MIDTERM QUIZ – PRACTICE

Instructions

- You have 1 hour to complete the quiz.
- You may use your written notes, but NO electronics (computer, calculator, etc.).
- Each problem is worth 3 points.
- Show all your work but **circle your final answer**.

Exercise 1 (Computations with data)**:** Given the data in the table below, compute

(a) $E[Y|X = 1]$

(b) $E[X|\{Y = 0\} \cap \{X < 3\}]$.

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

Exercise 2 (Python code)**:** Consider the following code which generates samples of variables X an Y.

```
> import numpy as np
> n = 1000
> x = np.random.choice([0,1],p=[0.5,0.5],size=n)
> y= np.random.normal(x,2,n)
```

(a) Write down a probability model that is simulated by this code, specifying either the marginal or conditional distribution for each variable. Is it a linear regression model?

(b) Write the code to compute $P(X = 1|Y > 0.3)$.

Exercise 3 (Estimators)**:** Suppose you are designing a study to investigate the effect of a supplement on performance in a 5k. Let Y denote the random variable representing someone's time (in minutes) and X represent whether they took the supplement ($X = 1$ if they took it). If your probability model is

$$Y|(X = x) \sim \text{Normal}(\mu_0 + \Delta x, \sigma^2(1 + x/4))$$

(1) Is this a linear regression model?

(2) If given numpy arrays x and y containing samples from this model, write down one way of estimating σ^2. Note that you don't have to use all the data.

Exercise 4 (Coefficient of determination)**:** Consider the linear regression model

$$Y|X \sim \text{Normal}(\beta_0 + \beta_1 X, \sigma^2)$$

For each of the following parameters, explain how changing it will effect the coefficient of determination and make sure to justify your answer.

(1) β_0

(2) β_1

(3) σ

(4) μ_X

(5) σ_X

Exercise 5 (Understanding continuous distributions)**:** Suppose

$$Y \sim \text{Uniform}(0, 1)$$

On the following diagram, label the probabilities/conditional probabilities which are given by the area of the green and blue rectangles. The vertical black line on the far right is at $Y = 1$, so the black is representing the density of Y.