

Analiza Danych Eksperymentalnych

Sprawozdanie - Laboratorium 2

Regresja liniowa, kwartet Anscombe'a

Michał KORDASZ 241289
Łukasz ŚMIERZCHAŁA 222276

31 października 2021

Termin zajęć:	Środa 11:15 TN
Data laboratorium:	20 października 2021
Prowadzący:	Dr inż. Krzysztof Halawa

Spis treści

1	Dane	3
2	Obliczenia wartości średniej, wariancji i współczynnika korelacji Pearsona	3
2.1	Excel	3
2.2	Python	3
3	Modele liniowe serii danych	4
3.1	Excel	4
3.2	Python	4
4	Wnioski	4

Tabela 1: Dane

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Tabela 2: Wartości Statystyczne

	I	II	III	IV
Pearson:	0.816	0.816	0.816	0.817
Wariancja:	3.752	3.752	3.748	3.748
Średnia:	7.501	7.501	7.500	7.501

1 Dane

Do przeprowadzenia ćwiczenia wykorzystane zostały 4 serie danych, określane także jako kwarter Anscombe’a. Dane przedstawione zostały w tabeli 1.

2 Obliczenia wartości średniej, wariancji i współczynnika korelacji Pearsona

Wykonano podstawą analizę danych poprzez obliczenie średniej wartości, wariancji i współczynnika korelacji Pearsona dla każdej z serii danych.

Wyniki zaprezentowano w tabeli 2

2.1 Excel

Średnią wartość obliczono przy pomocy funkcji AVERAGE.

Wariancję obliczono przy pomocy funkcji VAR.P.

Współczynnik korelacji Pearsona obliczono przy pomocy funkcji PEARSON.

2.2 Python

Wartość średnią oraz wariancję obliczono przy pomocy biblioteki *numpy* korzystając z funkcji *mean* oraz *var*. Współczynnik korelacji Pearsona obliczono przy pomocy biblioteki *scipy* korzystając z funkcji *stats.pearsonr*.

Tabela 3: Współczynniki funkcji liniowej

I		II		III		IV	
β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_0
0.500	3.000	0.500	3.001	0.500	3.002	0.500	3.002

3 Modele liniowe serii danych

Dla każdej z serii danych zdefiniowano model liniowy w postaci

$$y = \beta_1 x + \beta_0$$

Współczynniki β_1 oraz β_0 zostały wyznaczone korzystając z metody najmniejszych kwadratów. Wartości współczynników zostały przedstawione w tabeli 3.

Wykresy danych oraz funkcji liniowych zostały przedstawione na rys. 1.

