Stat 390, Spring, Test 1, April 16, 2021; Marzban

READ the following paragraph!!

BY SUBMITTING THIS TEST, I HEREBY PLEDGE ON MY HONOR THAT I HAVE TAKEN IT IN PERSON AND WITHOUT ASSISTANCE FROM ANY OTHER PERSON. I ACKNOWLEDGE THAT THE PENALTY FOR VIOLATING ACADEMIC INTEGRITY IS MOST SEVERE.

10+4+6=20

- This is an online test given during the Corona Virus outbreak.
- For questions that have only 1 correct response, the options will appear as circles.
- For questions that may have multiple response, the options will appear as squares.
- It is open book/web/hw/solutions/past_tests/calculator/etc, but closed collaboration.
- Googling will simply waste your time. Instead, just make sure you are organized.
- The questions are presented to students in random order. So, if/when you want to ask about a problem, type (in chat to me) the first few words of the problem.
- You have the option of changing your answer to an already answered question.
- The list of questions and the remaining time for the whole test appear at the bottom of each page.
- Questions 1-10 are worth 1 point and do not require much calculation or writing.
- Questions 11-13 are worth 2 points each, and require a bit more work.
- Questions 14-15 are worth about 2 or 3 points and require varying levels of calculation, all of which are to be done on paper (or Tablet), and saved/scanned/photographed, and uploaded to canvas before 3:20. For these non-multiple-choice question, SHOW WORK. NO CREDIT FOR

- With file-upload questions, I am giving you several choices: You may
1) upload individual files for each question, or
2) submit a single file containing answers to all of your file-upload questions, or
3) email all your files to me (marzban@uw.edu), but ONLY IF all else has failed.

sp17/1, sp17/1

1 **1.** The following table shows the mean error (in some units) for 4 different machine learning algorithms, on a given task. One can say that the data used in the study contains

Support Vector Machine	1.234
Neural Network	2.134
Decision Tree	1.314
Random Forest	3.141

algorithm

a) one quantitative and one categorical random variable
b) 4 categorical random variables
c) one quantitative and 4 categorical random variables
d) none of the above

mean error

test2, sp16/1

1 **2.** Select the random variable(s) in the following list:
a) The battery usage on your smartphone at 10AM every day.
b) The accuracy of a malware detection algorithm.
c) The true number of planets in the Universe, at a given time.
d) The efficacy (success rate) of a vaccine.

Even though sample means/medians are different, there is too much overlap between the boxplots.

1 2 3
1 4 7
1 5 9

mean median

sp20/10

1 **3.** The following table shows three prediction errors (in some units) for three different algorithms. There is evidence from this data that (select all correct answers)

Algorithm A	2 1 3	2
Algorithm B	1 4 7	4
Algorithm C	9 5 1	5

a) A is better than B b) B is better than C c) A is better than C d) None of the above.

wi17/3

1 **4.** Let x denote the distance (m) that an animal moves from its birth site to the first territorial vacancy it encounters. The most appropriate distribution for describing x is
a) Exponential b) Binomial c) Poisson d) Normal

fits the template of inter-arrival time.

many

1 **5.** Suppose we have made a relative frequency histogram of a discrete random variable x, and find it to be bell-shaped. Then, in general, $pr(x < A)$ is equal to (select all correct statements)

a) the sum of the heights of the bars for all $x < A$.

b) the area under the histogram for all $x < A$. ← Discrete doesn't have area.

c) the area in Table 1 to the left of $x = A$, after we standardize. ← Table 1 is for N(0,1)

d) $pr(x \le A)$. ← for discrete $pr(x=A) \neq 0$

many

1 **6.** Select all cor ect statements.

a) If $x \sim Poisson(\lambda)$, then $pr(x < A) = \sum_{x=0}^{A} e^{-\lambda}\lambda^x/x!$ ↳ A-1

b) If $x \sim Exp(\lambda)$, then $pr(x = A) = \lambda e^{-\lambda A}$ ← for continuous, $f(x) \neq pr(x)$

c) If $x \sim Binom(n, \pi)$, then $pr(x \le n) = 1$ ← default distr.

d) If $x \sim Bernoulli(\pi)$, then $pr(x = 1) = 1/2$. ↳ π

wi17/6, su17/4, su16/8

1 **7.** The 5^{th} percentile of the exponential distribution with $\lambda = 5$ is 0.01. It is also known that for such a variable, the proportion of times that x is between 0.01 and 0.46 is 85%. Therefore, (Hint: draw a figure)

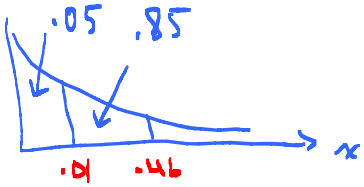
.05+.85

a) The 90^{th} percentile of the distribution is 0.46.

b) The 85^{th} percentile of the distribution is 0.46.

c) The 90^{th} percentile of the distribution is 0.01.

d) The 95^{th} percentile of the distribution is 0.99.



