Write **clearly** and **in the box**:

CSCI 3022
Midterm Exam
Spring 2021

Name:
Student ID:
Section number:

Read the following:

- **RIGHT NOW**! Write your name, student ID and section number on the top of your exam. If you're handwriting your exam, include this information at the top of the first page!
- You may use the textbook, your notes, lecture materials, and Piazza as resources. Piazza posts should not be about exact exam questions, but you may ask for technical clarifications and ask for help on review/past exam questions that might help you. You may not use external sources from the internet or collaborate with your peers.
- You may use a calculator or Python terminal to check numerical results.
- If you print a copy of the exam, clearly mark answers to multiple choice questions in the provided answer box. If you type or hand-write your exam answers, write each problem on their own line, clearly indicating both the problem number and answer letter.
- Mark only one answer for multiple choice questions. If you think two answers are correct, mark the answer that **best** answers the question. No justification is required for multiple choice questions. For handwriting multiple choice answers, clearly mark both the number of the problem and your answer for each and every problem.
- For free response questions you must clearly justify all conclusions to receive full credit. A correct answer with no supporting work will receive no credit.
- The Exam is due to Gradescope by midnight on Friday, March 12.
- When submitting your exam to Gradescope, use their submission tool to mark on which pages you answered specific questions. Submitting your exam properly is worth 1/100 points. The other problems sum to 99.

Multiple choice problems: Write your answers in the boxes if using a printed version of the exam.

1. (3 points) Consider the data set: $[4, 10, 9, 19, 0, x]$, where $x \in \mathbb{R}$ is an unknown quantity. What is the smallest set of possible values that the *median* of this data set **must** belong to?

   A. $(-\infty, \infty)$

   B. $[6.5, 9.5]$

   C. 9

   D. $[0, 19]$

   E. $\{4, 9, 10\}$

   F. $\emptyset$

   B

2. (3 points) Suppose Zach has a list consisting of all the first generation Pokémon and their types (Water, Ground, Fighting, etc.). He is conducting a study of how many of them are actually stronger than Mudkip - the cutest Pokémon ever - by drawing a sample from his Pokédex, which he has sorted alphabetically.

   He writes a loop that will randomly pick two Pokémon *of each type* and compares those selected Pokémon's statistics to Mudkip's.

   What type of sample did Zach collect?

   A. Simple random sample

   B. Systematic sample

   C. Census sample

   D. Stratified sample

   E. Free samples, all you can eat!

   D

3. (3 points) Consider performing a simulation experiment where we record whether or not the events A and B happened. We perform the experiment n times, and store whether or not the events occurred each time into logical vectors of length n. What does the code below calculate?

```
def CountOutcomes(eventA,eventB):
  return np.sum(np.logical_and((eventA==True),(eventB==True))) / np.sum((eventA==True))
```

   A. $P(A \cup B)$

   B. $P(A \cap B)$

   C. $P(A|B)$

   D. $P(B|A)$

   E. P(both flips are heads)

   D

4. (3 points) We're considering a random variable to describe major earthquakes (above 8 Richter). We want to *count* how many such earthquakes occur per year. What variable is most appropriate?

   A. Binomial

   B. Negative binomial

   C. Uniform

   D. Normal

   E. Poisson

   F. Exponential

   E

Use the following information for Problems 5 – 7, which may build off of each other.

Ani (A) has run out of interesting games to play over quarantine, and now is stuck playing a rather bland variant of Snakes and Ladders with Blaine (B). In this game, each player tries to escape a maze.

Suppose that in general, Ani escapes 40% of the time and Blaine escapes 35% of the time.

5. (3 points) Both players are able to escape in 1/4 of the games played. What is the probability that neither escape?

   A. 1/10
   B. 3/20
   C. 1/6
   D. 1/5
   E. 1/4
   F. 1/2
   G. 3/4
   H. 4/5
   I. 5/6
   J. 17/20
   K. 19/20

   F

6. (3 points) Suppose that Ani **fails** to escape. Now what is the probability that Blaine escapes?

