

Topik 7

Eksplorasi dan Visualisasi Data

Metode Grafis

Ukuran Pemusatan Data

Ukuran Variabilitas

Box Plot

Mendeskripsikan Data Lebih dari satu Variabel

Referensi

- Ott, Lyman. (2001). An introduction to statistical methods and data analysis. 5th ed. Duxbury Thomson Learning.
- Hadjar, Ibnu. (2019). Statistika. untuk ilmu Pendidikan, sosial, dan humaniora. Bandung: PT Remaja Rosdakarya.

Visualisasi data adalah representasi grafis dari data dan informasi. Dengan menggunakan elemen-elemen visual seperti diagram, grafik dan peta, komunikasi data dan informasi menjadi lebih mudah. Selain untuk menyampaikan ide, visualisasi data juga mempermudah untuk menarik kesimpulan dengan melihat tren atau pola yang ada. Mengetahui keberadaan pencilon dan berusaha memahaminya juga dimungkinkan dengan adanya visualisasi data. Selain itu, visualisasi data juga sangat membantu dalam menyampaikan informasi lintas disiplin dan keahlian.

Ketika infografis semakin mudah dipahami bukan terbatas pada isi dan manfaat, visualisasi data menjadi semakin menarik karena memberikan desain dan tampilan yang membuat orang lebih tertarik untuk membaca karena tampilan yang tidak membosankan.

Tujuan pokok visualisasi data adalah untuk memberikan atau mengkomunikasikan data secara jelas, mudah, dan efektif melalui sarana grafis. Visualisasi yang efektif membantu pengguna dalam menganalisa dan penalaran tentang data dan bukti. Ia membuat data yang kompleks bisa diakses, dipahami dan berguna. Pengguna bisa melakukan pekerjaan analisis tertentu, seperti melakukan perbandingan atau memahami kausalitas, dan prinsip perancangan dari grafik contohnya, memperlihatkan perbandingan atau kausalitas mengikuti pekerjaan tersebut. Tabel pada umumnya digunakan saat pengguna akan melihat ukuran tertentu dari sebuah variabel, sementara grafik dari berbagai tipe digunakan untuk melihat pola atau keterkaitan dalam data untuk satu atau lebih variabel.

7.1 Metode Grafis

Saat ini terdapat beberapa teknik membuat visualisasi data, teknik-teknik tersebut diantaranya adalah:

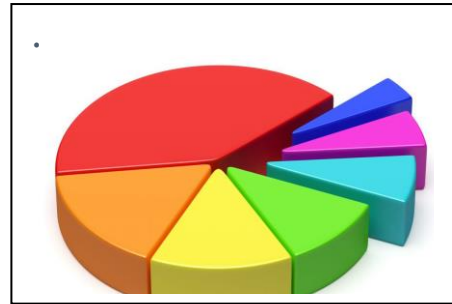
- plotting data mentah (pie chart, histogram, bar chart,)
- plotting statistik sederhana seperti (boxplot)

A. Pie Chart

Diagram lingkaran digunakan secara luas di berbagai bidang untuk mempresentasikan proporsi dari setiap klasifikasi yang berbeda dan untuk membandingkan berbagai klasifikasi berdasarkan busurnya.

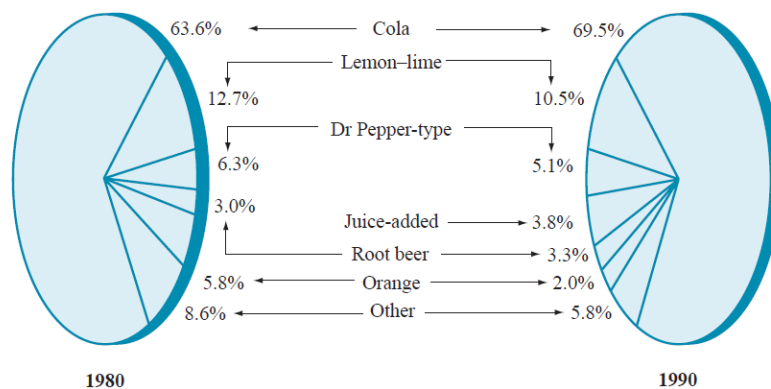
Jika sebuah grafik garis tidak cocok dipakai untuk data yang memiliki banyak klasifikasi, karena dengan

banyaknya klasifikasi, perbedaan ukuran pie tiap klasifikasi menjadi tidak jelas. Pie chart lebih baik digunakan untuk kategori dengan jumlah maksimal 5 atau 6 kategori. Karena terlalu banyak kategori akan menyulitkan interpretasi. Jika memungkinkan, persentase dalam pie chart dibuat terurut, menaik atau menurun.



