CSCI 3022
Midterm Exam
Summer 2018

Name: ______________________

By writing my name I promise to abide by the Honor Code

Read the following:

- **RIGHT NOW**! Write your name on the top of your exam.
- You are allowed one 8 1/2 × 11in sheet of **handwritten** notes (both sides). No magnifying glasses!
- You may use a calculator provided that it cannot access the internet or store large amounts of data.
- You may **NOT** use a smartphone as a calculator.
- Clearly mark answers to multiple choice questions on the provided answer line.
- 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 free response questions you must clearly justify all conclusions to receive full credit. A correct answer with no supporting work will receive no credit.
- If you need more space for free-response questions, there are blank pages at the end of the exam. If you choose to use the extra pages, make sure to **clearly** indicate which problem you are continuing.

Page	Points	Score
2	15	
3	12	
4	9	
5	9	
6	18	
7	18	
8	18	
For Luck!	1	1
Total	100	

1. (3 points) Halfway through the semester, an instructor emails every third person from the course roster, to ask about their favorite kind of donut. Identify the type of sample.

   A. Census

   B. Simple

   C. Stratified

   D. Systematic

   E. Delicious

1. **D**

2. (3 points) You are sampling the weights of various puppies from a population with a known mean of 10 pounds and variance of 9 pounds^2. You obtain a measurement of $X = 19$ pounds. What is the corresponding value of the *standardized* normal random variable, Z?

   A. 1

   B. 3

   C. 4

   D. 9

   E. 19

2. **B**

3. (3 points) Chris and Dan are playing a nice summer game of catch. How fun! They are rather clumsy, however. The probability that Dan drops the ball is 0.5, the probability that Chris drops the ball is 0.4, and the probability that neither of them drop the ball is 0.4. What is the probability that **both** Chris **and** Dan drop the ball?

   A. 0.1

   B. 0.2

   C. 0.3

   D. 0.9

   E. 1

3. **C**

4. (3 points) Let f be a probability density function for a random variable X defined by $f(x) = cx^2$ for x in $[0, 1]$ and $f(x) = 0$ otherwise. What is c?

   A. 1

   B. 2

   C. 3

   D. 4

   E. 5

4. **C**

5. (3 points) Let a probability density function be given by $f(x) = 2x$ for $0 \leq x \leq 1$ and 0 otherwise. What is the cumulative distribution function $F(x)$, for $0 \leq x \leq 1$?

   A. $F(x) = x^2$

   B. $F(x) = 2x$

   C. $F(x) = 1 - 2x$

   D. $F(x) = 1 - x^2$

5. **A**

6. (3 points) Tony, arriving at a bus stop, just misses the bus. Suppose that he decides to walk if the (next) bus takes longer than 5 minutes to arrive. Suppose also that the time in minutes between the arrivals of buses at the bus stop is a continuous random variable with a $U(4, 9)$ distribution. What is the probability that Tony will end up walking?

   A. 0

   B. 1/5

   C. 3/5

   D. 4/5

   E. 1

6. **D**

7. (3 points) Suppose you're playing a game with your friends, in which in each turn, a player rolls five fair six-sided dice all at once. Let X be a random variable representing the number of turns before someone rolls all 6s. What distribution does X follow?

   A. Binomial

   B. Poisson

   C. Geometric

   D. Uniform

   E. Exponential

7. **C**

8. (3 points) Suppose you are playing a slot machine where the probability of winning on any given round is independent of the other rounds and constant at p. You decide to only play 100 rounds, and then leave. Let X be the random variable representing the number of rounds you win. What distribution does X follow?

   A. Binomial

   B. Poisson

   C. Geometric

   D. Exponential

   E. Negative binomial

8. **A**

9. (3 points) You are still playing the slot machine from #8, and have calculated that you manage to win on average 20 rounds per hour. Let X be a random variable that represents the amount of time between your wins. Which distribution would be most appropriate for X to follow?

   A. Poisson

   B. Uniform

   C. Binomial

   D. Exponential

   E. Negative binomial

9. **D**

10. (3 points) Consider the following function related to hits at a website. What distribution does the return value of the function belong to?

```
def not_doing_homework_dot_com(par):
    x = 0
    t = np.random.exponential(1/par)
    while t <= 1:
        x += 1
        t += np.random.exponential(1/par)
    return x
```

A. Binomial

B. Geometric

C. Poisson

D. Uniform

E. Exponential

10. **C**

11. (3 points) Consider the following function related to a series of coin flips with a biased coin that lands Heads with probability p. What distribution does the return value of the function belong to?

```
def flippy_flip(p):
    x = 0
    cnt = 5
    while cnt > 0:
        flip = np.random.choice([0,1], p=[1-p, p])
        if flip==1:
            cnt += -1
        x += 1
    return x
```

A. Binomial

B. Geometric

C. Poisson

D. Negative binomial

E. Exponential

11. **D**

12. (3 points) Consider simulating the roll of a fair, six-sided die. Which of the following quantities does the following function estimate?

```
def rolly_polly(num_samples):
    rolls = np.random.choice([1,2,3,4,5,6], size=num_samples)
    tot = np.sum(rolls == 6) + np.sum(rolls < 3) - np.sum(np.logical_and(rolls == 6, rolls < 3))
    return tot/num_samples
```

