

Thomas Pattara

02.05.17

Homework 6

i. Recall the OURDATA3 from Homework 5. In Homework 6, we will modify this data further by using the various functions discussed in Chapters 11 and 12. Create the following additional variables USING SAS PROGRAM and call the new dataset **OURDATA4**. This new dataset will also contain all the previous variables.

a) **LNAMEUp**: A new variable where all last names (LNAME) are in upper cases.

b) **FLLNAME**: Use the concatenation function (Section 12.5) to put together FRSTNAME and LNAMEUp

c) **FLLNAMEL**: This is the length of FLLNAME.

d) **RAND_DOB**: This variable will contain a random date-of-birth. Hope you will agree and this is just a game! Create a random DOB using the following DO loop in your DATA statement. Good luck:

```
do j = 1 to 31;  
  RAND_DOB = int(2000*Ranuni(0) + '01jan1998'd);  
  output;  
end;  
drop j;  
format RAND_DOB mmddyy10.;
```

e) **RAND_AGE**: Using the RAND_DOB variable, create a random age variable (*in days*) on Thursday October 5th, 2017.

f) **InVVNAMEC**: This will be our first attempt to try to correct the strings in InVVNAME to make them valid variable names. Use the COMPRESS function to weed out as many unwanted characters as possible from the entries of InVVNAME. For example, you can remove blank, hyphen, hash, dollar sign, question mark, and more. The idea is to solve the problems we created!

g) **SASstrng**: Undoubtedly, you all created interesting valid as well as invalid variable names. Thanks again! That is what motivated me to dig the discussion data dipper. Now let's search for words in a string. You can find

further details in Section 12.10. Create a variable name called SASstrng that will be "yes" if "SAS" is in the string for VVNAME1 (the first valid variable name) and "no" otherwise.

```
data ourdata4;

    merge ourdata3 seem;

    by LASTNAME;

    length LNAMEUp $25;

    length FLLNAME $30;

    length InVVNAMC $ 30;

    LNAMEUp= upcase(LASTNAME);

    FLLNAME = cats(FRSTNAME,LNAMEUp);

    FLLNAMEEL = max(length(FLLNAME));

    InVVNAMC = compress(InVVNAME,'()-.%&$!');
```

```
do j = 1 to 1;

    RAND_DOB = int(2000*Ranuni(0) + '01jan1998'd);

    output;

end;

drop j;

format RAND_DOB mmddyy10.;
```

```
RAND_AGE=intck('days',RAND_DOB,'10oct2017'd);
```

```
output;
```

```
run;
```

```
data classSimple;
```

```
set ourdata4;
```

```
orig_obs = _n_;
```

```
if mod(_n_,2) eq 0 then output;
```

```
run;
```

```
proc print data=classSimple;
```

```
var FRSTNAME LASTNAME MAJOR VVNAME1 SASstrng VVNAME2  
InVVNAME InVVNAMC LNAMEUp FLLNAME FLLNAMEL RAND_DOB  
RAND_AGE;
```

```
run;
```

O b s	FRS TN AM E	LAS TN AM E	MAJ OR	VVNA ME1	SA Sst rn g	VVNAM E2	InVVN AME	InVVN AMC	LNA ME Up	FLLNA ME	FLL NA ME L	RA ND_ DO B	RA ND_ AG E
1	Cou rtne y	And ers on	Athl etic Trai	Height_ in_centi meters	No	LastNa me	Price per Pound	Pricep erPou nd	AN DER SON	Courtn eyAND ERSON	16	11/ 18/ 200	544 0