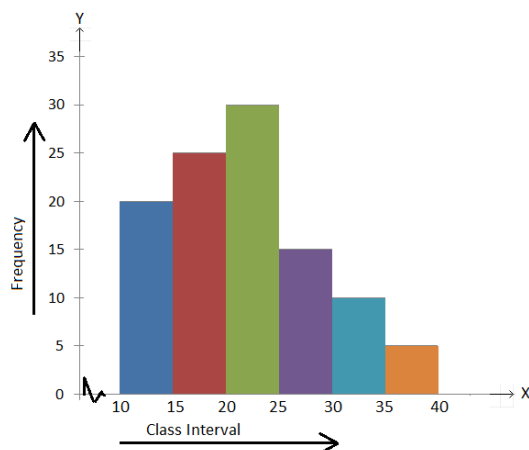
Contoh

Pie chart berikut menampilkan perbandingan persentase konsumsi berbagai jenis minuman tahun 1980 dan 1990



Terjadi peningkatan jumlah konsumsi untuk beberapa jenis minuman, diantaranya jenis Cola dan Rootbeer, selain penurunan beberapa minuman yang lain.

B. Histogram



Histogram adalah sajian tabel distribusi frekuensi dengan menggunakan gambar berbentuk persegi panjang yang saling berhimpit.

Langkah-langkah membuat histogram:

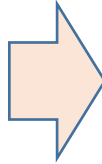
1. Membuat sumbu datar dan sumbu tegak yang saling berpotongan. Sumbu datar menyatakan kelas interval dan sumbu tegak menyatakan frekuensi
2. Buat skala menggunakan batas-batas kelas interval (dalam distribusi frekuensi)
3. Bentuk batang sesuai dengan batas kelas interval dan frekuensi. Sisi-sisi dari batang kelas interval harus berdekatan dan berhimpit

Histogram mengelompokkan angka ke dalam rentang. Ketinggian setiap batang menunjukkan jumlah (frekuensi) dari setiap rentang.

Contoh:

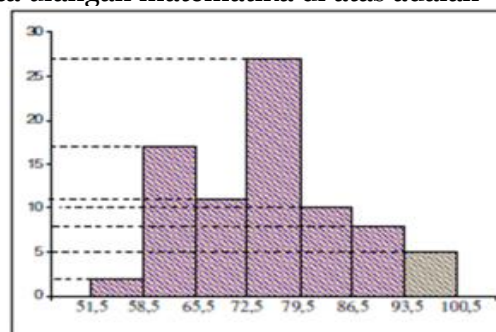
Data nilai ulangan matematika dari 80 siswa SMP di Bandung yang telah disajikan dalam tabel distribusi frekuensi.

75 84 68 82 68 90 62 88 93 76
88 79 73 73 61 62 71 59 75 85
75 65 62 87 74 93 95 78 72 63
82 78 66 75 94 77 63 74 60 68
89 78 96 97 78 85 60 74 65 71
67 62 79 97 78 85 76 65 65 71
73 80 65 57 88 78 62 76 74 53
73 67 86 81 72 65 76 75 77 85



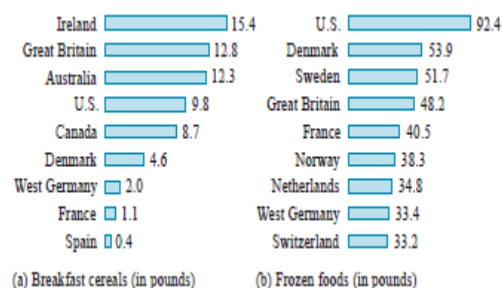
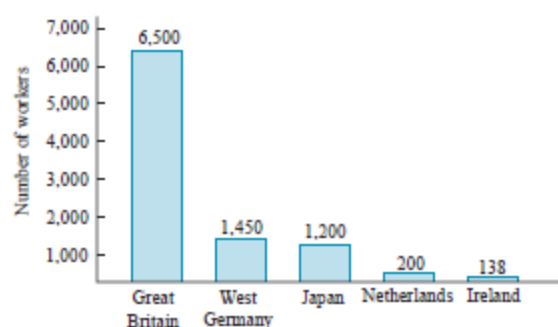
Kelas	Turus	Frekuensi	Batas bawah	Batas atas	Tepi bawah kelas	Tepi atas kelas
52 – 58	II	2	52	58	51,5	58,5
59 – 65	IIIIII	17	59	65	58,5	65,5
66 – 72	IIII	11	66	72	65,5	72,5
73 – 79	IIIIIIIIII	27	73	79	72,5	79,5
80 – 86	IIII	10	80	86	79,5	86,5
87 – 93	IIII	8	87	93	86,5	93,5
94 – 100	II	5	94	100	93,5	100,5
Jumlah		80				

