Name ____

Spring 2018 **STAT 401C** **Exam I**

(100 points)

Instructions:

- Full credit will be given only if you show your work.
- The questions are not necessarily ordered from easiest to hardest.
- You are allowed to use any resource except aid from another individual.
- Aid from another individual, will automatically earn you a 0.

1. A diagnostic test for disease D has a sensitivity of 0.95 and a specificity of 0.9. The prevalance of the disease is 0.02. (20 points)

   (a) Define notation for the following events (1 point each).

   Answer: Notation will defer, so really anything will work here. The notation below is what I will use.

   - having the disease
     Answer: D
   - not having the disease
     Answer: D^c
   - testing positive
     Answer: $+$
   - testing negative
     Answer: $-$

   (b) Use the notation in the previous step to define the following probabilities (2 points each).

   - sensitivity
     Answer: $P(+|D)$
   - specificity
     Answer: $P(-|D^c)$
   - prevalence
     Answer: $P(D)$

   (c) If an individual tests positive for the disease, what is the probability they actually have the disease? (10 points)

   Answer:

$$
\begin{aligned}
P(D|+) &= \frac{P(+|D)P(D)}{P(+|D)P(D)+P(+|D^c)P(D^c)} \\
&= \frac{P(+|D)P(D)}{P(+|D)P(D)+[1-P(-|D^c)][1-P(D)]} \\
&= \frac{0.95\times 0.02}{0.95\times 0.02+[1-0.9][1-0.02]} \\
&= 0.16
\end{aligned}
$$

2. Let X be a random variable with the following probability mass function:

x	-10	-5	0	5	10
P(X=x)	0.3	0.2	0.1	0.2	0.2

(a) Is $P(X = x)$ a valid probability mass function? Why or why not? (5 points)
Answer: Yes because $P(X = x) \geq 0$ for all x and the sum of the probabilities is

$$0.3 + 0.2 + 0.1 + 0.2 + 0.2 = 1$$

(b) Calculate $E[X]$. (5 points)
Answer:

$$\begin{aligned} E[X] &= -10 \times 0.3 + -5 \times 0.2 + 0 \times 0.1 + 5 \times 0.2 + 10 \times 0.2 \\ &= -10 \times 0.3 + 10 \times 0.2 = -3 + 2 = -1 \end{aligned}$$

(c) Let $Y = |X|$ what is the probability mass function for Y? (5 points)

Answer:

x	0	5	10
P(X=x)	0.1	0.4	0.5

(d) Find $E[|X|]$. (5 points)
Answer:

$$\begin{aligned} E[|X|] &= |-10| \times 0.3 + |-5| \times 0.2 + |0| \times 0.1 + |5| \times 0.2 + |10| \times 0.2 \\ &= 3 + 1 + 0 + 1 + 5 = 10 \end{aligned}$$

or, using the result from part c

$$\begin{aligned} E[|X|] &= E[Y] \\ &= |0| \times 0.1 + |5| \times 0.4 + |10| \times 0.5 \\ &= 2 + 5 = 7 \end{aligned}$$

3. Answer the following questions based on this joint distribution for the random variables X and Y.

X	Y = 1	2	3
-1	0.1	0.2	0.1
0	0.1	0.1	0.1
1	0.1	0.1	0.1

(a) What is the image for the random variable Y? (2 points)

Answer: 1, 2, 3

(b) Find the marginal probability mass function for X. (6 points)

Answer:

x	-1	0	1
P(X=x)	.4	.3	.3

(c) Find $P(Y > X)$. (6 points)

Answer:

$$\begin{aligned} P(Y > X) &= 1 - P(Y \leq X) \\ &= 1 - P(Y = 1, X = 1) \\ &= 1 - 0.1 = 0.9 \end{aligned}$$

(d) Are X and Y independent? Why or why not? (6 points)

Answer: No. You need to show that $P(X = x, Y = y) \neq P(X = x)P(Y = y)$ for some value of x and y. Here is one option

$$0.1 = P(X = 1, Y = 1) \neq P(X = 1)P(Y = 1) = 0.3 \times 0.3 = 0.09.$$

4. A warehouse has 46 high-intensity light bulbs and over the coming year the probability of each light failing is 5%. Assume light bulb failures are independent.