A. $P(X = 6 \cap X < 3)$

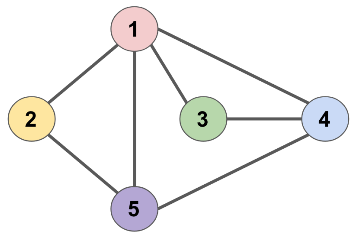
B. $P(X = 6 \cup X < 3)$

C. $P(X = 6 \mid X < 3)$

D. $P(X < 3 \mid X = 6)$

12. **B**

13. (3 points) Consider moving randomly on the following graph. A move on the graph consists of moving from the current node along an edge to one of the connected nodes with equal probability. You must always move, and you may move "backwards". Starting from Node 1, what is the probability that you visit Node 5 in two or fewer moves?



A. 0
B. 2/24
C. 5/24
D. 11/24
E. 1

13. **D**

14. (3 points) Suppose you are sampling random variable X, which is normally distributed with mean 2 and standard deviation 1. Which of the following would be appropriate to calculate the probability $P(3 < X < 4)$? Φ denotes the **standard** normal cumulative distribution function.

A. $1 - \Phi(2)$
B. $1 - \Phi(4)$
C. $\Phi(1) + \Phi(2)$
D. $\Phi(3) + \Phi(4)$
E. $\Phi(2) - \Phi(1)$
F. $\Phi(4) - \Phi(3)$

14. **E**

15. (3 points) Let $X \sim Pois(\lambda)$. Which of the following should you use to compute $P(X > 3)$?

A. $\displaystyle\sum_{k=0}^{3} \frac{\lambda^k e^{-\lambda}}{k!}$
B. $\displaystyle\frac{\lambda^3 e^{-\lambda}}{3!}$
C. $\displaystyle 1 - \sum_{k=0}^{3} \frac{\lambda^k e^{-\lambda}}{k!}$
D. $\displaystyle 1 - \frac{\lambda^3 e^{-\lambda}}{3!}$

15. **C**

16. (18 points) Consider the following data: 1, 2, 5, 6, 6, 7, 7, 7, 8, 10, 10

(a) Compute the quartiles Q_1, Q_2, and Q_3 as well as the IQR for this data set. Use the conventions discussed in lecture.

Solution:
The values above are in order. There are $n = 11$ items.
$\Rightarrow$ Q2 (also called the median) is then the 6th item. So $\boxed{\text{Q2} = 7}$

We now split the data set into the upper and lower halves, and **duplicate the median, placing it in both halves**: $\Rightarrow$ lower half: $\{1, 2, 5, 6, 6, 7\}$, and top half: $\{7, 7, 7, 8, 10, 10\}$

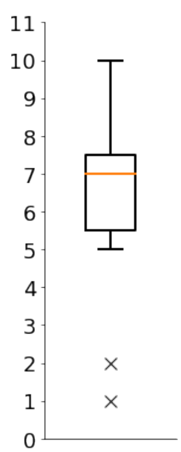
$\Rightarrow$ Q1 is the median of the lower half, so $\boxed{\text{Q1} = \frac{5+6}{2} = 5.5}$, and

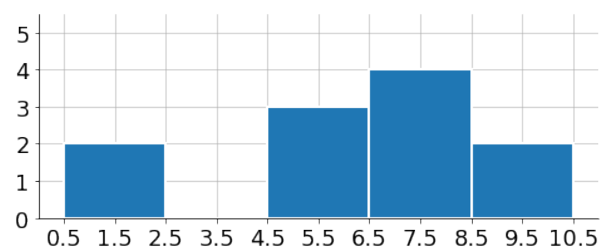
$\Rightarrow$ Q3 is the median of the upper half, so $\boxed{\text{Q3} = \frac{7+8}{2} = 7.5}$

Finally, the IQR is the difference between Q1 and Q3: $\boxed{\text{IQR} = \text{Q3-Q1} = 7.5 - 5.5 = 2}$

(b) Using the axes on the left, draw a box-and-whisker plot for the data set. Using the axes on the right, and bin edges at 0.5, 2.5, 4.5, 6.5, 8.5 and 10.5, draw a frequency histogram for the data. Again, use the conventions discussed in lecture.

Solution:
- Box from Q1 to Q3, 5.5 to 7.5, with median line at 7
- Whiskers extend from Q1 and Q3 *at most* $1.5 \times \text{IQR} = 3$
- Bottom whisker extends as far down as 2.5 but the nearest data point inside is at 5. Therefore, the whisker extends down only to 5.
- Top whisker extends as far up as 10.5 but the nearest data point inside is at 10. Therefore, the whisker extends up only to 10.
- This makes 1 and 2 outliers (or fliers), and they must be marked in some clear way.





(c) Classify the distribution of the data as symmetric, positively skewed, or negatively skewed. Clearly justify your response.