Maka histogram untuk data ulangan matematika di atas adalah



C. Bar Chart

Bar Chart (disebut juga Grafik Batang) adalah tampilan grafik data menggunakan batang dengan ketinggian menunjukkan frekuensi. Bar chart merupakan grafik untuk data diskrit dan nominal (tidak bisa untuk data kontinu). Bar chart digunakan jika kita ingin membandingkan data antar kategori dan melihat frekuensi atau besaran dari kategori/variabel tersebut. Bar chart dapat disajikan dalam bentuk vertikal maupun horizontal.

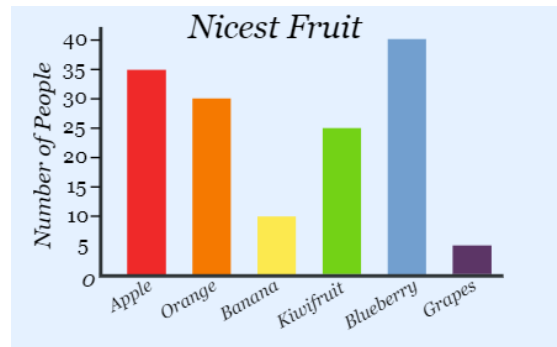


Contoh :

Sebuah survey dilakukan terhadap 145 orang dengan pertanyaan : Buah apakah yang paling disukai?

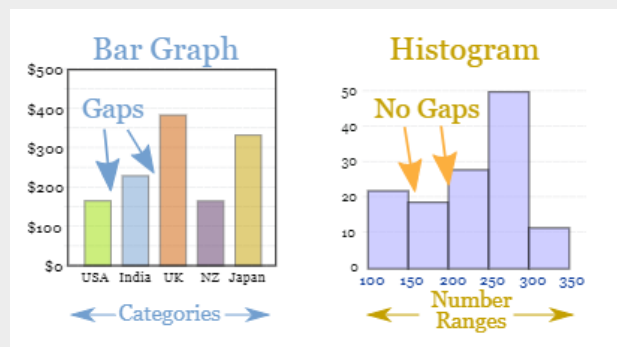
Buah:	Apple	Orange	Banana	Kiwifruit	Blueberry	Grapes
Jumlah	35	30	10	25	40	5

Dan berikut adalah grafiknya



Pie Chart vs Bar Chart vs Histogram

- ✓ Pie chart dan bar chart digunakan untuk menampilkan frekuensi data dari **variable kualitatif**.
- ✓ Histogram digunakan untuk menampilkan frekuensi data dari **variable kuantitatif**.

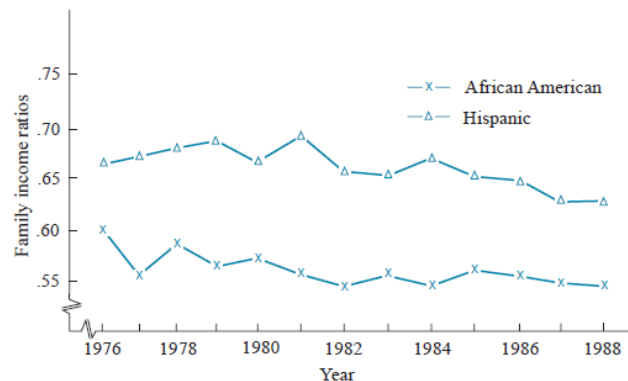


D. Grafik Garis

Line chart atau grafik garis merupakan chart yang pada umumnya mengilustrasikan trend atau menggambarkan hubungan bagaimana data berubah dalam periode waktu tertentu (*Time series*). Digunakan untuk menampilkan perubahan suatu variabel dalam jangka waktu tertentu. Kita dapat melihat pola dan fluktuasi dalam data, membandingkan variabel, dan membuat proyeksi dari data kita. Tipe data yang dapat digunakan dalam line chart umumnya adalah kontinu.

