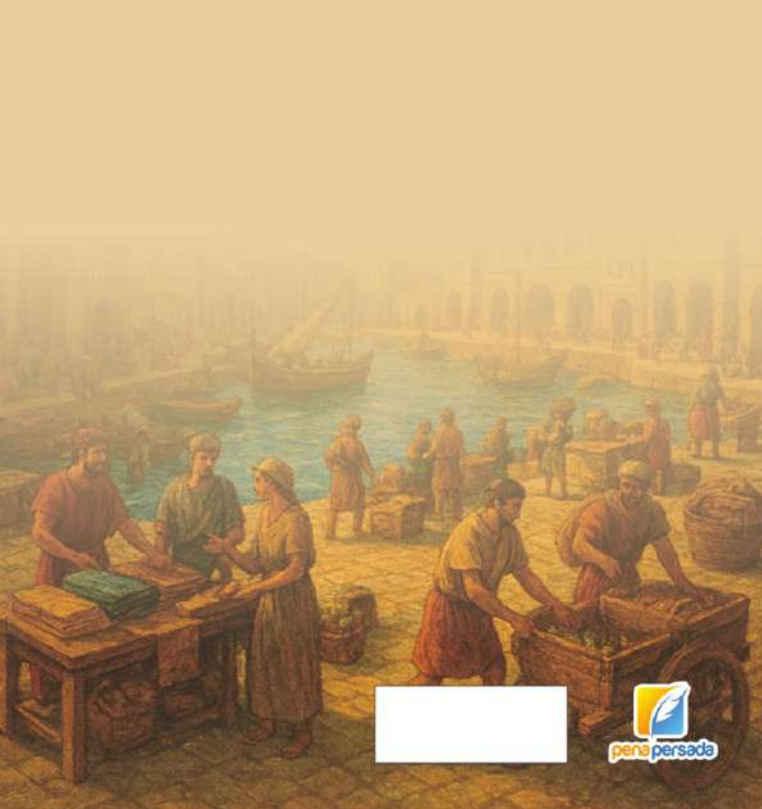


STATISTIK DAN PENGAMBILAN
KEPUTUSAN BISNIS



STATISTIK DAN PENGAMBILAN KEPUTUSAN BISNIS

Dr. Achmad Fauzi, S.E, M.M

**STATISTIK
DAN PENGAMBILAN
KEPUTUSAN BISNIS
DALAM ORGANISASI**

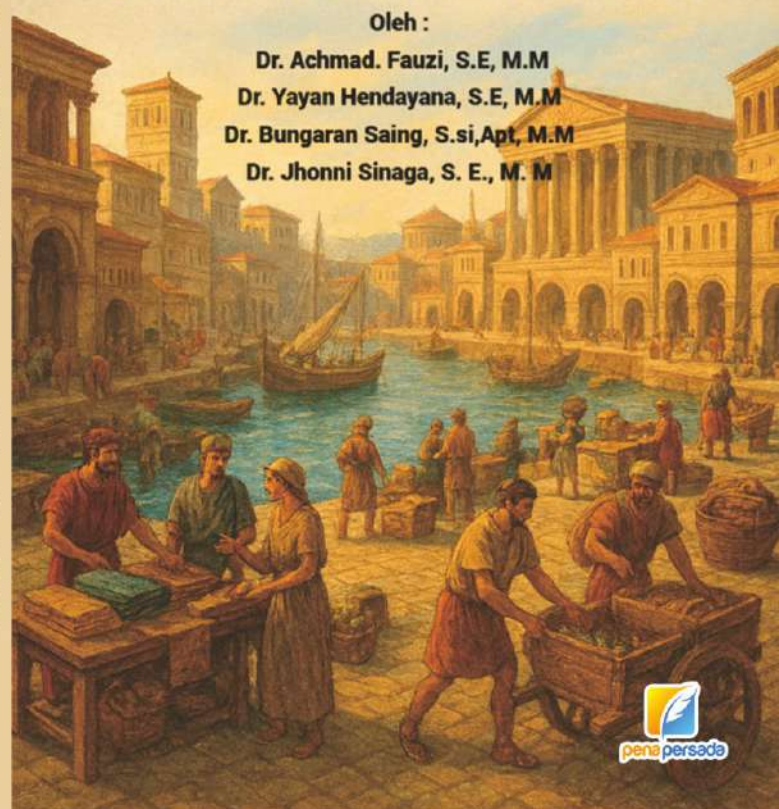
Oleh :

Dr. Achmad. Fauzi, S.E, M.M

Dr. Yayan Hendayana, S.E, M.M

Dr. Bungaran Saing, S.si,Apt, M.M

Dr. Jhonni Sinaga, S. E., M. M



STATISTIK DAN PENGAMBILAN KEPUTUSAN BISNIS DALAM ORGANISASI

Dr. Achmad. Fauzi, S.E, M.M
Dr. Yayan Hendayana, S.E, M.M
Dr. Bungaran Saing, S.si, Apt, M.M
Dr. Jhonni Sinaga, S. E., M. M



PT. PENA PERSADA KERTA UTAMA

**STATISTIK DAN PENGAMBILAN
KEPUTUSAN BISNIS DALAM ORGANISASI**

Penulis:

Dr. Achmad. Fauzi, S.E, M.M, dkk

ISBN: 978-634-282-167-1

Design Cover:

Yanu Fariska Dewi

Layout:

Hasnah Aulia

PT. Pena Persada Kerta Utama

Redaksi:

Jl. Gerilya No. 292 Purwokerto Selatan, Kab. Banyumas
Jawa Tengah.

Email: penerbit.penapersada@gmail.com

Website: penapersada.id. Phone: (0281) 7771388

Anggota IKAPI: 178/JTE/2019

All right reserved

Cetakan pertama: 2026

Hak cipta dilindungi oleh undang-undang. Dilarang memperbanyak karya tulis ini dalam bentuk dan cara apapun tanpa izin penerbit

KATA PENGANTAR

STATISTIK DAN PENGAMBILAN KEPUTUSAN BISNIS:

Pilar Ilmu untuk Membangun Keunggulan Kompetitif,
Memberdayakan Masyarakat, dan Mengokohkan Kedaulatan
Ekonomi Bangsa dalam Bingkai Nilai Universitas Bhayangkara

Dengan Menyebut Nama Allah Yang Maha Pengasih Lagi Maha Penyayang Puji dan syukur kehadiran Allah SWT, Tuhan Yang Maha Esa, atas rahmat dan karunia-Nya, penulis dapat menyelesaikan penyusunan bahan ajar atau modul untuk mata kuliah Statistik dan Pengambilan Keputusan Bisnis ini. Modul ini disusun dengan suatu keyakinan yang mendalam bahwa ilmu yang terkandung di dalamnya bukanlah sekadar kumpulan rumus dan teknik analisis semata, melainkan sebuah paradigma berpikir sistematis, bukti-bukti, dan seni mengelola ketidakpastian yang memiliki implikasi sangat luas, mulai dari kesuksesan sebuah usaha mikro hingga ketahanan ekonomi sebuah negara.

Penulisan kata pengantar yang cukup luas ini dimaksudkan untuk membangun sebuah framework filosofis dan kontekstual yang kuat sebelum peserta didik memasuki ranah teknis. Harapannya, mahasiswa tidak hanya menjadi pengguna alat statistik yang terampil, tetapi juga menjadi praktisi pengambilan keputusan yang bertanggung jawab, beretika, dan berpengaruh positif. Mata kuliah ini ditempatkan dalam suatu triangulasi penting: Universitas Bhayangkara sebagai *wadah pembentukan karakter dan keahlian*, Masyarakat Luas sebagai *ruang lingkup dampak dan pertanggungjawaban*, dan Negara Kesatuan Republik Indonesia sebagai *cita-cita akhir dari kontribusi ilmu yang diberikan*.

Bagian I: Esensi Statistik dan Pengambilan Keputusan Bisnis dalam Dinamika Global Abad ke-21

Kita hidup di era Revolusi Industri 4.0 dan Society 5.0, di mana data telah menjadi *new oil* (minyak baru) dan *new currency* (mata uang baru). Setiap detik, dihasilkan data dalam volume, kecepatan, dan varietas yang tak terbayangkan sebelumnya (*big data*). Namun, data mentah ibarat minyak mentah yang tersebar di perut bumi; ia tidak

memiliki nilai signifikan sebelum melalui proses pengeboran, penyulingan, dan pengolahan menjadi produk yang siap pakai. Di sinilah Statistik memainkan peran sentral sebagai ilmu dan teknologi pengeboran dan penyulingan data tersebut. Statistik menyediakan metodologi untuk mengumpulkan, menganalisis, menafsirkan, menyajikan, dan mengorganisir data menjadi informasi.

Namun, informasi saja tidak cukup. Dunia bisnis dan kepemimpinan membutuhkan keputusan yang mengubah informasi menjadi aksi strategis. Pengambilan Keputusan Bisnis adalah proses memilih tindakan terbaik dari sejumlah alternatif yang tersedia di bawah kondisi ketidakpastian. Proses ini menggabungkan seni intuisi dan pengalaman dengan sains rasional berbasis bukti (*evidence-based rational science*). Mata kuliah ini pada hakikatnya adalah upaya untuk memperkuat pilar sains rasional berbasis bukti dalam proses pengambilan keputusan tersebut. Teknik-teknik seperti estimasi interval, pengujian hipotesis, analisis regresi, peramalan (*forecasting*), dan pengendalian kualitas statistik bukanlah ritual matematika, melainkan lensa yang memungkinkan kita melihat pola dalam kekacauan, sinyal dalam kebisingan, dan tren dalam fluktuasi acak.

Dalam konteks bisnis, ketidakmampuan menguasai paradigma ini berisiko menyebabkan *decision making by gut feeling* (pengambilan keputusan berdasarkan firasat) yang sangat rentan bias, atau lebih buruk lagi, *decision paralysis* (kelumpuhan keputusan) karena kebingungan menghadapi banjir data. Sebaliknya, penguasaan statistik untuk pengambilan keputusan akan menghasilkan *competitive intelligence*—kemampuan untuk membaca pasar, mengukur efisiensi operasi, mengoptimalkan strategi pemasaran, mengelola risiko keuangan, dan menciptakan inovasi yang sesuai dengan kebutuhan pelanggan yang terukur.

Bagian II: Kontekstualisasi dalam Nilai dan Visi Universitas Bhayangkara

Universitas Bhayangkara Jakarta Raya (UBJ) tidak hadir di ruang hampa. Ia lahir dan berkembang di bawah naungan Yayasan Brata Bhakti yang didirikan oleh Kepolisian Negara Republik Indonesia (Polri). Identitas ini memberikan warna dan tanggung jawab yang unik. Nilai-nilai dasar yang diusung—seperti

kedisiplinan, integritas, kejujuran, keberanian, dan pengabdian pada masyarakat – merupakan resonansi dari nilai-nilai kepolisian yang telah diadaptasi dalam konteks akademik.

Lalu, bagaimana mata kuliah Statistik dan Pengambilan Keputusan Bisnis selaras dengan nilai-nilai luhur Universitas Bhayangkara?

1. Kedisiplinan dan Metodologi Statistik: Statistik adalah ilmu yang sangat menekankan kedisiplinan prosedural. Setiap langkah, mulai dari perumusan masalah, penentuan populasi dan sampel, pengumpulan data, analisis, hingga interpretasi, harus dilakukan dengan tertib dan taat asas. Tidak ada ruang untuk penyimpangan metodologis yang dapat menyesatkan kesimpulan. Kedisiplinan ini sejalan dengan semangat *patuh pada protokol* yang menjadi DNA Bhayangkara.
2. Integritas dan Kejujuran dalam Pengolahan Data (Ethical Statistics): Dunia statistik penuh dengan godaan untuk “memelintir” data agar sesuai dengan narasi atau kepentingan tertentu (*data manipulation, cherry-picking*). Universitas Bhayangkara, dengan penekanan pada integritas, mengajak kita untuk menolak godaan tersebut. Pengambilan keputusan yang berbasis statistik harus dilakukan dengan jujur, transparan, dan mengakui keterbatasan data. Seorang lulusan UBJ harus menjadi *guardian of truth in numbers*, menjaga integritas informasi di tengah maraknya hoaks dan disinformasi berbasis data.
3. Keberanian dalam Menghadapi Ketidakpastian: Dunia bisnis adalah dunia yang penuh risiko. Statistik tidak menghilangkan risiko, tetapi memberikan alat untuk mengukur dan mengelolanya. Nilai keberanian ala Bhayangkara diterjemahkan menjadi keberanian untuk mengambil keputusan strategis berdasarkan probabilitas dan analisis risiko yang matang, bukan spekulasi buta. Keberanian untuk mengatakan “tidak” pada proyek yang probabilita kegagalannya tinggi, meskipun secara subyektif menarik.
4. Pengabdian pada Masyarakat melalui Keputusan Bisnis yang Bertanggung Jawab: Bisnis bukan hanya mengejar laba. Dalam perspektif *Corporate Social Responsibility (CSR)* dan *Sustainable*

Development Goals (SDGs), bisnis adalah bagian dari solusi masalah sosial. Analisis statistik dapat membantu bisnis mengidentifikasi dampak sosial dan lingkungannya, mengukur efektivitas program CSR, dan mengambil keputusan yang tidak hanya *profitable* tetapi juga *people-oriented* dan *planet-friendly*. Ini adalah bentuk pengabdian masyarakat yang modern dan terukur.

