



INTRODUCTION TO STATISTICS

HISTOGRAM

Histogram adalah tampilan informasi statistik yang menggunakan persegi panjang untuk menunjukkan frekuensi item data dalam interval numerik berurutan dengan ukuran yang sama.

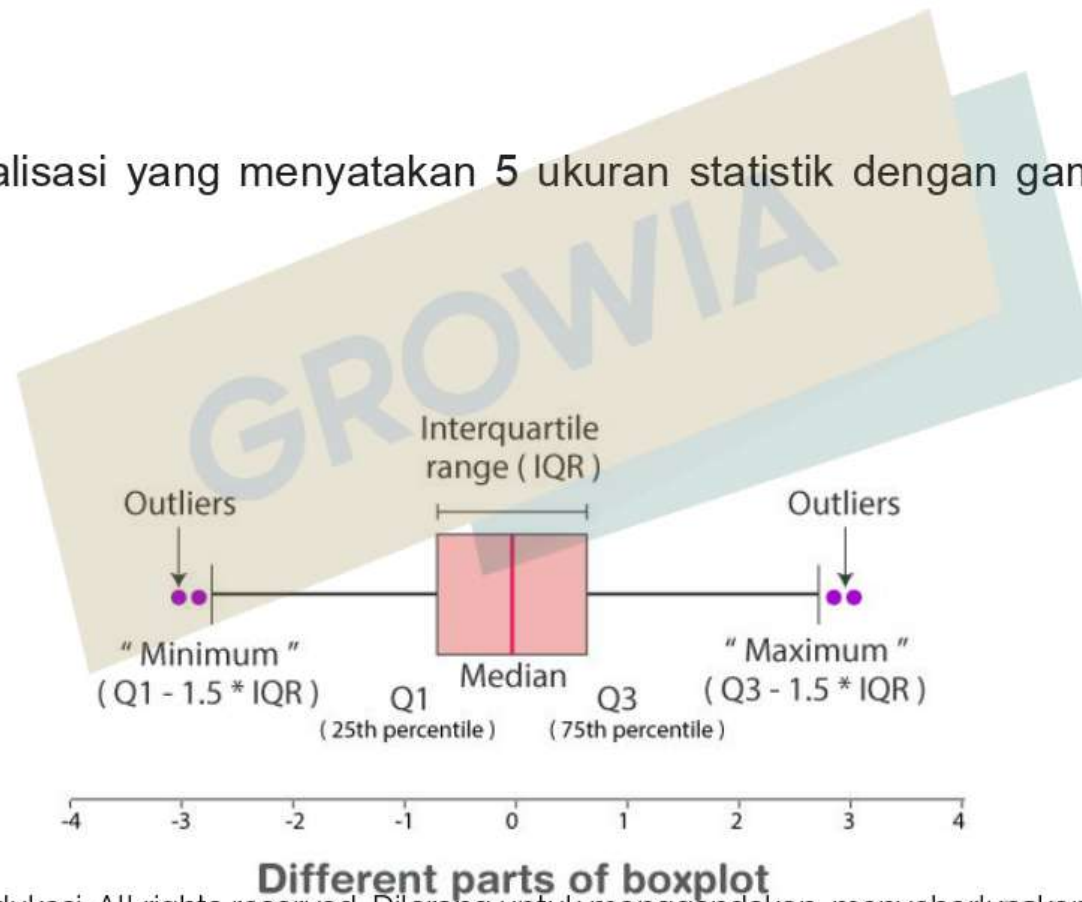


Transcription Visualisasi Statistik

Time	Transcription
00:03–01:10	<p>Tadi kita telah mempelajari hitung-hitungan pada statistika deskriptif. Sekarang kita akan belajar cara menampilkan data dengan visualisasi dimulai dengan histogram. Histogram adalah tampilan informasi statistik yang menunjukkan frekuensi item dalam interval numerik berurutan dengan ukuran yang sama ini adalah contoh dari histogram (perhatikan gambar). Bagaimana cara membacanya? Ada 6 murid yang memiliki nilai ujian di kisaran 66 hingga 72, 18 murid dinilai 73 hingga 79, 39 murid di 80 hingga 86 dan seterusnya. Semakin tinggi batangnya artinya frekuensinya pun semakin tinggi. Setiap batang dari histogram memiliki ukuran atau lebar yang sama yakni 6 angka dan nilai rendah di kiri dan semakin ke kanan maka nilainya semakin tinggi itulah yang dimaksud dengan interval numerik berurutan dengan ukuran yang sama.</p>

