Section: _____ Name:_____

Read the following directions carefully. DO NOT turn to the next page until the exam has started.

Write your name and section number at the top right of this page:

Class	**Section Number**
Miranda 8:50	001
Sahifa 12:05	002
Miranda 9:55	004

As you complete the exam, write your initials at the top right of each other page.

When the exam start time is called, you may turn the page and begin your exam.
If you need more room, there is a blank page at the end of the exam, or we can give you some scratch paper.

Some multiple choice questions are "Select ONE" while others are "Select ALL that apply". Pay attention to the question type and only mark one option if it says "Select ONE". Fill in the circles completely.

If you finish early, you can hand your exam to your instructor or TA and leave early.

Otherwise, stop writing and hand your exam to your instructor or TA when the exam stop time is called.

1. (a) (3 points) For the three random variables listed below, identify its distribution, which will be either binomial or normal.

   (i) A multiple-choice test has 15 questions, each of which has five choices. A specific unprepared student takes the test and has a 80% chance of getting a question wrong. Suppose X denotes the number of answers that the student gets right in the test. Find the distribution of X.
   $X \sim Binom(15, 0.2)$

   (ii) Suppose that grade points of undergraduate students at a university have a distribution with a mean of 2.63 and a variance of 0.15. Let X denote the average grade points of 100 undergraduate students. Find the distribution of X.
   $X \sim N\left(2.63, \sqrt{\frac{0.15}{100}}\right)$

   (iii) An experimental medication was given to 3000 patients with a certain medical condition. It is previously known that 1.6% people show severe side effects to a component in the medication. Let X denote the number of patients who develop severe side effects. Find the distribution of X.
   $X \sim Binom(3000, 0.016)$

   (b) (3 points) Out of the ones binomially distributed in (a), which of those can be approximated by a normal distribution? What will be the normal approximation for that X?
   The RV in (a.iii) is approximately normal since we have a very large sample size n, even though the probability p is close to 0. The normal approximation is

$$X \dot{\sim} N\left(3000(0.016), \sqrt{3000(0.016)(1-0.016)}\right)$$

2. Assume that the monthly electricity consumptions of all households in a certain region is approximately normally distributed with a mean of 1200 kilowatt-hours and a standard deviation of 110 kilowatt-hours.

   (a) (2 points) John Smith recieved a notice informing him that his household electricity consumption lies 2 standard deviation above the mean. What is John Smith's household electricity consumption? **You don't have to simplify your answer, you can leave it as an expression.**

   $$x = 1200 + 2(110)$$

   (b) (3 points) Find the value of average electricity consumption in a locality of 100 houses such that it lies 2 standard deviation below the mean of the sampling distribution of sample means. **You don't have to simplify your answer, you can leave it as an expression.**
   The mean electricity consumption of 100 housees is $\bar{X} \dot{\sim} N(1200, \frac{110}{\sqrt{100}}) = N(1200, 11)$. A value of the sample mean two SDs below average is

   $$\bar{x} = 1200 - 2(11)$$

   (c) (4 points) Using the *d*, *p* or *q* function for the normal distribution, fill in the blanks below such that you get R code for the interquartile range of the monthly electricity consumptions of all households in that region.

   [The interquatile range is the number that gives the difference between Q3 and Q1. The first quartile (Q1) is the 25th percentile and the third quartile (Q3) is the 75th percentile of the data/distribution.]

   $IQR =$ `qnorm( 0.75 , 1200, 110) - qnorm( 0.25 , 1200, 110)`

3. Four different bird species, with different levels of rarity, live in a conservation area. Scientists regularly walk through the area and record the count of unique of bird species they observe (out of 4). A distribution for X, the number of species observed, is given below.

x	0	1	2	3	4
$P(X = x)$	0.2	0.4	0.25	0.1	0.05

(a) (4 points) For each of the four BINS assumptions, comment on whether the assumption is met or not met for X. Is X a binomial random variable? Here, each "trial" is a species of bird.