Solution:
The two outliers on the low side and the fact that Q2 is quite close to Q3 demonstrate that the distribution is **negatively skewed** (or left skewed). The histogram also displays this left-skew clearly.
You could also compare the median ($\tilde{x} = 7$) against the mean ($\bar{x} \approx 6.27$), and note that $\tilde{x} > \bar{x}$ means the distribution is likely left-skewed.

17. (18 points) You are an analyst charged with the task of gauging support for a new ballot measure. You find that the probability of a Democrat supporting this ballot measure is 0.2, the probability of a Republican supporting this ballot measure is 0.8, and the probability of an Independent supporting the measure is 0.4. Furthermore, you know that in your area, 60% of voters are registered Democrats, 30% are registered Republicans, and 10% are registered Independents.

    (a) You interview a voter at random. What is the probability that they support this ballot measure?

    Solution:
    Let D = event that voter is a Democrat, R = ... Republican, I = ... Independent.
    Let S = event that voter supports the ballot measure.
    Then we want $P(S)$. We can use the Law of Total Probability:

$$\begin{aligned} P(S) &= P(S \mid D)P(D) + P(S \mid R)P(R) + P(S \mid I)P(I) \\ &= 0.2 \cdot 0.6 + 0.8 \cdot 0.3 + 0.4 \cdot 0.1 \\ &= 0.12 + 0.24 + 0.04 \\ &= \boxed{0.4} \end{aligned}$$

    (b) You interview someone at random and find out that they support this ballot measure. Given this information, what is the probability that they are a Republican?

    Solution:
    We want $P(R \mid S)$. Here, we can use our answer from Part (a) and Bayes' theorem, one of the greatest gifts ever bestowed upon humanity:

$$\begin{aligned} P(R \mid S) &= \frac{P(S \mid R)P(R)}{P(S)} \\ &= \frac{0.8 \cdot 0.3}{0.4} \\ &= 2 \cdot 0.3 \\ &= \boxed{0.6} \end{aligned}$$

    (c) Are the events "voter is a Republican" and "voter supports this ballot measure independent? Justify your answer **using math**.

    Solution:
    We can answer this by checking one of three possible statements:

    - $P(R \mid S) \stackrel{?}{=} P(R)$
    - $P(S \mid R) \stackrel{?}{=} P(S)$
    - $P(S \cap R) \stackrel{?}{=} P(S)P(R)$

    Since we have both of the ingredients for the first one, we use that.
    From Part (b), $P(R \mid S) = 0.6$. From the problem set-up, $P(R) = 0.3$. These are **not equal**, so the two events are **not independent**.

18. (18 points) The probability distribution of a discrete random variable X is given by

$$P(X=-1)=\frac{1}{10},\quad P(X=0)=\frac{1}{10},\quad P(X=1)=\frac{3}{10},\quad P(X=2)=\frac{5}{10}$$

(a) Compute $E[X]$

Solution:

$$\begin{aligned}
E[X] &= \sum_x x\cdot P(X=x) = -1\cdot P(X=-1)+0\cdot P(X=0)+1\cdot P(X=1)+2\cdot P(X=2)\\
&= -1\cdot\frac{1}{10}+1\cdot\frac{3}{10}+2\cdot\frac{5}{10}\\
&= \frac{2}{10}+\frac{10}{10}=\frac{12}{10}=\boxed{\frac{6}{5}=1.2}
\end{aligned}$$

(b) Let Y be the random variable $Y=X^2+1$. Write down the probability distribution of Y.

Solution:
The possible values of Y are 1, 2 and 5, as shown below:

$$(-1)^2+1=2,\quad (0)^2+1=1,\quad (1)^2+1=2,\quad (2)^2+1=5$$

The probabilities are:
$P(Y=1)=P(X=0)=\frac{1}{10}=0.1$
$P(Y=2)=P(X=1)+P(X=-1)=\frac{1}{10}+\frac{3}{10}=\frac{4}{10}=0.4$
$P(Y=5)=P(X=2)=\frac{5}{10}=0.5$

(c) Compute $E[Y]$

Solution:

$$\begin{aligned}
E[Y] &= \sum_y y\cdot P(Y=y) = 1\cdot P(Y=1)+2\cdot P(Y=2)+5\cdot P(Y=5)\\
&= 1\cdot\frac{1}{10}+2\cdot\frac{4}{10}+5\cdot\frac{5}{10}\\
&= \frac{1+8+25}{10}=\frac{34}{10}=\boxed{\frac{17}{5}=3.4}
\end{aligned}$$

(d) Compute $\mathrm{Var}(X)$

Solution:
There is no way to use the definition of Variance to calculate this, because we do not actually have any sense of what the data are.
So we use the formula:

$$\mathrm{Var}(X)=E[X^2]-E[X]^2$$

And we can obtain $E[X^2]$ from the linearity of $E[Y]$:

$$E[Y] = E[X^2 + 1] = E[X^2] + 1$$

$$E[X^2] = E[Y] - 1 = 3.4 - 1 = 2.4$$

$$\begin{aligned} \mathrm{Var}(X) &= E[X^2] - E[X]^2 \\ &= 2.4 - (1.2)^2 \\ &= 2.4 - 1.44 \\ &= \boxed{0.96} \end{aligned}$$