BOXPLOT

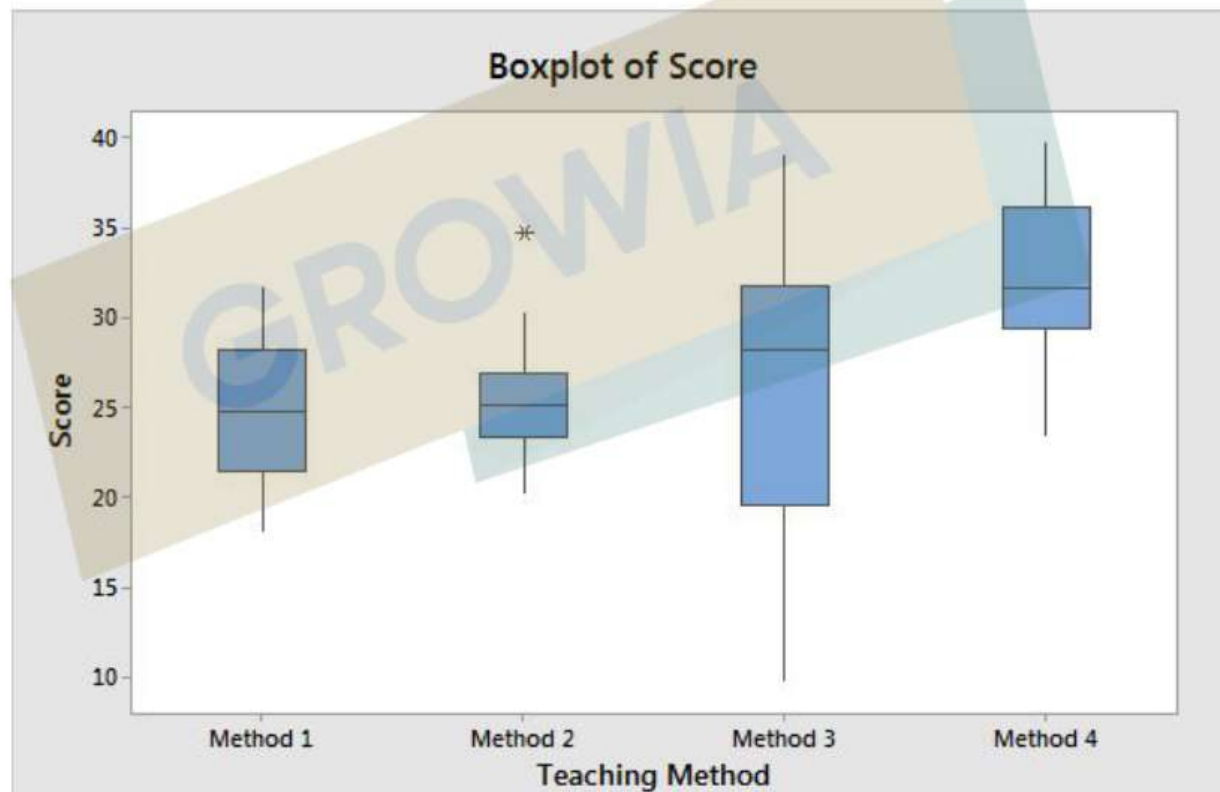
Boxplot adalah visualisasi yang menyatakan 5 ukuran statistik dengan gambar persegi dan garis



Transcription Visualisasi Statistik

Time	Transcription
01:11–02:00	Berikutnya adalah boxplot, boxplot bisa dianggap sebagai versi lebih ringkas dari histogram karena boxplot hanya menampilkan 5 statistik penting dari data ini. Nilai di bawah ini adalah nilai lower whisker merupakan nilai batas bawah yang dianggap wajar yang dihitung dengan formula berikut lokasi dari sisi pertama persegi di sini (lihat gambar) adalah kuartil 1, garis titik tengah adalah median dan akhir dari persegi adalah kuartil 3 dilanjutkan dengan lokasi whisker atas nilai di luar boxplot akan digambarkan sebagai titik yang merupakan indikasi bahwa sampel data merupakan port layer.

