

Penyajian Data Statistik

Pada penulisan **Kedua** tentang **Statistika Elementer** ini, penulis akan memberikan bahasan mengenai **Penyajian Data Statistik** yang bertujuan untuk menjelaskan bagaimana cara memilih dan menentukan bentuk penyajian data pengamatan dalam **Tabel dan Diagram Statistik**. Beberapa orang terkadang masih bingung bagaimana menerapkan penyajian data statistik yang benar dan sesuai dengan jenis data pengamatannya.

Penyajian Data Statistik pada Data Pengamatan.

Penyajian Data Statistik pada penulisan ini diberikan berupa Tabel dan Diagram Statistik yang digunakan untuk menggambarkan karakteristik dari data pengamatan. Tabel Statistik yang digunakan adalah **Tabel Distribusi Frekuensi** dan **Tabel Tabulasi Silang**. Dari kedua tabel tersebut, masing-masing akan diberikan visualisasi kedalam **Diagram Statistik** yang menjadi pelengkap penyajian data. Pada penyajian data, Tabel dan Diagram disesuaikan berdasarkan dua jenis data pengamatan yaitu **data kualitatif** dan **data kuantitatif** (diskrit & kontinu). Dua jenis data pengamatan ini sudah dijelaskan pada penulisan **Statistika Elementer** yang pertama (**Pengantar Statistika**).

Pengertian Distribusi Frekuensi.

Distribusi Frekuensi adalah suatu pengaturan baris data kedalam bentuk tabel dengan menggunakan kelas dan frekuensi^[1]. Terdapat dua tipe distribusi frekuensi yang sering digunakan, yaitu **Distribusi Frekuensi Kategori** dan **Distribusi Frekuensi Kelompok**. Distribusi Frekuensi Kategori digunakan untuk **data kualitatif** atau **data kategori** (berskala **nominal** dan **ordinal**); sedangkan Distribusi Frekuensi Kelompok digunakan untuk **data kuantitatif** atau **data numerik** (berskala **interval** dan **rasio**) ketika jangkauan data (**range** = $x_{MAX} - x_{MIN}$) bernilai **besar**, sehingga data pengamatan harus dikelompokkan penyebarannya kedalam kelas atau interval tabel agar penyajian data menjadi lebih ringkas.

Bagaimana Menyajikan Tabel Distribusi Frekuensi Kategori ?

Penyajian Tabel Distribusi Frekuensi Kategori untuk **data kualitatif** diberikan dengan tahapan^[1] sebagai berikut:

1. Menentukan **Kelas/interval** kategori.
2. Menghitung **Frekuensi** setiap kategori.
3. Menghitung **Persentase** dari nilai dalam setiap kelas dengan menggunakan rumusan **Persentase = $f/n * 100%$** , dengan f adalah frekuensi setiap kelas dan n adalah nilai total atau banyaknya pengamatan.
4. Membuat Tabel Distribusi Frekuensi.

Contoh Tabel Distribusi Frekuensi Kategori untuk Data Kategori :

Misalkan diberikan studi kasus **penentuan tipe darah**. Dipilih 20 orang untuk diambil sampel darahnya. Hal ini dilakukan untuk mengetahui tipe darah peserta, sehingga diperoleh sampel darah sebagai berikut:

A B B AB O O O B AB B B B O A O A O O O AB

Dari kasus ini, maka akan dihasilkan Tabel Distribusi Frekuensi Kategori dengan ketentuan **Kelas** kategori yaitu semua tipe darah (A,AB,B,O), selanjutnya dilakukan perhitungan frekuensi dan persentase frekuensi untuk setiap kategori sehingga dihasilkan Tabel Distribusi Frekuensi Kategori seperti ini.

Tabel 1. Tabel Distribusi Frekuensi Kategori untuk Studi Kasus Penentuan Tipe Darah

Kelas Kategori (Tipe Darah)	Frekuensi (f)	Persentase
A	3	15%
AB	3	15%
B	6	30%
O	8	40%
Total =	20	100%

Dari sampel yang diamati, diperoleh hasil studi bahwa dari dua puluh orang responden terdapat 40% orang yang mayoritas memiliki tipe darah O, di urutan kedua adalah tipe darah B sebesar 30% responden, dan tipe darah AB dan A masing-masing sebanyak 15% responden.

