with an INFILE statement, store a permanent SAS dataset on your diskette, and print the dataset. There should be five variables (three character and two numeric) and 20 observations.

```
LIBNAME TAN '\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\';
DATA TAN.AIRPORTS;
LENGTH CITY STATE ABBREV $32;
INFILE '\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\airport.CSV'
FIRSTOBS = 2 DLM = "," DSD MISSEVER;
INPUT CITY $ STATE $ ABBREV $ PASS95 PASS85;
PROC PRINT DATA = HW1.AIRPORTS;
TITLE "AIRPORTS";
RUN;
```

AIRPORTS						
Obs	CITY	STATE	ABBREV	ABREV	PASS95	PASS85
1	Chicago	IL		ORD	29885987	21510371
2	Atlanta	GA		ATL	27556894	20678095
3	Dallas/Ft. Worth	TX		DFW	25963950	17715224
4	Los Angeles	CA		LAX	21072273	15957127
5	San Francisco	CA		SFO	15013265	10948098
6	Denver	CO		DEN	14328457	13862996
7	Phoenix	AZ		PHX	13557883	6713293
8	Detroit	MI		DTW	13293568	7163840
9	St. Louis	MO		STL	12736080	9555195
10	Las Vegas	NV		LAS	12657051	4627078
11	Miami	FL		MIA	12030812	7717685
12	Newark	NJ		EWK	11899633	14272558
13	Minneapolis/St. Paul MN	MN		MSP	11835783	7250302
14	Houston	TX		IAH	10950826	6307582
15	Seattle/Tacoma	WA		SEA	10731233	5709488
16	Boston	MA		BOS	10507611	9112901
17	New York	NY		LGA	9682171	9613913
18	Charlotte	NC		CLT	9588900	5102703
19	New York	NY		JFK	9283314	10052007
20	Pittsburgh	PA		PIT	9209903	7002343

2. Refer to the RYAN dataset. Write a SAS program to read the data file from your USB with an INFILE statement and print the dataset. There should be four variables (one character and three numeric) and 21 observations.

```

DATA RYAN;
LENGTH TITLE $24;
INFILE '\\Client\C$\Users\tanay\Documents\Sem2\BusinessAnalytics\RYAN.TXT'
FIRSTOBS =2 DLM ='09'X MISOVER;
INPUT TITLE $ ROGER NYTIMES USAT;
RUN;
PROC PRINT DATA = RYAN;
TITLE "RYAN";
RUN;

```

RYAN				
Obs	TITLE	ROGER	NYTIMES	USAT
1	Addicted to Love	5	3	7
2	Amityville 3: The Demon	.	.	.
3	Armed and Dangerous	.	.	.
4	City of Angels	8	3	5
5	Courage Under Fire	9	.	9
6	D.O.A.	.	.	.
7	Flesh and Bone	.	.	.
8	French Kiss	5	3	6
9	I.Q.	.	.	.
10	Innerspace	.	.	.
11	Joe Versus the Volcano	.	.	.
12	Prelude to a Kiss	.	2	.
13	Promised Land	.	.	.
14	Restoration	.	8	.
15	Rich and Famous	.	.	.
16	Sleepless in Seattle	.	8	.
17	The Doors	.	8	.
18	The Presidio	.	.	.
19	Top Gun	.	.	.
20	When a Man Loves a Woman	.	3	.
21	When Harry Met Sally	.	.	.

- Refer to the HOCKEY dataset. Write a SAS program to read the data file from your USB with an INFILE statement and print dataset. Combine the month, day, and year into one variable representing the date. There should be six variables (three character and three numeric) and 37 observations.

```

LIBNAME TAN '\\Client\C$\Users\tanay\Documents\Sem2\BusinessAnalytics\';
FILENAME
 '\\Client\C$\Users\tanay\Documents\Sem2\BusinessAnalytics\hockey.csv';
DATA TAN.HOCKEY;
LENGTH TEAM CITY STATE $34;
INFILE HOCKEY FIRSTOBS=2 DLM=",";
INPUT DATE : MMDDYY10. TEAM $ CITY $ STATE $ OSU OPP;
FORMAT DATE MMDDYY10.;

```

```

RUN;
PROC PRINT DATA=TAN.HOCKEY;
TITLE "HOCKEY";
RUN;

