STAT340 Midterm - in class portion

First (given) name:

Write here: ______________________________

Last (family) name:

Write here: Powers Solutions

Lecture section:

Circle one: **Bi's section** **Brian's section**

Rules:

- You must show work for all computations (unless otherwise specified) to receive full credit.
- You do NOT need to simplify any expressions you write down.
- Note some of the multiple choice are **choose ONE** and some are **choose ALL that apply**, please pay attention to the instruction and select the appropriate number of responses!

Points:

MC1-3 (/6)	MC4-6 (/6)	SA1 (/4)	SA2 (/4)	Total (/20)

Multiple choice questions 2pts each

MC1

Which variables **always** satisfy P(X=0) > 0? **Choose ALL that apply!**

~~a.~~ Normal ← continuous
~~b.~~ Uniform ← not always
~~c.~~ Exponential
d. Poisson
e. Geometric

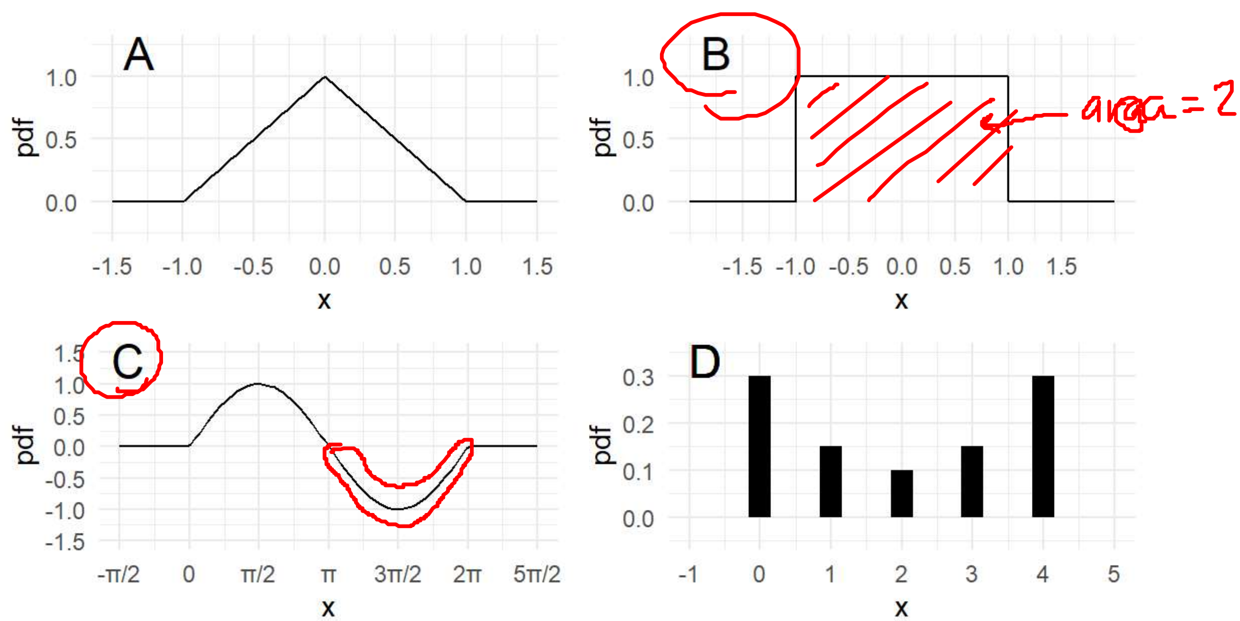
MC2

Let `X=pnorm(rnorm(n))` for some large sample size n, what distribution will X have? **Choose ONE!**

a. Normal
b. Uniform
c. Exponential
d. Poisson
e. Geometric

MC3

Which of the following are NOT valid probability density functions? **Choose ALL that apply!**



OR E. ALL of the above are valid probability density functions!