Diagram Statistik yang digunakan untuk kasus ini adalah **Diagram Batang (Bar Chart)** dan **Diagram Lingkaran (Pie Chart)**. **Diagram Batang** adalah diagram yang menggambarkan data dengan menggunakan batang vertikal atau horizontal yang memiliki tinggi atau panjang

yang menunjukkan frekuensi data, sedangkan **Diagram Lingkaran** adalah suatu lingkaran yang dibagi menjadi bagian-bagian atau irisan-irisan berdasarkan persentase dari frekuensi dalam setiap kategori dari suatu distribusi^[1]. Diagram Batang dan Diagram Lingkaran digunakan untuk menggambarkan penyebaran suatu data pengamatan yang bertipe kategori. Berikut hasil output Diagram Batang dan Diagram Lingkaran untuk studi kasus ini.

Diagram Batang untuk Studi Kasus Penentuan Tipe Darah

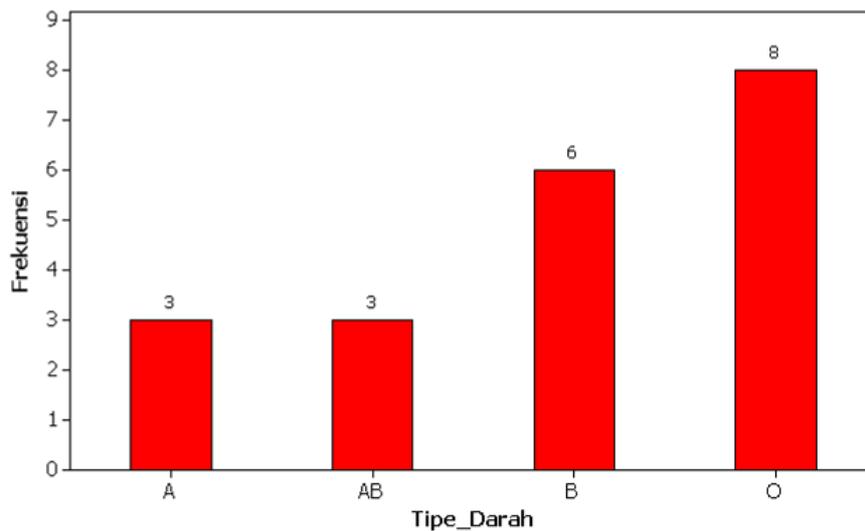
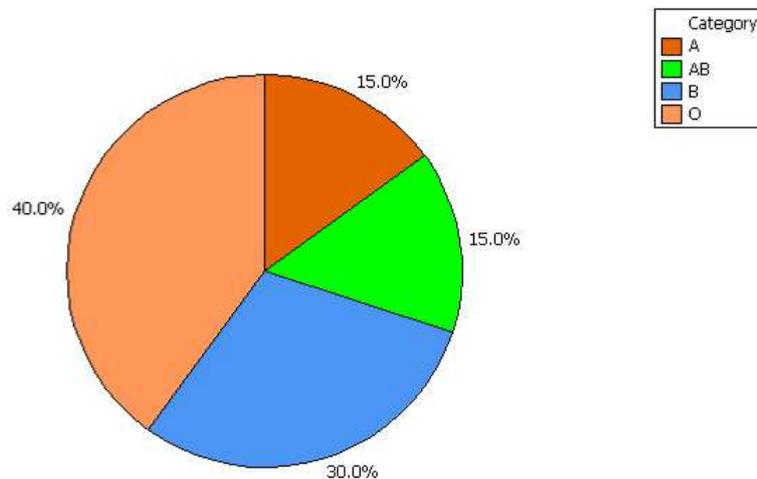


Diagram Lingkaran untuk Studi Kasus Penentuan Tipe Darah



Bagaimana Menyajikan Tabel Distribusi Frekuensi Kelompok ?

Penyajian Tabel Distribusi Frekuensi Kelompok untuk **data kuantitatif** (diskrit & kontinu) diberikan dengan tahapan sebagai berikut:

1. **Mengurutkan data** dengan tujuan mempermudah perhitungan frekuensi dari setiap kelas/interval.