973

1 **8.** Suppose the 10th percentile of a sample is approximately equal to the 10th percentile of some distribution. Similarly, the 20th percentile of the sample is approximately equal to the 20th percentile of that distribution, etc, for the 30th, ..., 100th percentiles. Then

a) $pr(x <$ 10th sample percentile$)$ is exactly equal to $pr(x <$ 10th distribution percentile$)$. $= 0.1$ $= 0.1$

b) the histogram and the distribution have similar shapes.

c) there is evidence that the sample has come from that distribution.

d) there is evidence that the sample has come from that distribution, only if the dist is Normal.

9.23, many

1 **9.** Hassan collects a sample of size 1000 for the variable x, while Xiaochuan collects a sample of size million. Then, Hassan's histogram will be ____ Xiaochuan's histogram.

a) narrower than b) wider than c) comparable to

hw_sect4-2

1 **10.** We use two different distributions to compute the probability of $x < A$. Then, the more appropriate distribution is the one that gives the ____ probability.

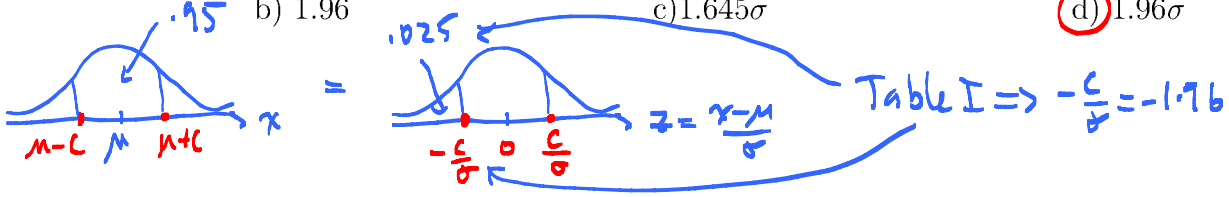
a) higher b) lower c) none of the above.

This $pr(x<A)$ must be compared with the $pr(x<A)$ from hist.

hw_sect6-1, 1.32

2 **11.** Assuming x has a Normal distribution with parameters μ, σ, what value c is such that there is a 95% probability that x falls in the interval $(\mu - c, \mu + c)$?

a) 1.645 b) 1.96 c) 1.645σ d) 1.96σ



hw_sect6-2

2 **12.** What is the probability of $x < 1$, if $e^x \sim N(4.72, 1)$? Let $e = 2.72$

a) 0.9772 b) 0.8413 c) 0.0228 d) 0.0013

$pr(x<1) = pr(e^x < e^1) = pr\left(\frac{e^x-\mu}{\sigma} < \frac{2.72-4.72}{1}\right) = pr(z < -2) = 0.0228$

↑ Standardize

hw-lect 5-2

3 **13.** In a hw you learned how to find the distribution of $z = \frac{x-\mu}{\sigma}$ if you know the distribution of x. Here, find the distribuion of $z = cx$ if $x \sim Unif(a, b)$. Show work, but the final answer must be of the form $f(z) = \cdots$ where $\cdots < z < \cdots$.

0.25 if something relevant.

$\int_a^b \frac{1}{b-a}\,dx = 1 \Rightarrow z = cx,\ dz = c\,dx$

(1) point. (0.5 if 1 missing)

$\int_{ac}^{bc} \frac{1}{b-a}\frac{1}{c}\,dz = 1$

$\int f(z)\,dz = 1$

$f(z) = \frac{1}{bc-ac}$ where $ac < z < bc$

(1) point (0.5 if $b/c, a/c$)

(1) point (0.5 if $f(z)$ is not written out)

hw-lect 6-3

3 **14.** How long/wide is the the box portion of the boxplot for an exponential distribution with parameter λ? Hint: this is a 2-step problem. Also, don't waste time on arithmetic; you may leave the answer in terms of numbers and λ.

n^{th} percentile.

$\int_0^u \lambda e^{-\lambda x}\,dx = \frac{n}{100} \Rightarrow y = \lambda x,\ dy = \lambda\,dx$

$\int_0^{\lambda u} e^{-y}\,dy = \frac{n}{100} \Rightarrow \frac{e^{-y}}{-1}\Big|_0^{\lambda u} = \frac{n}{100}$

$1 - e^{-\lambda u} = \frac{n}{100} \Rightarrow e^{-\lambda u} = 1 - \frac{n}{100} \Rightarrow -\lambda u = \ln\left(1 - \frac{n}{100}\right) \Rightarrow u = -\frac{1}{\lambda}\ln\left(1 - \frac{n}{100}\right)$

width = $75^{th} - 25^{th}$ percentile $= -\frac{1}{\lambda}\ln\left(1 - \frac{75}{100}\right) + \frac{1}{\lambda}\ln\left(1 - \frac{25}{100}\right)$

$= -\frac{1}{\lambda}\left[\ln\frac{1}{4} - \ln\frac{3}{4}\right] = -\frac{1}{\lambda}\ln\frac{1}{3} = \frac{\ln 3}{\lambda}$

Nearly everyone got this correctly. Well done! As a result, we did not even make rubric.