

B. Regresi Linear

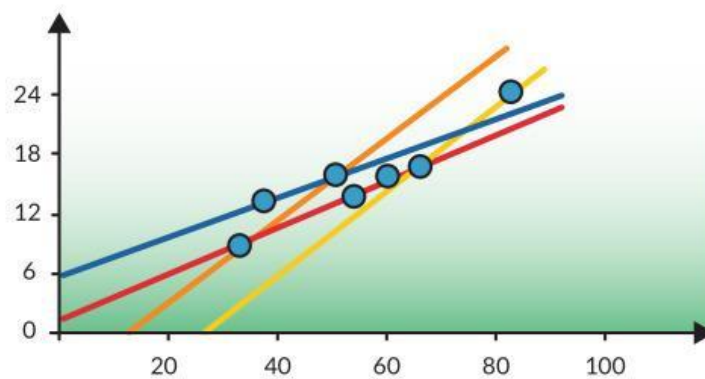
1. Pengertian

Ketika dua variabel kuantitatif pada suatu diagram pencar sudah menunjukkan adanya korelasi, kita dapat menggambar suatu garis yang paling tepat untuk mewakili semua data yang ada.



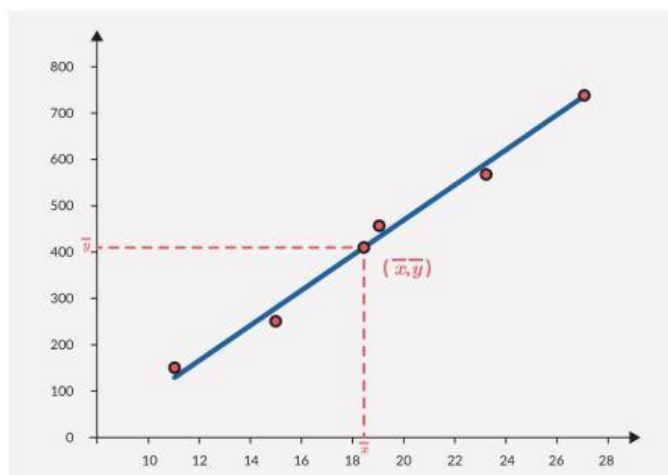
Ayo Berpikir Kritis

Perhatikan dan pikirkanlah garis mana yang paling tepat untuk mewakili data pada diagram pencar di bawah ini. Jika menurut kalian tidak ada garis yang tepat, buatlah garis kalian sendiri. Kemukakan alasan yang mendasari pilihan garis kalian.



Gambar 3.5 Diagram Pencar dan Berbagai Kemungkinan Garis Lurus

Di antara semua garis yang mungkin dibentuk, hanya ada satu garis yang paling tepat yang disebut sebagai garis *best-fit*. Garis ini merupakan model linear yang memperkirakan hubungan antara dua variabel kuantitatif pada diagram pencar tersebut. Model regresi yang memberikan hubungan garis lurus antara dua variabel ini disebut **regresi linear**.



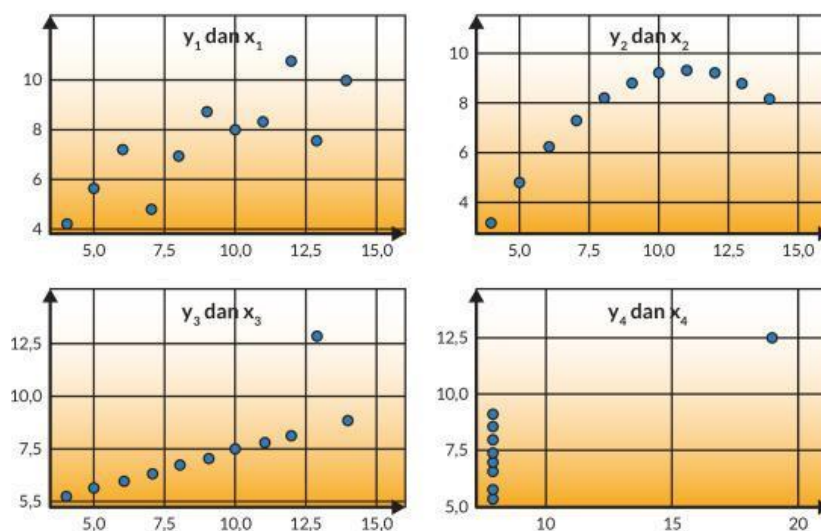
Gambar 3.6 Contoh Regresi Linear

Gambar 3.6 memberikan contoh bagaimana suatu garis *best-fit* digambar di antara titik-titik pada diagram pencar. Garis ini tidak harus melalui titik-titik tersebut karena hanya bersifat estimasi. Namun, bisa saja garis melewati satu titik atau lebih pada saat penggambaran. Tapi satu hal yang pasti adalah garis tersebut selalu melewati pasangan titik koordinat rata-rata nilai x dan y , $(\bar{x}, \bar{y})$.