O b s	FRS TN AM E	LAS TN AM E	MAJ OR	VVNA ME1	SA Sst rn g	VVNAM E2	InVVN AME	InVVN AMC	LNA ME Up	FLLNA ME	FLL NA ME L	RA ND_ DO B	RA ND_ AG E
			ning									2	
2	Alli son	Bod vig	Com pute r Scie nce	total_co st	No	itemCos t	total%	total	BOD VIG	Allison BODVI G	13	03/ 22/ 199 9	677 7
3	Alli son	Bod vig	Mat hem atic s	total_co st	No	itemCos t	total%	total	BOD VIG	Allison BODVI G	13	08/ 28/ 200 2	552 2
4	We sley	Bo wen	Mat hem atic s	_South Dakota	No	South_ Dakota	Hello World	Hello World	BO WE N	Wesley BOWE N	11	02/ 09/ 200 0	645 3
5	Aud rey	Bun ge	Mat hem atic s	Newton s	No	Unit_Of _Measu rement	% Yield	Yield	BU NGE	Audrey BUNGE	11	06/ 29/ 199 9	667 8
6	Josh	But tke	Mat hem atic s	PricePe rSquare Foot	No	Price_P er_Squa re_Foot	Price per square foot	Pricep ersqua refoot	BUT TKE	JoshBU TTKE	10	10/ 16/ 199 8	693 4
7		Citt erm a			No				CIT TER MA	CITTE RMA	8	12/ 14/ 199 8	687 5
8	Wal ter	Citt erm an	Mat hem atic s	Off_Top ic		Always Open54	Rooste r Teeth Shows	Rooste rTeeth Shows	CIT TER MA N	Walter CITTE RMAN	15	11/ 07/ 200 1	581 6
9	Tay lor	Deu tsch	Mat hem atic s	Miles_p er_Hou r	No	MilesPe rHour	miles- per- hour	milespe rhour	DEU TSC H	Taylor DEUTS CH	13	10/ 05/ 199 8	694 5
1	Rile	Hau	Mat hem	Riley_H	No	GoDuke	S-D-S-	SDSU	HA	RileyH	9	01/ 10/	575

O b s	FRS TN AM E	LAS TN AM E	MAJ OR	VVNA ME1	SA Sst rn g	VVNAM E2	InVVN AME	InVVN AMC	LNA ME Up	FLLNA ME	FLL NA ME L	RA ND_ DO B	RA ND_ AG E
0	y	g	atics	aug2		2	U		UG	AUG		200 2	2
1 1	Jon ath an	Hed ma n	Com pute r Scie nce	_Jonath an_Hed man	No	Numbe r_of_Cla sses	12_Peo ple	12_Peo ple	HED MA N	Jonath anHED MAN	14	09/ 20/ 200 1	586 4
1 2	Kor y	Hei er	Mat hem atic s	_college	No	meerka t	Chapte rs 1 & 2	Chapte rs12	HEI ER	KoryH EIER	9	07/ 17/ 199 9	666 0
1 3	Paig e	Hin ton	Mat hem atic s	SAS_Ma ster	Ye s	_Hello_	SAS Amate ur	SASAm ateur	HIN TON	PaigeH INTON	11	03/ 02/ 200 0	643 1
1 4	Han nah	Hus s	Mat hem atic s	total_re venue	No	August 29	%total	total	HUS S	Hanna hHUSS	10	06/ 03/ 200 1	597 3
1 5	Sam uel	Iva nec ky	Com pute r Scie nce	Univers ityNam e	No	_Favorit eColor	- favorit eSong	favorit eSong	IVA NEC KY	Samuel IVANE CKY	14	03/ 21/ 200 1	604 7
1 6	Sam uel	Iva nec ky	Mat hem atic s	Univers ityNam e	No	_Favorit eColor	- favorit eSong	favorit eSong	IVA NEC KY	Samuel IVANE CKY	14	08/ 25/ 199 9	662 1
1 7	Der ek	Joh ans on	Mat hem atic s	Meters_ per_Sec ond	No	Jackrab bits	50page s	50pag es	JOH ANS ON	DerekJ OHANS ON	13	08/ 27/ 200 0	625 3
1 8	Nic ole	Kne ip	Mat hem atic s		No				KNE IP	Nicole KNEIP	11	01/ 18/ 199 9	684 0

