Studio 1 Birthday Matches
18.05, Spring 2022
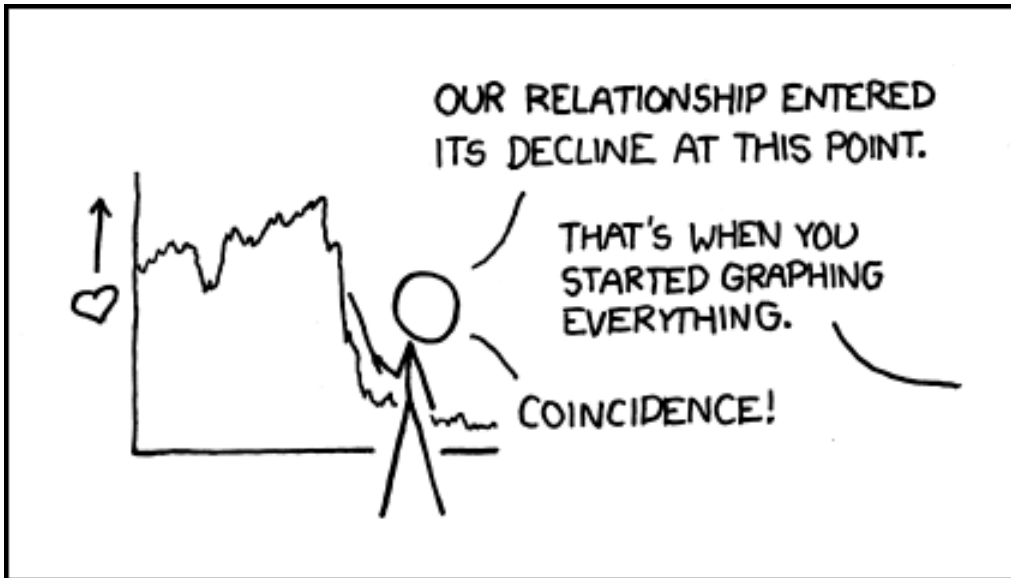
https://xkcd.com/523
Image courtesy of xkcd. License: CC BY-NC.

Overview of the studio

Suppose n people are gathered in a room. What is the probability that at least 2 of them will have the same birthday?

- We will use an R simulation to estimate this for various n.

- Our code will work for years of any given length. So Martians or Jovians could use our code.

- We will explore how the number of trials affect the variability of our estimates.

Download the zip file

- You should have downloaded the file `mit18_05_s22_studio1.zip` from our MITx site.

- Create an 18.05 studio folder somewhere on your computer and unzip this file there. You should see the following R files:
  `mit18_05_s22_studio1.r`
  `mit18_05_s22_studio1-start-here.r`
  `mit18_05_s22_studio1-start-here-include.r`
  `mit18_05_s22_studio1-samplecode.r`
  `mit18_05_s22_studio1-test.r`
  `mit18_05_s22_colMatches.r`

and the following other files:
`mit18_05_s22_Hints-RStudio.pdf`
`mit18_05_s22_studio1-test-answers.html`

General R instructions

If you've never used R Studio, take a minute to read the Hints-RStudio file in this packet. It is less than a page.

If in doubt: ask one of the staff!

Working on the assignment

- Detailed instructions on each problem are below.

- Do your work in the file `mit18_05_s22_studio1.r`. The structure of the file should be pretty clear. (We hope!)

- Do not ignore variables or print and cat statements we've included. They are meant to be used.

- For each problem, you will be putting your code inside a *wrapper* function. Look carefully at the arguments of that function. You must use them as described for each problem.

- The test cases give these arguments certain values for you to check your code.

- When we grade your code, we will use other values. So if you don't use the wrapper function arguments, we will not see correct results

General administrative instructions

1. Put your code answering the R questions in the file `mit18_05_s22_studio1.r`

2. Read all instructions carefully!

3. If code is given with a question it is meant to be used. Don't just ignore it. Likewise for print or cat statements.

4. **This is really important:** Before uploading the the code, read the sections (below) called *Before uploading your code* and *Upload your code*

5. **Due date**
The goal is to upload your work by the end of class.

If you need extra time, you can upload your work any time before 10 PM ET the day after the studio.

6. **Solutions uploaded**
Solution code will be posted on MITx at 10 PM the day after the studio.

Prepping R Studio

- Open R-studio

- From the *File* menu find and open `mit18_05_s22_studio1-start-here.r`

- From the Session menu, choose 'Set Working Directory' and pick 'To Source File Location'. This will let R find the files it needs.

- From the *Code* menu choose *Source* –note the keystroke equivalent. You should get the following message

  Congratulations: your working directory is set correctly!

  and see the following plot
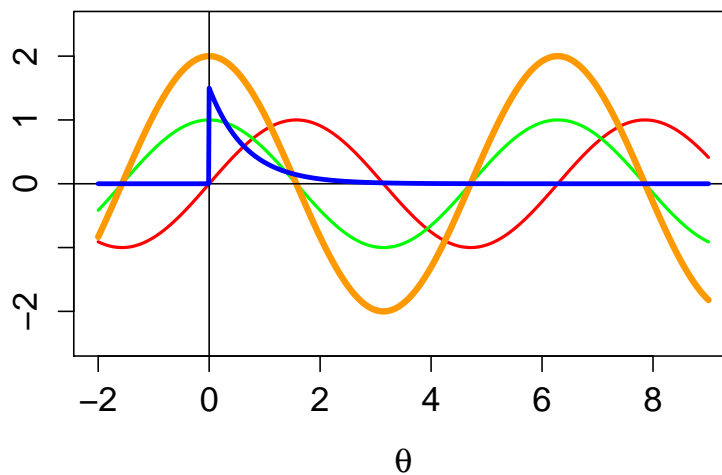
  If you get an error message beginning

  Error in file(filename, "r", encoding = encoding) :
  cannot open the connection

  it means you didn't set your working directory correctly. See the Hints-RStudio file for how to do that.