Ayo Berpikir Kritis

Dapatkah kalian menggunakan model regresi linear sebagai model dari suatu data bivariat yang mempunyai bentuk tren yang tidak menunjukkan bentuk linear? Gunakan gambar diagram pencar di bawah ini untuk membantu berpikir.



Gambar 3.7 Tren Data pada Diagram Pencar

2. Metode Kuadrat Terkecil

Di dalam proses analisis regresi linear, kita mencoba untuk mencari garis lurus yang paling tepat terhadap titik-titik yang ada pada diagram pencar. Garis lurus itu akan memberikan deskripsi terbaik mengenai hubungan antara variabel independen dan dependen. Ayo lakukan Eksplorasi 3.3 sebagai dasar berpikir mengenai metode ini.

**Eksplorasi 3.3**

Suatu sekolah menerapkan program rajin menabung pada seluruh siswanya. Mona tertarik untuk melihat bagaimana hubungan antara uang jajan yang diperoleh teman-temannya dan besar uang yang ditabung. Dia memilih satu kelas dan dari kelas tersebut diambil sampel delapan siswa untuk memperoleh data mengenai uang jajan yang diterima dan uang yang ditabung. Data yang diperoleh adalah sebagai berikut.

Uang jajan	Uang yang ditabung
Rp10.000,00	Rp2.000,00
Rp40.000,00	Rp11.000,00
Rp25.000,00	Rp8.000,00
Rp50.000,00	Rp14.000,00
Rp15.000,00	Rp5.000,00
Rp35.000,00	Rp10.000,00
Rp30.000,00	Rp7.000,00
Rp45.000,00	Rp15.000,00

1. Gambarlah diagram pencar dari data di atas.
2. Gambarlah prediksi garis *best-fit* dari hubungan antara uang jajan dan uang yang ditabung.
3. Jelaskan alasan mengapa kalian menggambar garis lurus seperti itu.
4. Bandingkan prediksi garis *best-fit* yang telah kalian buat dengan prediksi garis *best-fit* yang digambar oleh teman-teman kalian serta bandingkan alasan kalian.
5. Tuliskan kesimpulan dari berbagai ide atau gagasan yang menurut kalian paling tepat untuk menggambar suatu garis *best-fit*.

Berdasarkan Eksplorasi 3.3, pasti banyak ide atau gagasan yang berbeda-beda mengenai cara menggambar garis lurus yang tepat. Namun pertanyaannya, apakah ada cara yang paling tepat dalam memutuskan apakah suatu garis sudah tepat atau

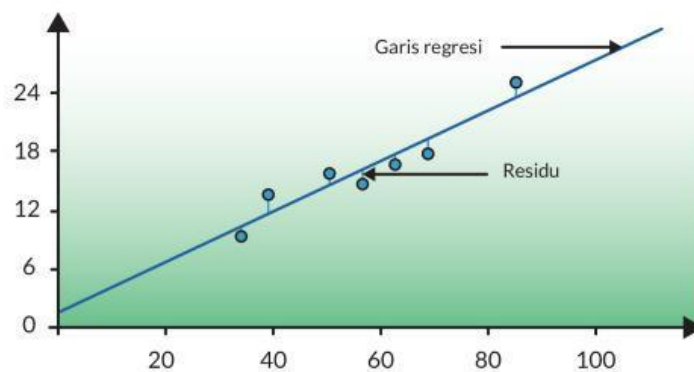
tidak? Ternyata ada loh, metode umum untuk menempatkan garis lurus ke data-data hasil observasi disebut sebagai **metode kuadrat terkecil**.



Tahukah Kamu?

