MATH 50 MIDTERM – V2

INSTRUCTOR: ETHAN LEVIEN

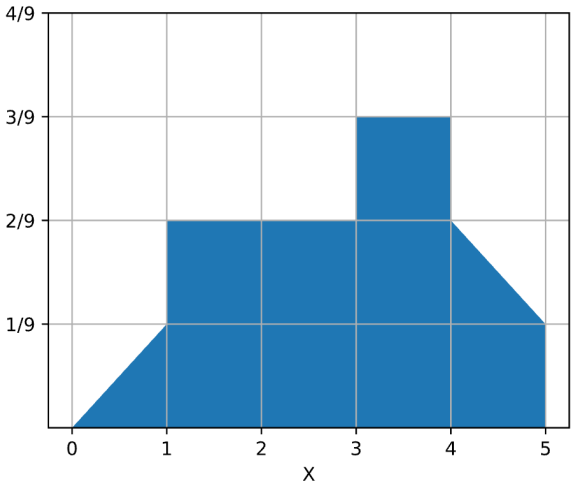
Instructions

- You have the entire class period to complete the exam.
- You may use one page of written notes, but NO electronics (computer, calculator, etc.).
- Each problem is worth 4 points.
- Show all your work but **circle your final answer**.
- Don't Cheat.

Exercise 1: Given the data in the table below, compute $\mathbb{E}[X^2|Y > 2]$.

i	X_i	Y_i
1	5	0
2	0	100
3	2	0
4	1	0
5	4	0
6	0	5
7	1	3
8	1	1
9	1	1

Exercise 2: Below is a histogram of a random variable X.



What is $P(X < 4|X > 2)$?

Exercise 3: For each of the following sets of conditions, state whether it is possible to come up with a model of two variables X and Y satisfying these conditions. If it is possible, sketch and $x - y$ plot with samples from such a model.

(a) (i) X and Y are independent (ii) their relationship is given by a linear regression model with Y as the response variable and (iii) $\rho^2 > 0$.

(b) (i) X and Y are not independent (ii) $\text{cov}(X, Y) = 0$ and (iii) the distribution of $Y|X$ is normal for all X values.

Exercise 4: Consider the linear regression model:

$$Y|X \sim \text{Normal}(1 + 2X, 1)$$

where X is assumed to be a binary predictor. Suppose that x and y are numpy arrays containing 100 samples from this model. Additionally, assume that there is an equal number of samples from each value of x (that is, 50 each).

(a) Consider the following line of code:

```
beta1_hat = np.mean(y[x==1]) - np.mean(y[x==0])
```

Why does this estimate β_1?

(b) What is the sample distribution of this estimate?

Exercise 5: Suppose X represents income of people in one generation at the age of 40 and Y represents income of their children at the same age. Income is measured in thousands of USD. Suppose X and Y are modeled by

$$X \sim \text{Normal}(50, 5)$$
$$Y|X \sim \text{Normal}(25 + X/2, 2)$$

(a) Is this a linear regression model?
(b) What is the marginal distribution of income in the second generation?

Score Sheet

Question	Points Earned
1	
2	
3	
4	
5	
Total	