STAT107 Data Science Discovery

Lab: Random Variable

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

Random story of the day

A story about copying and letting others to copy homework during my undergrad.

Practical experience of the day

A discrete random variable X is like a relationship between the possible value k and the associated probability (mass function) $\mathbb{P}(X = k)$. Therefore, given the probability (mass function), we can generate any arbitrary number of discrete random variables using sample with replacement.

Comment: Similarity

ILLINOIS

- Common/potential mistakes
  - 2.1/2.2: did not print 5 random rows correctly. I will deduct once only
  - 2.3: `.plot.hist()` is not accepted (you can see the output is strange)
  - 2.6: forgot to store the imputed columns in `df`/do the reflection
    - I let you go if you stored separately
    - some of you argued that removing the NA rows were better. It depends on the assumption in practice but fine for lab
    - filling with an arbitrary constant like 0 or -1 may not be a good choice

Comment: Similarity

ILLINOIS

- Common/potential mistakes
  - 4.1: it should not be none but this is open-ended
    - comments in 4.2 can be seen as response to 4.1
    - many of you forgot to do this one
  - 4.2 (worth 2 points): not using the given formula
  - 5.1: not making a tuple
  - 5.2: missing the `axis` or `args` argument
  - 5.3: did not sort the whole `df`
- Running the test cases successfully do not imply full score

Today's lab: Random Variable

ILLINOIS

- Main page
- Retrieve the lab using git
- Complete the notebook
  - hints are available by double clicking the question cells
  - 2.1: this is a sample without replacement problem (many possible solutions)
  - 2.4: edit the cell directly to not mess up total number of cells
  - 3.1: just guess (reasonable) numbers from the histogram
  - 3.2/3.3: (intuition, not hint) population quantity $\approx$ sample quantity
  - 3.4: edit the cell directly to not mess up total number of cells
- Submit your work. Feel free to:
  - ask us questions
  - leave whenever you finish the lab

Checking completion

Default total number of cells: 48

- 1.1 in cell 6
- 1.2 in cell 9
- 2.1 in cell 12–13
- 2.2 in cell 16
- 2.3 in cell 18
- 2.4 in cell 20 (reflection)
- 3.1 in cell 23–26, 28, 30
- 3.2 in cell 33, 35, 37
- 3.3 in cell 40, 42, 44
- 3.4 in cell 46 (reflection)