|  |
| --- |
| Image result for birthday partyImage result for birthday party<https://www.google.com/search?rlz=1C1CHBF_enNZ793NZ794&biw=1032&bih=878&tbm=isch&sa=1&ei=NfLsXJ20LsTTz7sPtLWF2AE&q=birthday&oq=birthday> |
| We can see the happiest moments of people’s lives from birthday photos, regardless of age, place, time or culture.  Last December I learnt that Gerrit and Patrick have the same birthday – it was quite surprising for me, as I knew the probability of the same birthday in a family is quite low.  Can you guess how often it happens?  Of course, it is absolutely depends on the size of the family – a family of 10 have more chance to have the same birthday than a family of 3.  Today, I would like to discuss a bit more about this topic - How often do we have the same birthday in a group of people? |
| I simulated this special event for the groups of 2 to 100 people with samples of 100 to 1000. The figure below shows the outcome as a percentage that any two people in the group have the same birthday. |
|  |
| The point at the left-bottom corner tells us the probability of the same birthday when the group size is two is the same as or close to zero. From this figure, we can read that the probability of Gerrit and Patrick’s family having the same birthday is approximately less than 5%. However, we can also catch that the probability dramatically increases for a group of 10 to 30.  Some might doubt about why the shape is not a straight line; it seems non-sense?!?!  Look at the figure below where I forced a line (called linear model) to fit the data points. It seems like someone who wears a suit which is both to large and to small; either way it does not look smart or elegant.  What’s going on here? |
|  |
| Image result for exponentialIf it is not linear, what can we call it? – yes, it is called EXPONENTIAL.  Some might remember their high school maths about exponentials. To recap, the graph of an exponential relationship between two values looks like this. It does not match with our data points either.  If you are curious enough about the mathematics behind this: the examples of exponentials are not only , but also . The graph of exponential plots going high on the right and plots flattening on the right.  You may be able to tell that the plots of to the two sets of exponentials above reach to different heights – the orange plot stops at 100 (of course, the probability can never be over that value), and the blue goes to infinitely.  Go back to our birthday matching story…  I am interested in finding out about:   * What size group has a probability of greater than 50% that any two of them have the same birthday? * What size group has a probability that any two of them have the same birthday at about 90%? * If there are any differences in the outcomes from different sample sizes?   To figure these out, I made a spaghetti plot (nick name of a trace plot) from the same data - data points from the same sample size are connected and shaped like spaghetti. |
|  |
| Do you feel this plot gives you more information than the earlier ones?   * Do you agree that a group of 23 people has 50% of chance that any two of them have the same birthday? * Do you agree that a group of 41 people has 90% of chance that any two of them have the same birthday? * Do you agree that a smaller sample size results have more unstable outcomes than a bigger sample size?   Great! You are very patient to read this boring story ☺ |
| Enjoy your Queen’s Birthday Weekend EXPONENTIALLY!! |