Obs	FRNAME	LASTNAME	MAJOR	VVNAME1	SAStng	VVNAME2	InVVNAME	InVVNAMEC	LNAMEUp	FLLNAME	FLLNAMEL	RANDOB	RANDAGE
19	Ali	Lacey	Mathematics	TimeOfClasses	No	Classes_to_take	5-classes	5classes	LACEY	AliLACEY	8	10/02/2001	5852
20	Jake	Larson	Mathematics	jake_larson	No	CarModel	% Of Injury	OfInjury	LARSON	JakeLARSON	10	06/30/2002	5581
21	Jacie	McDonald	Mathematics	FirstName	No	DATE	First Name	FirstName	MC DONALD	JacieMCDONALD	13	09/13/1998	6967
22	Drue	Miller	Mathematics	Gender	No	Drue_Miller	79elephant	79elephant	MILLER	DrueMILLER	10	08/29/2000	6251
23	Shea	Olson	Mathematics	CAR_PART_NUMBER	No	PricePerPound	37highway	37highway	OLSON	SheaOLSON	9	09/15/1999	6600
24	Thomas	Pattara	Mathematics	Nvidia	No	Thomas_Pattara	4%3--1	431	PATTARA	ThomasPATTARA	13	11/18/2001	5805
25	Amanda	Peterson	Mathematics	Parts	No	Cost_per_Pound	8_is_enough	8_is_enough	PETERSON	AmandaPETERSON	14	07/05/2000	6306
26	Nathan	Remmich	Mathematics	This_is_a_valid_name	No	_this_is_2	1_not_valid	1_not_valid	REMMICH	NathanREMMICH	13	09/17/1999	6598
27	Elizabeth	Rust	Mathematics	ValidVariable	No	Ice_Cream	Ice-cream	Icecream	RUST	ElizabethRUST	13	05/21/2003	5256

O b s	FRS TN AM E	LAS TN AM E	MAJ OR	VVNA ME1	SA Sst rn g	VVNAM E2	InVVN AME	InVVN AMC	LNA ME Up	FLLNA ME	FLL NA ME L	RA ND_ DO B	RA ND_ AG E
2 8	Jon ath an	Sax	Mat hem atic s	X12Y34 Z56	No	This_is_ Rando m	%is- bad	isbad	SAX	Jonath anSAX	11	09/ 11/ 200 2	550 8
2 9	Zac har y	Shr oed er	Mat hem atic s	Zachar y_Schro eder21	No	GolfClu b	Track & Field	TrackF ield	SHR OED ER	Zachar ySHRO EDER	15	02/ 18/ 200 3	534 8
3 0	Tim oth y	Slav ik	Mat hem atic s	Weight	No	Length	Weight in pound s	Weight inpoun ds	SLA VIK	Timoth ySLAVI K	13	09/ 21/ 200 2	549 8
3 1	Linc oln	Sto ebn er	Mat hem atic s	Mod_Re q	No	Height_ MM	\$Mone yMayw eather	Money Maywe ather	STO EBN ER	Lincoln STOEB NER	15	08/ 01/ 199 9	664 5
3 2	Nat han	Thi rste n	Mat hem atic s	SASPro gram1	Ye s	_llikepr ogram ming	Go-Jack s	GoJack s	THI RST EN	Nathan THIRS TEN	14	11/ 18/ 199 9	653 6
3 3	Ale xan der	Wa de	Mat hem atic s	Blue_ca r	No	Feet_Pe r_Secon d	92bloc ks	92bloc ks	WA DE	Alexan derWA DE	13	12/ 13/ 199 8	687 6
3 4	Ale x	Wie sele r	Mat hem atic s	MiamiD olphins	No	are_the _best	NFL- team!	NFLtea m	WIE SEL ER	AlexWI ESELE R	12	03/ 03/ 200 1	606 5
3 5	Apr il	Zha ng	Acc oun ting	AprilZh ang	No	April_Z hang	April Zhang	AprilZ hang	ZHA NG	AprilZ HANG	10	10/ 05/ 199 8	694 5

ii. Suppose that we want to form a committee containing five students from our class. Use the PROC SURVEYSELECT (on textbook page 200) to do this. It makes use of a simple random sample (SRS) procedure that gives equal chance for all students to be included in the committee. Use PROC PRINT to print only the **FULLNAME** and **Major** of the students in the committee you formed.

```
data subset;
```

```
length FLLNAM $40;
```

```
if ranuni(1347564) le .1;
```

```
run;
```

```
proc surveyselect data=ourdata4
```

```
    out=subset
```

```
    method=srs
```

```
    samsize=34;
```

```
run;
```

```
proc print data=subset(obs=5);
```

```
var major FLLNAME;
```


run;

Table of Contents

The SURVEYSELECT Procedure

Selection Method	Simple Random Sampling	
Input Data Set	OURDATA4	
Random Number Seed	375657646	
Sample Size	34	
Selection Probability	0.485714	
Sampling Weight	2.058824	
Output Data Set	SUBSET	

Obs	MAJOR	FLLNAME
1	Athletic Training	CourtneyANDERSON
2	Computer Science	AllisonBODVIG
3	Computer Science	AllisonBODVIG
4	Mathematics	AllisonBODVIG
5	Mathematics	WesleyBOWEN