Metode kuadrat terkecil ditemukan oleh Carl Friedrich Gauss (1777–1855) dari Jerman, Adrien-Marie Legendre (1752–1833) dari Prancis, dan Robert Adrain (1775–1843) dari Irlandia sekitar 200 tahun yang lalu yang masing-masing bekerja secara terpisah.

Wow... nama metodenya seperti istilah operasi matematika sehari-hari yang sering kita dengarkan, ya. Sesuai dengan nama metode itu nanti kalian perlu melakukan operasi kuadrat dan tujuannya adalah untuk menemukan suatu nilai terkecil. Nilai apa yang perlu kita kecilkan hingga sekecil mungkin? Mari kita lihat gambar diagram berikut ini sebagai visualisasi sederhana dari metode ini.



Gambar 3.8 Garis Regresi dan Residu

Gambar 3.8 memberikan gambaran bahwa ada selisih antara nilai variabel dependen (y) dari data asli dengan nilai variabel dependen ($\hat{y}$ yang dibaca y topi) dari garis regresi. Selisih antara nilai variabel dependen yang diamati (y) dan nilai variabel dependen yang diprediksi ($\hat{y}$) disebut sebagai **residu** (ε yang dibaca epsilon). Maka dari itu, rumus residu ditulis sebagai berikut.

$$\text{Residu } (\varepsilon) = y - \hat{y}$$

Menurut kalian, jumlah nilai mutlak residu yang semakin kecil atau semakin besar yang akan membuat suatu model garis regresi semakin tepat? Ya, benar, semakin kecil jumlah nilai mutlak residu ini, maka garis semakin dekat ke data asli yang artinya semakin tepat pula garis yang digambar.

Meskipun demikian, perhitungan yang dilakukan ternyata tidak cukup hanya menggunakan konsep jumlah nilai mutlak residu, namun harus menggunakan konsep jumlah kuadrat dari nilai residu (rumus tertulis di bawah paragraf ini) seperti pada konsep perhitungan varians suatu data. Konsep jumlah kuadrat dari nilai residu dapat memberikan karakteristik khusus untuk membedakan setiap garis regresi yang mungkin terbentuk dari suatu kumpulan data yang tidak dapat diberikan oleh konsep jumlah nilai mutlak residu.

$$\text{Kuadrat residu } (\varepsilon^2) = (y - \hat{y})^2$$

$$\text{Jumlah kuadrat residu } (\sum \varepsilon^2) = \sum (y - \hat{y})^2$$

Perlu diingat bahwa setiap kumpulan data mempunyai jumlah kuadrat residu terkecil yang dapat dicapai oleh model garisnya. Dasar inilah yang digunakan dalam penurunan rumus untuk mencari persamaan garis regresi. Akan tetapi, hal ini tidak memungkinkan untuk diajarkan saat ini karena memerlukan ilmu kalkulus lanjutan. Karena itulah buku ini akan berusaha menjelaskannya secara deskriptif. Namun bagi kalian yang tertarik dan ingin belajar lebih lanjut, kalian dapat menemukannya di berbagai buku matematika untuk tingkat universitas atau sumber yang tepat dan baik dari internet.



Ayo Bereksplorasi

Ayo kita kembali pada Eksplorasi 3.3 dan melanjutkan lagi aktivitasnya. Kali ini kalian bisa menggunakan kalkulator untuk mempermudah perhitungan.

6. Dari garis *best-fit* yang telah kalian gambar, carilah persamaan garisnya dengan mencari gradien terlebih dahulu dan titik potong sumbu y kemudian lakukan substitusi ke dalam persamaan garis lurus $\hat{y} = mx + c$ di mana m adalah gradien dan c adalah titik potong sumbu y .
7. Gunakan hasil nomor 6 untuk mencari $\hat{y}$ dan lengkapi tabel berikut ini untuk menghitung jumlah kuadrat residu.

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
10.000	2.000			
40.000	11.000			
25.000	8.000			

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
50.000	14.000			
15.000	5.000			
35.000	10.000			
30.000	7.000			
45.000	15.000			
Jumlah kuadrat residu				

8. Bandingkan hasil jumlah kuadrat residu kalian dengan teman-teman kalian. Prediksi garis *best-fit* yang mempunyai jumlah kuadrat residu terkecil adalah prediksi garis *best-fit* yang lebih tepat.

Latihan 3.3

1. Tabel berikut menunjukkan banyak tempat duduk terhadap biaya per jam dari tiga model pesawat terbang yang digunakan oleh maskapai Garuda Indonesia.