```

(Only 36 observations were present in the dataset)

HOCKEY						
Obs	TEAM	CITY	STATE	DATE	OSU	OPP
1	Toronto	Columbus	Ohio	10/10/1997	5	0
2	Miami	Oxford	Ohio	10/18/1997	0	3
3	Merrimack	Columbus	Ohio	10/24/1997	2	7
4	Merrimack	Columbus	Ohio	10/26/1997	5	3
5	Clarks on	Pots dam	New York	10/31/1997	1	1
6	Clarks on	Pots dam	New York	11/01/1997	6	2
7	Western Michigan	Columbus	Ohio	11/07/1997	1	3
8	Notre Dame	Columbus	Ohio	11/08/1997	3	2
9	Michigan State	Columbus	Ohio	11/21/1997	2	1
10	Michigan	Columbus	Ohio	11/23/1997	3	2
11	Northern Michigan	Marquette	Michigan	11/28/1997	5	1
12	Northern Michigan	Marquette	Michigan	11/29/1997	5	4
13	Alask a-Fairbanks	Columbus	Ohio	12/05/1997	8	3
14	Alask a-Fairbanks	Columbus	Ohio	12/06/1997	4	0
15	Lak e Superior	Saulte Ste. Marie	Michigan	12/12/1997	4	3
16	Lak e Superior	Saulte Ste. Marie	Michigan	12/13/1997	4	2
17	Michigan	Ann Arbor	Michigan	01/02/1998	4	2
18	Michigan	Ann Arbor	Michigan	01/03/1998	6	0
19	Lak e Superior	Columbus	Ohio	01/09/1998	7	0
20	Ferris State	Columbus	Ohio	01/10/1998	5	3
21	Bowling Green	Columbus	Ohio	01/18/1998	4	2
22	Northern Michigan	Columbus	Ohio	01/24/1998	2	0
23	Notre Dame	Columbus	Ohio	01/25/1998	5	3
24	Western Michigan	Kalamazoo	Michigan	01/30/1998	4	2
25	Michigan State	Columbus	Ohio	02/06/1998	4	2

26	Alaska-Fairbanks	Columbus	Ohio	02/07/1998	4	4
27	Notre Dame	South Bend	Indiana	02/13/1998	5	3
28	Michigan State	East Lansing	Michigan	02/14/1998	4	1
29	Miami	Columbus	Ohio	02/26/1998	5	2
30	Lake Superior	Columbus	Ohio	03/13/1998	2	1
31	Lake Superior	Columbus	Ohio	03/14/1998	6	0
32	Michigan	Detroit	Michigan	03/20/1998	4	2
33	Michigan State	Detroit	Michigan	03/21/1998	3	2
34	Yale	Ann Arbor	Michigan	03/27/1998	4	0
35	Michigan State	Ann Arbor	Michigan	03/28/1998	4	3
36	Boston College	Boston	Massachusetts	04/02/1998	4	1

- Refer to the HANKS dataset. Write a SAS program to read the data file from your USB with an INFILE statement and print the dataset. There should be 11 variables (two character and nine numeric) and 22 observations.

```

DATA HANKS;
INFILE "\\Client\C$\Users\Tanay\Documents\Sem2\BusinessAnalytics\HANKS.TXT"
FIRSTOBS = 2;
INPUT TITLE $ 1-25 YEAR 26-29 LENGTH 34-36 MPAA $ 42-46 ACTION 50-51 DRAMA
58-59 HUMOR 66-67 SEX 74 VOIL 82-83 SUSP 90 OFFB 98;
RUN;
PROC PRINT DATA = HANKS;
TITLE "HANKS";
RUN;