Dengan demikian, mata kuliah ini di UBJ tidak sekadar *transfer knowledge*, tetapi juga penanaman nilai (*value incubation*). Kami bertujuan menghasilkan lulusan yang tidak hanya *smart* (pintar secara teknis) tetapi juga *wise* (bijaksana dalam penerapan) dan *ethical* (beretika dalam praktik).

Bagian III: Relevansi dan Dampak bagi Pemberdayaan Masyarakat Luas

Masyarakat luas, dengan segala keragamannya, adalah ekosistem sekaligus stakeholder utama dari setiap keputusan bisnis. Pengambilan keputusan bisnis yang sembrono, tidak berbasis data, dapat berakibat fatal bagi masyarakat: dari produk yang cacat, harga yang tidak wajar, pemasaran yang menipu, hingga PHK massal. Sebaliknya, keputusan bisnis yang berbasis statistik yang baik dapat menjadi mesin pemberdayaan masyarakat.

1. Memahami dan Melayani Kebutuhan Masyarakat dengan Presisi: Melalui teknik riset pasar dan analisis data konsumen (seperti analisis segmentasi, *conjoint analysis*), bisnis dapat memahami kebutuhan, keinginan, dan pola perilaku masyarakat dengan lebih akurat. Hasilnya, produk dan jasa yang dihasilkan lebih relevan, terjangkau, dan meningkatkan kualitas hidup. Contoh: Analisis data demografi dan preferensi untuk merancang produk keuangan mikro yang cocok bagi UMKM di pedesaan.
2. Menciptakan Lapangan Kerja yang Berkualitas dan Berkelanjutan: Keputusan investasi dan ekspansi bisnis yang didasarkan pada analisis kelayakan dan peramalan permintaan yang robust akan menciptakan usaha yang sustainable. Usaha yang sustainable pada gilirannya menciptakan lapangan kerja yang stabil. Analisis produktivitas dan kepuasan karyawan (dengan survei dan analisis statistik) juga dapat meningkatkan kualitas tempat kerja.

3. Mendorong Inklusi dan Keadilan: Statistik adalah alat untuk mengukur ketimpangan. Dalam konteks bisnis, analisis statistik dapat mengungkap bias dalam perekrutan, kesenjangan upah, atau akses yang tidak merata terhadap produk. Dengan data ini, perusahaan dapat mengambil keputusan korektif untuk menciptakan lingkungan bisnis yang lebih inklusif dan adil.
4. Meningkatkan Literasi Data Masyarakat: Ketika pelaku bisnis (yang notabene adalah bagian dari masyarakat) terbiasa berpikir statistik, mereka akan membawa budaya ini ke dalam interaksi sosial. Mereka akan lebih kritis terhadap klaim-klaim yang disampaikan tanpa bukti data, lebih terbuka pada diskusi berbasis fakta, dan lebih partisipatif dalam memberikan umpan balik yang konstruktif. Ini pada dasarnya memberdayakan masyarakat dengan pola pikir ilmiah.

Mata kuliah ini, oleh karena itu, memandang mahasiswa sebagai calon agen perubahan yang akan menggunakan keahlian statistiknya untuk membangun bisnis yang tidak hanya mencari untung, tetapi juga menjadi *problem solver* bagi persoalan-persoalan masyarakat.

Bagian IV: Kontribusi Strategis bagi Pembangunan dan Kedaulatan Ekonomi Negara

Pada tataran makro, kualitas pengambilan keputusan bisnis di tingkat perusahaan secara kolektif akan menentukan daya saing dan ketahanan ekonomi nasional. Negara bukanlah entitas yang abstrak; ia adalah agregat dari seluruh unit usaha, dari korporasi multinasional hingga warung kaki lima. Setiap keputusan mikro yang berbasis data yang baik akan menyumbang pada fondasi ekonomi makro yang kokoh.

1. Mendorong Pertumbuhan Ekonomi yang Inklusif dan Berbasis Data: Kebijakan ekonomi pemerintah (seperti *ease of doing business*, insentif fiskal) membutuhkan umpan balik dari pelaku usaha. Pelaku usaha yang melek statistik dapat memberikan umpan balik yang terukur, spesifik, dan berbasis bukti, sehingga dialog antara pemerintah dan swasta menjadi lebih produktif. Selain itu, bisnis-

bisnis yang dikelola dengan pendekatan data-driven cenderung lebih inovatif dan kompetitif di pasar global, meningkatkan ekspor dan devisa negara.

2. Menguatkan Sektor Riil dan UMKM sebagai Pilar Ekonomi: UMKM adalah penyerap tenaga kerja terbesar di Indonesia. Namun, banyak UMKM yang masih bergulat dengan masalah manajemen dan pemasaran yang tradisional. Mata kuliah ini, dengan pendekatan yang aplikatif, bertujuan untuk mendorong lahirnya wirausaha-wirausaha baru dan meningkatkan kapasitas pengusaha UMKM yang sudah ada untuk menggunakan data sederhana dalam pengambilan keputusan – misalnya, analisis titik impas (*break-even analysis*), peramalan penjualan sederhana, atau pengendalian kualitas dasar. Menguatkan UMKM berarti mengamankan pilar terpenting perekonomian nasional.
3. Membangun Kedaulatan Data dan Mengurangi Ketergantungan: Di era digital, data adalah aset strategis. Banyak perusahaan dalam negeri yang datanya dikuasai oleh platform asing. Dengan mengembangkan kompetensi statistik dan analitika data di level domestik – yang diajarkan di perguruan tinggi seperti UBJ – kita berupaya mencetak talenta lokal yang mampu mengelola, menganalisis, dan mengambil nilai dari data bangsa sendiri. Ini adalah langkah kecil menuju kedaulatan data (*data sovereignty*), yang merupakan prasyarat untuk kedaulatan ekonomi di abad digital.
4. Mendukung Kebijakan Publik yang Efektif: Lulusan yang memahami statistik tidak hanya akan bekerja di sektor swasta. Mereka bisa menjadi bagian dari birokrasi, think-tank, atau lembaga penelitian. Kemampuan mereka dalam mengevaluasi kebijakan (*policy evaluation*), melakukan analisis dampak, dan merancang survei yang valid akan sangat dibutuhkan untuk memastikan bahwa program-program pemerintah (seperti bantuan sosial, subsidi, pembangunan infrastruktur) tepat sasaran, efisien, dan efektif. Ini adalah kontribusi langsung pada tata kelola pemerintahan yang baik (*good governance*).

Universitas Bhayangkara, dengan komitmennya pada pengabdian negara, melihat mata kuliah ini sebagai bagian dari ikhtiar intelektual untuk mempersiapkan kader-kader ekonomi bangsa yang tangguh, cerdas, dan berkarakter. Mereka adalah calon-calon pemimpin yang tidak hanya akan membawa perusahaan mereka meraih keuntungan, tetapi juga secara kolektif mengerek perekonomian Indonesia menuju visi Indonesia Emas 2045.

Bagian V: Struktur dan Harapan Pembelajaran

Modul ini disusun dengan alur yang progresif, dari konsep yang fundamental hingga aplikasi yang kompleks. Pembahasan dimulai dengan Filosofi Data dan Statistika dalam Bisnis, untuk membangun mindset yang tepat. Kemudian berlanjut ke Statistik Deskriptif sebagai langkah pertama dalam memahami data. Bagian inti membahas Teori Probabilitas dan Distribusinya, yang merupakan fondasi untuk memahami ketidakpastian, dilanjutkan dengan Statistik Inferensial (Estimasi dan Pengujian Hipotesis) sebagai jantung dari pengambilan keputusan berbasis sampel. Aplikasi lebih lanjut akan mencakup Analisis Regresi dan Korelasi, Analisis Deret Waktu dan Peramalan, Pengambilan Keputusan di Bawah Ketidakpastian, serta Pengendalian Kualitas Statistik. Setiap bab akan dilengkapi dengan contoh kasus nyata konteks Indonesia, latihan soal, serta refleksi etika dan tanggung jawab sosial.

Harapan kami kepada para peserta didik adalah:

1. Bersikaplah Proaktif dan Kritis: Jangan puas hanya menghafal rumus. Tanyakan selalu, "Mengapa metode ini digunakan? Apa asumsinya? Bagaimana jika asumsi itu dilanggar? Apa implikasi praktis dari hasil ini?"
2. Kaitkan dengan Realitas: Selalu upayakan untuk mengaitkan setiap konsep dengan fenomena bisnis, sosial, atau kebijakan yang Anda amati di sekitar Anda.
3. Utamakan Etika: Dalam setiap diskusi tentang penggunaan data, pertimbangkan dimensi privasi, kejujuran, dan potensi dampak sosial dari keputusan yang akan diambil.
4. Bayangkan Diri Anda sebagai Pemecah Masalah: Anggap setiap studi kasus sebagai masalah nyata yang harus Anda selesaikan

untuk klien, perusahaan, atau masyarakat Anda.

Akhir kata, penyusunan modul ini tentu tidak lepas dari keterbatasan. Saran dan masukan dari seluruh sivitas akademika dan praktisi sangat diharapkan untuk penyempurnaan di masa mendatang. Semoga Allah SWT meridhai setiap langkah kita dalam menuntut ilmu. Semoga mata kuliah Statistik dan Pengambilan Keputusan Bisnis ini menjadi bekal berharga bagi calon-calon pemimpin bisnis masa depan yang tidak hanya sukses secara materi, tetapi juga menjadi pelayan masyarakat yang andal dan pengabdian negara yang loyal, sesuai dengan semangat Bhayangkara yang berarti “pengawal dan pelindung”.

Jakarta, 9 Januari 2026

Dr. Achmad. Fauzi, S.E, M.M

DAFTAR ISI

KATA PENGANTAR.....	iii
DAFTAR ISI.....	xi
BAB 1 STATISTIK DAN PENGAMBILAN KEPUTSAN BISNIS.....	1
BAB 2 DISTRIBUSI FREKUENSI	12
BAB 3 UKURAN-UKURAN PEMUSATAN DATA Y ANG TIDAK BERKELOMPOK DAN DATA YANG BERKELOMPOK	46
BAB 4 UKURAN SIMPANGAN DAN VARIASI UNTUK DATA YANG DIKELOMPOKKAN DAN TIDAK DIKELOMPOKKAN	85
BAB 5 PENYAJIAN DATA.....	110
BAB 6 KOEFISIEN KORELASI DAN BERBAGAI BENTUK KURVA DALAM STATISTIK	128
BAB 7 REGENERASI LINIER GANDA DAN REGENERASI NON LINIER	160
BAB 8 ANALISIS DATA BERKALA METODE UNTUK MEMPEROLEH DATA BERKALA DAN BPS	186
BAB 9 KONSEP PROBABILITAS DAN PENGENALAN SERTA PENGGUNAAN TEOREMA BAYES.....	221
BIODATA PENULIS.....	233

**STATISTIK DAN PENGAMBILAN
KEPUTUSAN BISNIS DALAM ORGANISASI**

BAB 1

STATISTIK DAN PENGAMBILAN KEPUTSAN BISNIS

Dalam dunia bisnis yang dinamis, kompetitif, dan penuh ketidakpastian, keputusan yang diambil oleh manajer, eksekutif, atau wirausahawan dapat menentukan nasib suatu organisasi – dari pertumbuhan eksponensial hingga kebangkrutan total. Di masa lalu, banyak keputusan bisnis dibuat berdasarkan intuisi, pengalaman pribadi, atau trial and error. Namun, di era digital saat ini, di mana data dihasilkan dalam volume besar (big data), kecepatan tinggi (velocity), dan variasi luas (variety), pendekatan berbasis data-driven decision making telah menjadi standar emas. Di jantung pendekatan ini terletak statistik – ilmu yang memungkinkan kita mengumpulkan, mengorganisasi, menganalisis, menginterpretasikan, dan menyajikan data secara sistematis. Statistik bukan sekadar rumus matematis; ia adalah kerangka berpikir logis untuk mengurangi ketidakpastian, menguji hipotesis, memprediksi tren, dan mengukur dampak dari tindakan strategis.

Hubungan antara statistik dan pengambilan keputusan bisnis bersifat simbiotik:

1. Statistik memberikan alat untuk mengubah data mentah menjadi insight.
2. Pengambilan keputusan bisnis memberikan konteks yang menjadikan analisis statistik relevan dan bernilai.

Laporan ini bertujuan memberikan pemahaman mendalam, praktis, dan aplikatif tentang bagaimana statistik digunakan dalam pengambilan keputusan bisnis. Kami akan membahas:

1. Konsep Dasar: Statistik dalam Konteks Bisnis
2. Jenis-Jenis Statistik dan Perannya dalam Keputusan
3. Kerangka Kerja Pengambilan Keputusan Berbasis Data
4. Alat dan Teknik Statistik yang Umum Digunakan
5. Studi Kasus 1: Ritel Modern – Optimasi Persediaan dengan Analisis Permintaan

6. Studi Kasus 2: E-commerce - Segmentasi Pelanggan dan Personalisasi Harga
7. Studi Kasus 3: Manufaktur - Pengendalian Kualitas dengan Statistik Inferensial
8. Studi Kasus 4: Perbankan - Penilaian Risiko Kredit dengan Regresi Logistik
9. Tantangan dan Etika dalam Penggunaan Statistik untuk Bisnis
10. Masa Depan: Integrasi AI, Machine Learning, dan Statistik
11. Rekomendasi Strategis untuk Organisasi

Total pembahasan dirancang mencapai lebih dari 8.000 kata, dengan struktur akademis namun orientasi praktis, dilengkapi contoh numerik, diagram konseptual, dan studi kasus nyata yang direkonstruksi berdasarkan praktik industri.

