ILLINOIS

STAT107 Data Science Discovery

LAB: RANDOM VARIABLE

Man Fung (Heman) Leung

Spring, 2022

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
- Attendance form
  - you can come up if you do not want to use this form
  - submit before you leave the lab

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.

- Check email for score decomposition
- Some of you write a very good point: some experiments are immoral to be carried out and "simulation" may be a good substitution for analysis. However, we should make sure that our analysis is thorough
- **1.5**: -0.5 if you think it is due to chance but do not elaborate. The graph shows that it is unlikely due to chance so you need more elaboration if you write against the graph
    - Some of you write both arguments for and against that it is due to chance. Excellent work!
- **1.6**: 0 if you forget to write down the expected value, or write down anything other than 8

- 1.9: 0 if you use the wrong dataframe
- 2.3: 0 if you create df4 outside the function
- **2.5**: -0.5 if you think there is no discrimination but do not elaborate. This is similar to 1.5.
- **3**: -0.5 to -1 if your answer is too short (≤ 5 sentences as stated in the question)

- Main page
- Hints
    - 2.1: this is a sample **without** replacement problem. One possible solution is `df.sample()`
    - 2.3: use `density=1, bins=[2,3,4,5,6]` for better visualization
    - 3.1: just guess (reasonable) numbers from the histogram
- Submit your work. Feel free to:
    - ask us questions
    - leave whenever you finish the lab