```

HANKS											
Obs	TITLE	YEAR	LENGTH	MPAA	ACTION	DRAMA	HUMOR	SEX	VOIL	SUSP	OFFB
1	Saving Private Ryan	1998	160	R	10	10	4	0	10	8	4
2	That Thing You Do!	1998	110	PG	3	6	4	1	1	2	0
3	Apollo 13	1995	140	PG	6	8	4	0	2	9	2
4	Toy Story	1995	80	G	8	2	10	0	1	5	3
5	Forrest Gump	1994	135	PG-13	4	8	7	2	3	2	2
6	Sleepless in Seattle	1993	100	PG	0	5	7	0	0	4	0
7	Philadelphia	1993	119	PG-13	0	8	2	1	0	3	2
8	A League of Their Own	1992	128	PG	5	6	7	3	1	2	5
9	Joe Versus the Volcano	1990	106	PG	5	5	7	2	3	5	7
10	Bonfire of the Vanities	1990	125	R
11	The 'Burbs	1989	101	PG	3	1	5	1	3	2	2
12	Turner and Hoock	1989	97	PG	4	3	5	1	3	3	2
13	Big	1988	102	PG	3	6	8	2	1	3	2
14	Punchline	1988	100	R	1	7	6	2	0	3	4
15	Dragnet	1987	106	PG-13	4	1	5	3	3	3	2
16	The Money Pit	1986	91	PG	5	2	6	1	1	2	0
17	Nothing in Common	1986	119	PG	2	7	6	7	3	4	2
18	Every Time We Say Goodbye	1986	97	R	4	5	3	3	2	4	2
19	The Man With One Red Shoe	1985	92	PG	6	5	6	4	3	4	4
20	Volunteers	1985	107	R	5	2	5	2	2	2	3
21	Bachelor Party	1984	111	R	3	1	6	7	2	2	2
22	Splash	1984	111	PG	3	4	8	2	1	4	2

5. Chapter 13: 13.6 and 13.8

13.6

```
DATA BIG;
DO I = 1 TO 1000000;
X = RANUNI(0);
OUTPUT;
END;
DROP I;
RUN;
DATA SMALL;
X = .5;
OUTPUT;
X = .6;
OUTPUT;
RUN;
DATA MERGE_D;
SET BIG SMALL;
RUN;
PROC APPEND BASE = BIG DATA = SMALL;
RUN;
```

The time used by the set statement:

```
DATA statement used (Total process time):
real time          0.21 seconds
cpu time           0.21 seconds
```

The time used by the append statement:

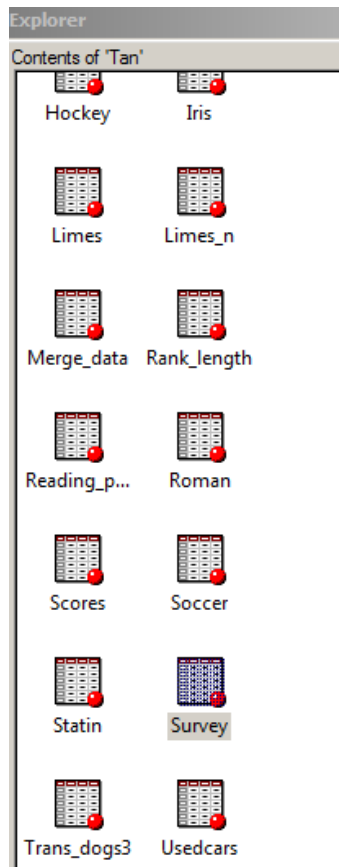
```
PROCEDURE APPEND used (Total process time):
real time          0.03 seconds
cpu time           0.03 seconds
```

The time taken by Append is less than Set Statement. This happens because executing the SET statement, SAS has to physically read in both datasets in order to create the output dataset.

13.8

```
LIBNAME TAN '\\Client\C$\Users\tanay\Documents\Sem2\BusinessAnalytics\';
OPTIONS FMTSEARCH = (TAN);
PROC FORMAT LIBRARY = TAN;
VALUE $FORMAT_SEX M = 'MALE'
F = 'FEMALE';
VALUE $FORMAT_Q '1' = 'YES'
'0' = 'NO';
RUN;
DATA TAN.SURVEY;
INFILE '\\Client\C$\Users\tanay\Documents\Sem2\BusinessAnalytics\survey.DAT'
FIRSTOBS=2;
INPUT GENDER $ Q1 $ Q2 $ Q3 $ Q4 $ Q5 $;
FORMAT GENDER $FORMAT_SEX.
Q1 $FORMAT_Q.
```

```
Q2 $FORMAT_Q. Q3 $FORMAT_Q. Q4 $FORMAT_Q. Q5 $FORMAT_Q.;
RUN;
```



VIEWTABLE: Tan.Survey						
	GENDER	Q1	Q2	Q3	Q4	Q5
1	MALE	YES	NO	YES	NO	YES
2	FEMALE	NO	NO	YES	YES	NO
3	MALE	YES	YES	YES	YES	YES
4	FEMALE	NO	NO	NO	YES	NO

