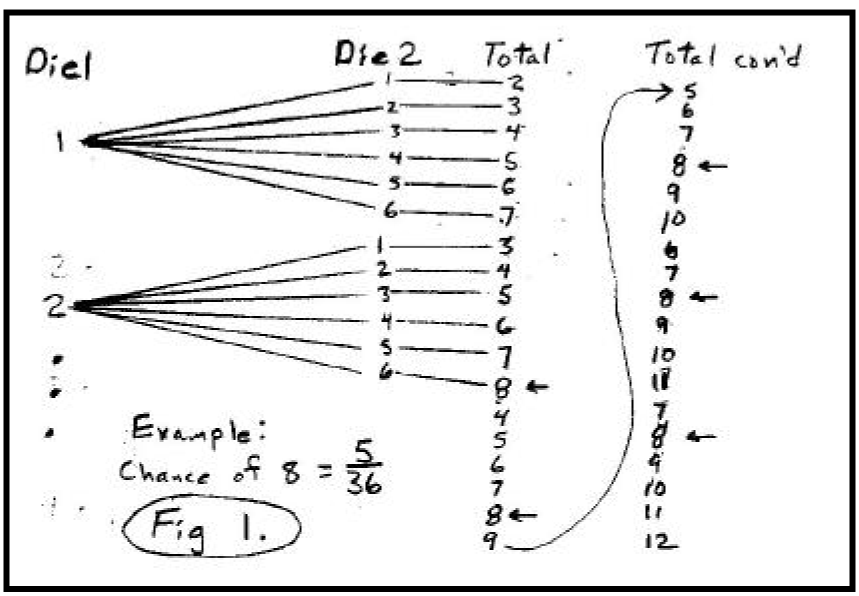
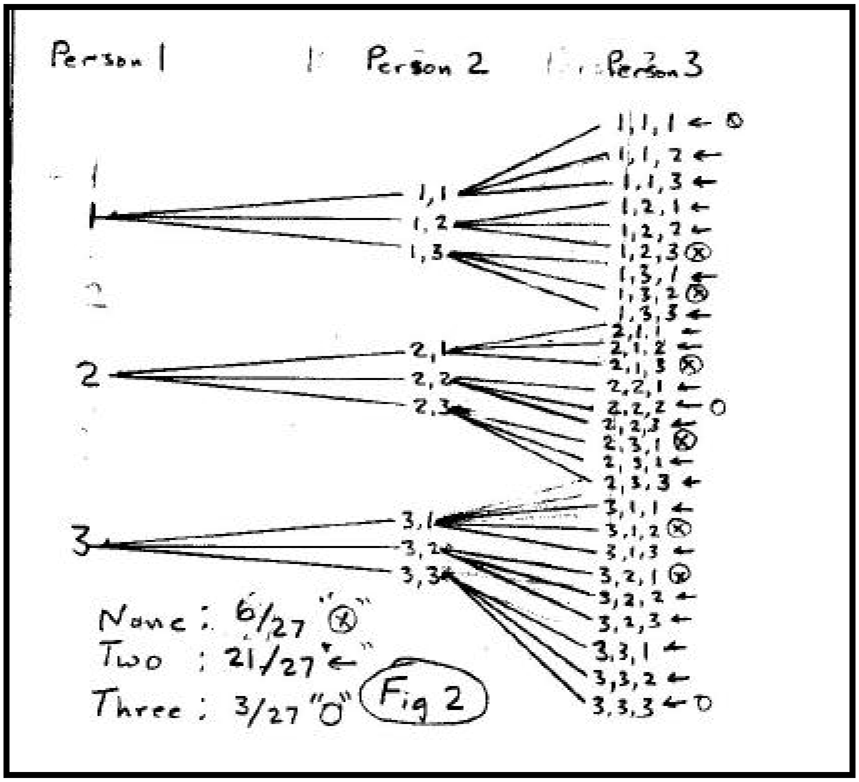
A little probability exercising

Most rigorous statisticians and statistics teachers would prefer that you have a lot of experience and consummate knowledge of discrete probability theory, including such atrocities as Baye's Theorem, before you have the privilege and excitement of inferencing from sample data. But, hey, what are the chances of that? The answer to that question falls into the category called subjective probability, which can involve judgment, experience, or even paranormal activity. We're into objective probability. So what we'll do is a few little exercises and some Excel and call it even.

