PPG prajabatan

BAHAN AJAR

STATISTIKA





Disusun oleh:
Prasetya Adika Putra
2398011279

Kurikulum Merdeka
Untuk SMA Kelas

XI

LIVEWORKSHEETS

A STATISTIKA



Banjir

Perhatikan gambar diatas!

Banjir merupakan sebuah bencana yang terjadi terutama saat musim hujan. Banjir terjadi disebabkan oleh curah hujan yang tingggi. Curah hujan yang sangat tinggi menyebabkan meluapnya air dan kondisi tanah atau daratan tidak dapat menyerap air dengan baik kedalam tanah. Selain karena curah hujan yang tinggi, banjir dapat juga disebabkan karena banyaknya sampah disungai, banyaknya penebangan hutan liar, dan banyaknya bangungan di daerah resapan. Untuk memastikan bahwa hal tersebut benar atau tidak, dapat ditentukan dengan menemukan hubungan atau korelasi antara dua variabel kuantitatif. Korelasi antara dua variabel dapat ditunjukkan menggunakan diagram pencar.

Pematik Belajar

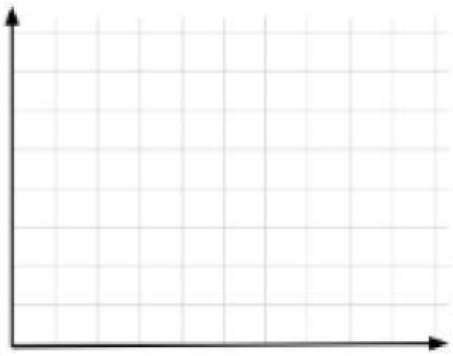
1. Apa yang dimaksud diagram pencar?
2. Apa yang dimaksud korelasi antara dua variabel?

LIVEWORKSHEETS

DIAGRAM PENCAR

URAIAN MATERI

Diagram pencar atau diagram scatter merupakan salah satu cara untuk menyajikan data yang terdiri dari dua variabel kuantitatif. Diagram ini dapat digunakan untuk melihat hubungan antara dua variabel. Diagram pencar memiliki bentuk yang serupa dengan diagram kartesius, yaitu gambar yang melibatkan sumbu horizontal, sumbu vertikal, serta titik yang berasal dari absis dan ordinat yang bersesuai dengan nilai pada datanya. Berikut contoh sumbu horizontal dan sumbu vertikal pada diagram pencar.



Sumbu horizontal menggambarkan nilai pada variabel independen, sedangkan sumbu vertikal menggambarkan nilai pada variabel dependen.

Pahami dulu, yuk!

Perhatikan, **rentang nilai pada sumbu diagram pencar** mengikuti rentang nilai data.

Berikut **cara menggambarkan** diagram pencar.

- ✓ **Tentukan variabel independen sebagai sumbu horizontal** dan **variabel dependen sebagai sumbu vertikal.**
- ✓ **Tentukan skala nilai pada diagram pencar.** Sebelum menentukan skala nilai, sebaiknya tentukan jangkauan dari setiap variabel dahulu.
- ✓ **Gambarkan diagram pencar** dengan menyatakan datanya sebagai bentuk pasangan titik koordinat. Nilai variabel independen sebagai absis dan nilai variabel dependen bersesuai sebagai ordinatnya.

LIVEWORKSHEETS

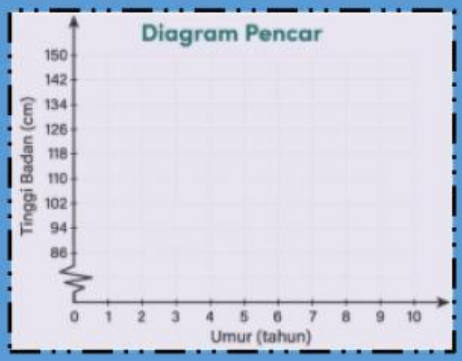
Contoh

Umur (tahun)	Tinggi Badan (cm)	Umur (tahun)	Tinggi Badan (cm)
1	89	5,8	116
1,5	90	6,2	118
2,4	94	6,5	120
2,5	87	7,2	126
2,7	98	7,5	124
3,3	102	7,8	128
3,8	105	8,3	134
4,2	110	8,7	125
4,6	102	9,2	130
5,3	118	9,4	134