MC4

P(A) = 0.5 and P(B) = 0.6. Which of the following MUST be true? **Choose ONE!**

a. A and B are independent
b. A and B are not independent
c. A and B are mutually exclusive
d. A and B are not mutually exclusive
e. None of the above must be true

$P(A \cap B) = 0$
implies
$P(A \cup B) = P(A) + P(B) = 1.1$
not possible

MC5

You have water temperatures from two different lakes in Wisconsin, and want to run a permutation test to see if there is evidence that the mean temperatures are different. You have 16 measurements from Lake A and 21 measurements from Lake B. Our test statistic is the difference of sample means. Why would you want to run the Monte Carlo simulation with 10,000 replicates rather than 100 to produce a distribution of test statistics? **Choose ONE!**

~~a.~~ 10,000 replicates lowers the standard error significantly
~~b.~~ 10,000 replicates eliminates sampling bias due to flawed data collection
~~c.~~ 10,000 replicates compensates for any difference in population variance
~~d.~~ 10,000 replicates decreases α while maintaining the power of the test
e. None of these are valid reasons

MC6

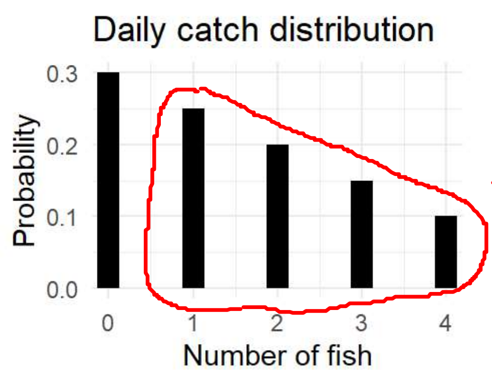
Which of the following are true of α, β and power? **Choose ALL that apply!**

~~a.~~ $\alpha + \beta = 1$
b. $\beta + \text{power} = 1$
c. As α increases, so does power
~~d.~~ As α decreases, so does β
e. None of the above are true

Short answer 4pts each

SA1

This winter in Madison, you decide to go ice fishing on lake Mendota every day of December (remember December has 31 days). Suppose each single day that you go fishing, you can model the number of fish you catch on a day using the following distribution:



sum = .7

At the end of each day, you count how many fish you caught and write this in your calendar. Let X = number of days in December you catch AT LEAST 1 fish. Assume each day is independent.

a. What variable is X?
b. What is P(X=0)? (i.e what is the probability you don't catch a single fish in the entire month?)
c. What are the expectation and variance of X?
d. Suppose you repeat the exact same routine every day in January (assume the daily catch has the same distribution and ALL fishing days are independent). Note January also has 31 days. Let Y = number of days in January you catch at LEAST 1 fish. You want to estimate how many MORE days you catch something in January compared to December. What are the mean and variance of (Y-X)?

a) $X \sim \text{Binomial}(31, .7)$

b) $P(X=0) = (.3)^{31}$

c) $E(X) = (.7)(31)$ $\quad Var(X) = (.3)(.7)(31)$

d) since $Y \sim \text{Binom}(31, .7)$
and X, Y independent

$E(X-Y) = E(X) - E(Y) = 0$

$Var(X-Y) = Var(X) + Var(Y) = 2(.3)(.7)(31)$

!!!

SA2

You are studying a rare disease that on average affects 2 out of every 1000 people in the general population. The current best available test for this disease has a type I error rate of about 5% and type II error rate of about 15%

a. If you test positive, what is your probability of having the disease?
b. If you test negative, what is your probability of not having the disease?

$P(\text{disease}) = .002 \Rightarrow P(\text{no disease}) = .998$

$P(\text{Pos} \mid \text{no disease}) = .05 \Rightarrow P(\text{Neg} \mid \text{no disease}) = .95$

$P(\text{Neg} \mid \text{disease}) = .15 \Rightarrow P(\text{Pos} \mid \text{disease}) = .85$

a) $$P(\text{disease} \mid \text{pos}) = \frac{P(\text{disease})P(\text{pos} \mid \text{disease})}{P(\text{disease})P(\text{pos} \mid \text{disease}) + P(\text{no disease})P(\text{pos} \mid \text{no disease})}$$

$$= \frac{(.002)(.85)}{(.002)(.85) + (.998)(.05)}$$

b) $$P(\text{no disease} \mid \text{neg}) = \frac{P(\text{no disease})P(\text{neg} \mid \text{no disease})}{P(\text{no disease})P(\text{neg} \mid \text{no disease}) + P(\text{disease})P(\text{neg} \mid \text{disease})}$$

$$= \frac{(.998)(.95)}{(.998)(.95) + (.002)(.15)}$$