Name ____

Fall 2020 **STAT 587-2** **Exam I**

(53 points)

Instructions:

- You are allowed to use any resource except aid from another individual.
- Aid from another individual, will automatically earn you a 0.
- All problems are worth 1 point except the last two that are worth 10 points.

1. Let $X \sim Ber(0.4)$.

   Answer:

   ```
   p = 0.4
   ```

   (a) Find the support for X.
   Answer: $\{0, 1\}$

   (b) Find $E[X]$.
   Answer: 0.4

   (c) Find $Var[X]$.
   Answer: 0.24

   (d) Find $P(X = 0)$.
   Answer: 0.6

2. Let $Y \sim Bin(10, 0.35)$.

   Answer:

```r
n = 10
p = 0.35
```

   (a) Find the support for Y.

   Answer:

```r
0:n
```

```
##  [1]  0  1  2  3  4  5  6  7  8  9 10
```

   (b) Find $E[Y]$.

   Answer:

```r
n*p
```

```
## [1] 3.5
```

   (c) Find $Var[Y]$.

   Answer:

```r
n*p*(1-p)
```

```
## [1] 2.275
```

   (d) Find $P(Y = 4)$.

   Answer:

```r
dbinom(4, size = n, prob = p)
```

```
## [1] 0.2376685
```

   (e) Find $P(Y \leq 5)$.

   Answer:

```r
pbinom(5, size = n, prob = p)
```

```
## [1] 0.9050659
```

3. Let $X \sim Unif(1, 4)$.

   Answer:

```
a = 1
b = 4
```

   (a) State the support for X.

   Answer: (1,4)

   (b) Find $E[X]$.

   Answer:

```
(a+b)/2

## [1] 2.5
```

   (c) Find $Var[X]$.

   Answer:

```
(b-a)^2/12

## [1] 0.75
```

   (d) Find $P(X = 3)$.

   Answer: 0, since X is continuous

   (e) Find $P(X \leq 2)$.

   Answer:

```
punif(2, min = a, max = b)

## [1] 0.3333333
```

4. Let $Y \sim N(-2, 9)$

Answer:

```
mu = -2
sd = sqrt(9)
```

(a) State the support for Y.

Answer: All real numbers: $(-\infty, \infty)$

(b) Find $E[Y]$.

Answer:

```
mu

## [1] -2
```

(c) Find $E[3Y + 2]$.

Answer:

```
3*mu+2

## [1] -4
```

(d) Find $Var[Y]$.

Answer:

```
sd^2

## [1] 9
```

(e) Find $SD[Y]$.

Answer:

```
sd

## [1] 3
```

(f) Find $Var[3Y + 2]$.

Answer:

```
3^2*sd^2

## [1] 81
```

5. For the following data scenarios, determine the most reasonable distribution (binomial, Poisson, uniform, or normal) to use to model these data.

   (a) Number patients out of 22 whose angioplasty stents have a failure.
   Answer: binomial

   (b) Measuring electrical current in the brain from a Transcranial Magnetic Stimulation.
   Answer: normal

   (c) Of the 623 individuals who received a nano vaccine, the number who still got the viral disease.
   Answer: binomial

   (d) Fluid flow rate through a microfluidic device.
   Answer: normal

   (e) The number of farmers who will install prairie strips on their farms in the next year.
   Answer: Poisson

6. A company has set a goal of having, on average, (no more than) 5 safety violations per month. For the following questions assume the number of safety violations each month follows a Poisson distribution with a mean of 5 and that the number in each month is independent of the number in all other months.

Answer:

```
rate = 5
```

(a) What is the probability we will observe exactly 5 safety violations in one month?

Answer: Let X be the number of number of safety violations in one month and assume $X \sim Po(5)$. Calculate $P(X = 5)$.

```
dpois(5, lambda = rate)

## [1] 0.1754674
```

(b) What is the probability that there are more than 5 safety violations in one month?

Answer: Calculate $P(X > 5) = 1 - P(X \leq 5)$.

```
1-ppois(5, lambda = rate)

## [1] 0.3840393
```

(c) How many safety violations do we expect in one year?

Answer: Let $Y = \sum_{i=1}^{1} 2X_i$ where $X \overset{ind}{\sim} Po(5)$. Then $Y \sim Po(12 * 5)$. Calculate $E[Y]$.

```
rate*12

## [1] 60
```

(d) What is the probability we will observe less than 50 safety violations in one year?

Answer: Calculate $P(X < 50) = P(X \leq 49)$.

```
ppois(49, lambda = rate*12)

## [1] 0.08440668
```

7. A virtual reality company is trying to design a new system that reduces motion sickness. With the new system, the company believes only 10% of individuals will get motion sickness. The company runs a trial with 60 volunteers. For the following questions, assume motion sickness among the participants is independent.

