S1201, Summer 2022

100 points **Midterm Exam** July 21 6:15-7:50PM EST

Name: **UNI:**

Instructions:

(i) Please submit a single PDF file on courseworks. You may scan your written solutions or directly write the solutions on your tablet. Please write your answers on the exam paper.

(ii) The exam is scheduled on July 21, 6:15-7:50 PM EST. Unless you have the permission from Ye or Yizi, please stop writing at 7:50 PM. You should scan and upload your solutions on Courseworks by 8:00 PM.

(iii) Please **keep your video on** during the exam. Make sure your computer camera is angled so I can see you writing at all times.

(iv) You are allowed to use a hand-held calculator. You can't use the calculator on your smartphone/iPad/computer.

(v) This is a **close-book/slides** exam. You can't access any course materials during the exam, but you are allowed to use a two-sided page of cheat sheet (of any size).

(vi) If you finish and want to submit your exam early, please message Ye first. Then you can submit your exam on Courseworks and quit from Zoom. I will mark the time stamp for your message and the submission.

(vii) During the exam, feel free to send a private chat to Ye for any questions!

(viii) Don't cheat! Cheating will lead to a straight zero, and will be reported to the university.

(ix) There are 100 points plus 10 bonus points, but the maximum score is 100.

(x) Good luck!

Please sign below to indicate your agreement with the Columbia College Honor Code, whether or not you are a student of Columbia College.

I affirm that I will not plagiarize, use unauthorized materials, or give or receive illegitimate help on assignments, papers, or examinations. I will also uphold equity and honesty in the evaluation of my work and the work of others. I do so to sustain a community built around this Code of Honor.

Signature:

1. **$2 \times 10 = 20$ points** TRUE/FALSE questions. No explanations are needed.

   (a) The probability density function, $f(x)$, must be less than or equal to 1 for any x.

   (b) If for events A and B, we have $\mathbb{P}(A) = 0.3$, $\mathbb{P}(B) = 0.7$, then it is possible for A and B to be mutually exclusive.

   (c) We can determine the skewness of the data from a boxplot.

   (d) If $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \times \mathbb{P}(B) \times \mathbb{P}(C)$, then events A, B and C are mutually independent.

   (e) The expectation and variance of a binomial distribution are always the same.

   (f) The total area of the bars in a density histogram is 1.

   (g) For any two events A and B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

   (h) For two independent events, knowing one event gives partial information of the other event.

   (i) Median and IQR are less sensitive to extreme values than mean and standard deviation.

   (j) If $X_1, X_2, \dots, X_n$ is a sample from some distribution with mean μ (not necessarily independent), then $\mathbb{E}\bar{X} = \mu$, where $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$.

2. **$6 \times 3 = 18$ points** A box in a supply room contains 15 compact fluorescent lightbulbs, of which 5 are rated 13-watt, 6 are rated 18-watt, and 4 are rated 23-watt. Suppose that three of these bulbs are randomly selected **without replacement**. The order of selection does **NOT** matter here.

   Do not calculate the combination numbers. Leave them as they are in the final answer.

   (a) What is the probability that all three of the bulbs are rated 18-watt?

   (b) What is the probability that one bulb of each type is selected?

   (c) What is the probability that exactly two of the selected bulbs are rated 23-watt? (Hint: The third one can be rated either 13-watt or 18-watt, so you have two different situations. Count them separately then add up the numbers of outcomes in two cases.)

3. **$10 \times 2 = 20$ points** Lab tests produce positive and negative results. Assume that for a lab test, 90% of patients with disease obtain positive results and only 1% of patients without disease obtain positive results. Assume that 2% of the population has the disease.

   (a) Pick one person at random. What is the chance that lab test will be positive?

   (b) Pick one person at random. The lab test shows positive result. What is the chance that the person really has the disease?

4. **$6 \times 3 = 18$ points** A contractor is required by a county planning department to submit one, two, three, four, or five forms (depending on the nature of the project) in applying for a building permit. Let Y = the number of forms required of the next applicant. The probability that y forms are required is known to be proportional to y — that is, $p(y) = ay$ for $y = 1, 2, \ldots, 5$.

   (a) What is the value of a for $p(y)$ to become a valid pmf?

   (b) Based on the value of a you obtained in part (a), what is the probability that at most three forms are required?

   (c) Based on the value of a you obtained in part (a), what is the probability that between two and four forms (inclusive) are required?

5. **$8 \times 3 = 24$ points** Some summary statistics for midterm scores (number of points; the maximum possible score was 50 points) from an intro stats class are

Min	Q1	Median	Q3	Max	Mean	SD
58	76	82	86	100	79	5.63

   (a) What are the IQR and range?

   (b) Would you expect the score distribution to be symmetric, right-skewed or left-skewed? Why?

   (c) Are any of the midterm score outliers? Please explain. (Here the outliers mean the suspected outliers as in the box plot.)

6. **Extra question with 10 bonus points**

   How do you think statistics can be applied in your field of expertise? Can you give any examples? (Full marks for any reasonable answers — I cannot help you more!!!)