STAT 350 Exam 1 Fall 2023

V1

Name: ________________________________ PUID ____________________________

Instructor (circle one): Anand Dixit Timothy Reese Halin Shin Khurshid Alam

Class Start Time: ○ 11:30 AM ○ 12:30 PM ○ 1:30 PM ○ 2:30 PM ○ 3:30 PM ○ Online

As a boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together - we are Purdue.

Instructions:

1. **IMPORTANT** Please write your **name** and **PUID clearly** on every **odd page.**
2. **Write your work in the box. Do not run over into the next question space.**
3. You are expected to uphold the honor code of Purdue University. It is your responsibility to keep your work covered at all times. Anyone caught cheating on the exam will automatically fail the course and will be reported to the Office of the Dean of Students.
4. It is strictly prohibited to smuggle this exam outside. Your exam will be returned to you on Gradescope after it is graded.
5. The only materials that you are allowed during the exam are your **scientific calculator**, **writing utensils, erasers, your crib sheet**, and **your picture ID**. If you bring any other papers into the exam, you will get a **zero** on the exam. Colored scratch paper will be provided if you need more room for your answers. Please write your name at the top of that paper also.
6. The crib sheet can be a handwritten or type double-sided 8.5in x 11in sheet.
7. Keep your bag closed and cellphone stored away securely at all times during the exam.
8. If you share your calculator or have a cell phone at your desk, you will get a **zero** on the exam.
9. The exam is only 60 minutes long so there will be no breaks (including bathroom breaks) during the exam. If you leave the exam room, you must turn in your exam, and you will not be allowed to come back.
10. You must show **ALL** your work to obtain full credit. An answer without showing any work may result in **zero** credit. If your work is not readable, it will be marked wrong. Remember that work has to be shown for all numbers that are not provided in the problem or no credit will be given for them. All explanations must be in complete English sentences to receive full credit.
11. All numeric answers should have **four decimal places** unless stated otherwise.
12. After you complete the exam, please turn in your exam as well as your table and any scrap paper that you used. Please be prepared to **show your Purdue picture ID**. You will need to **sign a sheet** indicating that you have turned in your exam.

Your exam is not valid without your signature below. This means that it won't be graded.

I attest here that I have read and followed the instructions above honestly while taking this exam and that the work submitted is my own, produced without assistance from books, other people (including other students in this class), notes other than my own crib sheet(s), or other aids. In addition, I agree that if I tell any other student in this class anything about the exam BEFORE they take it, I (and the student that I communicate the information to) will fail the course and be reported to the Office of the Dean of Students for Academic Dishonesty.

Signature of Student: ____

You may use this page as scratch.

The following is for your benefit only; we will not use this for grading:

Question Number	**Total Possible**	**Your points**
Problem 1 (True/False) (2 points each)	12	
Problem 2 (Multiple Choice) (3 points each)	12	
Problem 3	18	
Problem 4	16	
Problem 5	16	
Problem 6	26	
Total	100	

1. (12 points, 2 points each) True/False Questions. Please indicate the correct answer by filling in the circle. *<u>If you indicate the correct answer by any other way, you may receive 0 points for the question</u>*.

1.1. Suppose that events **A** and **B** belong to the same sample space **Ω**, and that the values of **P(A)** and **P(B)** are **non-zero** and known.

Ⓣ or Ⓕ If $\boldsymbol{P(A \cap B) = P(A)}$, then $\boldsymbol{P(A' \cap B') = P(A')}$.

1.2. A research group is studying a population known to be normally distributed but can only afford a sample of size 5 due to high sampling cost. One of the researchers states that the sampling distribution of the mean will be normally distributed. Another colleague in the research group states that this is incorrect as the CLT requires a sample size of at least 30 (n >= 30) for the sampling distribution of the mean to be normally distributed.

Ⓣ or Ⓕ Is the researcher correct in stating that the distribution for the sampling distribution of the mean will be normally distributed?

1.3. Let **X** satisfy the conditions to be distributed as a Poisson distribution.

Ⓣ or Ⓕ Since **X** follows a Poisson distribution it follows that $\boldsymbol{\sigma_X^2 = E[X]}$.

1.4. Suppose **X** follows a normal distribution with mean **μ** and standard deviation **σ**.

Ⓣ or Ⓕ It follows that $\boldsymbol{P(\mu - \sigma < X < \mu + \sigma) \approx 0.75}$.

