

■ RICK HESSE, Feature Editor, Mercer University

Simulating the Birthday Problem

by Rick Hesse, Mercer University

I want to share a nice spreadsheet template that simulates the "birthday problem," which is presented in most statistics books when discussing probability. Most students are surprised that there is better than a 70% chance that in a group of 30 people randomly chosen, at least two share the same birth date (not necessarily the same year). Shown in Figure 1 is the complete spreadsheet template, which will be explained in detail.

For this example run, there are two people born on November 3rd (day 307) and thus a match. For this run, the original birthdays are in column B while column C contains the same birthdays in ascending order.

Column A contains the number of people you wish to simulate in the group, which in this case is 25. Cells C2..D2 give the range of the birthday number (1-365, excluding leap year) to simply show the data ranges. Starting in cell B4 is the simulated result of a random birthday:

Lotus: @INT((\$D\$2-\$C\$2)*@RAND+1)

Excel: =TRUNC((\$D\$2-\$C\$2)*RAND()+1)

Of course this assumes an equal chance of being born on any of the 365 days, which is probably not correct, but close enough to being uniform for our purposes. Once these birthday numbers have been calculated, we need to determine if there are any matches. The easiest way to do this is to sort column B so that if there are two or more birthdays on the same day, these numbers will be adjacent to each other. But when this sorting is done, the numbers get recalculated (even with recalculation turned off, there are problems later on)! Therefore we use column C to do a Paste Special Value (or Range Value in classic Lotus) to "freeze" or save these numbers. Then we can sort these "frozen" values in column C and check to see if there are any matches in column D. The logic

formula checks to see if the previous date is the same as the current one.

Lotus (D5): C5=C4

Excel (D5): =(C5=C4)*1

It is necessary in Excel to change the logic result of True/False to a number by multiplying the logic statement in parentheses by 1. Even if 0 or 1 is showing, Excel doesn't recognize it as numeric unless you multiply by 1. With either spreadsheet it now is a simple matter to check if the sum of column D is 0. If it is, there is at least one birthday match. In Figure 1, there were two birthday matches. The formula in D4 is:

Lotus (D4):

@IF(@SUM(D5:D33)>0,"YES","NO")

Excel (D4):

=IF(SUM(D5:D33)>0,"YES","NO")

The actual probability of there being at least two people with the same birthday is given in E4 with the calculations of "n" people not having the same birthday shown from E5 on down.

E4: 1-E28.

The math formula for the probability of no one with the same birthday out of

See CLASSROOM, page 22

	A	B	C	D	E
1	BIRTHDAY PARTY		LOW	HIGH	
2	BIRTHDAY.WK1		1	365	
3	PERSON	BIRTHDAY	FREEZE	MATCH?	PROB
4	1	312	22	YES	56.87%
5	2	307	33	0	0.9973
6	3	153	36	0	0.9918
7	4	307	40	0	0.9836
8	5	162	75	0	0.9729
9	6	150	82	0	0.9595
10	7	75	96	0	0.9438
11	8	40	118	0	0.9257
12	9	96	150	0	0.9054
13	10	220	153	0	0.8831
14	11	203	162	0	0.8589
15	12	36	175	0	0.8330
16	13	175	182	0	0.8056
17	14	218	203	0	0.7769
18	15	363	218	0	0.7471
19	16	33	220	0	0.7164
20	17	246	246	0	0.6850
21	18	22	265	0	0.6531
22	19	334	284	0	0.6209
23	20	265	307	0	0.5886
24	21	182	307	1	0.5563
25	22	316	312	0	0.5243
26	23	118	316	0	0.4927
27	24	82	334	0	0.4617
28	25	284	363	0	0.4313

Figure 1: Birthday template.

will disappear by the end of the century for the simple reason that they will lose money at it. My advice to university presses is: get out of this area now. The trade publishers in New York City are very, very good at what they do, and it is naive to believe that a university press in Berkeley or Lincoln or even Cambridge can compete with them.

It is the scholarly monograph that hurts so much to contemplate. Does it have a future? Print runs for some titles are now down to as few as 500 to 700 copies; the market continues to contract. Inasmuch as there is a long tradition of subsidizing monographs, I am sure some titles will continue to be published, but the outlook is not good. Nor does electronic publishing seem very promising, for two reasons: the discursive text of a typical scholarly monograph works best in print, and in any event, it is the fixed, not the variable, costs that are undermining the monograph, and the fixed costs do not vary much between print and electronics. It is my view that the notion, dear to college administrators, that university presses can be self-sustaining is a pipe dream.

This leaves us with the quality segment of the trade book business. Here I am optimistic. Although the superstore chains such as Barnes & Noble and Borders are

primarily being built to sell remainders and bestsellers, stocking a broad selection of books is important to their image. Good books will find their way onto the shelves. And they will continue to be published, both by small independent presses and by sneaky editors in the large commercial houses. There is also little reason to believe that interactive media will make any significant inroads with this audience (at least outside the workplace) because the closely reasoned text that they enjoy is precisely what digital media is not good at. For some time to come, we can expect books to be the thinking person's medium.

There is so much hype surrounding electronic media that it is good to slow down once in a while for a reality check. We have already looked at the myth that electronic publishing alone will destroy print, but perhaps more insidious is the idea that multimedia is somehow superior to text. If this were true, all music would be opera. Except for kids, most publishing will continue to be text publishing for the simple reason that words can do things that images and sound cannot. The text may be hypertext, and we may be seeing more and more of it on a computer screen, but the primacy of text is just about the one thing that publishers can count on in the future.

What is for me the most intriguing question, whose answer may not be known for 10, 20, maybe 100 years, is how digital media will change our understanding of what an idea is. A paragraph is a creature of the print medium (obviously). But is the consciousness that creates paragraphs and thinks in paragraphs also a function of print? A closely reasoned argument proceeds step by step, word by word. Is such an argument an outgrowth of the linearity of print? I don't know the answer to that question, but what is clear is that if the future of publishing is increasingly going to be an electronic future, we will have to develop a poetics of new media. This is one publishing project whose time has come. The one outstanding question is whether we will publish it in print or electronic form. ■

Dr. Andrew Ruppel
 Monroe Hall
 University of Virginia
 Charlottesville, VA 22903
 Voice mail: 804-924-3867
 Fax: 804-924-7074
 e-mail: acr2y@virginia.edu

CLASSROOM, continued from page 16

"n" people is $P(n) = (366-n)/365 * P(n-1)$ with $P(1) = 1$. The formula for this is:

E5: $\$D\$2-A4)/\$D\2

E6: $E5*(\$D\$2-A5)/\$D\2 and copy down

The keyboard macro shown in G2..G5 is for classic Lotus. This macro calculates a new set of random birthdays in column B, copies the values over to column C, sorts it, and calculates any matches in column D. This macro will also work in Lotus 4, but must be rewritten to work in Excel. The macro is named \R (for recalculate) and

its range is defined as G2..G5. The text can be all in one cell or broken down into successive cells as shown in Figure 2.

As a class or lab exercise, students can run this 20 times or so and determine if

the probability comes close to the expected value. The template can also be easily changed for a different number of people in the group. ■

	F	G
2	\R	{CALC}
3		/RVB4..B28~C4~
4		/DSG
5		{CALC}

Figure 2: Classic Lotus macro.

Dr. Rick Hesse
 Industrial and Systems Engineering
 Department
 Mercer University
 Macon, GA 31207