   A. 1/10
   B. 3/20
   C. 1/6
   D. 1/5
   E. 1/4
   F. 1/2
   G. 3/4
   H. 4/5
   I. 5/6
   J. 17/20
   K. 19/20

   C

7. (3 points) What is the probability that exactly one of them escapes?

   A. 1/10
   B. 3/20
   C. 1/6
   D. 1/5
   E. 1/4
   F. 1/2
   G. 3/4
   H. 4/5
   I. 5/6
   J. 17/20
   K. 19/20

   E

8. (3 points) The average high temperature for Boulder, CO in March is 57° Fahrenheit with a standard deviation of 12 degrees. If the temperature C in Celsius is calculated from the temperature in Fahrenheit F by $C = \frac{5}{9}(F - 32)$, what is the *standard deviation* of the temperature in Boulder on March 7 in degrees Celsius?

   A. $\frac{5}{9} \cdot (57 - 32)$
   B. $12^2 \cdot \frac{5^2}{12^2}$
   C. $12 \cdot \frac{5}{9}$
   D. $\left(\frac{5}{9}\right)^2 25^2$
   E. $12^2 \cdot \frac{5}{9}$
   F. $12 \cdot \left(\frac{5}{9}\right)^2$

   C

9. (3 points) Suppose when you get to the cafeteria, you and your best friend have a competition to determine who gets the sushi rolls. You each type NP.RANDOM.RAND() into a terminal, and the person who gets the higher number wins and gets a sushi roll. You play the game until all 7 available sushi rolls are claimed. What is the probability that you get 5 or more of them?

   A. $\binom{7}{5}(0.5)^2(0.5)^5$

   B. $\binom{7}{2}(0.5)^2(0.5)^7$

   C. $\binom{7}{5}(0.5)^2(0.5)^5$

   D. $1 - \sum_{i=4}^{7}\binom{7}{i}(0.5)^i(0.5)^{17-i}$

   E. $\sum_{i=5}^{7}\binom{i}{7}(0.5)^i(0.5)^{7-i}$

   F. $1 - \sum_{i=0}^{4}\binom{7}{i}(0.5)^i(0.5)^{7-i}$

F

10. (3 points) Suppose we have a real-valued random variable with pdf/pmf f and cdf F. Consider each function where we input an outcome x. Which of the following can be said about the *relative* magnitudes of $f(x)$ and $F(x)$?

   A. $f(x) \geq F(x)$ for both discrete and continuous random variables.

   B. $f(x) \geq F(x)$ for discrete but not continuous random variables.

   C. $f(x) \geq F(x)$ for continuous but not discrete random variables

   D. $F(x) \geq f(x)$ for both discrete and continuous random variables.

   E. $F(x) \geq f(x)$ for discrete but not continuous random variables.

   F. $F(x) \geq f(x)$ for continuous but not discrete random variables.

   G. None of the above are *always* true.

E

11. (3 points) Suppose we have a discrete rv X with pdf f. We then decide to take X and compute $Z = X^2 + 2X + 1$. Which of the following represent the expected value of Z?