1.5. An application in senior technology aims to help improve the overall health, balance, and flexibility of the elderly by tracking several variables. One of these variables is the zip code of the participant.

Ⓣ or Ⓕ This variable is a discrete numerical variable.

1.6. A manufacturing company, aiming to meet quality standards, assesses the lifespan of its light bulbs. The company claims that the mean lifespan of the bulbs is 1200 hours with a standard deviation of 150 hours. To verify, this a large retailer tests a sample of 215 bulbs, finding a sample mean of 1180 hours with a sample standard deviation of 157 hours.

Ⓣ or Ⓕ The correct symbol to represent the 1200 hours is **x̄**.

2. (12 points, 3 points each) Multiple Choice Questions. Please indicate the correct answer by filling in the circle. *<u>If you indicate the correct answer by any other way, you may receive 0 points for the question</u>*. For each question, there is only one correct option given.

2.1. Suppose **X** is a continuous random variable with mean **μ = 5** and standard deviation **σ = 5**. What is the value of **P(X = 10)**?

Ⓐ **P(X = 10) = 0.68**

Ⓑ **P(X = 10) = 0.8413**

Ⓒ **P(X = 10) = 0.95**

Ⓓ **P(X = 10) = 0**

Ⓔ Not enough information to calculate it.

2.2. Suppose $X_1, X_2, \ldots, X_n$ is a random independent sample coming from the same (identically distributed) but unknown distribution with finite non-zero variance σ^2. **Identify the incorrect statement**.

Ⓐ For any **n**, $E[\bar{X}] = E[X_1]$.

Ⓑ $\mathrm{Var}(X_1 - X_2) \neq 0$

Ⓒ For any **n**, $\bar{X}$ will be approximately normally distributed.

Ⓓ As **n** increases the random variable $(\bar{X} - \mu) / \sigma$ becomes approximately normally distributed with mean 0 and standard deviation 1.

Ⓔ For any **n**, $\mathrm{Var}(\bar{X}) \leq \mathrm{Var}(X_1)$.

2.3. The measure of spread which is the most likely to be influenced by outliers in the dataset is

Ⓐ sample mean

Ⓑ sample median

Ⓒ sample standard deviation

Ⓓ Inner quartile range (IQR)

Ⓔ more than one of the above.

2.4. Let **X** and **Y** be two normal random variables with means μ_X and μ_Y, and standard deviations σ_X and σ_Y, respectively. Select the **correct statement** regarding the relationship between the parameters ($\mu_X, \mu_Y, \sigma_X, \sigma_Y$) when the **normal curve associated** with **X** is **more peaked** than the **normal curve associated** with **Y**.

Ⓐ $\sigma_X < \sigma_Y$

Ⓑ $\sigma_X > \sigma_Y$

Ⓒ $(\mu_X / \sigma_X) > (\mu_Y / \sigma_Y)$

Ⓓ $\mu_X < \mu_Y$

Ⓔ $\mu_X > \mu_Y$

3. (18 points) Halin goes fishing along the river in her area every weekend. Once she begins fishing, it is known that she needs to wait 20 minutes on average to catch a fish.

a) **(2 pts)** Define the random variable **X** that represents the number of fish Halin catches in **two hours** of fishing. Write the name of its distribution and provide the value of the parameter, **λ**.

There is an average of 1 fish caught every 20 minutes so there is an average of 3 fish caught per one hour therefore $\boldsymbol{\lambda} = \frac{\mathbf{1}}{\mathbf{20}} \times \mathbf{120} = \mathbf{6}$.

$$\boldsymbol{X} \sim \textbf{Poisson}(\boldsymbol{\lambda} = \mathbf{6})$$

b) **(4 pts)** What is the probability that Halin catches exactly 8 fish in two hours?

