

## SAS Program Notes

### Biostatistics: A Guide to Design, Analysis, and Discovery

#### Chapter 4: Probability and Life Tables

#### Note 4.1 – Simulation study to find probability of common birthdays

SAS will create random samples in a **DATA** step. We use the **ARRAY** statement to indicate to SAS that the variable **class** has locations set up to hold the 10 birthdates. The **RETAIN** command tells SAS not to lose the values stored in **class** or the **SEED** value as we process one observation to the next. The **DO** statement indicates how many times the statements included between the **DO** statement and its closing **END** statement are to be performed. We are using the function, **RANUNI**, to obtain a random number between 0 and 1. The value of the **SEED** variable must be initialized before **RANUNI** is called. By multiplying the number created by **RANUNI** by 1000, we obtain a number between 0 and 1000. The **INT** function truncates the number to an integer. If the integer is 0 or greater than 365, we delete it and draw another random number. If the integer is between 1 and 365, we store it in the **j-th** location of **class**. After obtaining 10 numbers, we use the **OUTPUT** statement to write them to the file named **BIRTH**. We repeat the process 29 more times and then print the resultant file.

#### SAS commands:

```
DATA BIRTH;
  RETAIN SEED;
  ARRAY class(10) class1-class10;
  SEED=0;
  DO student=1 TO 30;
    DO J=1 TO 10;
      REPL:CALL RANUNI(SEED,X);
      X=1000*X;
      X=INT(X);
      IF X = 0 OR X > 365 THEN GOTO REPL;
      class(J)=X;
    END;
  OUTPUT;
  END;
RUN;
PROC PRINT;
  ID student;
  VAR class1-class10;
RUN;
```

As we have already stated, the program begins by assigning a random number between 1 and 365 to the first student in each of the 10 classes. It then repeats this process 29 times. In the output below, we see that the variable **student** is followed by the variables **class1** to **class10**. The variable **student** is used to identify each of the 30 students in each of the 10 classes, and the variables **class1** to **class10** store birthdays for each student.

**SAS output:**

## The SAS System

student	class1	class2	class3	class4	class5	class6	class7	class8	class9	class10
1	85	87	159	34	236	131	238	293	222	242
2	113	3	305	12	38	347	279	132	102	148
3	253	304	152	32	282	328	244	250	147	13
4	257	242	364	313	204	229	63	178	132	324
5	98	235	28	192	9	303	358	309	153	197
6	160	252	234	300	253	119	309	90	122	110
7	193	212	82	223	164	158	120	334	257	138
8	285	314	261	273	335	336	114	122	288	131
9	239	218	336	136	280	335	79	161	335	112
10	184	198	24	21	22	158	317	38	208	271
11	101	197	240	351	55	206	301	160	163	218
12	178	124	229	127	38	274	234	139	23	208
13	293	169	213	295	35	351	293	180	79	354
14	365	273	32	303	286	38	112	339	348	10
15	136	68	92	280	71	187	256	3	312	261
16	233	1	36	156	158	217	94	175	5	91
17	349	39	52	219	78	106	361	65	238	55
18	10	256	314	169	153	275	71	204	279	365
19	179	72	15	238	361	110	128	139	344	148
20	279	351	155	286	155	39	74	235	2	81
21	51	347	39	182	263	128	295	173	179	336
22	296	152	199	175	323	117	251	46	351	315
23	48	282	144	221	31	115	168	141	205	202
24	71	234	37	238	223	214	165	258	342	41
25	89	35	310	39	352	130	170	76	152	125
26	203	225	314	80	237	201	162	301	284	39
27	272	137	18	341	78	261	27	121	192	166
28	121	34	82	138	223	340	49	361	38	62
29	94	4	289	279	93	86	78	49	297	327
30	307	95	107	186	147	237	176	312	166	253