

Bivariate Statistics Exploration

Jacob Bruner

September 22, 2022

1 Calculating Mean of X, Y for Datasets

1.1 Method

The mean is a commonly used measure of center for a given statistic. To calculate the mean, we can use the following formula: for a given indexed set of values, the mean is written:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

Where N is the length of the data set. In other words, the mean is the sum of each value divided by the number of values for a given parameter.

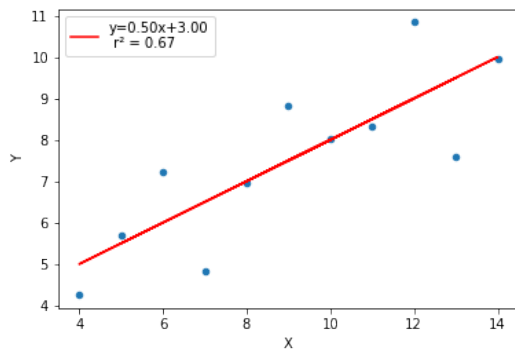
Calculating these values for each dataset we obtain:

Data set	I		II		III		IV	
Parameter	x	y	x	y	x	y	x	y
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5

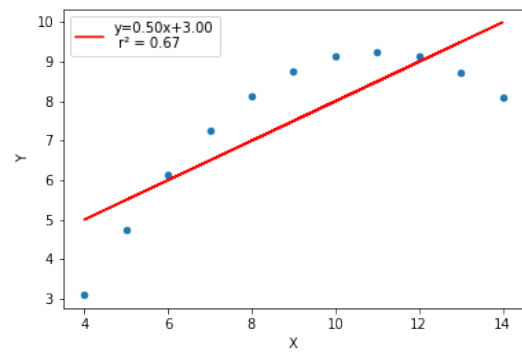
1.2 Interpretation

After computing the mean for each, it's clear that the means for each x, y of the respective datasets are equal. Without any other prior knowledge, one might infer that these datasets have a similar center, save variation or spread. This is because the mean doesn't provide information about the variation or spread of the data. Despite this, however, one might assume it to be a reasonable assumption that these datasets and parameters might be highly similar, especially because they match in two variables.

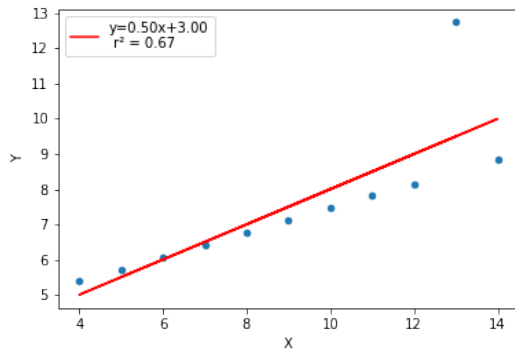
2 Plots



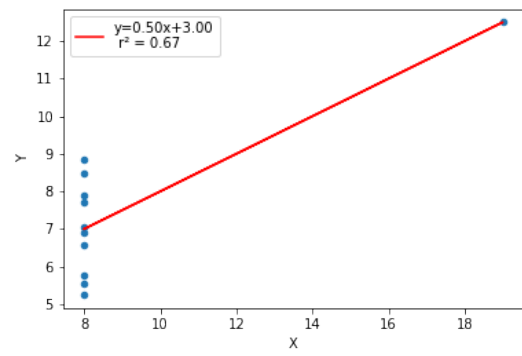
(a) Dataset I



(b) Dataset II



(c) Dataset III



(d) Dataset IV

Figure 1: Plot of Datasets I-IV — X against Y with Lin. Reg.