INTERPRETASI BOXPLOT



Copyright © PT Growia Online Edukasi, All rights reserved. Dilarang untuk menggandakan, menyebarkan dan memperjualbelikan materi ini tanpa seijin GROWIA. Apabila ada penyalahgunaan maka akan diproses sesuai dengan ketentuan hukum yang berlaku.

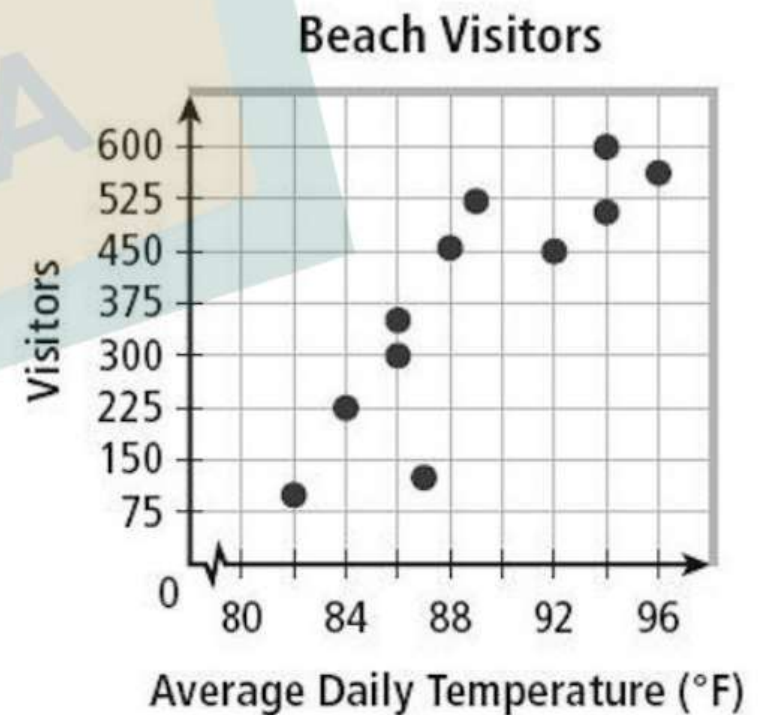
Transcription Visualisasi Statistik

Time	Transcription
02:01–02:46	Berikutnya adalah contoh cara menginterpretasikan boxplot misalkan kita ingin menguji 4 metode pengajaran yang berbeda dan ingin melihat metode mana yang mampu memberikan nilai yang paling tinggi. Di sini (lihat pada gambar) kita bisa melihat bahwa metode 3 adalah metode dengan variabilitas yang paling tinggi. Nilai ujian yang didapat bisa rendah sekali hingga tinggi sekali. Sementara metode 2 memberikan hasil yang lebih konsisten karena kotaknya lebih sempit meskipun konsisten terdapat satu murid yang nilainya tinggi di sekitar 35. Sedangkan metode 4 memberikan hasil yang paling tinggi karena jarak kuartil 1, 2 dan 3 nilainya lebih tinggi daripada 3 metode lainnya.

SCATTER PLOT

Visualisasi yang menunjukkan hubungan antar 2 variabel numerik.