2. Menghitung **Banyaknya Kelas (m)** dengan menggunakan rumus^[3] $m = n^{1/2}$ atau menerapkan **aturan Sturges** $m = 1 + 2^{\log(n)} = 1 + [\log(n) / \log(2)]$. Ada beberapa buku yang menyarankan untuk menentukan banyaknya kelas^[4] m antara 5 dan 20. Nilai m dibulatkan keatas agar bernilai bilangan bulat.
3. Menghitung **Panjang Kelas (p)** dengan menggunakan rumus $p = range/m$ dengan $range = x_{MAX} - x_{MIN}$. Nilai p juga dibulatkan keatas agar bernilai bilangan bulat. Dari tahapan 1 sampai 3, diharapkan terbentuknya suatu Kelas yang mengakomodasi semua data (**Mutually Exhaustive**). Dengan kata lain, tidak ada data pengamatan yang tidak memiliki kelas/interval.
4. Menghitung **Nilai Batas Bawah pada interval/kelas pertama** dengan rumusan $x_{L,1} = x_d$ untuk data diskrit dan $x_{L,1} = x_{MIN} - c$ untuk data kontinu. Pada penulisan ini, nilai x_d ditentukan sebarang dengan syarat $x_d \leq x_{MIN}$; Sedangkan nilai c adalah faktor koreksi yang dipilih seperti: 0,5 , 0,05 , 0,005,
5. Menghitung **Nilai Batas Bawah pada kelas selanjutnya** dengan rumusan $x_{L,i} = x_{L,(i-1)} + p$ untuk $i = 2,3,\dots,m$, sedangkan **Nilai Batas Atas dari setiap kelas** diatur dan disesuaikan berdasarkan **Nilai Batas Bawah pada kelas selanjutnya** agar suatu titik/data tidak dapat masuk kedalam dua kelas/interval (**Mutually Exclusive**).
6. Menghitung **Frekuensi (f)** dan **Frekuensi Relatif (f/n)**.
7. Membuat Tabel Distribusi Frekuensi.

Contoh Tabel Distribusi Frekuensi Kelompok untuk Data Diskrit :

Misalkan diberikan **studi kasus untuk data diskrit**, yaitu **proses produksi sekrup baja**. Seorang peneliti melakukan pengamatan proses produksi sekrup baja dengan bertujuan untuk mengetahui distribusi frekuensi dari sekrup baja yang cacat. Diambil 50 sampel dari 1000 sekrup baja yang diproduksi. Hal ini dilakukan sebanyak 30 kali produksi sehingga diperoleh banyaknya produk yang cacat dari setiap pengambilan sampel produksi sebagai berikut:

13 6 10 6 9 12 7 11 10 13 13 6 7 8 7 12 6 8 14 9 16 14 12 9 12 5 11 8 12 14 .

Penyelesaian:

Tabel Distribusi Frekuensi Kelompok dibentuk berdasarkan data terurut agar mempermudah perhitungan frekuensi dari setiap kelas. Berikut diberikan data yang sudah terurut:

5 6 6 6 6 7 7 7 8 8 8 9 9 9 10 10 11 11 12 12 12 12 12 13 13 13 14 14 14 16 .