   (a) What is the probability that no light bulbs fail? (6 points)
   Answer: Let Y be the number of light bulbs that fail and assume $Y \sim Bin(n, \theta)$ with $n = 46$.
   $$P(Y = 0) = .95^{46} \approx 0.094.$$
   In R,

```r
n <- 46
p <-.05
dbinom(0,n,p)

## [1] 0.09446824
```

   (b) What is the probability that more than 2 light bulbs fail? (6 points)
   Answer:
   $$P(Y > 2) = 1 - P(Y \leq 2) = 1 - \sum_{y=0}^{2} \binom{46}{y} .05^y (1 - .95)^{46-y} =\approx 0.406$$
   In R,

```r
1-pbinom(2,n,p)

## [1] 0.4059753
```

   (c) If each light bulb costs \$500 to replace, what is the expected expense due to light bulb replacement over the next year? (6 points)
   Answer: The expected expense is
   $$E[\$500Y] = \$500E[Y] = \$500 \times 46 \times .05 = \$1150.$$

   (d) Name one reason light bulb failures would not be independent. (2 points)
   Answer: One reason is due to power surges that could cause multiple lights to fail at one time.

5. A positive displacement pump is used to fill an ethanol tanker. The pump measures 1 gallon of ethanol at a time with a mean of 1 gallon and a standard deviation of 0.01 gallons and independently of all other measurements. The pump repeats this process 30,001 times.

   (a) What is the approximate probability that the true amount of ethanol in the tanker is less than 30,000 gallons? (10 points)

   Answer: Let Y be the actual amount of ethanol for n extra gallons. By the CLT, $Y \dot{\sim} N(30000+n, [30000+n] \times 0.01^2)$ and thus the approximate probability for $n = 1$ is

```
sd <- 0.01
n <- 1; pnorm(30000, 30000+n, sd * sqrt(30000+n))

## [1] 0.2818547
```

   (b) The company wants to ensure the amount in the tanker is greater than 30,000 gallons with 99% probability. How many gallons above 30,000 should the pump measure to ensure with 99% probability that the true amount is greater than 30,000 gallons. (10 points)

   Answer: The 0.01 quantile of a standard normal random variable is

```
qnorm(0.01)

## [1] -2.326348
```

   Thus, we need to find n such that

$$
\begin{aligned}
0.99 &= P(Y \geq 30,000) = 1 - P(Y < 30,000) \\
0.01 &= P(Y < 30,000) \\
&= P\left(\frac{Y-(30000+n)}{0.01\sqrt{30000+n}} < \frac{30000-(30000+n)}{0.01\sqrt{30000+n}}\right) \\
&= P\left(Z < \frac{30000-(30000+n)}{0.01\sqrt{30000+n}}\right)
\end{aligned}
$$

   One approach is to set

$$
\frac{30000-(30000+n)}{0.01\sqrt{30000+n}} = -2.3263479
$$

   and solve for n (and then round up). This is a bit difficult as n appears in both the numerator and denominator (in the square root).

   An alternative approaches uses trial-and-error by plugging in different values of n. Using this approach, we can find that $n = 5$ extra gallons suffices while 4 extra gallons is (just barely) insufficient.

```
n <- 5; 1-pnorm(30000, 30000+n, sd * sqrt(30000+n))

## [1] 0.9980523

n <- 4; 1-pnorm(30000, 30000+n, sd * sqrt(30000+n))

## [1] 0.9895351
```

   When this exam was graded, I believe I gave full credit to either 4 or 5 since 4 is so close.