The pmf for a Poisson is $\boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x}) = \frac{\boldsymbol{\lambda}^{\boldsymbol{x}} \boldsymbol{e}^{-\boldsymbol{\lambda}}}{\boldsymbol{x}!}$

$$\boldsymbol{P}(\boldsymbol{X} = \mathbf{8}) = \frac{\mathbf{6}^{\mathbf{8}} \boldsymbol{e}^{-\mathbf{6}}}{\mathbf{8}!} = \mathbf{0.1033}$$

c) **(8 pts)** While talking with her friends, Halin recalls that on a lucky weekend, she had caught somewhere between 7 and 9 fish (inclusive) in two hours. Given this information, what is the probability that Halin actually caught 8 fish that weekend?

$$\boldsymbol{P}(\boldsymbol{X} = \mathbf{8} | \mathbf{7} \leq \boldsymbol{X} \leq \mathbf{9}) = \frac{\boldsymbol{P}(\{\boldsymbol{X} = \mathbf{8}\} \cap \{\mathbf{7} \leq \boldsymbol{X} \leq \mathbf{9}\})}{\boldsymbol{P}(\mathbf{7} \leq \boldsymbol{X} \leq \mathbf{9})} = \frac{\boldsymbol{P}(\boldsymbol{X} = \mathbf{8})}{\boldsymbol{P}(\boldsymbol{X} = \mathbf{7}) + \boldsymbol{P}(\boldsymbol{X} = \mathbf{8}) + \boldsymbol{P}(\boldsymbol{X} = \mathbf{9})} = \frac{\mathbf{0.1033}}{\mathbf{0.1377} + \mathbf{0.1033} + \mathbf{0.0688}} = \mathbf{0.3333}$$

$$\boldsymbol{P}(\boldsymbol{X} = \mathbf{7}) = \frac{\mathbf{6}^{\mathbf{7}} \boldsymbol{e}^{-\mathbf{6}}}{\mathbf{7}!} = \mathbf{0.1377}$$

$$\boldsymbol{P}(\boldsymbol{X} = \mathbf{9}) = \frac{\mathbf{6}^{\mathbf{9}} \boldsymbol{e}^{-\mathbf{6}}}{\mathbf{9}!} = \mathbf{0.0688}$$

d) **(4 pts)** It is known that 35% of the fish caught in this river are largemouth bass. Halin now remembers that she did catch exactly 8 on that lucky weekend. What is the probability that 3 of the eight she caught were largemouth bass? Hint: Each fish she caught was either a largemouth bass or it was not.

Notice that she knows 8 fish were caught but we do not know how many were largemouth bass. Each fish is either largemouth bass or some other fish (**B**inary), each catch can be considered **I**ndependent of other catches, and there is a fixed sample size $\boldsymbol{n} = 8$, and the probability of a fish being largemouth bass is $p = 0.35$ (**S**uccess). The situation is **BInS** (Binomial experiment).
Let Y denote the number of largemouth bass caught $\boldsymbol{Y} \sim \boldsymbol{Bin}(\boldsymbol{n} = \mathbf{8}, \boldsymbol{p} = \mathbf{0.35})$.

$$\boldsymbol{P}(\boldsymbol{Y} = \mathbf{3}) = \binom{\mathbf{8}}{\mathbf{3}} \mathbf{0.35}^{\mathbf{3}} \mathbf{0.65}^{\mathbf{8-3}} = \mathbf{56} \times \mathbf{0.042875} \times \mathbf{0.11603} = \mathbf{0.2786}$$

4. (16 points) Over the past decade, the dean of admissions at a leading university has observed a consistent pattern in the mathematics placement exam scores of incoming freshmen. The data reveals that the scores follow a normal distribution with a mean of 82 and a variance of 64.

a) **(6 pts)** Consider the scenario where a random sample of 35 incoming freshman students is selected. In this context, describe the sampling distribution of the sample mean $\bar{\mathbf{X}}$. Clearly state the name of the distribution, identify any parameter(s), and determine the value of the parameter(s). (pts)

Since $X_1, X_2, \ldots, X_{35} \sim N\left(\mu_X = 82, \sigma_X = \sqrt{64} = 8\right)$ would be an independent sample from the same normal distribution. It follows that the sampling distribution of $\bar{X}$ is normally distributed as below.

$$\bar{X} \sim N\left(\mu_{\bar{X}} = 82, \sigma_{\bar{X}} = \frac{8}{\sqrt{35}}\right).$$

Notice that this does not require central limit theorem since the data was originally normally distrbiuted.

b) **(4 pts)** What is the probability that the mean score of a sample of 35 students is at least 80?

We just need to standardize to a standard normal distribution and use the table.

$$P(\bar{X} > 80) = P\left(Z > \frac{80 - 82}{\frac{8}{\sqrt{35}}}\right) = P(Z > -1.48) = \Phi(1.48) = 0.9306$$

Where we used symmetry for $P(Z > -1.48) = \Phi(1.48)$.

c) **(6 pts)** What value of $\bar{\mathbf{x}}$ would represent the lower **12.51st percentile** of the average score for a sample of 35 students?