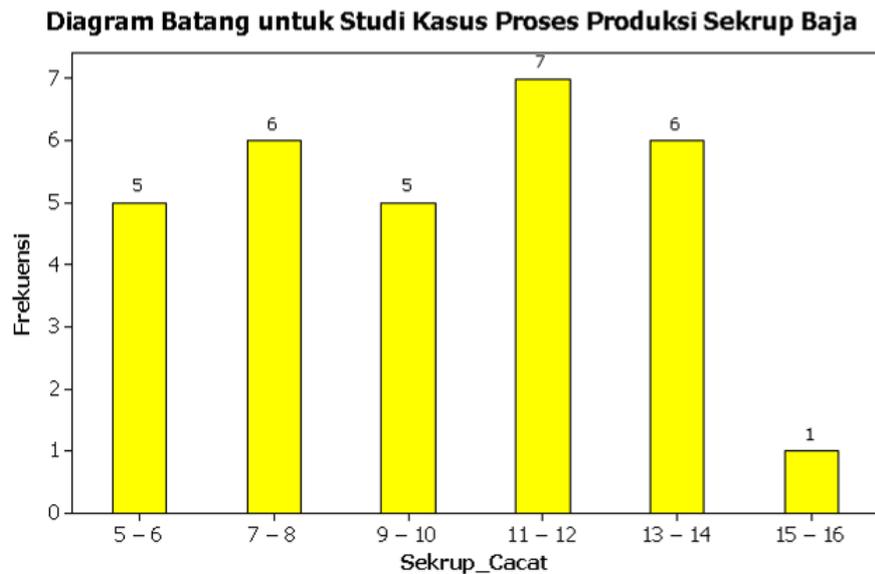
Selanjutnya, dilakukan penentuan **Nilai Batas Bawah pada kelas pertama** (x_d) sama dengan 5, selanjutnya diperoleh **Banyaknya Kelas** $m = 1 + (\log n / \log 2) = 1 + (\log 30 / \log 2) = 5,91 \approx 6$ dan **Panjang Kelas** $p = (x_{MAX} - x_{MIN})/m = (16 - 5)/6 = 1,83 \approx 2$. Demikian hingga, dihitung **Nilai Batas Bawah pada kelas selanjutnya** dengan rumusan $x_{L,i} = x_{L,(i-1)} + p$ untuk $i = 2,3,\dots,m$ sedangkan **Nilai Batas Atas dari setiap kelas** diatur dan disesuaikan berdasarkan **Nilai Batas Bawah pada kelas selanjutnya**. Misalkan **Nilai Batas Bawah pada kelas kedua** dengan rumusan $x_{L,2} = x_{L,1} + p = 5 + 2 = 7$ sedangkan **Nilai Batas Atas pada kelas pertama** disesuaikan menjadi 6; begitu juga seterusnya. Pada tahapan terakhir, dilakukan perhitungan **frekuensi & frekuensi relatif** dalam tabel berikut:

Tabel 2. Tabel Distribusi Frekuensi Kelompok untuk Studi Kasus Proses Produksi Sekrup Baja

Banyaknya Sekrup Cacat	Frekuensi	Frekuensi Relatif
5 – 6	5	0,17
7 – 8	6	0,20
9 – 10	5	0,17
11 – 12	7	0,23
13 – 14	6	0,20
15 – 16	1	0,03
Total	30	1

Pada proses produksi sekrup baja, banyaknya cacat yang sering terjadi terletak pada interval 11 sampai 12 dengan frekuensi sebanyak 7 dan frekuensi relative sebesar 0,23.

Diagram Statistik yang dapat digunakan untuk kasus ini adalah **Diagram Batang** dengan menentukan bahwa kelas/interval sebagai kategori. Berikut hasil output Diagram Batang untuk studi kasus ini.



Contoh Tabel Distribusi Frekuensi Kelompok untuk Data Kontinu :

Misalkan diberikan **studi kasus untuk data kontinu**, yaitu **proses pengukuran tinggi temperature ($^{\circ}\text{F}$)**^[1]. Diperoleh data pengamatan temperature dari 50 negara bagian yang diukur dan dicatat sebagai berikut:

112 100 127 120 134 118 105 110 109 112 110 118 117 116 118 122 114 114 105 109 107 112 114
115 118 117 118 122 106 110 116 108 110 121 113 120 119 111 104 111 120 113 120 117 105 110
118 112 114 114 .

Penyelesaian:

Tabel Distribusi Frekuensi Kelompok dibentuk berdasarkan data terurut seperti penyelesaian contoh sebelumnya. Berikut diberikan data yang sudah terurut:

100 104 105 105 105 106 107 108 109 109 110 110 110 110 110 111 111 112 112 112 112 113 113
114 114 114 114 114 115 116 116 117 117 117 118 118 118 118 118 118 119 120 120 120 120 121
122 122 127 134 .

Ditentukan **Nilai Batas Bawah pada kelas pertama** adalah $100 - 0,5 = 99,5$. Selanjutnya diperoleh hasil perhitungan **Banyaknya Kelas** $m = 1 + (\log n / \log 2) = 1 + (\log 50 / \log 2) = 6,64 \approx 7$ dan **Panjang Kelas** $p = (x_{MAX} - x_{MIN})/m = (134 - 100)/7 = 4,86 \approx 5$. Demikian hingga, dihitung **Nilai Batas Bawah pada kelas selanjutnya** dengan rumusan $x_{L,i} = x_{L,(i-1)} + p$ untuk $i = 2,3,\dots,m$ sedangkan **Nilai Batas Atas dari setiap kelas** diatur dan disesuaikan berdasarkan **Nilai Batas Bawah pada kelas selanjutnya**. Misalkan **Nilai Batas Bawah pada kelas kedua** dengan rumusan $x_{L,2} = x_{L,1} + p = 99,5 + 5 = 104,5$ sedangkan **Nilai Batas Atas pada kelas pertama** disesuaikan menjadi kurang dari 104,5; begitu juga seterusnya. Pada tahapan terakhir, dilakukan perhitungan **frekuensi & frekuensi relatif** dalam tabel berikut:

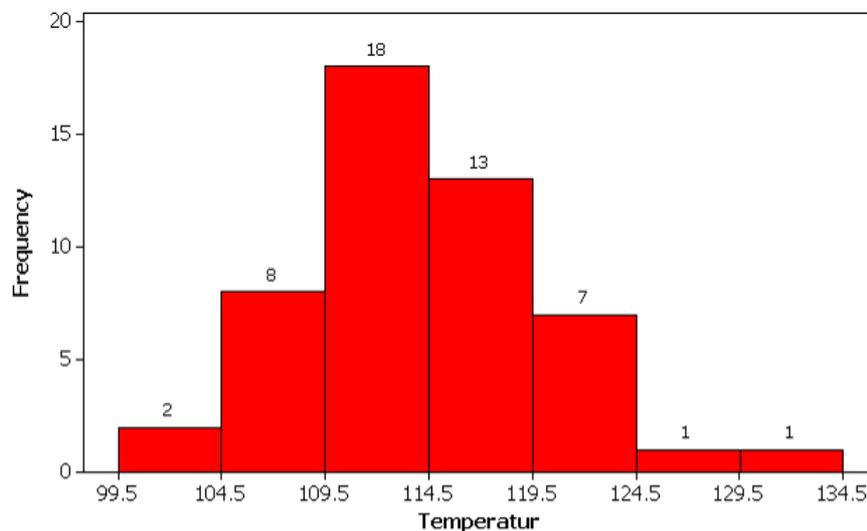
Tabel 3. Tabel Distribusi Frekuensi untuk Proses Pengukuran Tinggi Temperatur

Tinggi Temperatur (dalam Fahrenheit)	Frekuensi (<i>f</i>)	Frekuensi Relatif
$99,5 \leq x < 104,5$	2	0,04
$104,5 \leq x < 109,5$	8	0,16
$109,5 \leq x < 114,5$	18	0,36
$114,5 \leq x < 119,5$	13	0,26
$119,5 \leq x < 124,5$	7	0,14
$124,5 \leq x < 129,5$	1	0,02
$129,5 \leq x \leq 134,5$	1	0,02
Total	50	1

Pada proses pengukuran tinggi temperatur, diperoleh hasil ringkasan tabel distribusi frekuensi dari 50 negara bagian yang menunjukkan bahwa terdapat 36% negara bagian yang mayoritas memiliki tinggi temperature ($^{\circ}\text{F}$) pada interval $109,5^{\circ}\text{F} - 114,4^{\circ}\text{F}$.

Diagram Statistik yang digunakan untuk kasus ini adalah **Histogram**. **Histogram** adalah suatu diagram yang menggambarkan data dengan menggunakan baris vertikal dengan tinggi yang bervariasi untuk menunjukkan frekuensi dari suatu kelas^[1]. **Histogram** digunakan untuk menggambarkan penyebaran suatu data pengamatan yang bertipe **kontinu**. Berikut hasil output Histogram untuk studi kasus ini.

Histogram untuk Studi Kasus Proses Pengukuran Tinggi Temperatur



Penyajian Tabel Tabulasi Silang.

Tabel Tabulasi Silang (*Cross-Tabulation Table*) digunakan untuk mengetahui distribusi frekuensi respon dari hubungan antara dua variabel penelitian dalam bentuk baris dan kolom. Data pengamatan yang digunakan pada dua variabel tersebut bersifat kategori,

sehingga dihasilkan respon berupa frekuensi dari hubungan antara dua variabel dalam tabel tersebut.

