STAT107 Data Science Discovery

Lab: Justice

Man Fung (Heman) Leung

Fall, 2021

University of Illinois at Urbana-Champaign

- Please work in a group of 2–4 students
    - collaboration is important in data science!
    - meet new friends and discuss :)
    - let us know if you have any questions

- Common/potential mistakes
  - some reflections are really interesting! I hope you enjoy the lab overall
  - simulations, reflections and 4.2 are worth 2 points
  - other probability computations are worth 1 point
    - putting the number directly results in 0 point
  - 2.3: you can check "$<=2$" directly
  - 3.4: finding the exact probability is possible but difficult (and out of scope in this course). What you obtained in simulation is NOT the exact probability NOR more accurate
  - 4.3: simulation cannot give exact answer/predict stock price in general. One point will be deducted if the answer is too brief/does not make sense (according to the rubric)
- Running the test cases successfully do not imply full score

- Overview
  - similar to a k-nearest neighbors algorithm
- Ideas for extra credit (some were discussed on Discord)
  - vectorization using np.mean (sections 4 and 6)
  - subsampling (section 7)
  - other distance measure, e.g., sum of absolute difference (section 7)
  - filter the best tile, e.g., brighter/darker (section 7)
  - increase k in k-nearest neighbors (section 7)

- Main page
- Retrieve the lab using git
- Complete the notebook
    - pre-lab and post-lab survey for 2 extra points each
    - if you do not pass a test case, try the simulation once more
    - 1.2: for each iteration, generate a random integer (1-100). If it is ≤ 8, increase count by 1. Repeat 100 times and return count
    - 2.3: check here (or previous lab) if you forget how to use `df.sample`
    - 3: to rigourously justify our claim, we should conduct a hypothesis test (you can ignore this comment when you do the lab. Just trying to tell you something more)
- Submit your work. Feel free to:
    - ask us questions
    - leave whenever you finish the lab

Default total number of cells: 63

- 1.1 in cell 8 (reflection)
- 1.2 in cell 11–12
- 1.3 in cell 14
- 1.4 in cell 17
- 1.5 in cell 20 (reflection)
- 1.6 in cell 22, 24 (reflection)
- 1.7 in cell 27, 30

- 1.8 in cell 33, 36 (reflection)
- 1.9 in cell 38
- 2.1 in cell 43 (reflection)
- 2.2 in cell 46
- 2.3 in cell 48
- 2.4 in cell 53, 55, 58 (reflection)
- 3 in cell 61