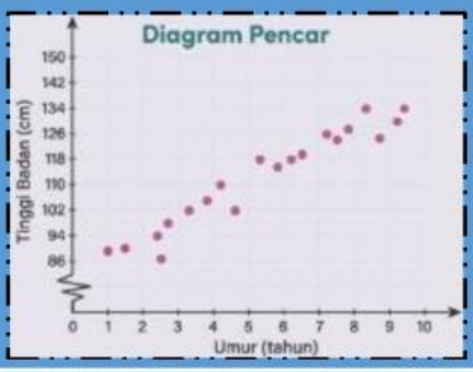
Berikut merupakan data penelitian yang digunakan untuk mengamati pengaruh umur terhadap tinggi badan anak. Gambarkan diagram pencar dari data tersebut!

Penyelesaian :

1. **Perhatikan, penelitian ini bertujuan untuk mengamati pengaruh umur terhadap tinggi badan, maka data umur sebagai variabel independen, sedangkan data tinggi badan sebagai variabel dependen. Untuk sumbu horizontal, karena rentang umur pada tabel berada pada nilai 1 sampai 9,4, maka skala nilai pada sumbu horizontal adalah 1 sampai 10 dengan skala 1 satuan. Untuk sumbu vertikal, karena rentang tinggi badan pada tabel berada pada nilai 87 sampai 134, maka skala nilai pada sumbu vertikal adalah 86 sampai 150 dengan skala 8 satuan. Sehingga, diperoleh skala diagram pencar sebagai berikut.**



2. **Gambarkan data ke dalam diagram pencar. Gambarkan pasangan data umur dan tinggi badan dalam bentuk pasangan titik koordinat (umur, tinggi badan), yaitu (1, 89), (1, 5,90), ..., (9, 4,134)**



LIVEWORKSHEETS

REGRESI LINEAR

VIDEO REGRESI LINEAR

SCAN ME

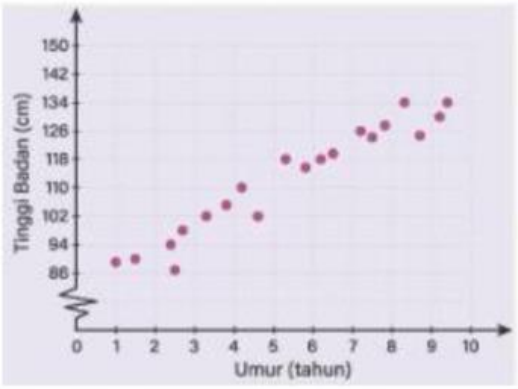
A. PERSAMAAN REGRESI LINEAR

URAIAN MATERI

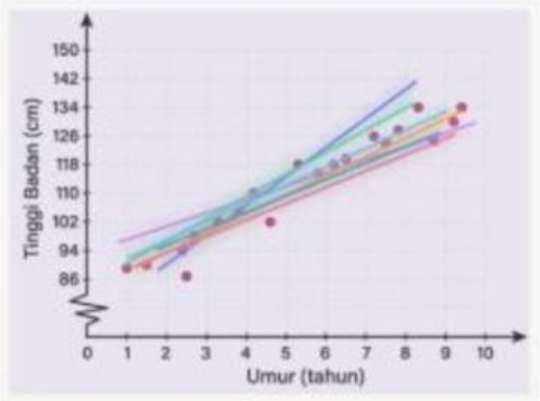
Regresi linear digunakan untuk memodelkan bentuk hubungan antara variabel dependen dan variabel independen pada diagram pencar yang memiliki korelasi linear. Saat data berkorelasi linear, kita dapat menggambarkan suatu garis yang paling tepat untuk mewakili semua data yang diberikan. Regresi linear akan memodelkan data tersebut menggunakan garis lurus yang mendekati data. Selanjutnya, kita bisa memprediksi nilai variabel dependen menggunakan regresi linear saat diberikan nilai variabel independennya.

