Mathematical and Statistical Foundations for Data Science(CMPINF 2105)
"Pen and Paper" Homework 2: Linear Regression & Random Samples (Modules 3 & 4)

1. Given the overdetermined system of linear equations:

$$2x_1 + 3x_2 = 5$$
$$4x_1 + 5x_2 = 11$$
$$6x_1 + 7x_2 = 17$$

   Use the Gram matrix approach to find the least squares solution to this system.

2. Consider the following data points for a simple linear regression problem:

x	y
1	2
2	3
3	5
4	7
5	8

   Find the best-fit line $\quad y = \beta_0 + \beta_1 x \quad$ by finding β_0 and β_1.

3. Suppose you have tabular data given by:

x_i	y_i
0	2.0
0.1	2.12
0.2	2.28
0.3	2.48
0.4	2.72
0.5	3.0
0.6	3.32
0.7	3.68
0.8	4.08
0.9	4.52

   Assume the data follows the model

$$y \approx \beta_0 + \beta_1 x + \beta_2 x^2$$

   Find the least squares estimates for β_0, β_1, and β_2.

4. Consider the following discrete random variable X with the probability distribution given below:

x	$P(X = x)$
1	0.2
2	0.5
3	0.3

Calculate the expected value $E[X]$ of the random variable X.

5. Suppose you have a population with an unknown distribution that has an expected value of $E[D] = 50$ and a standard deviation of $\sigma_D = 10$. You draw a sample s of size $n = 100$ from this population. According to the Central Limit Theorem, what is the expected distribution of the sample mean $E[s]$?

6. A factory produces light bulbs that have lifetimes following a distribution with $E[D] = \sigma_D = 800$ hours. If you take a sample s of 100 light bulbs, what is the probability that their average lifetime $E[s]$ is between 720 and 880 hours?