Name: _____

Homework 1

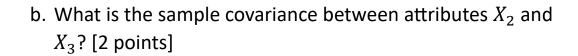
Show your work include any code snippets that you used to generate answers. Complete this assignment individually.

1) What are the two main types of attributes typically find in data. [2 points]

2) Consider the following matrix D; and answer all the questions. [14 points]

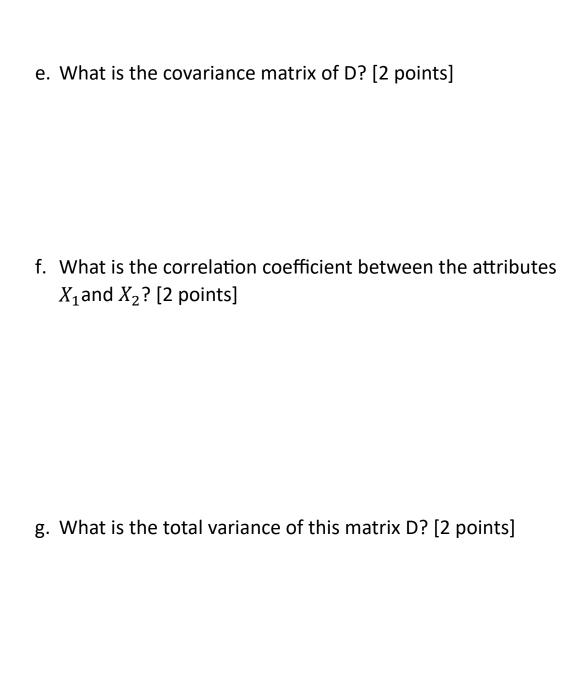
$$D = \begin{array}{ccccc} & X_1 & X_2 & X_3 \\ x_1 & 1.2 & 3 & 1.7 \\ x_2 & 2.4 & 5 & 2.4 \\ x_3 & 4.8 & 35 & 1.2 \\ x_4 & 6.6 & 60 & 3.1 \\ x_5 & -0.5 & 24 & 3.3 \\ x_6 & 3.4 & 32 & 8.4 \\ x_7 & 2.1 & 1 & 6.5 \end{array}$$

a. What is the sample mean of this data attribute X_1 ? [2 points]



c. What is the sample multi-dimensional mean $\hat{\mu}$ of this data matrix? (Your answer should be a vector) [2 points]

d. What is the sample covariance $\hat{\sigma}_{12}$ of the attribute X_1 ? [2 points]



3) Consider the following 5-dimensional vectors: [6 points]

$$a = (1.2 -2.3 4 7.1 -3.12)$$

 $b = (23.2 3 1.2 -3.21 5)$
 $c = (8.2 -4.6 2 1 -2)$

a. What is the $||a - c||_2$? [2 points]

b. What is the $||b - a||_1$? [2 points]

c. What is the angle between the vectors a and c? [2 points]

4) Consider the following matrix D.

$$D = \begin{pmatrix} X_1 & X_2 & X_3 \\ x_1 & 36.6 & Mild & 32 \\ x_2 & 38 & Severe & 21 \\ x_3 & 37 & Extreme & 67 \\ x_4 & 39 & Extreme & 11 \\ x_5 & 27 & Moderate & 71 \end{pmatrix}$$

a. Use the One-Hot encoding method to transform the categorical data into numerical data in the following matrix. You can assume that the attribute X_2 can only contain 4 values: $\{Mild, Moderate, Severe, Extreme\}$ [2 points]

b.	What is the data matrix)	_	$ x_3 - x_5 $ (based on the transformeds)		

5) The following questions reference the Heart Disease data set from the UCI Machine Learning Repository:

https://archive.ics.uci.edu/ml/datasets/Heart+Disease

Answer the following questions about the dataset.

a. How many rows (entities/instances) are there in this dataset? [1 point]

b. How many attributes are in this dataset? [1 point]

c. What is kind of data is stored in the attribute "cigs"? [1 point]