Model Pesawat	Banyak tempat duduk	Biaya (rupiah/jam)
A	50	1.100.000,00
B	100	2.100.000,00
C	150	2.700.000,00

Persamaan garis regresi mana yang lebih tepat untuk memprediksi banyak tempat duduk terhadap biaya?

$$\hat{y} = 367000 + 16000x \text{ atau } \hat{y} = 300000 + 16000x$$

2. Seorang siswa menyelidiki hubungan antara harga (y rupiah) dari 100 gram coklat dan persentase kandungan coklat (x %). Data yang diperoleh disajikan pada tabel berikut.

Merek Cokelat	x (% cokelat)	y (rupiah)
A	10	3.500,00
B	20	5.500,00
C	30	4.000,00
D	35	10.000,00
E	40	6.000,00
F	50	9.000,00
G	60	11.000,00
H	70	13.000,00

- Gambarlah diagram pencar dari data tabel tersebut.
Jika diketahui bahwa persamaan garis regresinya adalah $\hat{y} = 1700 + 154x$.
- Gambarlah garis regresinya pada diagram pencar.
Siswa tersebut melihat bahwa ada satu merek cokelat yang harganya terlalu tinggi.
- Merek cokelat mana yang dimaksud oleh siswa tersebut? Jelaskan alasannya.
- Siswa tersebut ingin memberikan saran harga yang cocok untuk cokelat tersebut. Berapakah prediksi harga yang cocok?

Eksplorasi 3.3 telah memberikan gambaran umum mengenai garis regresi. Sekarang, mari kita melihat bagaimana pendekatan yang dilakukan untuk memperoleh persamaan garis regresi yang memenuhi syarat dari metode kuadrat terkecil supaya kalian dapat menentukannya sendiri. Sama halnya untuk mendapatkan persamaan garis lurus pada umumnya, persamaan garis regresi sering dituliskan dalam bentuk umum berikut ini.

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

Bentuk persamaan di atas dibaca sebagai regresi y atas x , di mana:

$\hat{y}$ adalah nilai variabel dependen yang diprediksi

x adalah nilai variabel independen

a adalah titik potong sumbu y

b adalah gradien garis regresi

Maka dari itu, hal yang perlu dicari adalah nilai a dan b , dan kemudian nilai-nilai tersebut disubstitusikan ke dalam persamaan garis regresi di atas.

Nilai b dapat dihitung menggunakan konsep jumlah kuadrat variabel-variabelnya. Ada dua jenis jumlah kuadrat variabel yang akan digunakan dan disingkat menjadi SS yang merupakan singkatan dari “*sum of squares*” yang berarti jumlah kuadrat, yaitu:

1. Jumlah kuadrat selisih variabel independen x terhadap rata-ratanya dan variabel dependen y terhadap rata-ratanya.

$$SS_{xy} = \sum (x - \bar{x})(y - \bar{y}) \text{ atau } SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

2. Jumlah kuadrat selisih variabel independen x terhadap rata-ratanya.

$$SS_{xx} = \sum (x - \bar{x})(x - \bar{x}) = \sum (x - \bar{x})^2 \text{ atau}$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$

Ingat, jika nilai SS_{xy} dan SS_{xx} berdiri sendiri masing-masing maka mereka tidak memiliki makna apa-apa. Mereka hanyalah perhitungan sementara yang digunakan untuk proses perhitungan berikutnya.

Nilai b dapat dihitung dengan menggunakan kedua jenis jumlah kuadrat di atas sehingga menjadi:

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

Tahap berikutnya kita perlu mencari nilai a . Untuk mencari ini, kita perlu mengetahui bahwa salah satu karakteristik garis regresi yang memenuhi syarat metode kuadrat terkecil adalah titik rata-ratanya $(\bar{x}, \bar{y})$ selalu dilalui garis regresi tersebut.

Oleh karena garis regresi linearnya adalah $\hat{y} = a + bx$ dan titik $(\bar{x}, \bar{y})$ dilalui garis tersebut maka dapat disubstitusikan sehingga menjadi:

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

$$\bar{y} = a + b\bar{x} \dots \dots \text{substitusi } (\bar{x}, \bar{y})$$

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

Sesuai dengan persamaan di atas, jadi untuk mencari nilai a kalian perlu mencari nilai $\bar{x}$, nilai $\bar{y}$ dan nilai b terlebih dahulu.

