

## Stata Program Notes

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

#### Chapter 4: Probability and Life Tables

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

Below we display some simple commands to create ten columns of simulated data. This could be considered a naïve way of handling this problem. We encourage the reader to incorporate a scenario where students may have been born during a leap-year (ie. a year containing one additional day).

```
set obs 30
gen student = _n
gen class1 = round(364*uniform()+1)
gen class2 = round(364*uniform()+1)
gen class3 = round(364*uniform()+1)
gen class4 = round(364*uniform()+1)
gen class5 = round(364*uniform()+1)
gen class6 = round(364*uniform()+1)
gen class7 = round(364*uniform()+1)
gen class8 = round(364*uniform()+1)
gen class9 = round(364*uniform()+1)
gen class10= round(364*uniform()+1)
list
```

	student	class1	class2	class3	class4	class5	class6	class7	class8	class9	class10
1.	1	98	126	357	330	192	40	70	172	79	89
2.	2	44	297	267	65	283	230	349	325	84	342
3.	3	141	173	152	308	344	194	276	4	291	355
4.	4	247	65	230	198	355	273	142	22	213	146
5.	5	110	362	164	120	281	346	285	54	273	316
6.	6	78	113	66	332	330	192	138	4	159	298
7.	7	297	96	207	146	21	20	127	3	134	74
8.	8	116	228	214	351	7	26	291	80	326	68
9.	9	356	58	86	97	174	301	344	56	40	8
10.	10	130	207	252	227	282	241	56	47	273	6
11.	11	325	182	299	213	192	310	361	73	113	77
12.	12	145	128	20	17	296	192	242	359	145	267
13.	13	112	321	55	287	277	83	81	95	172	36
14.	14	90	102	46	199	362	65	3	144	194	140
15.	15	256	40	38	9	337	335	78	53	167	296
16.	16	274	259	188	146	290	270	212	362	317	85
17.	17	164	272	202	119	282	93	11	228	232	323
18.	18	257	349	243	268	288	199	233	218	120	85
19.	19	119	58	362	81	221	3	229	61	260	338
20.	20	173	125	115	295	228	320	147	273	284	202
21.	21	117	87	172	28	29	266	198	333	12	150
22.	22	278	127	138	174	293	114	313	256	338	230
23.	23	201	79	28	234	157	232	185	30	56	68
24.	24	109	105	176	234	77	292	201	306	115	130
25.	25	304	38	333	54	326	223	44	30	179	357
26.	26	60	67	196	102	226	141	294	282	91	60
27.	27	218	124	279	77	207	187	114	18	171	335
28.	28	300	111	233	246	236	100	93	362	325	265

29.	29	118	73	240	318	284	80	301	220	204	310
30.	30	307	1	144	221	298	2	89	271	303	230

In this example, we record 7 out of 10 trials that have duplicates. Because the command **set seed** was not used, every time we run the commands, we get a different set of random numbers. However, using the command `set seed` as indicated below, you will produce the same set of random numbers each time. Run more simulations to see how the results change if you have smaller classroom sizes.

```

set obs 30
gen student = _n
set seed 2
gen class1 = round(364*uniform()+1)
gen class2 = round(364*uniform()+1)
gen class3 = round(364*uniform()+1)
gen class4 = round(364*uniform()+1)
gen class5 = round(364*uniform()+1)
gen class6 = round(364*uniform()+1)
gen class7 = round(364*uniform()+1)
gen class8 = round(364*uniform()+1)
gen class9 = round(364*uniform()+1)
gen class10= round(364*uniform()+1)
list

```

	student	class1	class2	class3	class4	class5	class6	class7	class8	class9	class10
1.	1	311	349	77	223	121	3	283	226	168	91
2.	2	20	39	17	20	59	220	299	207	284	219
3.	3	230	275	300	21	68	74	208	172	127	272
4.	4	255	135	223	63	302	325	87	46	124	294
5.	5	190	183	224	254	165	50	101	344	335	102
6.	6	123	8	323	330	260	310	212	156	172	57
7.	7	65	268	49	351	160	5	274	253	81	130
8.	8	332	342	123	214	201	25	108	253	57	237
9.	9	142	276	89	344	356	150	330	357	198	77
10.	10	150	25	3	204	98	240	357	305	334	329
11.	11	350	184	128	259	102	233	262	139	257	258
12.	12	57	148	104	209	358	110	243	187	64	258
13.	13	239	237	6	271	167	184	200	30	278	22
14.	14	156	270	2	266	274	173	87	343	173	124
15.	15	154	11	134	214	52	324	206	232	173	353
16.	16	266	303	191	328	222	196	157	314	49	72
17.	17	156	260	344	184	29	79	332	156	61	335
18.	18	327	87	252	26	80	290	306	295	108	37
19.	19	34	181	191	131	299	192	169	328	96	221
20.	20	79	348	67	94	216	251	293	63	195	154
21.	21	206	141	132	361	304	314	344	197	345	263
22.	22	158	43	51	327	54	76	219	16	300	105
23.	23	353	333	214	302	52	96	242	239	35	96
24.	24	359	82	79	291	203	213	75	36	215	133
25.	25	199	354	145	80	204	210	116	284	117	74
26.	26	300	134	345	10	133	178	27	18	110	266
27.	27	186	254	12	348	324	343	192	270	348	298
28.	28	223	248	321	51	357	131	321	130	320	17
29.	29	277	114	62	151	325	74	47	354	65	281
30.	30	99	289	357	285	361	358	28	108	149	118

How to cite this page:

Stata Program Notes for Chapter 4. Biostatistics: A Guide to Design, Analysis, and Discovery. from <http://www.biostat-edu.com/ProgramNotes/> (accessed January 1, 2013).