Econometrics Economics/ Introduction to Econometrics
Problem Set 1 (Probability and Statistics Concepts)

Question 1: Expectation and Variance

Consider the equation $y = 3 - x_1 + 2x_2 + e$ with $x_1 \sim Ber(0.5)$, $x_2 \sim N(3, 4)$, and $e \sim N(0, 1)$. Also assume that $\text{Corr}[x_1, x_2] = 0.5$ and $e \perp (x_1, x_2)$. Let $\mathbf{x} \equiv (1, x_1, x_2)'$.

(a) Compute $\text{E}[y]$ and $\text{Var}[y]$.

(b) Compute $\text{E}[\mathbf{x}'\mathbf{x}]$, $\text{E}[\mathbf{x}\mathbf{x}']$ and $\text{E}[\mathbf{x}e]$.

(c) Compute $\text{E}[y \mid x_1, x_2]$.

Question 2: Independence

Consider a population with 8 individuals. The following table summarizes all information about the distribution of (y, x_1, x_2):

Individuals	y	x_1	x_2
1	6	1	1
2	6	1	1
3	3	1	1
4	4	1	0
5	5	0	1
6	3	0	0
7	3	0	0
8	0	0	0

Consider a random draw from the population.

(a) For $k_1 \in \{0, 1\}$ and $k_2 \in \{0, 1\}$, compute the marginal probabilities $\text{P}[x_1 = k_1]$ and $\text{P}[x_2 = k_2]$, the joint probabilities $\text{P}[x_1 = k_1, x_2 = k_2]$, and the conditional probabilities $\text{P}[x_2 = k_2 \mid x_1 = k_1]$.

(b) Based on the results in (a), determine whether x_1 and x_2 are independent.

(*Remark: Recall that two discrete random variables x_1 and x_2 with values in $\mathcal{X}$ are independent, if $\text{P}[x_1 = k_1, x_2 = k_2] = \text{P}[x_1 = k_1]\text{P}[x_2 = k_2]$ for each $k_1 \in \mathcal{X}$ and $k_2 \in \mathcal{X}$. Provided $\text{P}[x_1 = k_1] > 0$, this definition is equivalent to $\text{P}[x_2 = k_2 \mid x_1 = k_1] = \text{P}[x_2 = k_2]$ for each $k_1 \in \mathcal{X}$ and $k_2 \in \mathcal{X}$. Alternatively, provided $\text{P}[x_2 = k_2] > 0$, this definition is also equivalent to $\text{P}[x_1 = k_1 \mid x_2 = k_2] = \text{P}[x_1 = k_1]$ for each $k_1 \in \mathcal{X}$ and $k_2 \in \mathcal{X}$.*)

(c) Compute $\text{E}[y \mid x_1 = k_1, x_2 = k_2]$. Use this to compute $\text{E}[y]$ by using Law of Iterated Expectation.

Question 3:

The `cps09mar` dataset contains annual survey data on total salary earnings as well as other self-reported characteristics from 50,742 individuals across the U.S. from 2009. See `cps09mar_description.pdf` on OLAT for the description of variables in `cps09mar`. We obtain the following rounded `average earnings` (measured in 1000 U.S. dollars) according to self-reported gender:

	Men	Women
Education=0 (less than 1st grade)	26	20
Education=6 (5th or 6th grade)	27	20
Education=11 (11th grade, no high school diploma)	35	24

(a) Summarize your conclusions from the numbers above.

(b) Consider the variables `age` and `marital status` in the data. Explain why you think the gender earnings gap might also differ for these variables (no calculation needed).

(c) Can you think of another variable for which the gender earnings gap differs strongly?

(d) Can you think of one for which the gender earnings gap might be smaller?