Contoh Tabel Tabulasi Silang untuk Data Kategori :

Seorang manajer marketing ingin mengetahui distribusi frekuensi responden mengenai Kepuasan Produk yang dipasarkan. Digunakan variabel *Gender* dengan kategori bernilai 1 untuk Pria dan 0 untuk Wanita, serta variabel Kepuasan dengan kategori bernilai 1 untuk Sangat Tidak Setuju, 2 untuk Tidak Setuju, 3 untuk Agak Setuju, 4 untuk Setuju, dan 5 untuk Sangat Setuju.

Tabel 4. Tabel Pengamatan untuk Responden suatu Produk

Responden	Gender	Kepuasan
1	0	2
2	0	2
3	1	1
4	1	3
5	1	1
6	1	5
7	0	2
8	0	4
9	1	2
10	1	4

Dari data pengamatan ini, dihasilkan Tabel Tabulasi Silang responden dari *Gender* versus Kepuasan terhadap Produk yang dipasarkan adalah.

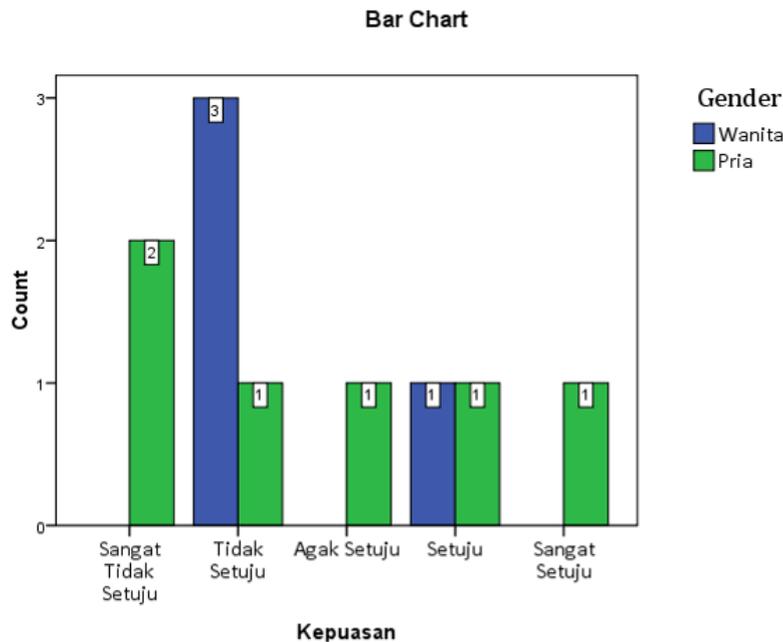
Tabel 5. Tabel Tabulasi Silang untuk Studi Kasus Kepuasan Suatu Produk

Frekuensi Responden		Gender		Total
		Wanita	Pria	
Kepuasan	Sangat Tidak Setuju	0	2	2
	Tidak Setuju	3	1	4
	Agak Setuju	0	1	1
	Setuju	1	1	2
	Sangat Setuju	0	1	1
Total		4	6	10

Sebagian besar responden **Tidak Puas** (4 responden) dengan produk yang dipasarkan, yaitu responden wanita 3 orang dan Pria 1 orang; serta terdapat 2 orang Pria yang **Sangat Tidak Puas** dengan produk tersebut.

Diagram Statistik yang digunakan untuk kasus ini adalah **Diagram Batang dengan Cluster**.

Berikut hasil output Diagram Batang dengan *Cluster* untuk studi kasus ini.



Penyajian *Scatter Plot*.

Scatter Plot (atau diagram pencar) digunakan untuk menggambarkan data kuantitatif yang berpasangan (x,y) dengan sumbu horizontal x dan sumbu vertikal y ^[4]. **Data berpasangan** merupakan suatu data pengamatan (x,y) yang dicatat secara bersamaan dan memiliki suatu keterkaitan yang tidak dapat dipisahkan. Pola dari titik-titik yang digambar dalam *Scatter Plot* ini sering membantu dalam menentukan pola hubungan antara dua variabel pengamatan.