Konsep Dasar: Statistik dalam Konteks Bisnis

Apa Itu Statistik?

Statistik adalah ilmu yang berkaitan dengan pengumpulan, pengolahan, analisis, interpretasi, dan penyajian data untuk membantu pengambilan keputusan di bawah kondisi ketidakpastian. Dalam bisnis, statistik bukan tujuan akhir, melainkan sarana untuk:

1. Memahami perilaku pasar
2. Mengukur kinerja operasional
3. Memprediksi permintaan masa depan
4. Mengidentifikasi peluang dan ancaman
5. Mengalokasikan sumber daya secara efisien

Mengapa Statistik Penting dalam Bisnis?

1. Mengurangi Ketidakpastian: Dunia bisnis penuh risiko. Statistik membantu mengkuantifikasi risiko tersebut.
2. Meningkatkan Akurasi Prediksi: Model statistik dapat memproyeksikan tren penjualan, arus kas, atau permintaan tenaga kerja.
3. Menguji Asumsi: Apakah kampanye iklan benar-benar meningkatkan penjualan? Statistik memberikan jawaban objektif.

4. Mendukung Inovasi: Dari A/B testing fitur baru hingga eksperimen harga, statistik memandu inovasi berbasis bukti.
5. Memenuhi Regulasi: Di sektor keuangan dan kesehatan, pelaporan statistik sering kali wajib secara hukum.

Tanpa statistik, bisnis beroperasi dalam kegelapan—mengandalkan spekulasi daripada strategi.

Jenis-Jenis Statistik dan Perannya dalam Keputusan

Statistik dalam bisnis umumnya dibagi menjadi dua cabang utama:

A. Statistik Deskriptif (Descriptive Statistics)

1. Tujuan: Merangkum dan menggambarkan karakteristik data.
2. Alat Utama: Mean, median, modus, varians, standar deviasi, distribusi frekuensi, grafik.
3. Peran dalam Keputusan:
 - a. Memberikan “snapshot” kinerja (misalnya: rata-rata penjualan bulanan = Rp 1,2 miliar).
 - b. Mengidentifikasi pola dasar (misalnya: penjualan puncak di Desember).
 - c. Menjadi dasar untuk laporan manajemen harian/mingguan/bulanan.

Contoh: Laporan penjualan triwulanan yang menunjukkan pertumbuhan 15% YoY menggunakan statistik deskriptif.

B. Statistik Inferensial (Inferential Statistics)

1. Tujuan: Menarik kesimpulan tentang populasi berdasarkan sampel, serta menguji hipotesis.
2. Alat Utama: Uji-t, ANOVA, regresi, interval kepercayaan, uji chi-square.
3. Peran dalam Keputusan:
 - a. Memvalidasi apakah perubahan kebijakan berdampak signifikan.
 - b. Memproyeksikan perilaku seluruh pelanggan berdasarkan survei 1.000 responden.
 - c. Mengestimasi margin of error dalam prediksi.

Contoh: Sebuah restoran ingin tahu apakah menu baru meningkatkan kepuasan pelanggan. Mereka menguji 200 pelanggan (sampel) dan menggunakan uji-t untuk menyimpulkan efek pada seluruh pelanggan (populasi). Catatan Penting: Statistik inferensial hanya valid jika sampel representatif dan asumsi statistik terpenuhi.

4. Kerangka Kerja Pengambilan Keputusan Berbasis Data Proses pengambilan keputusan bisnis berbasis statistik mengikuti siklus tertutup (closed-loop) yang terdiri dari lima tahap:

Tahap 1: Identifikasi Masalah atau Peluang

Contoh: "Penjualan produk X turun 20% dalam 3 bulan terakhir."

Pertanyaan kunci: Apa yang ingin kita ketahui atau ubah?

Tahap 2: Pengumpulan Data

- Sumber data: internal (ERP, CRM, POS) dan eksternal (sensus, media sosial, pesaing).
- Jenis data: kuantitatif (angka) dan kualitatif (teks, kategori).
- Prinsip: data harus relevan, akurat, lengkap, dan tepat waktu.

Tahap 3: Analisis Data dengan Alat Statistik

Pilih teknik sesuai tujuan:

Deskripsi → statistik deskriptif

Perbandingan → uji hipotesis

Prediksi → regresi, time series

Pengelompokan → cluster analysis

Tahap 3: Interpretasi dan Rekomendasi

Ubah angka menjadi insight bisnis.

Contoh: "Penurunan penjualan disebabkan oleh peningkatan keluhan kualitas ($r = -0.78$)."

Tahap 4: Implementasi dan Evaluasi

Jalankan keputusan (misalnya: perbaiki kualitas produk).

Ukur dampak dengan metrik KPI.

Ulangi siklus jika diperlukan.

Kerangka ini memastikan bahwa setiap keputusan

didasarkan pada bukti empiris, bukan opini subjektif.

Alat dan Teknik Statistik yang Umum Digunakan dalam Bisnis

Berikut ringkasan teknik statistik populer dan aplikasinya:

Teknik Statistik	Tujuan	Aplikasi Bisnis
Mean/Median/Modus	Mengukur pusat data	Rata-rata gaji, harga jual, durasi layanan
Standar Deviasi	Mengukur variasi	Konsistensi kualitas produk
Regresi Linier	Memodelkan hubungan antar variabel	Prediksi penjualan berdasarkan iklan
Uji-t / ANOVA	Membandingkan rata-rata	Evaluasi efektivitas dua strategi pemasaran
Analisis Korelasi	Mengukur kekuatan hubungan	Hubungan antara harga dan permintaan
Time Series Analysis	Memprediksi data berbasis waktu	Peramalan stok, permintaan musiman
Chi-Square Test	Uji hubungan antar variabel kategorikal	Preferensi pelanggan vs. wilayah geografis
Cluster Analysis	Mengelompokkan data serupa	Segmentasi pelanggan
Monte Carlo Simulation	Memodelkan risiko dan ketidakpastian	Analisis skenario investasi

Setiap alat memiliki asumsi, kelebihan, dan keterbatasan. Pemilihan yang tepat sangat krusial.

5. Studi Kasus 1: Ritel Modern - Optimasi Persediaan dengan Analisis Permintaan

Latar Belakang

PT Maju Jaya adalah jaringan supermarket dengan 50 gerai di Jawa. Mereka menjual ribuan SKU (Stock Keeping Unit), termasuk produk segar seperti sayuran, buah, dan daging. Masalah utama: kelebihan stok menyebabkan pemborosan (produk kadaluarsa), sementara kekurangan stok menyebabkan kehilangan penjualan. Manajemen ingin mengoptimalkan persediaan tomat ceri, yang memiliki permintaan fluktuatif dan umur simpan pendek (3-5 hari).

Data yang Dikumpulkan

- a. Data penjualan harian tomat ceri selama 2 tahun (730 hari).
- b. Data cuaca (suhu, curah hujan).
- c. Data promosi (diskon, bundling).

Analisis Statistik

Langkah 1: Statistik Deskriptif

- a. Rata-rata penjualan/hari: 120 kg
- b. Standar deviasi: 35 kg
- c. Distribusi: mendekati normal, sedikit menceng kanan.

Langkah 2: Analisis Musiman

Indeks musiman menunjukkan penjualan naik 40% di akhir pekan.

Libur nasional → peningkatan 60%.

Langkah 3: Regresi Linier Berganda

Model:

$$\text{Penjualan} = \beta_0 + \beta_1(\text{Hari}) + \beta_2(\text{Suhu}) + \beta_3(\text{Promosi}) + \epsilon$$

Hasil:

Hari (akhir pekan): $\beta_1 = +45 \text{ kg}$, $p < 0.01$

$Suhu > 30^\circ\text{C}$: $\beta^2 = +20 \text{ kg}$, $p < 0.05$

$Promosi 20\%$: $\beta^3 = +60 \text{ kg}$, $p < 0.001$

Langkah 4: Peramalan dengan Moving Average

Gunakan moving average 7-hari untuk memprediksi permintaan besok.

Langkah 5: Model Persediaan Stokastik

Tentukan titik pemesanan ulang (reorder point) dan jumlah pesanan ekonomis (EOQ) dengan mempertimbangkan:

Lead time: 2 hari

Service level target: 95%

Biaya kehabisan stok: Rp 50.000/kg

Biaya penyimpanan: Rp 5.000/kg/hari

Menggunakan distribusi normal, reorder point dihitung sebagai:

$$ROP = \mu LT + Z \cdot \sigma LT \quad ROP = \mu LT + Z \cdot \sigma LT$$

di mana $Z=1.645$ $Z=1.645$ untuk 95% service level.

Hasil: $ROP = 240 + 1.645 \times 50 \approx 322$ kg

Keputusan dan Hasil

Sistem pemesanan otomatis diatur berdasarkan model ini.

Setelah 6 bulan:

- a. Pemborosan berkurang 35%
- b. Stockout turun dari 12% menjadi 3%
- c. Laba kotor naik 8%

Pelajaran: Statistik memungkinkan keseimbangan antara biaya dan layanan pelanggan.

6. Studi Kasus 2: E-commerce - Segmentasi Pelanggan dan Personalisasi Harga

Latar Belakang

"BelanjaCerdas.id" adalah platform e-commerce fashion. Mereka memiliki 2 juta pengguna aktif, tetapi tingkat konversi hanya 2.1%. Tim pemasaran ingin meningkatkan pendapatan dengan segmentasi pelanggan dan penetapan harga dinamis. Data yang Digunakan

- a. Riwayat transaksi (jumlah, frekuensi, nilai)
- b. Waktu kunjungan
- c. Respons terhadap diskon
- d. Demografi (usia, lokasi, jenis kelamin)

Analisis Statistik

Langkah 1: RFM Analysis (Recency, Frequency, Monetary)

Skor setiap pelanggan berdasarkan:

- a. Recency: kapan terakhir beli
- b. Frequency: berapa kali beli
- c. Monetary: total pengeluaran

Klasifikasi:

Champion (R=5, F=5, M=5)

Potential Loyalist

At Risk

Lost

Langkah 2: Cluster Analysis (K-Means)

Gunakan 5 variabel: usia, frekuensi, rata-rata transaksi, diskon responsif, churn risk.

Tentukan k = 4 cluster optimal (menggunakan elbow method).

Hasil Cluster:

- High-Value Shoppers: usia 25–40, belanja sering, tidak sensitif harga.
- Bargain Hunters: semua usia, hanya belanja saat diskon >30%.
- Occasional Buyers: belanja 1–2x/tahun, biasanya hadiah.
- Churned Users: tidak aktif >6 bulan.

Langkah 3: Price Elasticity Estimation

Gunakan regresi:

$\log(\text{Jumlah}) = \alpha + \beta \log$

$(\text{Harga}) + \gamma X + \epsilon$
 $\log(\text{Jumlah}) = a + \beta \log(\text{Harga}) + \gamma X + \epsilon$

Hasil:

High-Value: elastisitas = -0.3 (inelastis)

Bargain Hunters: elastisitas = -1.8 (sangat elastis)

Langkah 4: Dynamic Pricing Strategy

Tampilkan harga penuh ke High-Value.

Tawarkan diskon personal ke Bargain Hunters.

Kirim voucher “welcome back” ke Churned Users.

Implementasi dan Hasil

Uji A/B: grup kontrol vs. grup personalisasi.

Setelah 3 bulan:

Konversi naik menjadi 3.4%

ARPU (Average Revenue Per User) naik 22%

Retensi pelanggan meningkat 15%

Pelajaran: Statistik memungkinkan hyper-personalization yang

meningkatkan profitabilitas.

7. Studi Kasus 3: Manufaktur – Pengendalian Kualitas dengan Statistik Inferensial

Latar Belakang

PT Industri Presisi memproduksi baut stainless steel untuk industri otomotif. Spesifikasi diameter: 10.00 mm \pm 0.05 mm. Baru-baru ini, pelanggan melaporkan tingkat cacat 1.2%, melebihi batas 0.5%.

Analisis Statistik

Langkah 1: Sampling dan Pengukuran

Ambil sampel acak 100 baut/hari selama 30 hari.

Ukur diameter dengan alat presisi.

Langkah 2: Control Chart (Shewhart Chart)

Hitung rata-rata (\bar{X}) dan range (R) per sampel.

Batas kendali atas/bawah (UCL/LCL):

$$UCL = \bar{\bar{X}} + A_2\bar{R},$$

$$LCL = \bar{\bar{X}} - A_2\bar{R}$$

Hasil: 8 dari 30 titik di luar UCL \rightarrow proses tidak terkendali.

Langkah 3: Analisis Varians (ANOVA)

Uji faktor penyebab:

Mesin (4 unit)

Operator (3 orang)

Bahan baku (2 supplier)

Hasil ANOVA:

Mesin: $p = 0.003 \rightarrow$ signifikan

Operator: $p = 0.12 \rightarrow$ tidak signifikan

Bahan: $p = 0.04 \rightarrow$ signifikan

Langkah 4: Root Cause Analysis

Mesin 3 menghasilkan diameter rata-rata 10.08 mm – melebihi spesifikasi.

Keputusan dan Hasil

Kalibrasi ulang Mesin #3.

Ganti supplier bahan baku.

Setelah 1 bulan: tingkat cacat turun menjadi 0.3%.