Setelah memperoleh nilai a dan b , substitusikan nilai-nilai tersebut ke dalam persamaan $\hat{y} = a + bx$. Maka akhirnya kalian akan mendapatkan persamaan garis regresinya.

Ayo kita gunakan tahapan dan rumus di atas untuk menyelesaikan permasalahan pada eksplorasi berikut ini serta menggunakan garis regresi yang diperoleh untuk interpretasi dan analisis lanjutan.



Ayo Bereksplorasi

Eksplorasi 3.4

Tabel berikut ini berisi informasi dari 12 siswa SMA mengenai rata-rata waktu yang digunakan per hari dalam menggunakan media sosial (Facebook, Twitter, dan lain-lain) dan internet untuk bersosialisasi dan hiburan, dan nilai mereka.

Waktu (jam per hari)	4,4	6,2	4,2	1,6	4,7	5,4	1,3	2,1	6,1	3,3	4,4	3,5
Nilai	81	55	78	92	68	55	90	82	67	72	68	84

1. Gambarlah diagram pencar dari data di atas.
2. Apakah diagram pencarnya memberikan indikasi bahwa ada hubungan linear antara rata-rata waktu untuk media sosial dan internet dengan nilai?
3. Tentukan persamaan garis regresinya. Ikutilah tahapan berikut ini.
 - a. Hitunglah nilai $\bar{x}$ dan $\bar{y}$
 - b. Hitunglah nilai SS_{xy} dan SS_{xx}
 - c. Hitunglah nilai b , gradien garis regresi, menggunakan hasil dari a) dan b).
 - d. Hitunglah nilai a , titik potong sumbu y , menggunakan hasil dari a) dan c).
 - e. Tentukan persamaan garis regresinya dengan menggunakan hasil dari c) dan d).
4. Interpretasikan masing-masing arti nilai a dan b yang ditemukan pada nomor 3.
5. Hitunglah prediksi nilai siswa yang menggunakan rata-rata waktu 3,8 jam per hari untuk media sosial dan internet menggunakan persamaan garis regresi yang ditemukan pada nomor 3.

6. Hitunglah prediksi nilai siswa yang menggunakan rata-rata waktu 16 jam per hari untuk media sosial dan internet menggunakan persamaan garis regresi yang ditemukan pada nomor 3. Berikan komentar mengenai hasil yang ditemukan.

Pada Eksplorasi 3.4 nomor 5 kalian telah melakukan suatu proses yang disebut interpolasi. **Interpolasi** adalah penggunaan hubungan antar variabel untuk memprediksi nilai yang berada di dalam jangkauan data. Sedangkan pada nomor 6 kalian telah melakukan suatu proses yang disebut ekstrapolasi. **Ekstrapolasi** adalah penggunaan hubungan antar variabel untuk memprediksi nilai yang berada di luar jangkauan data dengan asumsi bahwa hubungan ini berlaku meskipun di luar jangkauan data. Tentunya, hasil interpolasi lebih dapat dipercaya dibandingkan dengan ekstrapolasi.



Ayo Berkomunikasi

Coba diskusikan dengan teman-teman kalian, menurut kalian kapan hasil ekstrapolasi cukup dapat dipercaya?

Latihan 3.4

1. Pada saat kondisi mendadak, para pengendara mobil memerlukan waktu yang berbeda-beda untuk dapat bereaksi untuk menginjak rem mobil. Jarak yang diperlukan hingga terjadi reaksi menginjak rem disebut sebagai jarak reaksi. Tabel berikut ini memberikan informasi mengenai jarak reaksi dari mobil yang melaju dengan kecepatan yang berbeda-beda.

Kecepatan (km/jam)	Jarak Reaksi (m)
20	4,1
30	6,2
40	8,3
50	10,1
60	12,4
70	14,5

- a. Gambarlah diagram pencar dari data di atas.
 - b. Apakah diagram pencarnya memberikan indikasi bahwa ada hubungan linear antara kecepatan dengan jarak reaksi?
 - c. Tentukan persamaan garis regresinya.
 - d. Interpretasikan nilai a dan b yang diperoleh pada bagian c).
 - e. Hitunglah prediksi jarak reaksi jika suatu mobil bergerak dengan kecepatan 35 km/jam.
 - f. Hitunglah prediksi jarak reaksi jika suatu mobil bergerak dengan kecepatan 55 km/jam.
2. Tabel berikut ini adalah data mengenai rata-rata tinggi badan anak perempuan yang berumur dari 2–14 tahun.

