The so-called 'birthday problem' is often found in the literature of statistics. On the next page there is a more mathematical discussion while this page shows some simulations.

Each one of the six diagrams has a square for each day of the year. Each row of squares is a month (top row is January) and each column is a day of the month. (1st of each month to the left). The crossed squares are non-existing dates.

Suppose that there is a sample of 30 persons and that their birthdays are recorded in the diagram. If there is more than one person in a square this is marked by a (slightly larger) red dot.

The probability of a sample of 30 persons having at least two persons sharing a birthday is 0.706. This high probability, in spite of the small number in the sample, is rather amazing and illuminates the difficulties of relying on intuition or common sense. We often need statistical methods together with knowledge of the process in order to use the result in the best way.

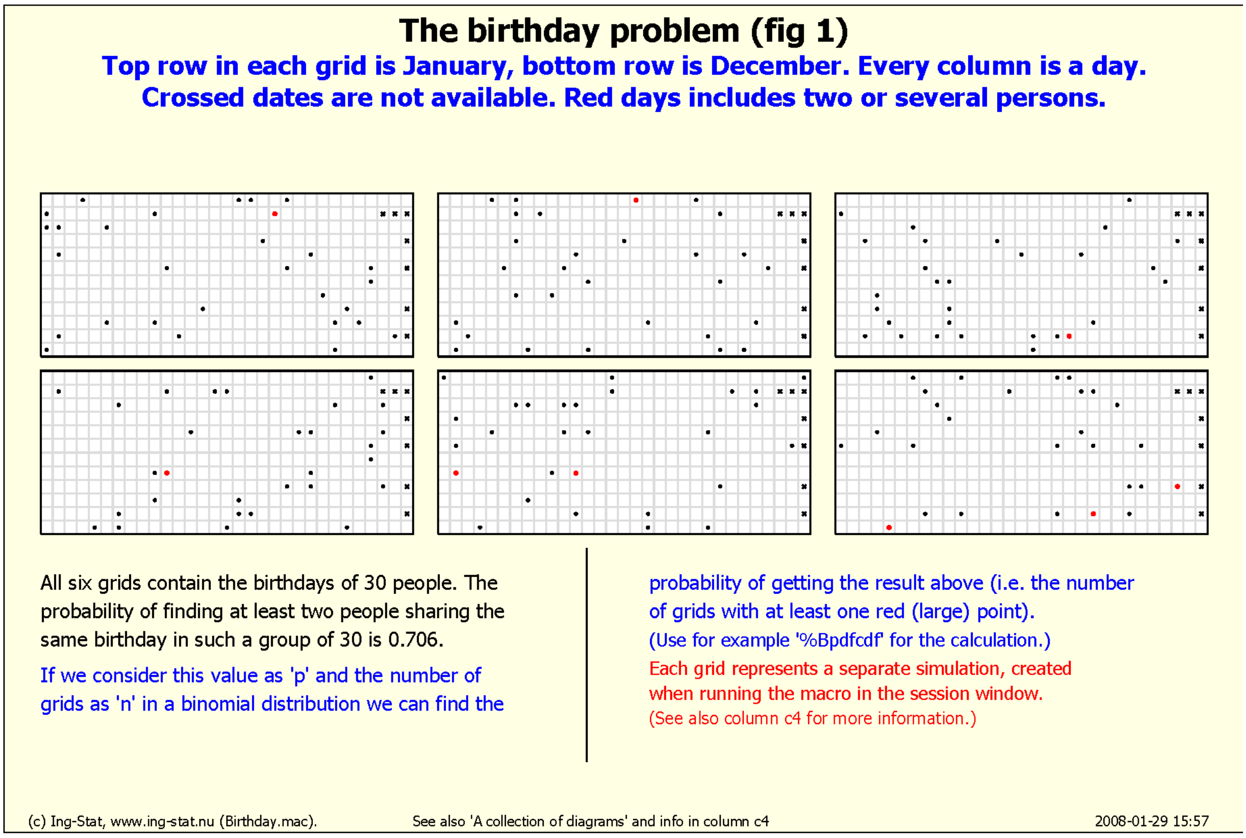
The sample of the six diagrams, with the probability of 0.706 of something happening (here at least two persons having the same birthday) is an example of the binomial distribution. The number of diagrams where the event has happened, is one observation from the Bin[6; 0.706]. The full distribution can be obtained via the %Bpdfcdf-macro.

(The most common way of exemplify the binomial distribution is otherwise 'we have inspected 50 items and found 3 incorrect items'.)

Macros and menus
%Bpdfcdf, %Birth

Data on the printout

There is no data on the printout



The birthday problem is often stated in books in mathematical statistics and is sometimes formulated in the following way:

"How many persons are needed in a group to have an approx 50 % chance of having at least two persons with the same birthday?"

We do not consider the year and we suppose that the birth rate is equal through the year. The answer is usually unexpected and is shown in the diagram and also in two data columns.

The result has been calculated for group sizes of 2 to 70 persons. This probability is calculated surprisingly easy. Firstly the probability of all persons having *different* birthdays. (of 2 persons, 3 persons, 4 persons...):

$$1 \cdot \frac{364}{365} \qquad 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \qquad 1 \cdot \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{362}{365}$$

Then we calculate (1 – the result) which gives the wanted probability. (1 minus the probability of *nobody* has the same birthday must mean that at least two birthdays are equal!)

Data on the printout

Group size (m)	=	Number of people per group
Number of grps (n)	=	Number of groups
Probability (p)	=	'Fault rate' (theoretical value)
Expected value (μ)	=	Expected value (μ)
Standard dev. (σ)	=	Standard deviation (σ)
Average ($\bar{x}$)	=	Mean 'faulty units'
Standard dev. (s)	=	Standard deviation calculated from data
Number of values	=	Number of sim.values (= number of 'runs')

Macros and menus
%Bpdfcdf, %Birth