Contoh

Rasio median pendapatan pada keluarga Afrika Amerika dan Hispanik di Anglo Amerika tahun 1976- 1988



7.2 Ukuran Pemusatan (*central tendency*).

Sekumpulan data biasanya digambarkan dengan suatu nilai pusat (*central tendency*). Nilai tersebut disebut dengan Ukuran Pemusatan Data. Nilai ukuran pemusatan data merupakan nilai yang mewakili suatu kumpulan data sehingga nilai tersebut harus memiliki sifat-sifat berikut. Merupakan penyederhanaan data untuk mempermudah peneliti membuat interpretasi dan mengambil suatu kesimpulan.

Terdapat tiga ukuran tendensi sentral yang sering digunakan, yaitu:

- ❖ Mean (Rata-rata hitung/rata-rata aritmetika)
- ❖ Median
- ❖ Mode

A. Modus (Mode)

Modus adalah nilai yang paling sering muncul dari sekumpulan data. Modus tidak dipengaruhi oleh nilai ekstrem atau pencilan (*outlier*). Biasanya modus hanya digunakan untuk tujuan deskriptif karena nilai modus tidak mempertimbangkan distribusi data. Jika nilai-nilai pengamatan sangat bervariasi dari nilai pusatnya, maka modus kurang cocok digunakan sebagai ukuran pemusatan data.

Karakteristik Mode

- ✓ Dalam satu data set bisa terdapat lebih dari satu mode
- ✓ Tidak terpengaruh oleh hasil pengukuran yang bernilai ekstrem (*outlier*)
- ✓ Mode dari himpunan-himpunan bagian tidak dapat dikombinasikan untuk menentukan mode dari keseluruhan data set
- ✓ Untuk kelompok data, nilainya dapat berubah bergantung kepada kategorisasi yang digunakan
- ✓ Dapat diaplikasikan untuk data kualitatif dan kuantitatif

Contoh:

Hitunglah nilai modus dari nilai ujian mahasiswa jurusan statistika berikut ini:

2; 4; 5; 6; 6; 7; 7; 7; 8; 9

Jawab:

Dari soal tersebut dapat diketahui

- Data 2, muncul 1 kali
- Data 4, muncul 1 kali
- Data 5, muncul 1 kali
- Data 6, muncul 2 kali
- Data 7, muncul 3 kali
- Data 8, muncul 1 kali
- Data 9, muncul 1 kali

Data yang paling banyak muncul adalah 7, yaitu sebanyak 3 kali. Dengan demikian modus dari sekumpulan data tersebut adalah 7.

B. MEAN /Nilai tengah (rata-rata)

Rata-rata adalah nilai yang mewakili himpunan atau sekumpulan data (*a set of data*). Rata-rata sangat baik digunakan apabila sebaran data merata atau nilai antara data yang satu dengan data yang lain tidak berbeda jauh (homogen).

Rata-rata hitung digunakan apabila:

- Jenis datanya adalah numerik (interval atau rasio). Jika datanya adalah kategorik ordinal maka sebaiknya menggunakan median dan jika datanya adalah numerik nominal, maka sebaiknya menggunakan modus.
- Nilainya terpengaruh oleh hasil pengukuran yang bernilai ekstrim, untuk itu pemotongan data dilakukan untuk menekan tingkat pengaruh nilai ekstrimnya.
- Mean dari himpunan-himpunan bagian dapat dikombinasikan untuk menentukan nilai mean dari keseluruhan data set
- Hanya dapat diaplikasikan untuk data kuantitatif.

$$\bar{x} = \frac{(x_1+x_2+x_3+\dots+x_n)}{n} = \frac{\sum x}{n}$$

Contoh

Hitunglah nilai rata-rata dari nilai ujian mahasiswa jurusan statistika berikut ini: 2; 4; 5; 6; 6; 7; 7; 7; 8; 9

Jawab:

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

$$= \frac{1}{10}(2 + 4 + 5 + 6 + 6 + 7 + 7 + 7 + 8 + 9) = 6.1$$

C. Median

Median adalah nilai yang terletak di bagian tengah dari suatu kumpulan data yang telah diurutkan. Nilai median dipengaruhi oleh banyaknya pengamatan dan tidak tergantung pada besarnya nilai pengamatan walaupun nilainya tersebut sangat ekstrem. Dengan demikian, median sangat cocok digunakan untuk mewakili data yang distribusinya tidak homogen.