Backwards problem:

Start from standard normal solve $\Phi(z^*) = 0.1251 \rightarrow z^* = -1.15$

Move this point to the distribution of $\bar{X} \rightarrow \bar{x}^* = \mu_{\bar{X}} + \sigma_{\bar{X}} * z^* = 82 + \frac{8}{\sqrt{35}}(-1.15) = 80.4449$

5. (16 points) Suppose a college student uses the bus 70% of the time to go to his first class of the week on the college campus. Further suppose the student chooses to walk whenever he chooses not to take the bus. Assume that the student can choose only one form of transportation at a time. If the student chooses to walk, he is late 30% of the time. Further, if he chooses to use the bus, he is late 20% of the time. Using this information, answer the following questions.

a) **(6 pts)** What is the probability that the student will be late when he travels to his first class of the week on the college campus?

Let L = the student is late, B = the student takes the bus, and W = the student walks.

Given:
$P(B) = 0.7$, $P(W) = 0.3$
$P(L|B) = 0.2$, $P(L|W) = 0.3$

Compute: Use the law of total probability

$$\boldsymbol{P(L) = P(L|W)P(W) + P(L|B)P(B) = 0.3 \times 0.3 + 0.2 \times 0.7 = 0.23}$$

b) **(4 pts)** What is the probability that the student will not be late when he travels to his first class of the week on the college campus?

Complement Rule:
$P(L') = 1 - P(L) = 0.77.$

c) **(6 pts)** What is the probability that the student chooses to walk, given that he was found to be on-time for his first class of the week on the college campus?

Compute: Bayes Rule

$$\boldsymbol{P(W|L') = \frac{P(L'|W)\,P(W)}{P(L')} = \frac{0.7 \times 0.3}{0.77} = 0.2727}$$

Where $P(L'|W) = 1 - P(L|W) = 1 - 0.3 = 0.7$

6. (26 points) At a beverage company, a vigilant quality control engineer has detected irregularities in the bottle filling process. The production line includes two distinct zones. In Zone A, bottles consistently leave the machine underfilled, with a deviation of 1 to 3 ml below the desired level. Conversely, in Zone B, bottles consistently exit overfilled, exceeding the target by 1 to 3 ml. To understand these variations, we utilize a functional relationship, denoted as $f_X(x)$ that characterizes the probabilities associated with the fill levels of these bottles. However, the normalizing constant must be established before any probability calculations can be performed.

$$\boldsymbol{f_X(x) = \begin{cases} k & -3 \leq x \leq -1 \\ k\,(-15x + 45) & 1 \leq x \leq 3 \\ 0 & \textbf{otherwise} \end{cases}}$$

a) **(4 pts)** Determine the value of $\boldsymbol{k}$ such that $\boldsymbol{f_X(x)}$ is a valid probability density function.

Clearly $k > 0$ sketch the function out if you cannot see this.

Now solve

$$\int_{-3}^{-1} k dx + \int_{1}^{3} k(-15x + 45)dx = 1$$

$$2k + 15k\left[-\frac{x^2}{2} + 3x\right]_1^3 = 1$$

$$3k + 30k = 1$$

$$\boldsymbol{k = \frac{1}{32}}$$

$$\boldsymbol{F_X(x) = \begin{cases} [\mathrm{A}] & x \leq -3 \\ [\mathrm{B}] & -3 \leq x < -1 \\ [\mathrm{C}] & -1 \leq x < 1 \\ -\frac{15}{64}x^2 + \frac{45}{32}x - \frac{71}{64} & 1 \leq x < 3 \\ [\mathrm{E}] & x \geq [\mathrm{D}] \end{cases}}$$

b) **(6 pts)** Determine the missing parts **[A, B, C, D, E]** of the cumulative distribution function.