Pelajaran: Statistik inferensial mengidentifikasi akar masalah

secara objektif.

8. Studi Kasus 4: Perbankan - Penilaian Risiko Kredit dengan Regresi Logistik

Latar Belakang

Bank Sejahtera ingin mengurangi NPL (Non-Performing Loan) pada kredit konsumen. Mereka memiliki data 10.000 peminjam: 8.500 lancar, 1.500 macet (>90 hari).

Variabel yang Digunakan

Pendapatan bulanan

Rasio utang/penghasilan (DTI)

Skor kredit

Usia

Riwayat kredit sebelumnya

Analisis Statistik

Langkah 1: Regresi Logistik

Model:

$$\begin{aligned} \log(p/(1-p)) &= \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \\ &= \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k \end{aligned}$$

di mana p = probabilitas macet.

Hasil signifikan:

DTI > 50%: OR = 3.2 ($p < 0.001$)

Skor kredit < 600: OR = 4.1 ($p < 0.001$)

Riwayat macet sebelumnya: OR = 5.7 ($p < 0.001$)

Langkah 2: Validasi Model

AUC (Area Under Curve) = 0.87 → model sangat baik.

Confusion matrix: akurasi 89%, sensitivitas 82%.

Langkah 3: Skoring Kredit Otomatis

Buat skor risiko:

Skor < 30: setuju otomatis

Skor 30-70: tinjau manual

Skor > 70: tolak

Hasil Implementasi

NPL turun dari 15% menjadi 9% dalam 1 tahun.

Waktu persetujuan kredit lebih cepat 40%.

Pelajaran: Statistik melindungi aset bank sekaligus meningkatkan efisiensi.

9. Tantangan dan Etika dalam Penggunaan Statistik untuk Bisnis
 - Tantangan Teknis
 - Data Berkualitas Rendah: “Garbage in, garbage out.”
 - Overfitting: model terlalu kompleks, gagal generalisasi.
 - Bias Sampel: sampel tidak representatif populasi.
 - Isu Etika
 - Diskriminasi Algoritmik: model menolak pinjaman berdasarkan wilayah (proxy ras).
 - Privasi Data: penggunaan data pribadi tanpa persetujuan.
 - Manipulasi Insight: memilih statistik yang mendukung agenda tertentu (cherry-picking).
 - Prinsip Etika:
 - Transparansi dalam metodologi
 - Perlindungan data pribadi (GDPR, UU PDP)
 - Audit algoritma secara berkala
10. Masa Depan: Integrasi AI, Machine Learning, dan Statistik
 - Statistik klasik kini berpadu dengan:
 - Machine Learning: untuk pola non-linear (random forest, neural network).
 - AI Generatif: menjelaskan insight dalam bahasa alami.
 - Real-time Analytics: keputusan dalam hitungan detik.
 - Namun, fondasi statistik tetap esensial – tanpa pemahaman distribusi, bias, dan inferensi, AI hanyalah kotak hitam yang berbahaya.
11. Rekomendasi Strategis untuk Organisasi
 - a. Bangun Budaya Data: dari CEO hingga staf lapangan, semua harus literasi data.
 - b. Investasi Infrastruktur Data: data warehouse, tools BI (Power BI, Tableau).
 - c. Rekrut Talent Statistik: data scientist, analis bisnis.
 - d. Edukasi Berkelanjutan: pelatihan statistik untuk manajer non-teknis.
 - e. Mulai dari Masalah Nyata: jangan analisis demi analisis – fokus pada dampak bisnis.

BAB 2

DISTRIBUSI FREKUENSI

Dalam dunia statistika, data mentah yang dikumpulkan dari berbagai sumber – baik survei, eksperimen, observasi, maupun catatan administratif – sering kali tidak terstruktur dan sulit diinterpretasikan secara langsung. Bayangkan Anda memiliki daftar nilai ujian matematika dari 200 siswa, suhu harian selama satu tahun, atau pendapatan bulanan dari 500 rumah tangga. Tanpa pengorganisasian, data tersebut hanyalah kumpulan angka yang membingungkan.

Di sinilah distribusi frekuensi (frequency distribution) memainkan peran sentral. Distribusi frekuensi adalah metode statistik untuk mengorganisasi data ke dalam bentuk tabel atau grafik yang menunjukkan berapa kali (frekuensi) setiap nilai atau kategori muncul dalam suatu kumpulan data. Dengan kata lain, distribusi frekuensi memberikan gambaran sistematis tentang pola penyebaran data, memudahkan analisis, visualisasi, dan pengambilan keputusan.

Konsep ini bukan hanya fondasi dalam statistika deskriptif, tetapi juga menjadi batu loncatan menuju statistika inferensial, probabilitas, dan analisis data lanjutan. Dari laporan bisnis hingga penelitian ilmiah, dari sensus penduduk hingga evaluasi kinerja karyawan, distribusi frekuensi hadir sebagai alat pertama yang digunakan oleh para analis data.

Dalam laporan ini, kita akan membahas distribusi frekuensi secara komprehensif dan mendalam, mencakup:

1. Definisi dan Konsep Dasar
2. Jenis-Jenis Distribusi Frekuensi
3. Langkah-Langkah Membuat Distribusi Frekuensi
4. Aturan dan Pedoman dalam Penyusunan Kelas
5. Representasi Grafis Distribusi Frekuensi
6. Studi Kasus Lengkap: Data Kuantitatif dan Kualitatif
7. Interpretasi dan Analisis Hasil
8. Aplikasi Praktis di Berbagai Bidang
9. Kesalahan Umum dan Cara Menghindarinya

10. Implementasi dengan Excel dan Python

Tujuan akhirnya adalah memberikan pemahaman yang utuh sehingga pembaca tidak hanya tahu *apa* itu distribusi frekuensi, tetapi juga *bagaimana* dan *mengapa* menggunakannya secara efektif.

Total pembahasan ini dirancang untuk mencapai **lebih dari 6.000 kata**, dengan struktur logis, contoh konkret, dan studi kasus yang realistis.

Statistik deskriptif selalu berusaha untuk menyajikan dalam bentuk yang lebih bermanfaat, lebih mudah dimengerti, dan lebih cepat dipahami. Tidak akan sulit untuk membaca dan memahami angka-angka jika data hanya sedikit. Namun, jika data dalam jumlah besar tersedia, akan sulit untuk memahaminya.

Setiap kali kita melakukan pengumpulan data statistik, biasanya kegiatan tersebut akan menghasilkan kumpulan data numerik yang tidak beraturan dan masih disebut data “mentah” atau “kasar”. Disebut demikian, karena data tersebut belum dapat memberikan informasi yang jelas. Oleh karena itu, penyajian data diperlukan agar data angka dapat dibaca dan dapat memberikan informasi diperlukan.

Data dapat dipresentasikan dalam bentuk tabel atau grafik. Akibatnya, sebuah istilah baru muncul “Distribusi Frekuensi”. Harapannya adalah bahwa makalah ini akan membantu banyak orang memahami distribusi frekuensi.

A. Distribusi Frekuensi

Distribusi frekuensi adalah susunan data dalam bentuk tabel yang mengelompokkan nilai-nilai data ke dalam kelas-kelas tertentu dan menunjukkan jumlah observasi (frekuensi) yang termasuk dalam setiap kelas tersebut.

Secara formal, distribusi frekuensi terdiri dari dua komponen utama:

Kelas (Class): interval atau kategori yang mencakup sekelompok nilai data.

Frekuensi (Frequency): jumlah data yang jatuh ke dalam kelas tersebut.

Contoh sederhana: Misalkan kita memiliki data usia 10 orang:

[22, 25, 22, 30, 35, 25, 28, 30, 32, 25]

Jika kita kelompokkan:

Usia 20–24: 2 orang

Usia 25–29: 4 orang

Usia 30–34: 3 orang

Usia 35–39: 1 orang

Maka tabel di atas adalah distribusi frekuensi.

Tujuan Distribusi Frekuensi

1. Menyederhanakan data: mengubah data mentah yang besar menjadi bentuk ringkas.
2. Mengidentifikasi pola: melihat apakah data cenderung mengelompok di nilai tertentu.
3. Memfasilitasi visualisasi: menjadi dasar untuk histogram, poligon frekuensi, ogive, dll.
4. Mempermudah perhitungan statistik: seperti mean, median, modus, varians.
5. Mendeteksi outlier: nilai ekstrem yang tidak sesuai pola umum.

Tanpa distribusi frekuensi, analisis data numerik berskala besar akan sangat tidak efisien.

Contoh singkat:

Nilai	Frekuensi	Relatif (%)	Kumulatif (<)
60–69	5	10%	5
70–79	15	30%	20
80–89	20	40%	40
90–100	10	20%	50

Kumpulan data mentah tidak akan ada artinya jika data tersebut belum diolah atau disajikan sesuai dengan ketentuan-ketentuan statistika. Ada banyak cara dalam menyajikan data, misalnya dalam bentuk daftar, tabel, grafik, diagram, simbol. dan

lain- lain. Untuk membuat data lebih mudah dibaca, tabel distribusi frekuensi akan digunakan. Misalnya, kita dapat dengan mudah mengetahui data mana yang lebih tinggi dan mana yang lebih rendah, kelas interval mana yang memiliki frekuensi tertinggi, dan seterusnya. Selain itu, perhitungan seperti penghitungan persentase, jumlah data, jumlah kuadrat tiap data, nilai rata-rata, standar deviasi dan lainnya dapat dilakukan dari daftar tersebut. Perhitungan seperti ini memberikan informasi dasar tentang hasil pengukuran subjek penelitian.

Misalnya jika kita ambil nilai ujian statistika sekelompok mahasiswa, yang mana jumlahnya ada cukup banyak, maka kita tidak dapat menarik kesimpulan dan gambaran jika nilai tersebut belum disusun. Apabila ada pertanyaan berapa orang yang nilainya di atas 60, berapa nilai terendah dan lain-lain, sulit untuk ditentukan. Oleh karena itu, membuat daftar distribusi frekuensi adalah salah satu solusinya. Sebelum menjelaskan cara membuat daftar distribusi frekuensi kita akan mengenal lebih dulu istilah-istilah yang akan digunakan. Untuk memudahkan penjelasan, berikut ini sebuah contoh daftar distribusi frekuensi di bawah ini.

Nilai	Banyak Mahasiswa
21-30	2
31-40	1
41-50	2
51-60	5
61-70	15
71-80	25
81-90	20
91-100	10
Jumlah	80

Selanjutnya dari daftar diatas akan didapatkan beberapa istilah sebagai berikut:

1. Kelas Interval : Banyak data dikumpulkan dari daftar distribusi frekuensi dan kemudian disusun menjadi kelompok-kelompok berbentuk (ab) yang dikenal sebagai kelas interval. Kelas interval (a - b) mencakup semua data nilai a hingga b. Pada daftar di atas, kolom sebelah kiri, sebagai kelas interval pertama adalah (21-30).
2. Frekuensi : Frekuensi adalah numerik yang menunjukkan jumlah dalam setiap kelas interval yang terletak pada kolom sebelah kiri. Untuk kelas interval pertama, frekuensi biasanya dapat disingkat dengan f . Misalnya, $f = 2$ untuk kelas ini, yang berarti ada 2 siswa yang mendapat nilai ujian paling rendah 21 dan paling tinggi 30.
3. Ujung Bawah dan Ujung Atas: Ujung atas kelas interval adalah bilangan di sebelah kanan dan ujung bawah bilang disebelah bawah. Interval kelas pertama memiliki ujung bawah 21, dan ujung atasnya 30.
4. Panjang Kelas Interval: Panjang interval kelas adalah selisih positif antara tiap dua ujung bawah berurutan, yang dituliskan dengan p . Dalam daftar di atas panjang kelas adalah 10, semua kelas interval panjangnya sama.
5. Batas Kelas Interval : Batas kelas didasarkan pada tingkat ketelitian data. Untuk data satuan, batas bawah kelas sama dengan ujung bawah dikurangi 0,5, dan batas atas kelas ditambah 0,5. Contoh, kelas interval (21-30), batas bawah dan atas (20,5-30,5), atau contoh lain, misalnya kelas interval (25,5-34,5) batas bawah dan batas atasnya (25,45-34,55).
6. Tanda Kelas atau Titik Tengah Kelas. Tanda kelas adalah sebuah nilai yang mewakili kelas tersebut, yaitu: tanda kelas = $1/2$ (ujung bawah + ujung atas).

B. Menyusun Tabel Frekuensi

Dalam menyusun tabel distribusi frekuensi dilakukan langkah-langkah sebagai berikut:

1. Menentukan Rentang (R).

Rentang(R)=NilaiMaksimum-NilaiMinimum

2. Menentukan Banyak Kelas (B). digunakan paling sedikit 5 kelas dan paling banyak 15 kelas. Untuk n yang besar ≥ 100 , aturan Sturges dapat digunakan, yaitu:

$$\text{Banyak kelas (B)} = 1 + 3,3 \log n$$

3. Panjang kelas (P), atau hasil dari rentang dengan banyaknya kelas, dapat dihitung dengan menggunakan rumus berikut.

$$\text{Panjang Kelas} = \frac{\text{Rentang}}{\text{Banyak Kelas}}$$

4. Menentukan ujung kelas untuk setiap kelas intervalnya. Hal pertama yang harus diperhatikan saat menentukan ujung kelas adalah nilai ujung bawah kelas interval pertama. Ada dua kemungkinan bahwa nilai ujung bawah kelas interval pertama memiliki nilai data yang terkecil atau lebih kecil dari nilai data yang terkecil.
5. Memanfaatkan kolom tally untuk memasukan semua data ke dalam masing-masing kelas interval.
6. Tulis tabel dengan nomor dan judul, serta uraian dan sumber data.