How about rolling fair dice (2 die)? What's the chance of rolling an 8? We can use a "tree diagram" like in sloppy hand-drawn Fig. 1 below to determine all possible outcomes, count the number of 8's, and divide by the total number of total outcomes, 36, to find the probability of rolling an 8. Note that all possible outcomes are equally probable or likely. Note also that the probability of not rolling an 8, 31 divided by 36, could be found the same way or by subtracting 5/36 from 1, 1 being the maximum probability, that of a sure thing.



Here's another similar problem using a tree diagram to generate all possible outcomes for a situation in which all the outcomes are equally as likely to occur. Suppose 3 people randomly choose one of three numbers. What is the probability that two and only two *(not shown on the drawing)* will choose the same number, that all three will chose different numbers, that all three will choose the same number, or that two or more will choose the same number. See Fig 2. below:



So the probability of 2 or more choosing the same number is 21/27. This is indicated in the Fig 2 as “Two”. Exactly two would be 18/27.

Seven brilliant geeks had mind-altering epiphanies (one each) on random days in June (30 of them). What is the probability that 2 epiphanies occurred on the same day? You’re not going to make a tree diagram of this, are you? Let’s try something else. You may decide that probability problems, like differential equations, are a “bag of tricks”.

The probability that the second person had his/her experience the same day as the first is 1/30, of course. The probability that that isn’t so is 29/30. The probability that the third person’s great day is not the same as either of the first two people’s (**given that the first two's days are not the same**) is 28/30. So the probability that none of the folks had their experience on the same day in June can be stated as “The probability that the second’s day is not the first’s <u>and</u> the third’s day is neither the first’s or second’s <u>and</u> the fourth’s is neither the first’s nor the second’s nor the third’s <u>and</u> the fifth’s is neither the first’s nor second’s nor third’s nor fourth’s...” The underlined “ands” indicate multiplication of probabilities. What do you think “ors” indicate? Hint: not division or subtraction. So the result in numbers for 7 people and thirty days is:

(29/30)*(28/30)*(27/30)*(26/30)*(25/30)*(24/30) = 0.469.

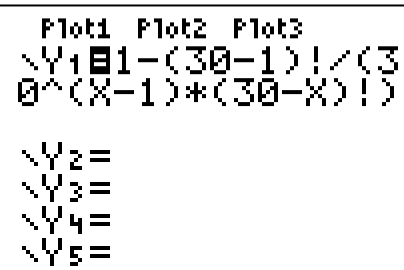
So the probability that two or more had their experience on the same day in June is:

1 – 0.469 = 0.531 or about 53%

It would be neat if you could come up with a function of the probability that two or more people have the same day out of the thirty with the number of people as the independent variable, a formula in terms of the number of people. Suppose I told you or you know about the factorial (designated “!”) function of a positive integer. By example, 3! = 3*2. And 5! = 5*4*3*2. Now that you have it, do you see a way to write the function for the probability of two or more people not having the same day in 30 by looking at the pattern of the above calculation? The numerator appears to be trying to be 29!, but is missing all the terms of 23!. So you can obtain the correct expression for the numerator with 29!/23!. The denominator of the expression, not including the 23! we just decided to put in, is clearly? 30^(x-1), where x is the number of people - in this case 7. Since 29! is (30-1)! and 23! is (30-x)!, what is the formula? How about (30-1)!/((30^(x-1)*(30-x)!)). You could even make the 30 a variable, say “y”, to make a function of two independent variables.

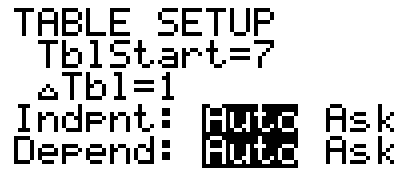
Don’t forget that this is probability that no two persons have the same day of the month and to get the probability that at least two do, you must subtract this expression from 1.

What could you do if someone asked you to find the number of epiphanizing geeks such that the probability is 80% or better that two or more had their experience on the same day of June? You could refuse. You could do trial and error until you found the answer. You could graph the function in the 83plus and trace to find the answer (if ! were defined for non-integers, which it’s not) or you could table the function to find it.