Umur (tahun)	2	3	4	5	6	7	8	9	10	11	12	13	14
Rata-rata tinggi (cm)	89	98	105	112	118	123	131	136	143	151	155	160	161

- a. Gambarlah diagram pencar dari data di atas.
- b. Apakah diagram pencarnya memberikan indikasi bahwa ada hubungan linear antara umur dengan rata-rata tinggi badan?
- c. Tentukan persamaan garis regresinya.
- d. Interpretasikan nilai a dan b yang diperoleh pada bagian c).
- e. Hitunglah prediksi tinggi badan anak perempuan yang berumur 5,8 tahun.
- f. Hitunglah prediksi tinggi badan seorang perempuan yang sudah berumur 30 tahun.
- g. Berikan komentar mengenai reliabilitas nilai perkiraan di bagian f).

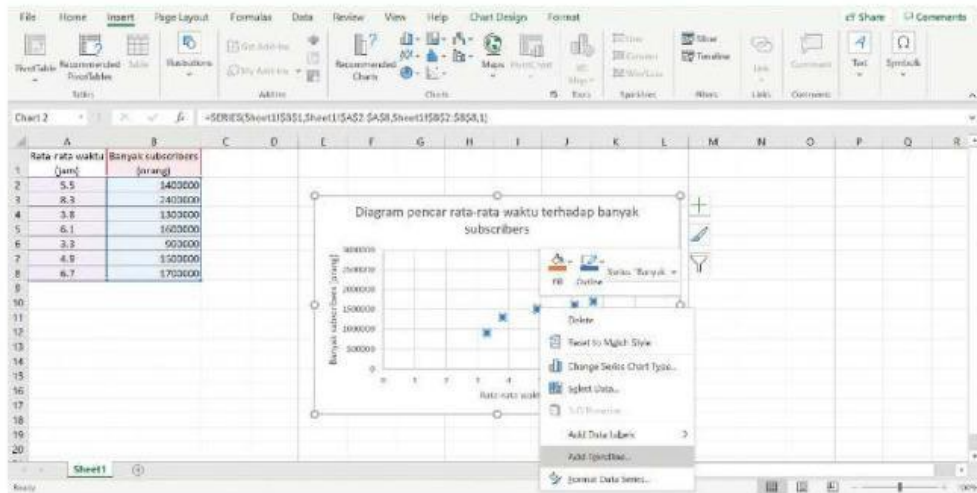


Ayo Berteknologi

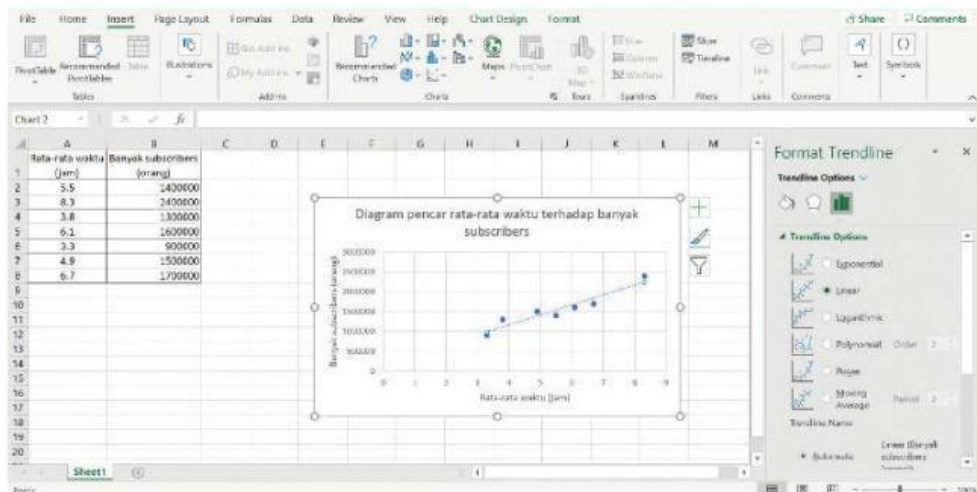
Apakah kalian sudah mahir menggunakan aplikasi *Microsoft Excel* untuk menggambar diagram pencar? Kalian bisa selalu kembali ke bagian panduan untuk melatih ulang keterampilan menggambar diagram pencar. Ayo, kita berlatih kembali.