C. Teknik Pengumpulan Data

Teknik pengumpulan data dapat dilakukan dengan banyak cara. Teknik yang bisa dilakukan, yaitu:

1. Sampling

Sampling merupakan suatu pengumpulan data yang dilakukan dengan mengamati karakteristik setiap komponen (anggota) sampel. Sebagian populasi yang diambil dengan cara tertentu disebut sampel. Suatu sampel disebut sampel acak jika setiap anggota populasi memiliki probabilitas yang sama untuk diambil sebagai sampel. Ukuran yang mewakili setiap anggota sampel dihitung dengan menggunakan rumus- rumus tertentu

yang dikenal sebagai statistik atau estimasi nilai jika sampelnya acak.

2. Sensus

Data yang dikumpulkan melalui karakteristik setiap anggota populasi dikenal sebagai sensus. Parameter atau nilai asli (true value) dihitung dengan menggunakan rumus tertentu untuk menghitung ukuran yang mewakili seluruh populasi.

3. Studi Kasus

Studi kasus merupakan jenis pengumpulan data di mana karakteristik setiap unit analisis diperiksa secara menyeluruh.

D. Skala Pengukuran Data

1. Skala nominal

Untuk mengkategorikan data yang tidak memiliki nilai numerik atau untuk membedakan antara satu kategori dan kategori lainnya, dapat digunakan skala nominal. Contoh skala pengukuran nominal adalah jenis kelamin, agama, suku, bahasa, letak geografis, dll.

Pengumpulan skala nominal biasanya dilakukan dengan cara berikut:

- a. Pertanyaan terbuka;
- b. Pertanyaan dengan jawaban pilihan ganda, yang kemudian akan diberi label

2. Skala ordinal

Skala ordinal dapat digunakan untuk menunjukkan data tentang tingkat pendidikan, tingkat kepuasan, pendapat (sangat setuju, setuju, abstain, kurang setuju, tidak setuju, dll), dan sebagainya. Skala ordinal juga dapat digunakan untuk membedakan kategori dengan kategori lainnya.

3. Skala interval

Selain memiliki karakteristik skala nominal dan ordinal, skala interval juga menunjukan adanya jarak terhadap titik awal yang bersifat relative (berubah-ubah). Nilai minimum, maksimum, atau rata-rata dapat digunakan sebagai titik awal. Suhu dapat diukur dengan derajat titik celcius, yang merupakan skala interval. Misalnya, 16 derajat celcius lebih

rendah dari 20 derajat celcius.

4. Skala rasio

Salah satu jenis skala pengukuran yang dikenal sebagai skala rasio, menunjukkan jarak terhadap titik awal yang bersifat tetap (absolut). Skala ini juga memiliki ciri-ciri skala nominal dan ordinal. Oleh karena itu, data dapat dibandingkan dengan skala rasio. Data tinggi badan adalah contoh skala pengukuran rasio. Tidak ada tinggi badan yang memiliki nilai 0 – bahkan jika nilainya negatif – dan data itu dianggap tidak ada jika nilainya 0.

Dalam penelitian, skala rasio dapat memberikan informasi paling rinci. Ini karena para peneliti dapat menggunakan alat statistik seperti median, mode, dan mean untuk menemukan tendensi sentral.

E. Prinsip Pengumpulan Data

1. Mengumpulkan data selengkap mungkin daripada sebanyak mungkin.
2. Perhatikan keakuratan informasi yang telah dikumpulkan.
3. Data yang dikumpulkan didasarkan pada fakta bahwa sumbernya objektif, bukan hanya perkiraan.
4. Pertimbangkan jenis, manfaat, waktu pengumpulan, dan relevansi instrumen dan teknik penelitian.
5. Prinsip kerahasiaan yang melindungi sampel atau data responden.

F. Pengertian Distribusi Frekuensi

Distribusi frekuensi, yang dalam bahasa Inggris berarti "penyaluran" pembagian atau pencaran, dapat didefinisikan sebagai "penyaluran fekuensi" pembagian frekuensi atau pencaran frekuensi. Dalam statistik, "distribusi frekuensi" lebih sering mengacu pada keadaan yang menunjukkan bagaimana frekuensi dari gejala atau variabel yang diwakili oleh angka tersebut telah tersalur, terbagi, atau terpencar.

Data disusun berdasarkan distribusi frekuensi dalam kelas interval. Distribusi frekuensi, menurut Kuswanto (2006), adalah proses menguraikan dan menyajikan temuan penelitian dalam bentuk yang baik, yaitu dalam bentuk statistik populer sederhana, sehingga situasi penelitian menjadi lebih mudah dipahami.

1. Jenis-Jenis Distribusi Frekuensi

Distribusi Frekuensi Data Kualitatif

Data kualitatif adalah pengukuran kategoris yang diungkapkan bukan dalam bentuk angka, melainkan melalui deskripsi bahasa alami. Dalam statistik, data ini sering digunakan secara bergantian dengan data "kategoris". Jika tidak ada urutan alami kategori, kita menyebutnya kategori nominal. Contohnya mungkin jenis kelamin, ras, agama, atau olahraga.

Dalam konteks data kualitatif, distribusi frekuensi digunakan untuk mengelompokkan data berdasarkan kategori atau label tertentu, lalu menghitung berapa kali setiap kategori muncul dalam data yang dikumpulkan.

Distribusi frekuensi pada data kualitatif berguna untuk memberikan gambaran yang lebih jelas tentang bagaimana data tersebar di berbagai kategori. Tanpa distribusi ini, akan sulit untuk melihat pola atau tren dalam data yang mungkin berisi banyak kategori berbeda. Dengan distribusi frekuensi, kita dapat dengan cepat melihat kategori mana yang paling dominan atau mana yang kurang umum. Misalnya, dalam survei tentang preferensi warna, distribusi frekuensi akan menunjukkan berapa banyak orang yang menyukai warna tertentu.

Penelitian kualitatif biasanya mengumpulkan data melalui observasi, wawancara, dan dokumentasi. Selain itu, jangan mengabaikan kemungkinan menggunakan sumber informasi non-manusia, seperti dokumen dan rekaman yang tersedia. Menciptakan rapport, memilih informan, dan mencatat data dan informasi hasil pengumpulan data adalah beberapa aktivitas tambahan yang dilakukan selama proses pengumpulan data ini.

Distribusi Frekuensi Data Relatif

Distribusi frekuensi relatif menunjukkan proporsi data dalam suatu kelas interval. Ini ditemukan dengan membagi frekuensi dengan total data dari pengamatan atau observasi.

Tabel frekuensi relatif juga disebut sebagai "tabel persentase" karena frekuensi yang disajikan di sini adalah persenan, bukan frekuensi yang sebenarnya.

Ada dua hal yang harus diketahui untuk menghitung distribusi frekuensi data relatif:

- a. Jumlah kejadian untuk suatu kategori
- b. Jumlah total kejadian

Perhitungan frekuensi relatif mengubah hitungan menjadi persentase dengan mengambil hitungan jenis kejadian tertentu dengan membaginya dengan jumlah total pengamatan. Rumusnya adalah sebagai berikut:

$$\text{Frekuensi relatif} = \frac{\text{hitungan kejadian}}{\text{total data}} \times 100\%$$

Distribusi Frekuensi Data Kumulatif

Menurut KBBI 'Akumulatif' mempunyai arti sebagai berkenaan dengan sesuatu yang terkumpul, tertimbun, terhimpun. Distribusi Frekuensi Akumulatif merupakan konsep statistik yang mempunyai fungsi untuk menyajikan data-data dalam bentuk tabel ataupun grafik yang menunjukkan sebuah akumulasi frekuensi data dalam interval tertentu. Dalam distribusi frekuensi kumulatif, jumlah atau persentase data yang kurang dari atau sama dengan batas tertentu dari interval kategori yang telah ditentukan sebelumnya. Ini membantu memahami sebaran data secara lebih luas daripada hanya melihat distribusi frekuensi biasa. Jumlah data data yang berada jatuh ke setiap kelas atau interval disebut frekuensi absolut. Ini adalah ukuran seberapa sering nilai tertentu muncul dalam data.

Cara menghitung frekuensi kumulatif:

a. Pengurutan Nilai Data

Mengurutkan data dari nilai-nilai yang terkecil hingga terbesar.

b. Menghitung Frekuensi Absolut

Hitunglah frekuensi absolut dari setiap nilai pada data, dimana Frekuensi absolut adalah jumlah angka yang dapat menunjukkan sejumlah besar data untuk kelompok tertentu.

c. Menghitung Frekuensi Kumulatif Nilai Pertama

Hitunglah frekuensi kumulatif untuk mendapatkan nilai pertama, dari nilai terkecil terlebih dahulu.

d. Menghitung Frekuensi Kumulatif Nilai Berikutnya

Carilah frekuensi kumulatif nilai selanjutnya lalu nilai frekuensi absolut dijumlahkan dengan nilai frekuensi kumulatif sebelumnya.

e. Proses diulang

Terus lakukan proses penjumlahan frekuensi kumulatif pada setiap nilai dari yang terkecil hingga terbesar sampai mendapatkan nilai terakhir.

f. Pemeriksaan Hasil Akhir

Nilai frekuensi kumulatif terakhir yang didapatkan seharusnya sama dengan hasil frekuensi absolut yang dijumlahkan.

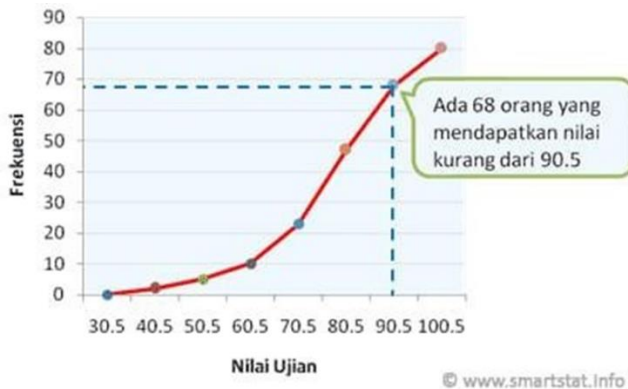
Distribusi Frekuensi Data Kumulatif mempunyai fungsi:

- a. Dalam analisis statistik, distribusi frekuensi kumulatif sering digunakan untuk menunjukkan sejauh mana data berkonsentrasi pada nilai-nilai tertentu dan untuk menemukan outlier atau nilai-nilai ekstrim.
- b. Menentukan frekuensi hingga ambang batas tertentu.
- c. Menemukan Outlier atau nilai nilai tertentu.

Jenis-Jenis Grafik dari Tabel Frekuensi

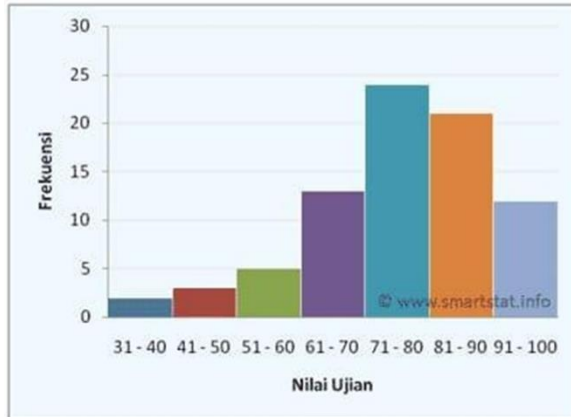
1. Kurva Ogive

Kurva ogive digambarkan sebagai diagram garis yang menunjukkan kombinasi antara frekuensi kumulatif dan interval kelas. Sumbu vertikal kurva menunjukkan frekuensi kumulatif, sedangkan sumbu horizontal menunjukkan tepi interval kelas. Kita dapat melihat frekuensi kumulatif sebagai nilai absolut dan nilai relatif pada tingkat atau interval tertentu dengan menggunakan kurva ogive.



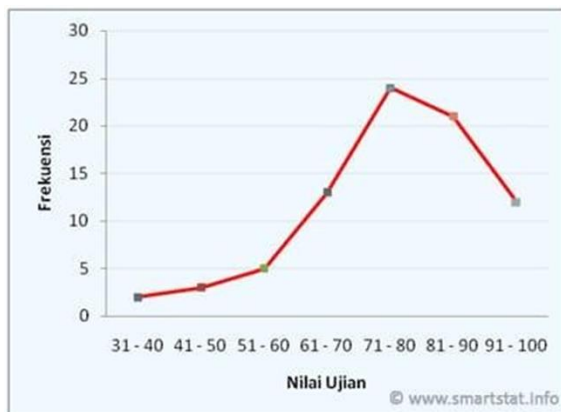
2. Grafik Histogram

Diagram balok yang digunakan untuk menunjukkan bentuk distribusi frekuensi disebut histogram. Histogram terhubung ke sumbu vertikal (Y) dan horizontal (X) interval kelas.



3. Polygon

Baik histogram maupun polygon sangat mirip, tetapi polygon memiliki garis yang menghubungkan koordinat antara nilai tengah kelas dan jumlah frekuensi kelas. Pada diagram polygon, sumbu horizontal merupakan nilai tengah kelas, dan sumbu vertikal merupakan jumlah frekuensi kelas. Selain itu, titik tengah kelas juga merupakan representasi dari karakter kelas, dan nilai tengah ini berfungsi sebagai pengganti posisi interval kelas pada diagram histogram.