Karakteristik Median :

- ✓ Merupakan nilai tengah 50% data ada di atasnya, dan 50% data ada dibawahnya
- ✓ Hanya terdapat 1 median dalam sebuah data set
- ✓ Tidak terpengaruh oleh hasil pengukuran yang bernilai ekstrim (outlier)
- ✓ Hanya berlaku untuk data kuantitatif (skala data numerik ordinal)

Contoh:

(menghitung median yang banyaknya data (n) ganjil)

Hitunglah median dari nilai ujian mahasiswa jurusan statistika berikut ini: 8; 4; 5; 6; 7; 6; 7; 7; 2; 9; 10

Jawab:

Langkah-langkah penyelesaiannya adalah sebagai berikut.

- Diketahui datanya adalah 8; 4; 5; 6; 7; 6; 7; 7; 2; 9; 10.
- Data tersebut diurutkan dari nilainya yang terkecil ke yang terbesar. Hasil setelah pengurutan adalah 2; 4; 5; 6; 6; 7; 7; 7; 8; 9; 10
- Banyaknya data (n) = 11.
- Posisi median ditentukan menggunakan rumus

$$\text{Posisi Median data ke } \frac{1}{2}(n + 1) = \frac{1}{2}(11 + 1) = \text{data ke 6}$$

Artinya median adalah data yang berada pada urutan keenam (x_6).

Dengan demikian nilai median data adalah 7.

Contoh:

(menghitung median yang banyaknya data (n) genap)

Hitunglah median dari nilai ujian mahasiswa jurusan statistika berikut ini: 8; 4; 5; 6; 7; 6; 7; 7; 2; 9

Jawab:

Langkah-langkah penyelesaiannya adalah sebagai berikut.

- Diketahui datanya adalah: 8; 4; 5; 6; 7; 6; 7; 7; 2; 9.
- Data tersebut diurutkan dari nilainya yang terkecil ke yang terbesar. Hasil setelah pengurutan adalah: 2; 4; 5; 6; 6; 7; 7; 7; 8; 9
- Banyaknya data (nn) adalah 10.

- Posisi median ditentukan menggunakan rumus

$$\text{Posisi Median} = \text{data ke } \frac{1}{2}(n + 1) = \frac{1}{2}(10 + 1) = \text{data ke } 5.5$$

Artinya nilai median data adalah data yang berada diantara data urutan kelima (x_5) dan keenam (x_6). Dengan demikian nilai median data adalah

$$\text{Median} = \frac{1}{2}(6 + 7) = 6.5$$

7.3 Ukuran Penyebaran (Variabilitas)

Ukuran penyebaran/variabilitas adalah data yang menggambarkan seberapa besar penyimpangan nilai-nilai data dengan nilai pusatnya. Ukuran keragaman menggambarkan bagaimana berpencarnya data atau menggambarkan seberapa jauh data menyebar dari rata-ratanya.

Ukuran tendensi sentral (*mean, median, mode*) merupakan nilai perwakilan dari suatu sampel, tetapi ukuran tersebut tidak memberikan gambaran informasi yang lengkap mengenai bagaimana penyebaran data pengamatan terhadap nilai sentralnya. Ukuran tendensi sentral saja tidak cukup untuk menggambarkan distribusi frekuensi. Selain itu kita harus memiliki ukuran persebaran data pengamatan.

Ukuran pemusatandata hanya menggambarkan pemusatan penyebaran skor pada suatu sampel, tidak memberikan informasi tentang kecenderungan perbedaan skor antar subjek dalam sampel.

Ukuran variabilitas memberikan informasi skor perbedaan antar subjek dalam sampel

Ukuran keragaman data yang sering digunakan antara lain: jangkauan/rentang (range), ragam (variansi), dan simpangan baku (standar deviasi).