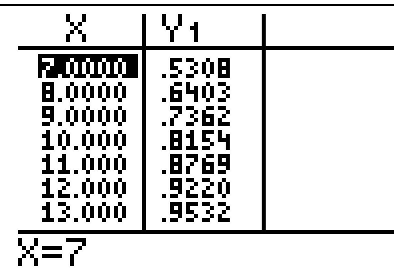
Y=. Enter the expression as shown. You get the ! by **Math**-*PRB*-4. Remember that X is generated with the “**X,T,?,n**” key.	

After quitting and clearing the home screen, you could get the value for a given x with **VARS**-*YVARS*-1-1, finishing with “(7)” so that “Y1(7)” appears on the home screen, and ENTER. You should get the 0.531. Now let’s table this function.

2nd-TBLSET (top row). Since you know that 7 isn’t enough for 85%, you can start with 7 and use *?Tbl*=1. Leave the Auto’s.



2nd-TABLE. You can see that 11 is the first number of geeks such that there is an 85% probability that 2 or more had their experience on the same day.



Unfortunately, your calculator won’t go over 69! Of course, that’s fine for 30 days but what if you wanted to go much larger on days and/or people? You could put this formula in Excel, but it can only handle factorials of up to 170. Well you learned function tabling anyway.

So what do we do to solve this problem for larger numbers? We could look for a more clever function that wouldn’t require such large factorials. We could write a program so that calculations would test (days-1) against (days-people) and stop the multiplying as soon as possible. Or we could seek to get the same effect with some more or less clever technique in Excel and maybe learn something about Excel as well. We’ll do the latter.

Suppose the number of days is 170. Below is a solution for X number of people such that the probability is 50% that two or more have the same day. Below that graphic is the solution with formulas shown in the cells.

1	170	1	
2	169	0.994118	0.005882
3	168	0.982422	0.017578
4	167	0.965085	0.034915
5	166	0.942377	0.057623
6	165	0.91466	0.08534
7	164	0.882378	0.117622
8	163	0.846045	0.153955
9	162	0.806231	0.193769
10	161	0.763548	0.236452

11	160	0.718634	0.281366
12	159	0.672134	0.327866
13	158	0.624689	0.375311
14	157	0.576919	0.423081
15	156	0.529408	0.470592
16	155	0.482695	0.517305
17	154	0.437265	0.562735

1	170	1	
=A1+1	=B1-1	=C1*(B1-1)/$B$1	=1-C2
=A2+1	=B2-1	=C2*(B2-1)/$B$1	=1-C3
=A3+1	=B3-1	=C3*(B3-1)/$B$1	=1-C4
=A4+1	=B4-1	=C4*(B4-1)/$B$1	=1-C5
=A5+1	=B5-1	=C5*(B5-1)/$B$1	=1-C6
=A6+1	=B6-1	=C6*(B6-1)/$B$1	=1-C7
=A7+1	=B7-1	=C7*(B7-1)/$B$1	=1-C8
=A8+1	=B8-1	=C8*(B8-1)/$B$1	=1-C9
=A9+1	=B9-1	=C9*(B9-1)/$B$1	=1-C10
=A10+1	=B10-1	=C10*(B10-1)/$B$1	=1-C11
=A11+1	=B11-1	=C11*(B11-1)/$B$1	=1-C12
=A12+1	=B12-1	=C12*(B12-1)/$B$1	=1-C13
=A13+1	=B13-1	=C13*(B13-1)/$B$1	=1-C14
=A14+1	=B14-1	=C14*(B14-1)/$B$1	=1-C15
=A15+1	=B15-1	=C15*(B15-1)/$B$1	=1-C16
=A16+1	=B16-1	=C16*(B16-1)/$B$1	=1-C17

The answer is 15 people. In case you're not already familiar with Excel, the columns are lettered A,B,C,… and the rows are numbered 1,2,3,…as you will see when you open a sheet. Only the first two rows will need to be typed in as you can drag a formula down the column. Highlight the cell to be dragged and drag the little square on the lower right corner. The $B$1 allows the B1 reference to be maintained during this drag down procedure. This is called absolute referencing. The other references are relative references that Excel will change appropriately as you drag down the column

Before leaving this exercise about probability, let’s revisit the concept of hypothesis(es) and null hypothesis. The hypotheses are outcomes of an experiment. Your outcomes should be mutually exclusive and exhaustive is another way of saying that the probabilities of your hypothesis(es) added to the probability of the null should equal one. If all possible outcomes are not covered with the set of hypotheses including the null, you must change your hypothesis or add alternatives. This was discussed before but not in terms of probabilities.