Kali ini kita akan menggunakan aplikasi yang sama untuk membantu menggambar garis regresi linear. Kalian juga bisa langsung tahu persamaan garis regresinya tanpa harus dihitung menggunakan rumus yang telah kalian pelajari sebelumnya. Sangat efisien, bukan? Yuk, perhatikan tahapan di bawah ini. Kita akan menggunakan data yang sama dengan latihan membuat diagram pencar sebelumnya.

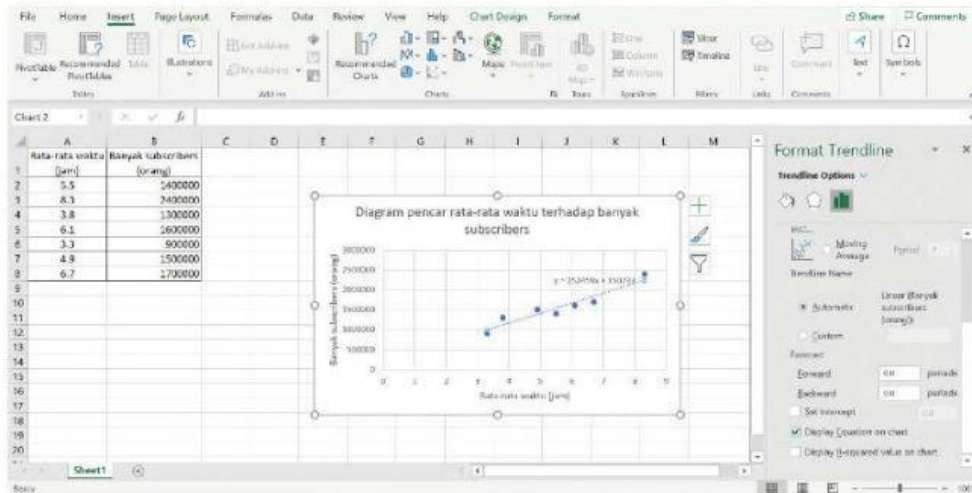
1. Klik salah satu titik data pada diagram, kemudian klik kanan dan pilih “Add Trendline.”



2. Menampilkan garis regresi linear: Setelah tahapan no.1, akan ada tampilan menu tambahan di sebelah kanan dan pilihlah “Linear” pada Trendline Options.



3. Menampilkan persamaan garis regresi linear: pada menu yang sama di sebelah kanan, pada bagian bawahnya ada pilihan untuk menampilkan persamaan garis regresi, pastikan centang “Display Equation on chart” maka kalian dapat melihat bahwa persamaan garis regresi sudah ditampilkan pada diagram.



Mudah, bukan? Jika sudah terbiasa, maka tahapan ini hanya akan memerlukan waktu hitungan detik saja untuk menggambar garis regresi dan menemukan persamaan garis regresinya.



Ayo Berpikir Kritis

Ayo bandingkan persamaan garis yang diberikan oleh *Microsoft Excel* dengan hasil perhitungan persamaan garis regresi menggunakan rumus.

1. Apakah jawaban kalian sama atau berbeda?
2. Jika jawabannya berbeda, apa penyebab perbedaan tersebut?
3. Interpretasikan nilai koefisien dari variabel x pada garis regresi yang kalian dapatkan dikaitkan dengan rata-rata waktu dan banyak *subscribers*.
4. Berapa banyak *subscribers* dari seorang *YouTuber* yang bekerja dengan rata-rata waktu 5 jam per hari?



Ayo Berkomunikasi

Jika memang penggunaan teknologi akan mempermudah segala hal baik dalam menggambar diagram pencar, menggambar garis regresi dan menemukan persamaan garis regresi, kenapa kalian masih harus melakukan semua hal di atas secara manual dengan menggunakan gambar tangan dan menggunakan rumus?



Ayo Berefleksi

Mari kita merefleksikan kembali hal-hal apa saja yang telah kita pelajari.

1. Apakah saya sudah bisa menentukan persamaan garis regresi?
2. Apakah saya sudah bisa memberikan interpretasi komponen garis regresi, yaitu gradien dan titik potong sumbu y ?
3. Apakah saya sudah bisa melakukan proses interpolasi dan ekstrapolasi?