A. Rentang/Range /Jangkauan

Ukuran penyebaran yang paling sederhana adalah **Range** (*Jangkauan/Rentang*, terkadang di beberapa literatur diterjemahkan dengan istilah **wilayah**). Range/Rentang adalah salah satu ukuran statistik yang menunjukkan jarak penyebaran antara skor (nilai) yang terendah (Lowest score) sampai skor nilai yang tertinggi (Highest Score). Jangkauan ini adalah selisih antara nilai maksimum dan nilai minimum yang terdapat dalam data. Range hanya memperhitungkan dua nilai, yaitu nilai maksimum dan nilai minimum dan tidak memperhitungkan semua nilai, sehingga sangat tidak stabil atau tidak dapat diandalkan sebagai indikator dari ukuran penyebaran. Hal ini terjadi karena range sangat dipengaruhi oleh nilai-nilai ekstrim.

$$\text{Range} = \text{nilai maksimum} - \text{nilai minimum}$$

Sebagai contoh, misal distribusi hasil panen dua varietas padi (kg per plot), masing-masing terdiri dari 5 plot. Andaikan distribusi datanya sebagai berikut:

Varietas I : 45 42 42 41 40

Varietas II : 54 48 42 36 30

Varietas III : 45 40 44 41 40

Misalnya, range untuk Varietas I pada tabel di atas adalah $45 - 40 = 5$ (45 adalah nilai maksimum dan 40 adalah nilai minimum). Seringkali kita mengatakan range dengan pernyataan seperti "hasil berkisar antara 40 - 45 kg per petak". Kisarannya lebih sempit dibandingkan dengan pernyataan "hasil berkisar antara 40 - 60 kg per petak". Pernyataan pertama menggambarkan bahwa variasi hasil padi tidak terlalu beragam, sedangkan pada pernyataan kedua, terjadi hal sebaliknya.

B. Varians dan standar deviasi

Varian dan standar deviasi (simpangan baku) adalah ukuran-ukuran keragaman (variasi) data statistik yang paling sering digunakan. Standar deviasi (simpangan baku) merupakan akar kuadrat dari varian.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad \text{atau} \quad s^2 = \frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n - 1)}$$

Nilai varian (s^2) yang dihasilkan merupakan nilai yang berbentuk kuadrat sehingga standar deviasi (simpangan baku) adalah ;

$$s = \sqrt{\frac{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}{n(n - 1)}}$$

Keterangan:

s^2 = varian

s = standar deviasi (simpangan baku)

x_i = nilai x ke- i

\bar{x} = rata-rata

n = ukuran sampel

Contoh

Misalkan dalam suatu kelas, tinggi badan beberapa orang siswa yang dijadikan sampel adalah sebagai berikut.

172, 167, 180, 170, 169, 160, 175, 165, 173, 170

Dari data tersebut diketahui bahwa jumlah data (n) = 10, dan ($n - 1$) = 9. Selanjutnya dapat dihitung komponen untuk rumus varian.

i	x_i	x_i^2
1	172	29584
2	167	27889
3	180	32400
4	170	28900
5	169	28561
6	160	25600
7	175	30625
8	165	27225
9	173	29929
10	170	28900
Σ	1701	289613

Dari tabel tersebut dapat diketahui:

$$\sum_{i=1}^n x_i = 1701 \qquad \sum_{i=1}^n x_i^2 = 289613 \quad \left(\sum_{i=1}^n x_i \right)^2 = 1701^2 = 2893401$$

jika dimasukkan ke dalam rumus varian, maka hasilnya adalah sebagai berikut.

$$\begin{aligned}
 s^2 &= \frac{(10) \cdot (289613) - (2893401)}{(10) \cdot (9)} \\
 &= \frac{2729}{90} \\
 &= 30,32
 \end{aligned}$$

Dari nilai tersebut bisa langsung diperoleh nilai standar deviasi (simpangan baku) dengan cara mengakarkuadratkan nilai varian.

$$s = \sqrt{30,32} = 5,51$$

Atau dengan menggunakan rata-rata

i	Data ke- i x_i	Rata-rata (\bar{x})	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
(1)	(2)	(3)	(4)	(5)
1	172	170.1	1.9	3.61
2	167	170.1	-3.1	9.61
3	180	170.1	9.9	98.01
4	170	170.1	-0.1	0.01
5	169	170.1	-1.1	1.21
6	160	170.1	-10.1	102.01
7	175	170.1	4.9	24.01
8	165	170.1	-5.1	26.01
9	173	170.1	2.9	8.41
10	170	170.1	-0.1	0.01
				$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 272.9$

$$\sum_{i=1}^{10} (x_i - \bar{x})^2 = 272,9$$

Sehingga

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} = \frac{272,9}{10 - 1} = 30,32$$