2. Studi Kasus

a. Distribusi Frekuensi Kualitatif Studi Kasus 1

Data berikut ini merupakan data 25 orang yang membeli motor dari 5 jenis merek perusahaan otomotif.

Honda	Piaggio	Kawasaki	Honda	Suzuki
Honda	Yamaha	Suzuki	Yamaha	Piaggio
Suzuki	Yamaha	Honda	Kawasaki	Honda
Kawasaki	Kawasaki	Piaggio	Suzuki	Suzuki
Piaggio	Honda	Kawasaki	Yamaha	Piaggio

Perusahaan	Frekuensi
Yamaha	4
Honda	6
Piaggio	5
Kawasaki	5
Suzuki	5
Jumlah	25

Studi Kasus 2

Data Berikut Merupakan Data 30 Orang Yang Membeli Air Meneral Dari 3 Jenis Merk Air Meneral

Aqua	Le Meneral	Aqua	Vit	Aqua
Le Meneral	Vit	Aqua	Aqua	Vit
Le Meneral	Vit	Vit	Aqua	Aqua
Aqua	Aqua	Le Meneral	Le Meneral	Aqua
Aqua	Vit	Le Meneral	Vit	Aqua

Le Meneral	Aqua	Vit	Le Meneral	Vit
Merk Air Meneral			Frekuensi	
Aqua			13	
Le Meneral			8	
Vit			9	
Jumlah			30	

b. Distribusi Frekuensi Relatif Studi Kasus 1

Data hasil ujian akhir Mata Kuliah Statistik dari 60 orang mahasiswa FEB

36	82	70	52	74	67	32	79	60	23
76	55	85	92	65	41	95	81	77	80
67	81	98	80	25	78	75	64	10	52
43	74	62	88	72	64	54	83	71	41
79	15	90	84	48	84	76	89	78	60
61	85	80	63	74	69	82	17	67	34

Penyelesaian :

Urutkan Data

10 32 43 55 62 67 72 76 79 81 84 89
15 34 48 57 63 67 72 76 79 81 84 90
17 36 52 60 64 69 72 77 80 82 85 92
23 41 52 60 64 70 72 78 80 82 85 95
25 41 54 61 65 71 75 78 80 83 88 98

Jangkauan (R)

$$R = X \text{ max} - X \text{ min}$$

$$= 98 - 10$$

$$= 88$$

Banyak Kelas (K)

$$K = 1 + 3,322 \log n$$

$$K = 1 + 3,322 \log 60$$

$$= 6,8 \approx 7 \text{ kelas (pembulatan)}$$

Lebar Interval (I)

$$I = \frac{R}{K}$$

K

88

$$I = \frac{88}{7}$$

7

$$= 12,5 \approx 13 \text{ (pembulatan)}$$

Interval Kelas

$$Ba = Bb + (P - 1)$$

$$Ba = 10 + (13 - 1)$$

$$Ba = 22$$

Interval kelas 1 - 7

10 - 22 → Interval pertama

23 - 35 → Interval kedua

36 - 48 → Interval ketiga

49 - 61 → Interval keempat

62 - 74 → Interval kelima

75 - 87 → Interval keenam

88 - 100 → Interval ketujuh

Kelas	Interval	Frekuensi Absolut	Frekuensi Relatif
1	10 - 22	3	5%
2	23 - 35	4	6,67%
3	36 - 48	4	6,67%
4	49 - 61	8	13,33%
5	62 - 74	12	20%
6	75 - 87	23	38,33%

7	88 - 100	6	10%
Jumlah		60	100%

$$Frekuensi\ relatif = \frac{Frekuensi\ Absolut}{n\ (total\ data)} \times 100\%$$

$$Interval\ kelas\ pertama \rightarrow Frelatif = \frac{3}{60} \times 100\% = 5\%$$

$$Interval\ kelas\ kedua \rightarrow Frelatif = \frac{4}{60} \times 100\% = 6,67\%$$

$$Interval\ kelas\ ketiga \rightarrow Frelatif = \frac{4}{60} \times 100\% = 6,67\%$$

$$Interval\ kelas\ keempat \rightarrow Frelatif = \frac{8}{60} \times 100\% = 13,33\%$$

$$Interval\ kelas\ kelima \rightarrow Frelatif = \frac{14}{60} \times 100\% = 20\%$$

$$Interval\ kelas\ keenam \rightarrow Frelatif = \frac{23}{60} \times 100\% = 38,33\%$$

$$Interval\ kelas\ ketujuh \rightarrow Frelatif = \frac{8}{60} \times 100\% = 10\%$$

Studi Kasus 2

Pada hari sabtu, UKM Badminton melakukan pengukuran tinggi badan kepada 30 anggotanya, berikut data hasil pengukuran tinggi badan :

172	168	164	160	156	152
153	157	161	165	169	172
173	170	166	162	158	154
155	159	163	171	152	156
160	153	157	154	156	155

Penyelesaian :

- Urutkan Data

152	154	157	160	164	170
152	155	157	160	165	171
153	155	158	161	166	172
153	156	158	162	168	172
154	156	159	163	169	173

Jangkauan (R)

$$R = X \text{ max} - X \text{ min}$$

$$= 173 - 152$$

$$= 21$$

Banyak Kelas (K)

$$K = 1 + 3,322 \log n$$

$$K = 1 + 3,322 \log 30$$

$$= 5,90 \approx 6 \text{ kelas (pembulatan selalu keatas)}$$

Lebar Interval (I)

$$I = \frac{R}{K}$$

$$21$$

$$6$$

$$I = \frac{21}{6}$$

$$3,5$$

$$= 3,5 \approx 4 \text{ (pembulatan)}$$

Interval Kelas

$$Ba = Bb + (P - 1)$$

$$Ba = 152 + (4 - 1)$$

$$Ba = 155$$

Kelas	Interval Kelas	Frekuensi Absolut	Frekuensi Relatif
1	152 – 155	8	$\frac{8}{30} \times 100\% = 26,67\%$
2	156 – 159	7	$\frac{7}{30} \times 100\% = 23,33\%$
3	160 – 163	5	$\frac{5}{30} \times 100\% = 16,77\%$
4	164 – 167	3	$\frac{3}{30} \times 100\% = 10\%$
5	168 – 171	4	$\frac{4}{30} \times 100\% = 13,33\%$
6	172 – 175	3	$\frac{3}{30} \times 100\% = 10\%$
Jumlah		30	100%

Studi Kasus 3

Biaya pengeluaran iklan 40 pada perusahaan logistik tahun 2021

Kelas	Pengeluaran Iklan (Juta Rupiah)	Banyak Iklan
1	241 – 250	5
2	251 – 260	8
3	261– 270	10
4	271– 280	6
5	281 – 290	4
6	291 – 300	4
7	301 – 310	3
Total		40

Tabel kelas pertama (241 – 250) memiliki frekuensi kelas 5 dan jumlah frekuensi semua kelas adalah 40 ($n = 40$). Frekuensi Relatif kelas ini adalah $\frac{\text{Frekuensi Absolut}}{n \text{ (total data)}}$

100% . Dengan demikian, tabel Distribusi Frekuensi Relatif pada tabel berikut :

Kelas	Pengeluaran Iklan (Juta Rupiah)	Frekuensi Relatif
1	241 – 250	$\frac{5}{40} \times 100\% = 12,5\%$
2	251 – 260	$\frac{8}{40} \times 100\% = 20,0\%$
3	261– 270	$\frac{10}{40} \times 100\% = 25,0\%$
4	271– 280	$\frac{6}{40} \times 100\% = 15,0\%$
5	281 – 290	$\frac{4}{40} \times 100\% = 10,0\%$
6	291 – 300	$\frac{4}{40} \times 100\% = 10,0\%$
7	301 – 310	$\frac{3}{40} \times 100\% = 7,5\%$
Total		100%

c. Distribusi Frekuensi Kumulatif

Studi Kasus 1

Pada hari jumat siswa kelas 8B menjalankan ujian mata pelajaran Bahasa Indonesia, Berikut adalah data nilai ujian Bahasa Indonesia siswa siswi kelas 8B:

50	56	62	63	74	68	80	88	53	56
58	55	61	91	63	74	68	80	65	75
85	72	79	71	81	67	75	65	76	67
66	76	68	82	70	84	70	78	67	77
67	77	77	69	70	68	69	82	83	67

NO.	Nilai Ujian B. Indonesia	Frekuensi Absolut	Frekuensi Kumulatif
1	50-55	3	3
2	56-61	4	7
3	62-67	9	16
4	68-73	13	29
5	74-79	11	40
6	80-85	8	48
7	86-91	2	50

1) Nilai Pertama Frekuensi Kumulatif = 3

2) $4 + 3 = 7$

3) $9 + 7 = 16$

4) $13 + 16 = 29$

5) $11 + 29 = 40$

6) $8 + 40 = 48$

$$7) 2 + 48 = 50$$

Hasil Akhir Frekuensi Kumulatif Pada Data Sudah Di Dapatkan Karena Nilai Frekuensi Kumulatif yang dihasilkan Sama Dengan Jumlah Hasil Frekuensi Absolut ($3 + 4 + 9 + 13 + 11 + 8 + 2 = 50$)

Studi Kasus 2

Dina ingin membuka toko sendiri, lalu ia membeli banyak barang atau produk. Produk dengan harga yang berbeda beda. Berikut adalah list harga yang sudah Dina beli untuk dijual Kembali ditokonya.

86.000	53.000	37.000	21.000	33.000	7.000	9.000	40.000	55.000	69.000
10.000	34.000	40.000	71.000	90.000	22.000	42.000	77.000	22.000	15.000
34.000	16.000	44.000	35.000	55.000	80.000	56.000	20.000	51.000	23.000
18.000	25.000	26.000	36.000	44.000	59.000	96.000	58.000	100.000	60.000
92.000	60.000	46.000	26.000	50.000	35.000	27.000	29.000	30.000	68.000
27.000	32.000	95.000	65.000	50.000					

NO	List Harga	Frekuensi Absolut	Frekuensi Kumulatif
1.	Rp5.000 - Rp20.000	7	7
2.	Rp21.000 - Rp36.000	18	25
3.	Rp37.000 - Rp52.000	10	35
4.	Rp53.000 - Rp68.000	10	45
5.	Rp69.000 - Rp84.000	4	49
6.	Rp85.000 - Rp100.000	6	55
TOTAL		55	

Mencari Nilai Frekuensi Akumulatif Setiap Nilai

1. Frekuensi Akumulatif pertama = Frekuensi Absolut pertama, yaitu 7
2. $18 + 7 = 25$
3. $10 + 25 = 35$
4. $10 + 35 = 45$
5. $4 + 45 = 49$
6. $6 + 49 = 55$

Hasil akhir Frekuensi Akumulatif sama dengan jumlah Frekuensi Absolut yaitu 55

Studi Kasus 3

PT. Sepatu Bata melakukan produksi sepatu terbaru dengan ukuran yang berbeda - beda. Berikut merupakan data ukuran sepatu produksi terbaru PT. Sepatu Bata;

23	27	31	35	39	23	27	31	42	38
30	26	33	36	39	33	28	23	31	35
37	41	38	32	28	24	34	36	40	38
25	28	32	35	39	32	29	25	38	42
41	26	30	32	34	39	35	38	37	34
40	37	35	42	25	29	33	31	36	40
26	30	41	36	40	37	41	36	37	33

NO	UKURAN SEPATU	FREK. ABSOLUT	FREK. KUMULATIF
1.	23 - 26	10	10
2.	27 - 30	10	20
3.	31 - 34	15	35
4.	35 - 38	20	55
5	39 - 42	15	70
	TOTAL	70	

Menghitung Frekuensi Kumulatif setiap data

1. Frekuensi Kumulatif pertama = Frekuensi Absolut pertama yaitu 10
2. Frekuensi Kumulatif data 27 - 30: $10 + 10 = 20$
3. Frekuensi Kumulatif data 31 - 34: $15 + 20 = 35$
4. Frekuensi Kumulatif data 35 - 38: $20 + 35 = 55$
5. Frekuensi Kumulatif data 39 - 42: $15 + 55 = 70$

Hasil akhir Frekuensi Kumulatif 70 sama dengan jumlah total dari Frekuensi Absolut yaitu $(10+10+15+20+15 = 70)$

Distribusi Frekuensi dengan Menggunakan Grafik Studi Kasus 1
Diketahui data sebagai berikut

18	26	31	36	30	19	35	40	23	41
26	34	27	40	38	30	46	26	38	22
32	17	37	26	33	35	31	27	42	43
48	34	31	37	31	17	22	23	40	35
16	22	38	36	35	24	39	22	44	34
32	36	38	35	31	37	21	23	26	34
23	32	32	29	50	29	36	23	45	45
19	28	37	32	38	50	27	25	30	37

Dari data di atas tentukan:

1. Tabel distribusi frekuensi
2. Tabel distribusi frekuensi relatif
3. Grafik histogram
4. Grafik polygon
5. Tabel distribusi frekuensi kumulatif kurang dari dan lebih dari
6. Grafik ogive kurang dari dan lebih dari

Jawab:

1. Langkah-langkah:

a. Mengurutkan data

16	17	17	18	19	19	21	22	22	22
22	23	23	23	23	24	25	26	26	26
26	26	27	27	27	28	29	29	30	30
30	31	31	31	31	31	32	32	32	32
32	33	33	34	34	34	34	35	35	35
35	35	36	36	36	36	37	37	37	37
37	38	38	38	38	38	39	40	40	40
41	42	43	44	45	45	46	48	50	50

b. Range atau jangkauan

$$R = X \max - X \min$$

$$R = 50 - 16 = 34$$

c. Banyak kelas

$$K = 1 + 3,322 \log n$$

$$K = 1 + 3,322 \log 80$$

$$= 7,32 \approx 7 \text{ kelas (dibulatkan ke bawah)}$$

d. Lebar interval kelas

$$I = \frac{R}{K}$$

$$I = \frac{34}{7}$$

$$= 4,8 \approx 5$$

e. Interval kelas

$$Ba = Bb + (P - 1)$$

$$Ba = 16 + (5 - 1)$$

$$Ba = 20$$

- Interval kelas 1-7
- Interval kelas ke - 1 → 16 - 20
- Interval kelas ke - 2 → 21 - 25
- Interval kelas ke - 3 → 26 - 30
- Interval kelas ke - 4 → 31 - 35
- Interval kelas ke - 5 → 36 - 40
- Interval kelas ke - 6 → 41 - 45
- Interval kelas ke - 7 → 46 - 50

f. Tabel distribusi frekuensi

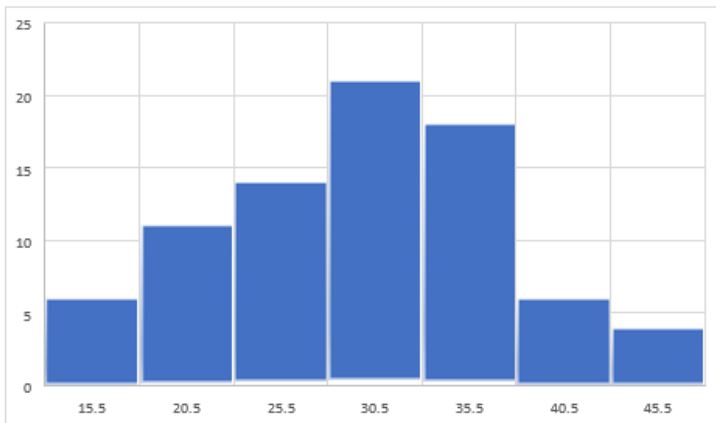
Kelas	Interval kelas	Frekuensi
1	16 - 20	6
2	21 - 25	11
3	26 - 30	14
4	31 - 35	21
5	36 - 40	18
6	41 - 45	6
7	46 - 50	4

2. Tabel distribusi frekuensi relatif

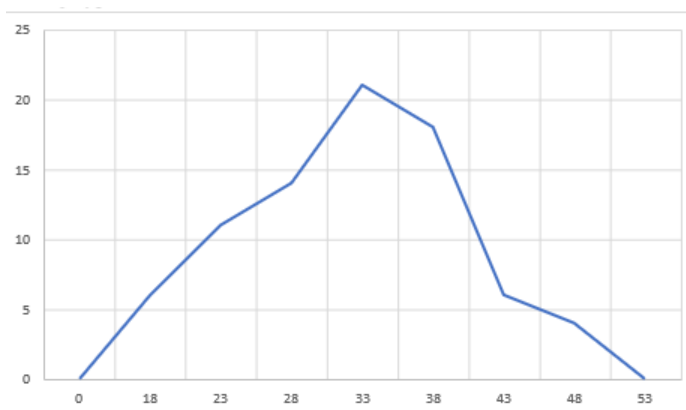
Kelas	Interval Kelas	Frekuensi Absolut	Frekuensi Relatif
1	16 - 20	6	$\frac{6}{80} \times 100\% = 7,5\%$
2	21 - 25	11	$\frac{11}{80} \times 100\% = 13,75\%$
3	26 - 30	14	$\frac{14}{80} \times 100\% = 17,5\%$
4	31 - 35	21	$\frac{21}{80} \times 100\% = 26,25\%$
5	36 - 40	18	$\frac{18}{80} \times 100\% = 22,5\%$
6	41 - 45	6	$\frac{6}{80} \times 100\% = 7,5\%$
7	46 - 50	4	$\frac{4}{80} \times 100\% = 5\%$
Jumlah		80	100%

3. Grafik histogram

Interval Kelas	Tepi Kelas	Titik Tengah	Frekuensi Absolut
16 – 20	15,5 – 20,5	18	6
21 – 25	20,5 – 25,5	23	11
26 – 30	25,5 – 30,5	28	14
31 – 35	30,5 – 35,5	33	21
36 – 40	35,5 – 40,5	38	18
41 – 45	40,5 – 45,5	43	6
46 – 50	45,5 – 50,5	48	4



4. Grafik polygon

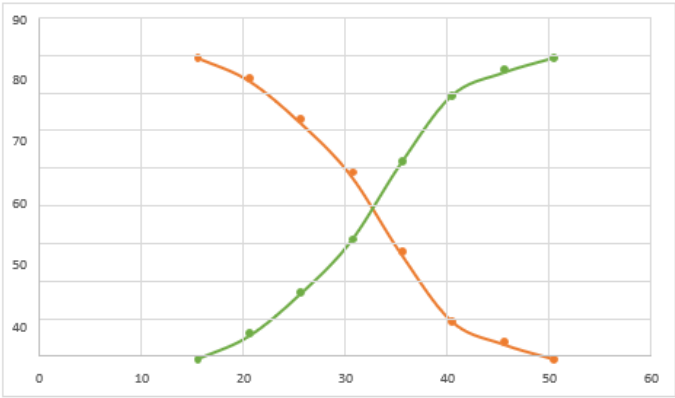


5. Tabel distribusi frekuensi kumulatif kurang dari dan lebih dari

Interval Kelas	Tepi Kelas	Frekuensi Absolut	Tepi Kelas Bawah	Frekuensi Kumulatif Kurang Dari ($f_k <$)
16 - 20	15,5 - 20,5	6	15,5	0
21 - 25	20,5 - 25,5	11	20,5	6
26 - 30	25,5 - 30,5	14	25,5	17
31 - 35	30,5 - 35,5	21	30,5	31
36 - 40	35,5 - 40,5	18	35,5	52
41 - 45	40,5 - 45,5	6	40,5	70
46 - 50	45,5 - 50,5	4	45,5	76
			50,5	80

Interval Kelas	Tepi Kelas	Frekuensi Absolut	Tepi Kelas Bawah	Frekuensi Kumulatif Lebih Dari ($f_k >$)
16 - 20	15,5 - 20,5	6	15,5	80
21 - 25	20,5 - 25,5	11	20,5	74
26 - 30	25,5 - 30,5	14	25,5	63
31 - 35	30,5 - 35,5	21	30,5	49
36 - 40	35,5 - 40,5	18	35,5	28
41 - 45	40,5 - 45,5	6	40,5	10
46 - 50	45,5 - 50,5	4	45,5	4
			50,5	0

6. Grafik ogive kurang dari dan lebih dari



G. Kesimpulan

Pengelompokan data ke dalam beberapa kategori, yang menunjukkan jumlah data yang ada di dalamnya, disebut juga Distribusi frekuensi. Distribusi frekuensi juga mengacu kepada keadaan yang menunjukkan bagaimana frekuensi gejala suatu variabel yang diwakili oleh angka telah tersalur, terbagi, atau terpecah.

Tabel distribusi frekuensi adalah salah satu alat yang dapat digunakan untuk menampilkan data statistik. Data ini dapat disajikan dalam bentuk grafik, yang menampilkan berbagai jenis macam dan membuatnya lebih menarik.

Dengan demikian, distribusi frekuensi adalah alat yang tidak ternilai dalam statistika ekonomi bisnis, esensial untuk analisis data, perencanaan strategis, dan pengambilan keputusan. Penggunaan efektif dari distribusi frekuensi dapat memberikan keuntungan kompetitif yang signifikan, mendukung bisnis dalam menghadapi tantangan pasar yang dinamis dan meningkatkan efektivitas operasional.

Pertanyaan :

Soal 1

Distribusi Frekuensi Kualitatif

Data berikut adalah jenis motor yang dibeli oleh 25 orang dari lima merek yang berbeda:

Honda	Piaggio	Kawasaki	Honda	Suzuki
Honda	Yamaha	Suzuki	Yamaha	Piaggio
Suzuki	Yamaha	Honda	Kawasaki	Honda
Kawasaki	Kawasaki	Piaggio	Suzuki	Suzuki
Piaggio	Honda	Kawasaki	Yamaha	Piaggio

Pertanyaan:

Buatlah tabel distribusi frekuensi kualitatif dari data di atas!

Jawaban:

Merek Motor	Frekuensi
Yamaha	4
Honda	6
Piaggio	5
Kawasaki	5
Suzuki	5
Total	25

Soal 2 -

Distribusi Frekuensi Relatif

Diberikan data tinggi badan 30 anggota UKM Badminton (dalam cm):

152, 154, 157, 160, 164, 170, 152, 155, 157, 160, 153, 155, 158, 161, 166, 172, 153, 156, 158, 162, 168, 172, 154, 156, 159, 163, 171, 155, 156, 155.

Pertanyaan:

Buatlah tabel distribusi frekuensi relatif dengan 6 kelas interval!

Jawaban:

1. Rentang (R) = $173 - 152 = 21$
2. Banyak Kelas (K) = 6
3. Lebar Interval (I) = $21 \div 6 \approx 4 \rightarrow 4$
4. Interval Kelas:
 - 152 - 155
 - 156 - 159
 - 160 - 163
 - 164 - 167
 - 168 - 171
 - 172 - 175

Kelas	Interval (cm)	Frekuensi Absolut	Frekuensi Relatif
1	152 - 155	8	26,67%
2	156 - 159	7	23,33%
3	160 - 163	5	16,67%
4	164 - 167	3	10,00%
5	168 - 171	4	13,33%
6	172 - 175	3	10,00%
Total		30	100%

Soal 3 -

Distribusi Frekuensi Kumulatif

Data berikut adalah ukuran sepatu dari 70 pasang sepatu produksi PT. Sepatu Bata:

NO	Ukuran Sepatu	Frekuensi Absolut
1	23 - 26	10
2	27 - 30	10
3	31 - 34	15
4	35 - 38	20
5	39 - 42	15

Pertanyaan:

Buatlah tabel distribusi frekuensi kumulatif "kurang dari" untuk data di atas!

Jawaban:

NO	Ukuran Sepatu	Frekuensi Absolut	Frekuensi Kumulatif (fk<)
1	< 27	10	10
2	< 31	10	20
3	< 35	15	35

NO	Ukuran Sepatu	Frekuensi Absolut	Frekuensi Kumulatif (fk<)
4	< 39	20	55
5	< 43	15	70

Keterangan:

Frekuensi kumulatif dihitung dengan menjumlahkan frekuensi absolut secara bertahap dari kelas pertama hingga kelas yang bersangkutan.

DAFTAR PUSTAKA

- Banahsan, N. (2013, 9). *Distribusi Frekuensi dan Grafik*. Retrieved from wordpress:
<https://nailabanahsan14.wordpress.com/about/distribusi-frekuensi-dan-grafik/>
- Dr. Bambang Widjanarto Otok, M., & Dewi Juliah Ratnaningsih, M. (n.d.). *Konsep Dasar dalam Pengumpulan dan Penyajian Data*. *pustaka.ut.ac.id*.
- Habib, F. K. (n.d.). *Tabel Distribusi Frekuensi, Grafik Histogram, Grafik Polygon, Grafik Ogive, dan Frekuensi Relatif dan Kumulatif*. Retrieved from id.scribd:
<https://id.scribd.com/doc/178683790/Tabel-Distribusi-Frekuensi-Grafik-Histogram-Grafik-Polygon-Grafik-Ogive-Dan-Frekuensi-Relative-Dan-Kumulatif>
- Hafidz. (2016, 3 20). *Relatif Dan Absolut*. Retrieved from scribd:
<https://id.scribd.com/doc/305380543/Relatif-Dan-Absolut>
- Informatika, B. S. (2022, 10 17). *Makalah Statistika Distribusi frekuensi*. Retrieved from studocu:
<https://www.studocu.com/id/document/universitas-bina-sarana-informatika/sistem-informasi/makalah-statistika-distribusi-frekuensi/40064133>
- Kumparan. (2022, 1 19). *5 Cara Mengumpulkan Data dalam Penelitian Kualitatif*. Retrieved from kumparan.com:
<https://m.kumparan.com/amp/kabar-harian/5-cara-mengumpulkan-data-dalam-penelitian-kualitatif-1xKwzMspFeV>
- Kumparan. (2023, 11 15). *Cara Mencari Distribusi Frekuensi Kumulatif*. Retrieved from Kumparan: <https://kumparan.com/berita-hari-ini/cara-mencari-frekuensi-kumulatif-dan-contoh-soalnya-21ZpSzOO2jz>
- Penyajian Data* . (2016, 9 5). Retrieved from edcyclopedia.com:
<https://edscyclopedia.com/penyajian-data-3/>
- Prof. Dr. H. Mudjia Rahardjo, M. (2011, 6 10). *Metode Pengumpulan Data Kualitatif*.