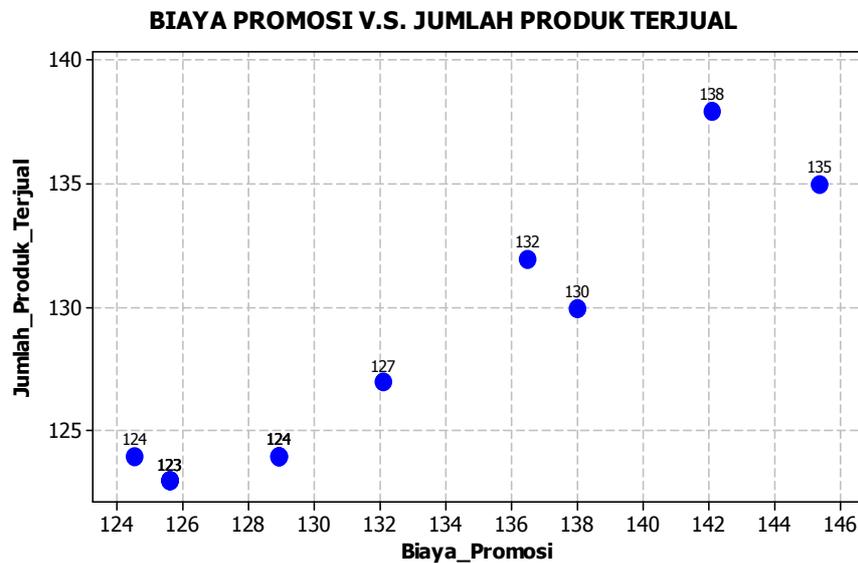
Contoh *Scatter Plot* untuk Data Berpasangan :

Seorang manajer pemasaran ingin melihat hubungan antara Biaya Promosi dan Jumlah Produk yang Terjual. Data pengamatan dicatat berdasarkan historis biaya promosi yang dikeluarkan bersamaan dengan jumlah produk yang terjual, sebagai berikut:

Tabel 6. Tabel Pengamatan untuk Menyajikan *Scatter Plot*

No.	Biaya Promosi	Jumlah Produk Terjual
1	125.6	123
2	128.9	124
3	132.1	127
4	145.4	135
5	124.5	124
6	136.5	132
7	138.0	130
8	142.1	138
9	125.6	123
10	128.9	124
11	125.6	123
12	128.9	124

Berikut diberikan hasil *Scatter Plot* untuk melihat hubungan antara Biaya Promosi dan Jumlah Produk yang Terjual:



Diperoleh hasil *Scatter Plot* bahwa semakin besar Biaya Promosi, maka semakin besar Jumlah Produk yang Terjual. Hal ini menunjukkan bahwa adanya pola hubungan antara Biaya Promosi dan Jumlah Produk Terjual yang bersifat positif (searah), sehingga terlihat output bahwa pengeluaran biaya promosi dapat membantu meningkatkan jumlah produk yang terjual.

Penyajian Diagram Garis.

Diagram Garis (*Line Chart*) digunakan untuk melihat trend/pergerakan suatu variabel data tertentu. *Line Chart* juga merupakan salah satu diagram yang digunakan untuk **data berpasangan**. Ketika Diagram Garis digunakan untuk menggambarkan suatu respon pengamatan terhadap suatu periode waktu, Diagram garis dapat disebut juga sebagai diagram runtun waktu (*Time Series Chart*)^[2] atau *Time Series Plot*.

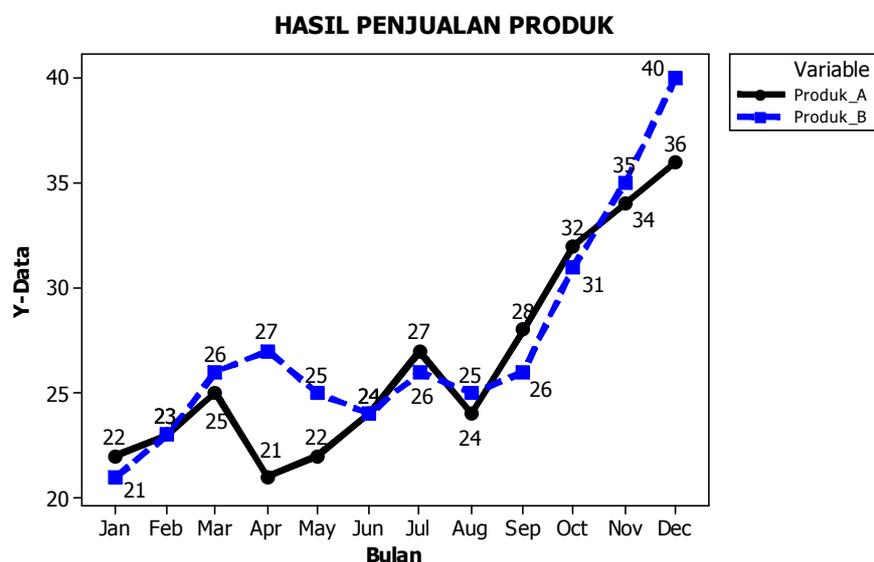
Contoh Diagram Garis untuk Data Berpasangan :