   A. $E[X]^2 + 2E[X] + 1$

   B. $E[(X+1)^2]$

   C. $Var[X]$

   D. $Var[X] + 2E[X] + 1$

   E. $Var[X] + (E[X]+1)^2$

   F. $Stdev[X] + 2E[X] + 1$

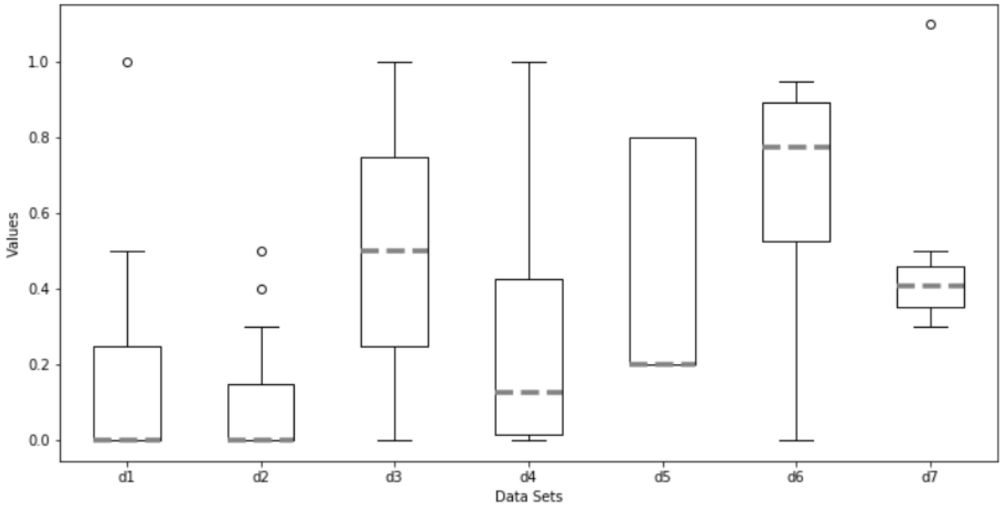
B or E

Free Response problems: Write your answers in the spaces following each prompt if possible. Make note if your work continues elsewhere!

12. (20 points) No justification is necessary for this problem. Consider the given seven box plots, each from a data set with exactly 15 elements. Three of the data sets were generated by:

```
A = np.linspace(0,1,15)
B = [x**3 for x in np.linspace(0,1, num=15)]
C = [0,0,0,0,0,0,0,0,0,.1,.2,.3,.4,.5,1.0]
```

(a) (5 points) Match each of the three given data sets to their boxplot (box-whisker plot). Use the conventions from lecture, and clearly mark the number corresponding to your choice of boxplot in the boxes below for each data set. No boxplot is used more than once, and some are not used.



A: d3	*B*: d4	*C*: d1

(b) (5 points) For the data sets labeled $d2, d3$ and $d6$ in the image above, characterize the data set as symmetric, right-skew, or left-skew. No justification needed.

$d2$: Right-skew	$d3$: Symmetric	$d6$: Left-skew

(c) (4 points) Some of the data sets in question have *repeated* observations, where the same data value occurs multiple times. Using only the boxplots shown, which of the 9 data sets **must** have repeated observations in them? Explain why.

(d) (3 points) Is the mean of $d1$ or $d2$ likely to be higher? Explain.

(e) (3 points) The data sets above were meant to represent possible observations of *probabilities* or *proportions*. Could any of the data sets above not have been proportion data? Why or why not?

13. (10 points) In order to get a little stress relief, we're volunteering for the humane society, which is currently sheltering 23 puppies (P) and 14 kittens (K). The adorable creatures usually put us in a good mood (G), to the degree that we end up having a good time if we're petting a puppy 85% of the time and we end up having a good time 70% of the times we're petting a kitten.

Suppose on a Monday we're assigned one of the available pets at random to play with when we enter the humane society.

(a) (6 points) After we're done, we reflect on our time and realize we didn't actually have a good time. What's the probability that the random pet we were assigned was a puppy?

(b) (4 points) Are the events $P(G|P)$ and $P(G)$ independent? Justify your response.

Solution:

(a) We want $P(P|\text{not } G)$ By Bayes theorem and the law of total probability:

$$P(P|\text{not } G) = \frac{P(\text{not } G|P)P(P)}{P(\text{not } G)} = \frac{0.15 * \frac{23}{37}}{0.15 * \frac{23}{37} + 0.3 * \frac{14}{37}}$$

(b) Awkward question. For one, *probabilities* aren't independent, events are. This is a valid answer. For two, the *events* $G|P$ and G are trivially not independent, since one is a subset of the other: once G happens either way, the other one is inherently determined. But the intended question of P and G being independent is also not true: we demonstrate that in A since $P(P|G) \neq P(P)$.