LIVEWORKSHEETS

Sebagai contoh, seorang peneliti ingin mengetahui pengaruh umur terhadap tinggi badan anak. Kemudian, data yang diperoleh digambarkan pada diagram pencar sebagai berikut.



Perhatikan, data tersebut menyerupai korelasi positif-linear. Sehingga, data tersebut dapat didekati dengan menggunakan garis lurus. Namun, terdapat banyak kemungkinan garis lurus yang dapat dibuat sebagai berikut.



Meskipun demikian, tentu **hanya terdapat satu garis yang paling tepat** untuk memodelkan hubungan kedua variabel pada diagram pencar. Garis yang paling tepat itu disebut garis ***best fit***. Perhatikan, garis best fit **tidak selalu** melalui semua titik data, karena garis ini **hanya** estimasi. Oleh karena itu, garis best fit bisa melalui satu atau lebih titik data.

Misal :

$x = variabel\ independen$

$y = variabel\ dependen$

LIVEWORKSHEETS

Garis *best fit* pasti melalui titik koordinat dengan absis sebesar rata-rata nilai x dan ordinat sebesar rata-rata nilai y. Rata-rata nilai x dinotasikan dengan $\bar{x}$ dan rata-rata nilai y dinotasikan dengan $\bar{y}$. Sehingga, garis best fit selalu melalui koordinat $(\bar{x}, \bar{y})$.

Garis ***best fit*** merupakan garis yang **paling tepat** untuk memodelkan hubungan dua variabel pada diagram pencar. Garis ini **tidak selalu** melewati semua titik data, tapi selalu melalui titik $(\bar{x},\bar{y})$



Persamaan garis regresi memiliki bentuk yang sama dengan bentuk persamaan garis lurus dengan kemiringan tertentu $y = a + bx$, yaitu $\hat{y} = a + bx$.

Dengan $\hat{y}$ adalah prediksi nilai y (variabel dependen) dan x adalah nilai sebenarnya dari x (variabel independen).

PERSAMAAN GARIS REGRESI

$$\hat{y} = a + bx$$

KETERANGAN:
$\hat{y}$: ***nilai prediksi variabel dependen***
x : ***nilai sebenarnya variabel independen***

LIVEWORKSHEETS

Menentukan Pers. Garis Regresi

Dengan menggunakan metode kuadrat terkecil, kita memperoleh koefisien dari persamaan garis regresi ($\hat{y} = a + bx$) sebagai berikut.

$$b = \frac{SS_{xy}}{SS_{xx}}$$

$$SS_{xy} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) \quad (i)$$

$$= \sum_{i=i}^{n}\left(x_i y_i\right) - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n} \quad (ii)$$

$$SS_{xx} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right) \quad (iii)$$

$$= \sum_{i=i}^{n}\left(x_i x_i\right) - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right)}{n} \quad (iv)$$

$$= \sum_{i=i}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} \quad (v)$$

Keterangan :

x_i : nilai variabel independen ke-n

y_i : nilai variabel dependen ke-n

n : banyak pengamatan

$\bar{x}$: rata-rata nilai variabel independen

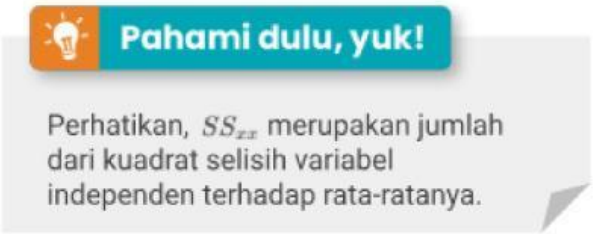
$\bar{y}$: rata-rata nilai variabel dependen

Pahami dulu, yuk!