- B - Each trial is binary, since we either see or do not see each species.
- I - We can safely assume that each bird species is independent.
- N - There are four trials/species.
- S - The probability of success is not the same, since some species are more or less common.

Since the same probability assumption is violated, X is not a binomial random variable.

(b) (3 points) What is the 75th percentile of X?

This 75th percentile is the smallest value x such that $P(X \leq x)$ is at least 0.75. We have $P(X \leq 1) = 0.6$ and $P(X \leq 2) = 0.85$ so the 75th percentile is 2.

(c) (3 points) Write an expression to calculate $E(X)$. **You don't have to simplify your answer, you can leave it as an expression.**

$$E(X) = 0(0.2) + 1(0.4) + 2(0.25) + 3(0.1) + 4(0.05)$$

4. A doctor records the cholesterol level of 20 of her patients (in mg/dL). He reports a 95% confidence interval for the average cholesterol:

$$(147.29,\ 215.71)$$

(a) (3 points) Identify the point estimate and margin of error of this CI. You do not need to simplify your answer.

The point estimate is the midpoint of the interval:

$$\text{PE} = \frac{147.29 + 215.71}{2}$$

To get the upper bound, we add the margin of error to the point estimate, and to get the lower bound, we subtract the margin of error from the point estimate. So, the MOE is the half-width of the interval. Here are three ways to calculate this:

$$\text{MOE} = \text{PE} - 147.29 = 215.71 - \text{PE} = \frac{215.71 - 147.29}{2}$$

(b) (2 points) Which of the following statements is the best interpretation of this CI? **Select ONE**.

- ○ If we repeated the CI procedure across multiple samples, 95% of the intervals would cover the cholesterol level for the patients in the sample.
- ● If we repeated the CI procedure across multiple samples, 95% of the intervals would cover the mean cholesterol level for all patients.
- ○ 95% of the 20 patients in this sample have a cholesterol level between 147.29 and 215.71.
- ○ 95% of all patients have a cholesterol level between 147.29 and 215.71.

(c) (2 points) The doctor wants to be especially sure that her confidence interval correctly covers the true mean cholesterol of her patients. What change should she make to the 95% interval above? **Select ONE**.

- ○ Choose a smaller confidence level, which gives a narrower interval.
- ○ Choose a larger confidence level, which gives a narrower interval.
- ○ Choose a smaller confidence level, which gives a wider interval.
- ● Choose a larger confidence level, which gives a wider interval.

5. A manager at a campus cafeteria is deciding what pizza toppings to offer. He thinks that if over 25% of his customers like pineapple on pizza, he will offer pineapple as a topping.

   He takes a sample of 45 customers and finds that 14 of them like pineapple on pizza.

   (a) (3 points) Identify null and alternative hypotheses to answer the manager's question. They should be in terms of p, the proportion of customers who like pineapple on pizza.

   $$H_0 : p = 0.25 \text{ or } p \leq 0.25$$

   $$H_A : p > 0.25$$

   (b) (2 points) The manager will model the number of customers who like pineapple as Binom$(45, p)$. Which line of R code below correctly calculates the p-value for the test? **Select ONE**.

   - ○ `pbinom(14, 45, 0.25)`
   - ○ `1 - pbinom(14, 45, 0.25)`
   - ○ `pbinom(14, 45, 14/45)`
   - ● `1 - pbinom(13, 45, 0.25)`
   - ○ `1 - pbinom(15, 45, 0.25)`
   - ○ `1 - pbinom(15, 45, 14/45)`

   (c) (5 points) The observed p-value is greater than the manager's chosen significance level $\alpha = 0.05$. Which of the following statements are true? **Select ALL that apply**.

   - ○ We are certain that 25% or less of all customers like pineapple on pizza.
   - ● It is possible that more than 25% of all customers like pineapple on pizza.
   - ● We do not have evidence that more than 25% of all customers like pineapple on pizza.
   - ○ We could have made a Type I error.
   - ● We could have made a Type II error.