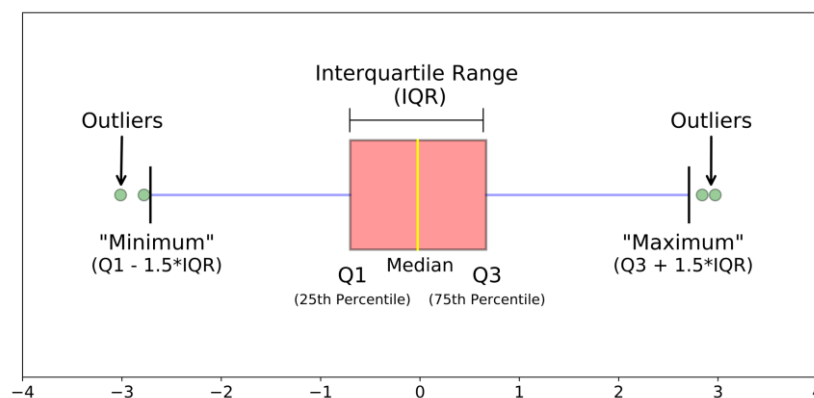
Sedangkan standar deviasi adalah akar kuadrat dari varian tersebut.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{272,9}{10 - 1}} = \sqrt{30,32} = 5,51$$

7.4 Boxplot

Box plot atau boxplot (juga dikenal sebagai diagram box-and-whisker) merupakan suatu box (kotak berbentuk bujur sangkar). Boxplot berfungsi untuk menyajikan informasi tentang ukuran pemusatan, distribusi, dan deteksi outlier. Boxplot adalah salah satu cara dalam statistik deskriptif untuk menggambarkan secara grafik dari data numeris melalui lima ukuran sebagai berikut:

- ✓ nilai **minimum**: nilai observasi terkecil
- ✓ **Q1**: kuartil terendah atau kuartil pertama
- ✓ **Q2: median** atau nilai pertengahan
- ✓ **Q3**: kuartil tertinggi atau kuartil ketiga
- ✓ nilai **maksimum**: nilai observasi terbesar.



- Garis bagian kiri box menyajikan kuartil pertama (Q1), sementara bagian kanan menyajikan kuartil ketiga (Q3). Bagian dari box adalah bidang yang menyajikan interquartile range (IQR), atau bagian pertengahan dari 50 % observasi. Panjang box ditentukan oleh IQR ini. IQR adalah ukuran yang terkenal untuk mengukur penyebaran data. Semakin lebar bidang IQR (box), menunjukkan data semakin menyebar.
- Garis tengah yang membagi dua box menyajikan median dari data. Median adalah ukuran /nilai pusat.
- Garis yang memperpanjang box dinamakan dengan whiskers. Whiskers menunjukkan nilai yang lebih rendah dan lebih tinggi dari kumpulan data yang berada dalam IQR (kecuali outlier). Panjang garis Whisker bagian atas ini adalah kurang dari atau sama dengan $Q3 + (1.5 \times IQR)$. Panjang garis Whisker bagian bawah ini adalah lebih besar atau sama dengan $Q1 - (1.5 \times IQR)$. Masing-masing garis whisker dimulai dari akhir box.
- Nilai yang berada di atas atau dibawah whisker dinamakan nilai outlier atau ekstrim.

7. Mendeskripsikan Data Lebih dari satu Variabel

A. Tabel Kontingensi

Tabel kontingensi merupakan **tabel** yang digunakan untuk mengukur hubungan (asosiasi) antara dua variabel kategorik dimana **tabel** tersebut merangkum frekuensi bersama dari observasi pada setiap kategori variabel. Tabel kontingensi pada dasarnya adalah sebuah format tampilan yang digunakan untuk menganalisis dan merekam hubungan antara dua atau lebih dari variabel kategori dimana tabel ini memiliki baris dan kolom berisi frekuensi bersama dari observasi untuk setiap kategori variabel.

Contoh

Tabel berikut memperlihatkan preferensi jaringan TV pada masyarakat perkotaan, pedesaan dan pinggiran kota. Survey dilakukan pada 1500 orang.