3.1 Excel

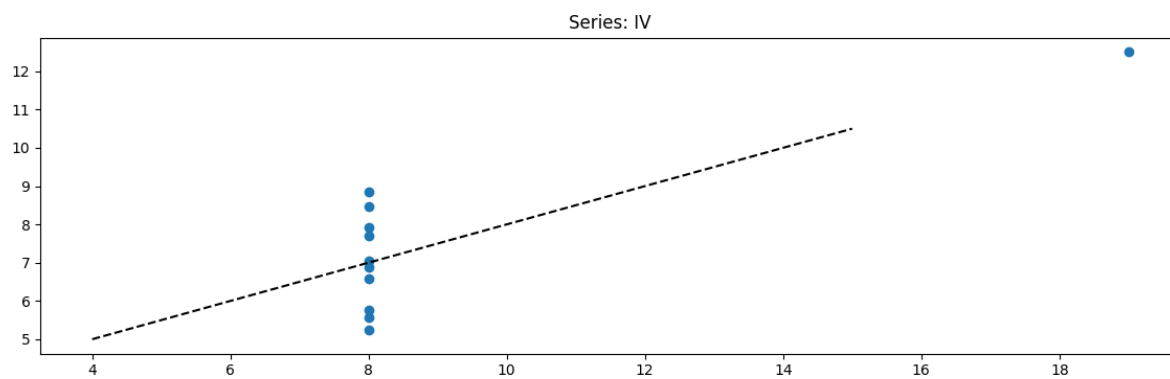
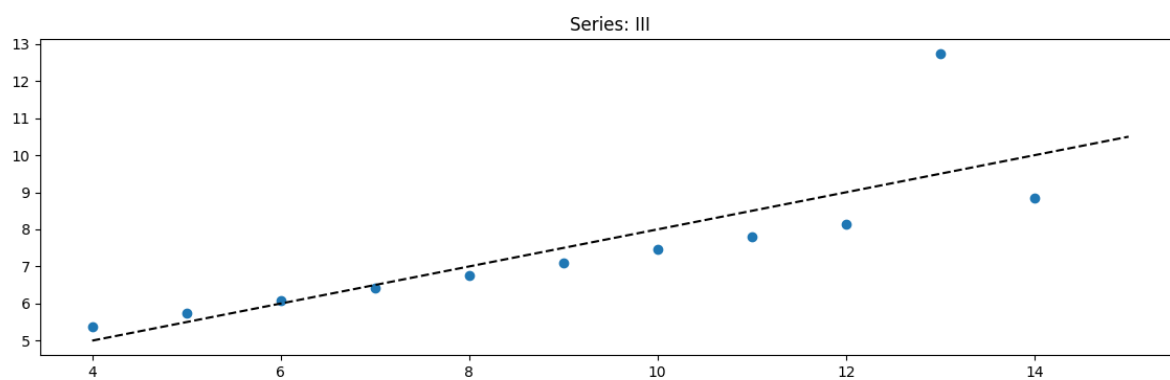
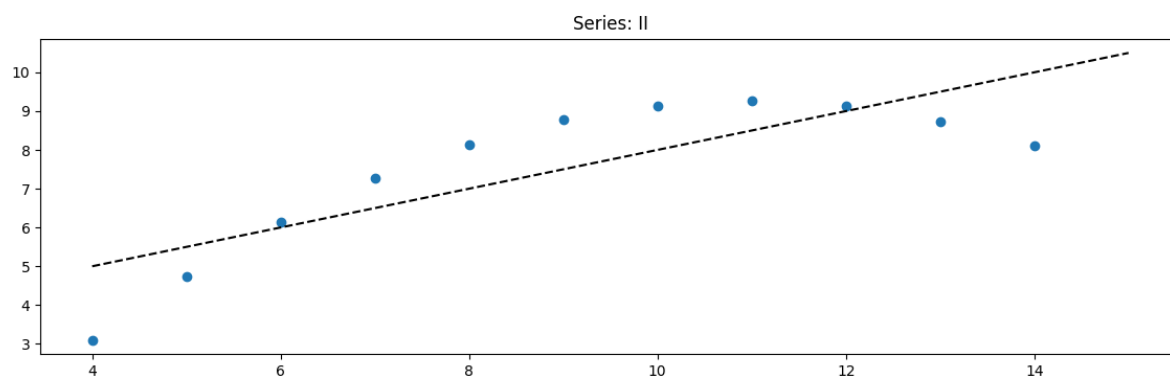
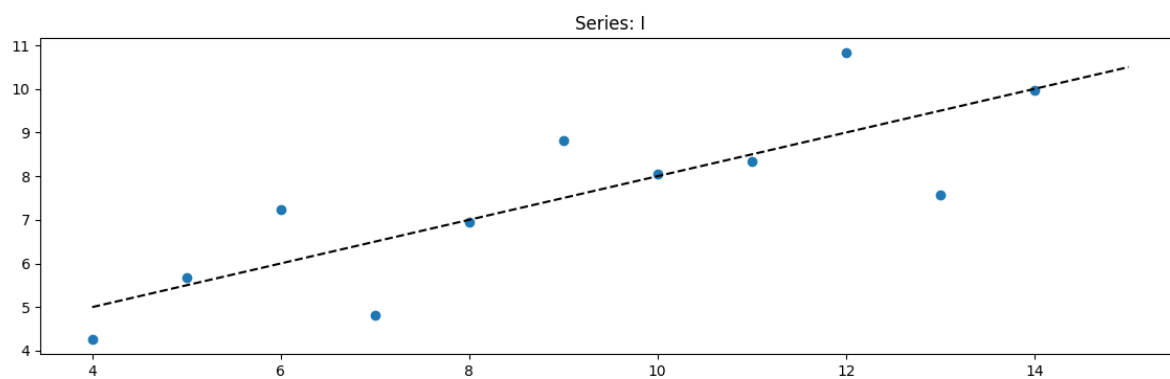
Współczynniki zostały wyznaczone z wykorzystaniem funkcji SLOPE.

3.2 Python

Współczynniki zostały wyznaczone przy pomocy biblioteki *scipy* korzystając z funkcji *linalg.lstsq*.

4 Wnioski

Zarówno wartości statystyczne (średnia, wariancja, współczynnika korelacji Pearsona) oraz modele liniowe dla wszystkich czterech zestawów danych, są w przybliżeniu identyczne. Mimo to, graficzne reprezentacje obrazują jak bardzo różnych są charakterystyki danych. Wartości statystyczne i modele nie zawsze są wystarczające do precyzyjnego opisanie danych, a graficzna reprezentacja danych nie powinna być pomijana podczas ich analizy.



Rysunek 1: Modele Liniowe