FORMAT NAME: \$FORMAT_SEX LENGTH: 6		
MIN LENGTH: 1	MAX LENGTH: 40	DEFAULT LENGTH: 6 FUZZ: 0
START	END	LABEL (VER. 9.4 24APR2015:21:31:16)
F	F	FEMALE
M	M	MALE

FORMAT NAME: \$FORMAT_Q LENGTH: 3		
MIN LENGTH: 1	MAX LENGTH: 40	DEFAULT LENGTH: 3 FUZZ: 0
START	END	LABEL (VER. 9.4 24APR2015:21:31:16)
0	0	NO
1	1	YES

6. Chapter 14: 14.8 and 14.10

Dataset PRICES for 14.10

```

DATA PRICES;
    INPUT PART_NUMBER QUANTITY PRICE @@;
DATALINES;
100 23 29.95 102 12 9.95 103 21 15.99 123 9 119.95 113 40 56.66
111 55 39.95 105 500 .59
;

```

14.8

```

DATA STUDENT;
    INPUT ID NAME & $30. @@;
DATALINES;
1 JOHN TORRES 5 ALEX ANTONIOU 3 THOMAS FRIEND
2 SHELDON GOLDSTEIN 11 JOANNE DIPIETRO 12 BILL MURRAY
21 JANET RENO 4 DEBORAH SMITH 6 DON DUBIN 7 ALICE FORD
8 DIANE FARLEY 9 LAURA EASTON 10 BRIAN FISHMANN
13 ERIC GARRETT 14 JAMES GALT 15 TONI GILMAN
;

DATA TEST;
    INPUT ID SCORE @@;
DATALINES;
15 95 1 80 3 98 21 75 4 87 14 67 13 91 11 85 12 57
29 93
;

PROC SORT DATA = STUDENT;
BY ID;
RUN;
PROC SORT DATA = TEST;
BY ID;
RUN;

DATA ROSTER;
    LENGTH LAST $8;
    UPDATE STUDENT TEST (IN = JOIN);
    BY ID;
    IF JOIN;
    IF NAME = "" THEN DO NAME = "NOT IN STUDENT FILE";
    END;
    IF NAME = "NOT IN STUDENT FILE" THEN LAST = "ZZZ";
    ELSE
    LAST = SCAN(NAME, -1);
    RUN;
PROC SORT DATA = ROSTER;
BY LAST;
RUN;
PROC PRINT DATA = ROSTER;
TITLE "STUDENTS ROSTER";
VAR ID NAME SCORE;
    RUN;

```


STUDENTS ROSTER			
Obs	ID	NAME	SCORE
1	11	JOANNE DIPIETRO	85
2	3	THOMAS FRIEND	98
3	14	JAMES GALT	67
4	13	ERIC GARRETT	91
5	15	TONI GILMAN	95
6	12	BILL MURRAY	57
7	21	JANET RENO	75
8	4	DEBORAH SMITH	87
9	1	JOHN TORRES	80
10	29	NOT IN STUDENT FILE	93

14.10

```

DATA PRICES;
INPUT PART_NUMBER QUANTITY PRICE @@;
DATALINES;
100 23 29.95 102 12 9.95 103 21 15.99 123 9 119.95 113 40 56.66
111 55 39.95 105 500 .59
;
DATA UPDATE_PRICES;
INPUT PART_NUMBER QUANTITY PRICE @@;
DATALINES;
103 . 18.99 111 45 29.95 113 35 . 123 . 129.96
;
PROC SORT DATA = PRICES;
BY PART_NUMBER;
RUN;
PROC SORT DATA = UPDATE_PRICES;
BY PART_NUMBER;
RUN;
DATA CHANGED_PRICE;
UPDATE PRICES UPDATE_PRICES;
BY PART_NUMBER;
RUN;

PROC PRINT DATA=CHANGED_PRICE;
TITLE "CHANGED PRICE";
RUN;

```

CHANGED PRICE			
Obs	PART_NUMBER	QUANTITY	PRICE
1	100	23	29.95
2	102	12	9.95
3	103	21	18.99
4	105	500	0.59
5	111	45	29.95
6	113	35	56.66
7	123	9	129.96