14. (16 points) Suppose whenever you roll a six-sided die and it comes up as a ”2”, you yell the word ”deuces” as loud as you can. You enjoy this activity so much that you manufacture a biased die. This die rolls a two 2/3 of the time, with each other other outcome equally likely.

   (a) (4 points) What is the probability mass function for the face of the die?

   (b) (4 points) What is the probability that it takes you 5 rolls before you can yell “deuces” twice?

   (c) (4 points) What is the expected value of the face of the die after rolling?

   (d) (4 points) What is the expected value of the square root of the face of the die after rolling?

Solution:

(a) $f(x) = \begin{cases} 2/3 & \text{for } x = 2 \\ 1/15 & \text{for each other face} \end{cases}$

(b) ”it takes you” was meant to be exact here, which is the negative binomial with $p = 1/2$, since that’s the rate of 2’s:

$$\binom{4}{1}(1/2)^2(1/2)^3$$

. Interpreting it as ”5 rolls or more” is fine too, but this is an infinite sum on the negative binomial and is easiest by doing one minus the outcomes of ”exactly 2 3 or 4”.

(c) We sum outcome times it’s probability:

$$E[X] = 2(2/3) + (1/15)(1 + 3 + 4 + 5 + 6) = 3$$

(d) We sum the root of outcome times it’s probability:

$$E[\sqrt{X}] = \sqrt{2}(2/3) + (1/15)(\sqrt{1} + \sqrt{3} + \sqrt{4} + \sqrt{5} + \sqrt{6}) = 3$$

15. (20 points) Consider a random variable X given by probability density function $f(x) = ax(3-x)^2$ on the region $x \in (0,3)$ with an unknown constant a.

    (a) (4 points) Sketch the pdf of X on a standard Cartesian axis, labeling any important features.

    (b) (4 points) Find the value of a that will make f a proper pdf. Use this value for the rest of the problem.

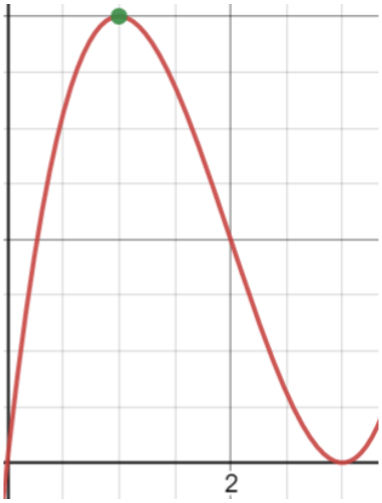
    (c) (3 points) Calculate $E[X]$.

    (d) (3 points) Calculate the cumulative density function for X.

    (e) (3 points) Calculate $Var[X]$.

    (f) (3 points) Is the median for x greater than or less than its mean? You may compute exactly or explain your result graphically.

Solution:



(a) Make a reasonable sketch!

(b) Check that it integrates to 1:

$$\int f(x) = a\int_0^3 9x - 6x^2 + x^3 = a\left(\frac{9x^2}{2} - 2x^3 + \frac{x^4}{4}\right)|_0^3 = a\frac{27}{4}$$

. So a must be 4/27.

(c) Now we integrate:

$$\int xf(x) = a\int_0^3 x\left(9x - 6x^2 + x^3\right) = \frac{6}{5}$$

(d) This is the same integral as in part b!

$$F(x) = \int_0^x f(z)\,dz = a\left(\frac{9z^2}{2} - 2z^3 + \frac{z^4}{4}\right)|_0^x = a\left(\frac{9x^2}{2} - 2x^3 + \frac{x^4}{4}\right)$$

(e) First we find

$$E[X^2] : \int x^2 f(x) = a\int_0^3 x^2\left(9x - 6x^2 + x^3\right) = \frac{9}{5}.$$

Then via shortcut formula, variance is $\frac{9}{5} - \left(\frac{6}{5}\right)^2 = \frac{9}{25}$

(f) Median is less than the mean since the given region is right-skewed. Or check from c and d!