Perhatikan, SS_{xy} merupakan jumlah dari perkalian antara selisih variabel independen terhadap rata-ratanya dan selisih variabel dependen terhadap rata-ratanya.

LIVEWORKSHEETS

Nilai SS_{xy} dapat ditentukan dengan menggunakan rumus (i). Dengan menggunakan penurunan secara matematis, rumus (i) dapat ditulis kembali menjadi rumus (ii). Dalam modul ini, rumus yang digunakan untuk menentukan nilai SS_{xy} adalah rumus , (ii) karena perhitungan dengan rumus ini lebih sederhana dibandingkan rumus (i).



Nilai SS_{xx} dapat ditentukan dengan menggunakan rumus (iii). Perhatikan, rumus SS_{xx} serupa dengan rumus SS_{xy}, hanya saja variabel dependen (y) harus diganti dengan variabel independen (x). Oleh karena itu, nilai SS_{xx} dapat dihitung dengan menggunakan rumus (v). Perhatikan, nilai koefisien pada persamaan regresi (ii) diperoleh dengan membagi nilai SS_{xy} dengan SS_{xx} . Setelah itu nilai ini digunakan untuk menentukan konstanta pada persamaan regresi. Sebelumnya, ingat kembali bahwa garis best-fit (garis regresi linear) selalu melalui titik koordinat $(\bar{x}, \bar{y})$, sehingga nilai $x = \bar{x}$ dan $y = \bar{y}$ memenuhi persamaan garis regresi $\hat{y} = a + bx$ sebagai berikut.

$$\bar{y} = a + b\bar{x}$$
$$a = \bar{y} - b\bar{x}$$

Perhatikan, nilai a dapat diperoleh dengan menggunakan nilai b, $\bar{x}$, dan $\bar{y}$

LIVEWORKSHEETS

Ayo Kita Mencoba

Masalah

Tabel berikut adalah data mengenai rata-rata berat badan anak usia 6-12 tahun

Usia (tahun)	6	7	8	9	10	11	12
Rata-rata Berat Badan (kg)	20	23	26	29	33	37	42

Tentukan persamaan garis regresi dari data tersebut!

x	y	xy	x^2
6	20		
7	23		
8	26		
9	29		
10	33		
11	37		
12	42		
$\Sigma x =$	$\Sigma y =$	$\Sigma xy =$	$\Sigma x^2 =$

$\bar{x} = \frac{\Sigma x}{n} =$

$\bar{y} = \frac{\Sigma y}{n} =$

$a = \bar{y} - b\bar{x} =$

$\hat{y} = a + bx$
$= \ldots + \ldots x$

$b = \frac{SS_{xy}}{SS_{xx}} =$

$SS_{xy} = \Sigma xy - n\bar{x}\bar{y} =$

$SS_{xx} = \Sigma x^2 - n\bar{x}^2 =$

PASANGKAN URUTAN YANG SESUAI UNTUK MENENTUKAN PERS. REGRESI LINEAR

1 [ ]
2 [ ]
3 [ ]
4 [ ]
5 [ ]
6 [ ]
7 [ ]

LIVEWORKSHEETS

REFLEKSI

Setelah mempelajari materi, peserta didik diharapkan mempu bernalar kritis, serta memiliki sifat yang baik. Sebagai refleksi diri isi tabel berikut dengan tanda (√) sesuai dengan keadaan sebenarnya.

No.	Karakter yang diharapkan	Mampu	Belum Mampu
1.	Saya mempu menjelaskan jenis data		
2.	Saya mampu menentukan korelasi dua variabel		
3.	Saya mampu menganalisis persamaan garis regresi linear		
4.	Saya mampu menganalisis nilai prediksi yang berada di dalam dan di luar rentang data		

LIVEWORKSHEETS