- From the *File* menu open `mit18_05_s22_studio1.r`. This is where you will work.

- Answer the questions in the detailed instructions just below.

- Solution code will be posted tomorrow at 10 pm

Detailed instructions for the studio

Before you get started and certainly **before** you upload your code, read the 'General administrative instructions' and 'General R instructions' sections above.

Problem 1. Use sections 1-5 of `mit18_05_s22_studio1-samplecode.r` as a tutorial to learn a little more about R and to find useful code for the later problems.

Problem 2a. This asks you to code a simulation of the birthday problem.

- You will put your code inside the function `studio1_problem_2a()`. The skeleton for this function is provided.

- The arguments are:
  `ndays_in_year` –This is the number of days in the year.
  `npeople` –This is the number of people in the room.
  `ntrials` –This is the number of times you run the simulation. Once the code is working, we will want this to be large enough to get an accurate estimate.

Your code should do the following

1. Randomly pick `npeople` birthdays (repetitions allowed) from a list of `ndays_in_year` birthdays.

2. Check if any 2 of the chosen birthdays are the same.

3. Repeat the previous two steps `ntrials` times and keep count of the number of trials in which at least one match is found.

4. Report the fraction of times at least one match was found.

After sourcing your code, look at studio1-test-answers.html. Find the function that was used to test problem 2a. Paste it into your console and run it. Compare your results to those shown.

Problem 2b. For this problem, take `ndays_in_year= 365` and `ntrials` large enough to get a good simulation (say 2000 or more)

Then, run the function `studio1_problem_2a()` with various values of `npeople` until you figure out the smallest number of people it takes to have at least a 0.5 probability of a match.

All the function studio1_problem_2b() does is print out the value of `npeople`. You don't need to write any code. Just set the value of `npeople` to the number you found.

Extra. Set `ndays_in_year= 365` and `npeople = 15`. Now run `studio1_problem_2a()` several times with `ntrials = 100`. Notice how the estimated probability varies. Repeat this with `ntrials = 500, 1000, 2000, 50`.

Problem 3. (Optional) Do this **only if you have time**. Use sections 6 and 7 in `mit18_05_s22_studio1-samplecode.r` to learn how to make a plot of the probability of a shared birthday as a function of npeople. Set `ndays_in_year= 365` and do a plot of estimated probability vs. npeople for `npeople = 1 to 100`.

After sourcing your code, look at `mit18_05_s22_studio1-test-answers.html`. Copy the function shown there that was used to test problem 3 into your console and run it. Compare your results to those shown.

Testing your code

For each problem, we ran the problem function with certain parameters. You can see the function call and the output in `mit18_05_s22_studio1-test-answers.html`. If you

call the same function with the same parameters, you should get the same results as in `mit18_05_s22_studio1-test-answers.html` – if there is randomness involved the answers should be close but not identical.

For your convenience, the file `mit18_05_s22_studio1-test.r` contains all the function calls used to make `mit18_05_s22_studio1-test-answers.html`.

Before uploading your code

1. Make sure all your code is in `mit18_05_s22_studio1.r`. Also make sure it is all inside the functions for the problems.

2. Clean the environment and plots window.

3. Source (run) the entire file.

4. Call each of the problem functions with the same parameters as the test file `mit18_05_s22_studio1-test-answers.html`.

5. Make sure it runs without error and outputs just the answers asked for in the questions.

6. Compare the output to the answers given in `mit18_05_s22_studio1-test-answers.html`.

Upload your code

Use the upload link on our MITx site to upload your code for grading.

Leave the file name as `mit18_05_s22_studio1.r`. (The upload script will automatically add your name and a timestamp to the file.)

You can upload more than once. We will grade the last file you upload.

Due date

Due date: The goal is to upload your work by the end of class.

If you need extra time, you can upload your work any time before 10 PM ET the day after the studio.

Solutions uploaded: Solution code will be posted on MITx at 10 PM the day after the studio.

Here's Johnny

The Tonight show in the 1970s and 80s was hosted by Johnny Carson. Here is his attempt to deal with the birthday question. Fashion back then was interesting.

Johnny Carson attempt 1

https://www.cornell.edu/video/the-tonight-show-with-johnny-carson-feb-6-1980-excerpt

Attempt 2: after getting hate mail from mathematicians

https://www.cornell.edu/video/the-tonight-show-with-johnny-carson-feb-7-1980-excerpt

Attempt 3:

https://www.cornell.edu/video/the-tonight-show-with-johnny-carson-feb-8-1980-excerpt

Here is the full NY Times article

https://opinionator.blogs.nytimes.com/2012/10/01/its-my-birthday-too-yeah/

MIT OpenCourseWare

https://ocw.mit.edu

18.05 Introduction to Probability and Statistics

Spring 2022