Seorang peneliti pemasaran ingin membandingkan hasil penjualan produk A dan produk B dalam satu tahun. Berikut diberikan data pergerakan historis dari produk A dan Produk B selama satu tahun:

Tabel 7. Data Pengamatan untuk Menyajikan Diagram Garis

Bulan	Produk A	Produk B
Januari	22	21
Februari	23	23
Maret	25	26
April	21	27
Mei	22	25
Juni	24	24
Juli	27	26
Agustus	24	25
September	28	26
Oktober	32	31
November	34	35
Desember	36	40

Berikut hasil output Diagram Garis untuk perbandingan hasil penjualan produk A dan B.

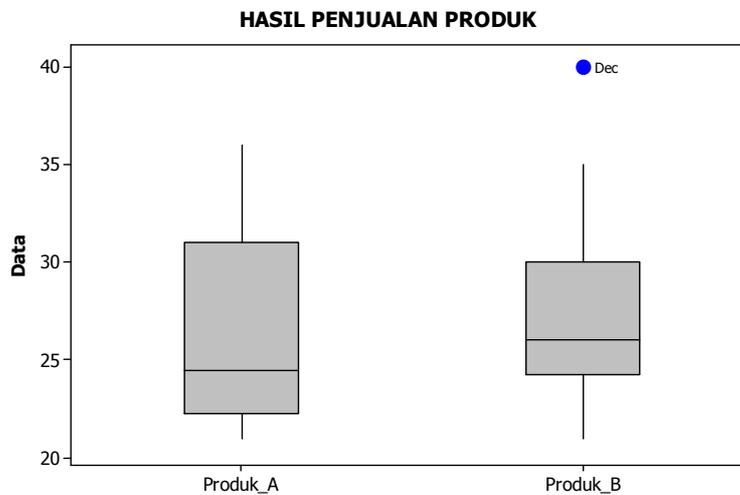


Hasil penjualan produk A dan produk B saling bersaing. Diawal bulan, penjualan produk A lebih baik dari produk B, namun diakhir bulan penjualan produk A tidak lebih baik dari produk B. Hal ini disebabkan penjualan produk B yang meningkat di bulan April yang bersamaan dengan penurunan penjualan produk A, selanjutnya diikuti dengan penurunan penjualan Produk A kembali di bulan Agustus.

Penyajian Box Plot.

Box Plot digunakan untuk melihat perbandingan distribusi/variasi antara kelompok variabel data dan mengidentifikasi terjadinya **Outliers** data. *Outliers* adalah suatu individu data yang nilainya menyimpang jauh dari penyebaran data pengamatan. Contoh Data yang digunakan

adalah data pengamatan seperti pada **Tabel 7**. Berikut hasil output *Box Plot* untuk studi kasus tersebut.



Hasil penjualan produk B lebih besar daripada produk A, serta produk B memiliki nilai varian yang lebih kecil daripada produk A dalam konsistensi penjualan. Selanjutnya terdapat lonjakan prestasi atau penyimpangan yang menguntungkan (*outlier*) dalam penjualan produk B dibulan Desember. Hal ini yang menjadi penjualan tertinggi produk B dalam bersaing dengan produk A.

REFERENSI

- [1] Bluman, A.G., (2012), *Elementary Statistics: A Step By Step Approach, Eighth Edition*, New York: McGraw-Hill.
- [2] Larson, R. dan Farber, B., (2012), *Elementary Statistics: Picturing The World, Fifth Edition*, Boston: Pearson Education.
- [3] Montgomery, D.C., (2009), *Introduction to Statistical Quality Control, Sixth Edition*, New Jersey: John Wiley & Sons.
- [4] Triola, M.F., (2012), *Elementary Statistics: Technology Update, 11th Edition*, Boston: Addison-Wesley.