   Answer: Let X be the number of individuals who get sick in the trial and assume $X \sim Bin(60, 0.1)$.

```
n = 60
p = 0.1
```

   (a) What is the most likely number of people to get sick in the trial?
   Answer: Calculate $E[X]$.

   ```
   n*p

   ## [1] 6
   ```

   You could also find the value x with the maximum $P(X = x)$, but you would find this is also $E[X]$.

   (b) What is the probability that nobody gets sick in the trial?
   Answer: Calculate $P(X = 0)$.

   ```
   dbinom(0, size = n, prob = p)

   ## [1] 0.00179701
   ```

   (c) The company decides to expand the trial to include 100 participants total. What is the probability that 7 or fewer individuals get sick?
   Answer:

   ```
   n = 100
   ```

   Answer: Let Y represent the number of individuals in this expanded trial that get sick and assume $Y \sim Bin(100, 0.1)$. Calculate $P(Y \leq 7)$.

   ```
   pbinom(7, size = n, prob = p)

   ## [1] 0.2060509
   ```

   (d) In the expanded trial, you are responsible for making sure you have enough supplies for those who get sick. If you want to ensure with at least 90% probability that you have enough supplies, what is the minimum number of individuals you should buy supplies for?
   Answer: Find y such that $P(Y \leq y) \geq 0.9$.

   ```
   ceiling(qbinom(0.9, size = n, prob = p))

   ## [1] 14
   ```

8. A machine learning algorithm has been constructed to diagnose a cold, a flu, or allergy. The algorithm has been provided the following prevalence and fever occurrence for cold, flu, and allergy.

	Cold	Flu	Allergy
Prevalence	0.2	0.1	0.7
Fever	0.1	0.8	0.5

Note that prevalence sums to 1 since these are the only maladies that the algorithm can diagnose. In contrast, fever does not sum to 1 because these are the probabilities that a fever occurs when an individual has a cold, flu, or allergy.

Given that a patient has a fever, calculate the three probabilities: that they have a 1) cold, 2) flu, or 3) allergy. Show all your work. (10 points)

Answer: Assume the following notation

- C: individual has a cold
- F: individual has a flu
- A: individual has an allergy
- V: individual has a fever

We know

$$P(C) = 0.2, P(F) = 0.1, P(A) = 0.7$$

and

$$P(V|C) = 0.1, P(V|F) = 0.8, P(V|A) = 0.5.$$

We will need $P(V)$ for the following calculations. We can calculate this using the law of total probability

$$\begin{aligned} P(V) &= P(V|C)P(C) + P(V|F)P(F) + P(V|A)P(A) \\ &= 0.1 \times 0.2 + 0.8 \times 0.1 + 0.5 \times 0.7 \\ &= 0.45 \end{aligned}$$

Finally we can calculate the desired probabilities

$$\begin{aligned} P(C|V) = P(V|C)P(C)/P(V) = 0.1 \times 0.2/0.45 = 0.04 \\ P(F|V) = P(V|F)P(F)/P(V) = 0.8 \times 0.1/0.45 = 0.18 \\ P(A|V) = P(V|A)P(A)/P(V) = 0.5 \times 0.7/0.45 = 0.78 \end{aligned}$$

9. A typical corn plant produces 0.34 lbs of corn with a standard deviation of 0.9 lbs. Suppose a farm has a single acre and they plant exactly 32,000 corn plants. What is the probability they will produce more than 200 bushels of corn? A bushel of corn is 56 lbs. Show all your work, name any assumptions you are making, and state any results you are using. (10 points)

Answer:

```
mu = 0.34
sd = 0.9
lbs_per_bushel = 56
n = 32000

x_mean = mu/lbs_per_bushel
x_var  = (sd/lbs_per_bushel)^2

y_mean = n*mu/lbs_per_bushel
y_var  = n*(sd/lbs_per_bushel)^2
y_sd = sqrt(y_var)
```

Let X_i be the yield in bushels for a single corn plant. Assume the X_i are independent of each with

$$E[X_i] = 0.34/56 = 0.0060714 \quad \text{and} \quad Var[X_i] = (0.9/56)^2 = 2.5829082 \times 10^{-4}.$$

Let $Y = X_1 + X_2 + \cdots + X_{3.2\times 10^4}$, so

$$E[Y] = n * E[X_i] = 194.2857143 \quad \text{and} \quad Var[Y] = n * Var[X_i] = 8.2653061.$$

By the CLT,

$$Y \dot\sim N(194.2857, 2.874945^2)$$

and

```
1-pnorm(200, mean = n*mu/lbs_per_bushel, sd = sqrt(n*(sd/lbs_per_bushel)^2))

## [1] 0.02342709
```