$\boldsymbol{[\mathrm{A}] = 0}$

$\boldsymbol{[\mathrm{D}] = 3}$

$\boldsymbol{[\mathrm{E}] = 1}$

$$\boldsymbol{[\mathrm{B}] = \int_{-3}^{x} \left(\frac{1}{32}\right) dt = \frac{1}{32}(x + 3)}$$

$\boldsymbol{[\mathrm{C}] = \int_{-3}^{-1} \left(\frac{1}{32}\right) dt = \frac{1}{32}(-1 + 3) = \frac{1}{16}}$ **(CDF is constant in the region $\boldsymbol{-1 \leq x < 1}$)**

$$F_X(x) = \begin{cases} 0 & x \le -3 \\ \frac{1}{32}(x+3) & -3 \le x < -1 \\ \frac{1}{16} & -1 \le x < 1 \\ -\frac{15}{64}x^2 + \frac{45}{32}x - \frac{71}{64} & 1 \le x < 3 \\ 1 & x \ge 3 \end{cases}$$

c) **(4 pts)** Determine the probability that the random variable **X** will be non-negative.

$$P(X > 0) = 1 - P(X \le 0) = 1 - F_X(0) = 1 - \frac{1}{16} = \frac{15}{16} = 0.9375$$

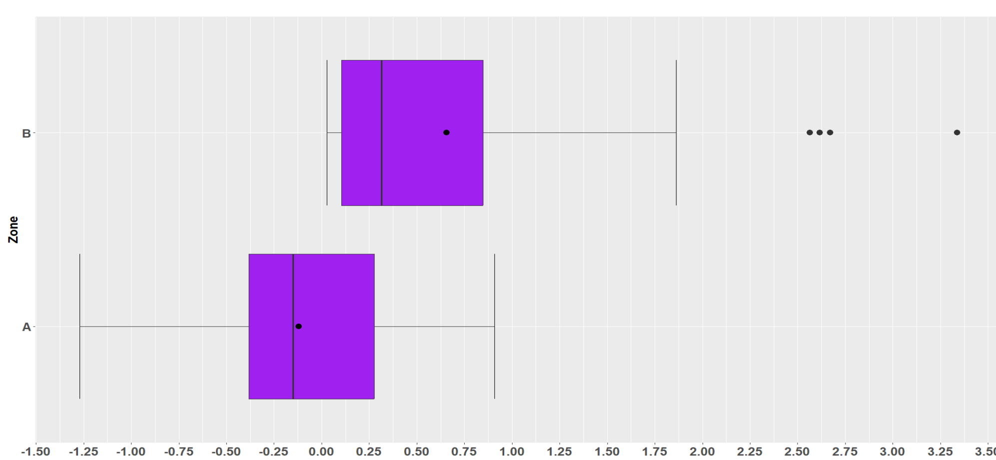
d) **(6 pts)** What value represents the 5th percentile of this distribution?

$F_X(x^*) = 0.05$

The 5th percentile must occur in the first region because we accumulate at least 1/16 or 0.0625 of the probability after -1.

$$\frac{1}{32}(x^* + 3) = 0.05$$
$$x^* = 32 \times 0.05 - 3 = -1.4$$

e) **(6 pts)** After implementing modifications to the machinery, the diligent quality control engineer has gathered a fresh sample of the production process. This time, the engineer has distinctly categorized the bottles passing through Zone A and Zone B. Presented below is a side-by-side boxplot illustrating a sample of 48 bottles from each zone. Approximate the minimum and maximum values for each boxplot. Furthermore, discuss any important characteristics and make a conclusion regarding if the production process has been fixed.



Zone A → 5-number summary approximately

$$[\mathbf{Min} = -\mathbf{1.26}, \boldsymbol{Q_1} = -\mathbf{0.35}, \boldsymbol{Q_2} = -\mathbf{0.2}, \boldsymbol{Q_3} = \mathbf{0.26}, \mathbf{Max} = \mathbf{0.9}]$$

Zone B → 5-number summary approximately

$$[\mathbf{Min} = \mathbf{0.05}, \boldsymbol{Q_1} = \mathbf{0.15}, \boldsymbol{Q_2} = \mathbf{0.35}, \boldsymbol{Q_3} = \mathbf{0.86}, \mathbf{Max} = \mathbf{3.4}]$$

Zone A looks almost symmetric around 0 with a slight positive skew because median is closer to Q_1. There aren't explicit points so there aren't any real outliers.

Zone B is right skewed with four high real outliers. The large gap between these points and the whisker indicates that they are real outliers.

Zone A seems to have been fixed as it is roughly centered around 0.

Zone B still seems to have issues as it is now right skewed and is still overfilling the bottles.