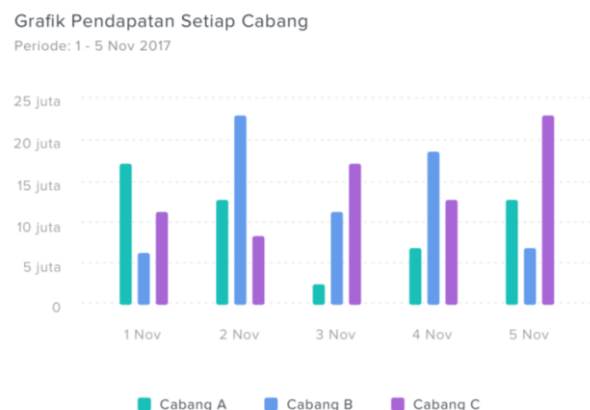
Network Preference	Residence			Total
	Urban	Suburban	Rural	
ABC	144	180	90	414
CBS	135	240	96	471
NBC	108	225	54	387
Other	63	105	60	228
Total	450	750	300	1,500

B. Clustered Column Chart

Digunakan untuk membandingkan berbagai data sejenis dalam satu item.

Contoh:

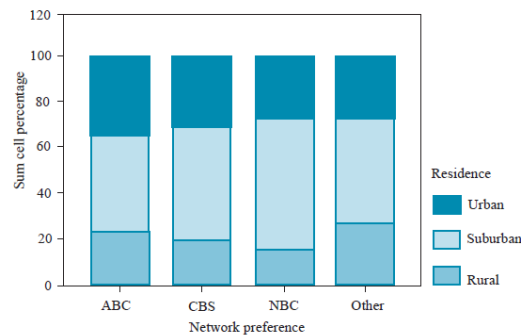
Seorang direktur pemasaran butuh membandingkan data pendapatan setiap cabang untuk periode 1 November hingga 5 November.



C. Stacked Column Chart

Mirip dengan Clustered Column Chart, tapi di Stacked column chart ini semua datanya disajikan dalam satu kolom. Sehingga kita bisa melihat jumlah totalnya, tetapi tetap dapat melihat komposisi perbandingan dalam satu kolom.

Merupakan ekstensi dari bar chart yang digunakan untuk menampilkan data dari sepasang variable kualitatif



Perbandingan distribusi dari jaringan Televisi dan domisili

Contoh:

Seorang direktur pemasaran membutuhkan data pendapatan untuk periode 1 November hingga 5 November. Dan Ia juga butuh melihat komposisi pendapatan setiap cabangnya.



D.Scatter Diagram

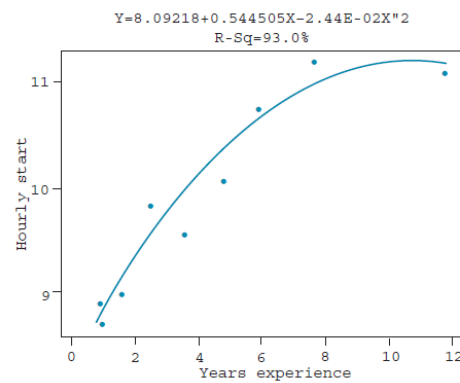
Scatter Diagram atau Diagram Tebar digunakan untuk menggambarkan relasi antara dua variable kuantitatif. Diagram ini dapat berfungsi untuk melakukan pengujian terhadap seberapa kuatnya hubungan antara 2 (dua) variabel serta menentukan jenis hubungan dari 2 (dua) variabel tersebut apakah hubungan Positif, hubungan Negatif ataupun tidak ada hubungan sama sekali. Bentuk dari Scatter Diagram adalah gambaran grafis yang terdiri dari sekumpulan titik-titik (point) dari nilai sepasang variabel (Variabel X dan Variabel Y).

Scatter Diagram sering disebut juga dengan Scatter Chart, Scatter plot, Scattergram dan Scatter graph.

Contoh

Scatter plot dari upah per jam dan pengalaman (tahun)

y (dollars)	8.90	8.70	9.10	9.00	9.79	9.45	10.00	10.65	11.10	11.05
x (years)	1.25	1.50	2.00	2.00	2.75	4.00	5.00	6.00	8.00	12.00



Grafik scatter plot memperlihatkan bahwa semakin lama pengalaman pekerja maka upah perjamnya akan semakin meningkat. Artinya terdapat hubungan positif antara upah per jam dengan pengalaman.