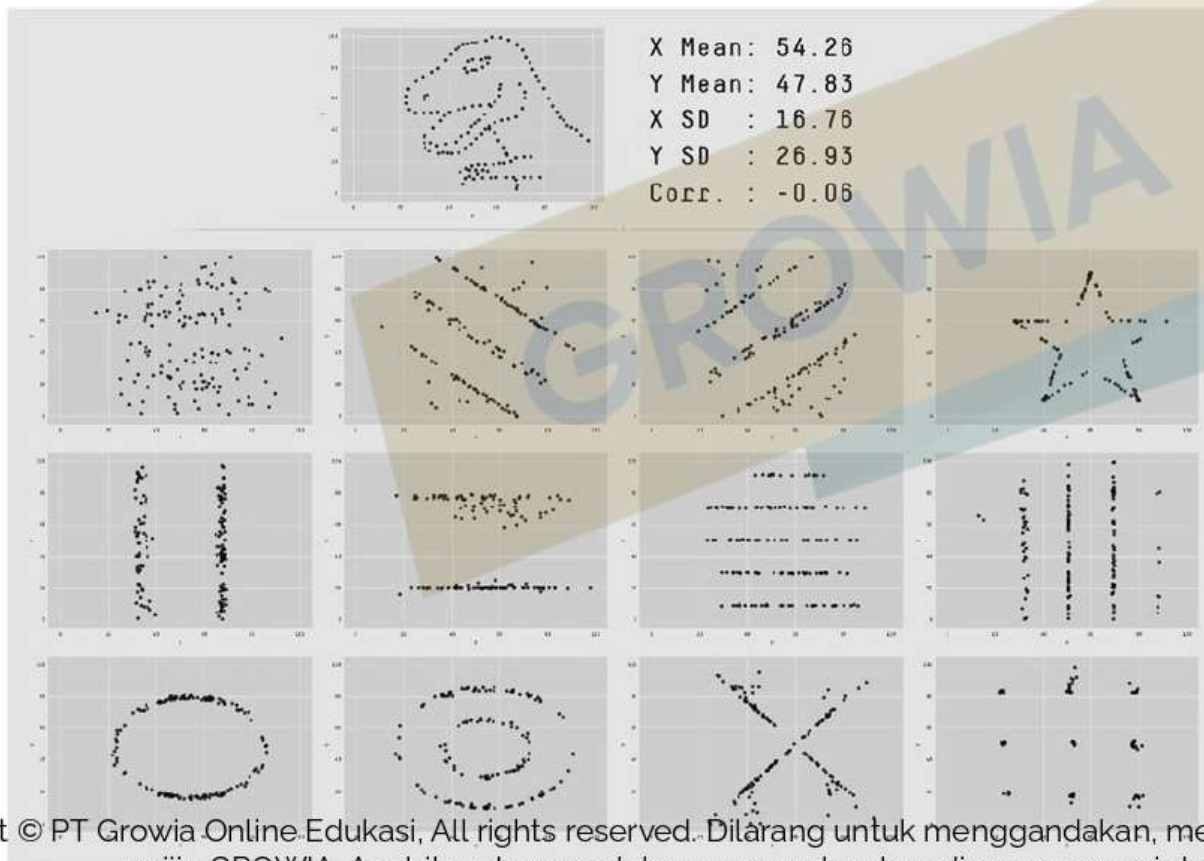
Setiap 1 titik merepresentasikan 1 pengamatan.



Transcription Visualisasi Statistik

Time	Transcription
02:49–03:34	<p>Bentuk visualisasi berikutnya yang bisa digunakan adalah scatter plot. Scatter plot menunjukkan hubungan antara dua variabel numerik, setiap satu titik pada merepresentasikan satu pengamatan. Misalkan pada data ini (lihat pada gambar) per hari ini kita mencatat suhu dan banyaknya pengunjung di pantai, pada hari dengan suhu 84 derajat ada 225 pengunjung, di sini kita melihat adanya hubungan korelasi positif bahwa ketika cuaca semakin panas jumlah pengunjung pun juga semakin meningkat, ini adalah beberapa contoh visualisasi sederhana yang sering digunakan dalam statistika deskriptif. Kita akan mempelajari teknik-teknik visualisasi data yang lebih dalam pada modul yang berbeda</p>

DATASAURUS DOZEN



Melihat angka saja tanpa visualisasi dapat membuat kita mencapai kesimpulan yang tidak akurat.

Transcription Visualisasi Statistik

Time	Transcription
03:34–end	<p>Di sini saya mau menjelaskan bahwa kita harus melakukan visualisasi data bersamaan dengan analisis secara numerik karena jika kita hanya terlalu mengandalkan angka kita bisa kehilangan gambaran mengenai kondisi dari data kita. Di sini ada scatter plot dari 13 data yang bentuknya sangat berbeda akan tetapi ke-13 data di sini memiliki suatu kesamaan yakni rata-rata dari kedua variabel semuanya sama di antara ke-13 scatter plot ini. Standar deviasi semuanya juga sama dan korelasinya antar dua variabel ini semuanya juga sama mendekati 0, data ini terkenal disebut sebagai datasaurus dozen. Jadi sering sekali angka saja tidak cukup untuk mendapatkan gambaran baru dari data kita. Melihat angka saja tanpa visualisasi dapat membuat kita mencapai kesimpulan tidak akurat dari data kita.</p>