Retrieved from uin-malang.ac.id: <https://uin-malang.ac.id/r/110601/metode-pengumpulan-data-penelitian-kualitatif.html>

Setiawan, A. (2023, 8 2). *Tabel Distribusi, Histogram, dan Poligon Frekuensi*. Retrieved from smartstat.info: <https://www.smartstat.info/materi/statistika/statistik-deskriptif/distribusi-frekuensi.html>

BAB 3

UKURAN-UKURAN PEMUSATAN DATA YANG TIDAK BERKELOMPOK DAN DATA YANG BERKELOMPOK

Dalam dunia statistika, salah satu pertanyaan paling mendasar yang sering diajukan adalah: “Di mana pusat data ini berada?” Jawaban atas pertanyaan ini tidak hanya membantu kita memahami karakteristik umum suatu kumpulan data, tetapi juga menjadi fondasi bagi analisis lebih lanjut mulai dari peramalan, pengambilan keputusan, hingga evaluasi kinerja.

Untuk menjawab pertanyaan tersebut, para ahli statistik mengembangkan apa yang disebut ukuran pemusatan data (measures of central tendency). Ukuran ini memberikan nilai tunggal yang merepresentasikan “titik tengah” atau “nilai tipikal” dari seluruh observasi dalam suatu dataset. Tiga ukuran utama yang paling sering digunakan adalah mean (rata-rata), median (nilai tengah), dan modus (nilai yang paling sering muncul).

Namun, cara menghitung ketiga ukuran ini tidak selalu sama. Perbedaan utama terletak pada bentuk data yang tersedia:

1. Data tidak berkelompok (ungrouped data): data mentah yang masih dalam bentuk individual, belum dikelompokkan ke dalam interval atau kategori.
2. Data berkelompok (grouped data): data yang telah diorganisasi ke dalam distribusi frekuensi, biasanya dalam bentuk interval kelas.

Meskipun konsep dasarnya sama, metodologi perhitungan, asumsi, dan interpretasi untuk kedua jenis data ini memiliki perbedaan signifikan. Mengabaikan perbedaan ini dapat menyebabkan kesalahan analisis yang berdampak pada keputusan bisnis, kebijakan publik, atau penelitian ilmiah.

Laporan ini bertujuan memberikan pemahaman komprehensif, sistematis, dan aplikatif tentang ukuran-ukuran pemusatan data, baik untuk data tidak berkelompok maupun berkelompok. Kami akan membahas:

1. Konsep Dasar Ukuran Pemusatan Data
2. Mean: Definisi, Rumus, Kelebihan, dan Kekurangan
3. Median: Konsep, Perhitungan, dan Ketahanan terhadap Outlier
4. Modus: Identifikasi dan Interpretasi dalam Berbagai Konteks
5. Perbandingan Mean, Median, dan Modus dalam Berbagai Bentuk Distribusi
6. Perhitungan untuk Data Tidak Berkelompok (Disertai Contoh Numerik)
7. Perhitungan untuk Data Berkelompok (Langkah demi Langkah)
8. Studi Kasus 1: Analisis Gaji Karyawan (Data Tidak Berkelompok)
9. Studi Kasus 2: Evaluasi Nilai Ujian Nasional (Data Berkelompok)
10. Studi Kasus 3: Survei Usia Pelanggan di E-commerce
11. Kesalahan Umum dan Cara Menghindarinya
12. Aplikasi Praktis dalam Dunia Nyata
13. Implementasi dengan Excel dan Python

Ukuran pemusatan data adalah nilai statistik yang menunjukkan lokasi pusat dari suatu distribusi data. Tujuan utamanya adalah menyederhanakan informasi: alih-alih melihat ratusan atau ribuan angka, kita cukup melihat satu angka representatif.

Namun, penting dipahami bahwa tidak ada satu ukuran pemusatan yang universal. Pilihan antara mean, median, atau modus tergantung pada:

1. Jenis data (numerik vs. kategorikal)
2. Bentuk distribusi (simetris vs. menceng)
3. Tujuan analisis
4. Kehadiran outlier (nilai ekstrem)

Sebagai analogi: jika Anda ingin tahu “di mana letak pusat kota Jakarta”, jawabannya bisa berbeda tergantung konteks:

1. Secara geografis → titik koordinat tengah (mirip mean)
2. Secara populasi → wilayah dengan kepadatan tertinggi (mirip modus)
3. Secara aksesibilitas → lokasi yang membagi kota menjadi dua bagian sama jauh (mirip median)

Ukuran sentralitas data adalah konsep penting dalam statistik yang sering digunakan saat menganalisis sekumpulan data untuk menentukan karakteristiknya. Nilai rata-rata kumpulan data ditampilkan melalui pemusatan data. Dua kategori utama tindakan pemusatan data adalah tindakan sentralisasi untuk data yang tidak dikelompokkan dan tindakan sentralisasi untuk data yang dikelompokkan. Karena ada banyak metode pemrosesan dan interpretasi data yang berbeda, pendidik dan peneliti harus memahami kedua kategori metrik konsentrasi ini.

Data terpilah tersedia secara terpisah dan tidak dikategorikan ke dalam kelas tertentu, sehingga memudahkan penghitungan dan evaluasi ukuran sentralitas. Mean, median, dan mode adalah contoh ukuran sentralisasi yang umum diterapkan pada data terpilah. Nilai median dan yang paling sering muncul dalam suatu data dapat dicari dengan menggunakan median dan modus, sedangkan mean digunakan untuk menghitung mean aritmatika dari semua data. Ketiga pengukuran ini sering digabungkan untuk mengkarakterisasi sebaran data. Sebaliknya, data yang dikelompokkan dipecah menjadi.

Ada beberapa kelas, masing-masing dengan nilai berbeda. Karena sejumlah asumsi harus dibuat, terutama ketika mencari tahu titik tengah kelas mana yang paling menggambarkan data, pemusatan metrik data berkelompok menjadi sedikit lebih rumit. Mean, median, dan modus adalah ukuran pemusatan lainnya yang sering digunakan dalam data berkelompok. Namun, penghitungan data kelompok memerlukan metode yang lebih kuantitatif, seperti memanfaatkan frekuensi untuk menghitung mean atau mode menggunakan rumus yang lebih tepat. Memahami perbedaan kedua kategori data ini satu sama lain sangat penting untuk lebih memahami distribusi data. Misalnya, data yang tidak dikelompokkan memudahkan untuk melihat nilai data individual, sedangkan data yang dikelompokkan mempersulit untuk melihat pola distribusi dan dapat menyembunyikan detail setiap angka. Untuk menjaga relevansi dan keterwakilan hasil analisis Anda, Anda harus menggunakan metode estimasi saat memproses data kelompok.

A. Memahami Pengukuran Untuk Pusat Data

Lebih mudah untuk menghitung dan mengevaluasi ukuran sentralitas karena data yang tidak dikelompokkan tersedia secara individual dan tidak diklasifikasikan ke dalam kelas tertentu. Ukuran sentralisasi yang sering digunakan pada data yang tidak dikelompokkan adalah mean, median, dan mode. Nilai tengah dan nilai yang paling sering muncul pada data dapat dicari dengan menggunakan median dan modus, sedangkan mean digunakan untuk menghitung rata-rata aritmatika seluruh data. Untuk mengkarakterisasi distribusi data, ketiga metrik ini sering digabungkan. Sebaliknya, data yang terkumpul dibagi menjadi beberapa kelas, masing-masing dengan rentang nilai. Pemusatan metrik data berkelompok menjadi sedikit lebih rumit karena banyak asumsi harus dibuat, terutama ketika mencari tahu titik tengah kelas mana yang paling baik menggambarkan data. Modus, median, dan mean adalah ukuran pemusatan lainnya yang sering digunakan.

Median, mean, dan mode adalah tiga metrik sentralitas yang paling sering digunakan. Semua pendekatan sentralisasi data ini memiliki kekurangan. Nilai pusat akan sangat dipengaruhi oleh nilai outlier. Untuk parameter populasi, mediannya sangat tidak pasti. Namun, opsi ini hanya dapat digunakan untuk dataset berukuran besar. Perbedaan utama antara ukuran pemusatan data yang terklaster dan tidak terklaster akan dibahas dalam penelitian ini.

1. Definisi Data yang Tidak Berkelompok

Data yang belum dikategorikan atau terstruktur disebut data tidak dikelompokkan atau data mentah. Contoh data tersebut antara lain tinggi badan, nilai ujian dari siswa, dan data lain yang diukur secara langsung, seperti 0-20, 20-40, dan seterusnya. Rata-rata, median, dan modus sering kali digunakan sebagai metrik untuk menentukan sentralitas data ini.

2. Definisi Data yang Berkelompok

Sebaliknya, data berkelompok adalah data yang disusun ke dalam kategori atau interval tertentu. Misalnya, tinggi badan dapat dikelompokkan ke dalam kategori mulai dari 150-160 cm hingga 161-170 cm. Untuk data berkelompok ini, ukuran pemusatan yang digunakan adalah median dan rata-rata kelompok.

3. Mean Pada Data Tidak Berkelompok

Untuk menghasilkan ringkasan nilai rata-rata dari kumpulan data yang tidak dikelompokkan, jumlahkan semua nilai dan bagi totalnya dengan jumlah titik data. Nilai tengah, bagaimanapun, dapat dipengaruhi oleh nilai ekstrem.

4. Mean Pada Data Berkelompok

Sebaliknya, data berkelompok adalah data yang disusun ke dalam kategori atau interval tertentu. Misalnya, tinggi badan dapat dikelompokkan ke dalam kategori seperti 150-160 cm atau 161-170 cm. Untuk data berkelompok, ukuran pemusatan yang digunakan adalah median dan rata-rata kelompok.

Median adalah nilai tengah dari kumpulan data yang tidak dikelompokkan. Sebelum data diurutkan, median dapat ditemukan. Ini sangat bagus untuk melihat pusat data, terutama ketika distribusi data asimetris.

5. Median Pada Data Berkelompok

Untuk data berkelompok, median dihitung dengan menggunakan frekuensi kumulatif. Ini melibatkan menentukan kelas median berdasarkan frekuensi, kemudian menghitung nilai median dengan mempertimbangkan panjang interval. Metode ini mungkin lebih kompleks dibandingkan dengan metode yang digunakan untuk data tidak berkelompok.

6. Modus Pada Data Tidak Berkelompok

Nilai yang paling sering muncul dalam data tidak berkelompok yang dikenal sebagai modus. Modus ini mencerminkan nilai yang paling umum dalam kumpulan data tersebut. Namun, penting untuk dicatat bahwa data tidak berkelompok bisa memiliki lebih dari satu modus atau bahkan tidak memiliki modus sama sekali.

7. Modus Pada Data Berkelompok

Nilai yang paling sering ditemukan dalam kelompok data dapat dihitung dengan menghitung modus kelas modal, yang merupakan kelas dengan frekuensi tertinggi..

8. Keuntungan Data Berkelompok

Data yang dikelompokkan memungkinkan analisis yang lebih mudah dan lebih terstruktur. Peneliti dapat melihat distribusi dan tren dalam data dengan cepat dan membuat visualisasi yang lebih informatif.

9. Kekurangan Data Berkelompok

Namun, pengelompokan data dapat mengaburkan detail penting, menyebabkan kesalahpahaman dalam analisis, terutama jika interval yang tidak sesuai digunakan.

10. Kegunaan dalam Analisis Statistik

Data tidak berkelompok umumnya digunakan ketika ukuran sampel kecil atau ketika detail nilai individu penting untuk dianalisis. Di sisi lain, data berkelompok lebih sering digunakan ketika ukuran sampel besar dan analisis perlu disederhanakan, misalnya dalam survei populasi atau penelitian pasar. Meskipun ada kehilangan detail dalam hal ini, metode ini membantu menyajikan data dengan lebih ringkas.

11. Contoh Kasus Penggunaan Data Tidak Berkelompok

Misalnya, dalam penelitian pendidikan, peneliti dapat menggunakan data tidak berkelompok untuk menganalisis nilai ujian siswa dengan lebih akurat, melihat nilai individu, dan menemukan area yang perlu diperbaiki.

12. Contoh Kasus Penggunaan Data Berkelompok

Sebaliknya, dalam penelitian epidemiologi, data berkelompok dapat digunakan untuk menganalisis distribusi penyakit dalam kategori usia, yang memudahkan pengenalan kelompok yang lebih rentan.

13. Fleksibilitas perhitungan

Menghitung data tidak fleksibel karena berbagai metode dapat digunakan tanpa kehilangan informasi. Sebaliknya, perhitungan data grup perlu dianggap sebagai data yang dirangkum. Fleksibilitas ini menyederhanakan analisis

terperinci data yang tidak dikelompokkan, terutama ketika diperlukan beberapa pengukuran pemusatan.

14. Kompleksitas Perhitungan

Perhitungan ukuran pemusatan pada data tidak berkelompok relatif lebih sederhana karena langsung bekerja dengan nilai asli dari data. Sementara itu, pada data berkelompok, perhitungan seringkali membutuhkan langkah tambahan seperti menentukan titik tengah kelas dan menghitung frekuensi kumulatif, sehingga lebih kompleks dan memakan waktu.

B. Macam-Macam Ukuran Pemusatan Data

1. Macam-Macam Ukuran Pemusatan

Dalam statistik, pengukuran konsentrasi data adalah nilai yang digunakan untuk mewakili pengumpulan data dalam bentuk angka yang dapat memberikan gambaran umum pusat data. Pengukuran konsentrasi ini penting untuk analisis data karena membantu untuk menyimpulkan karakteristik umum kelompok induk atau sampel tanpa melihat semua data. Ukuran sentralitas yang paling umum digunakan meliputi mean, median, dan mode.

a. Mean (rata-rata)

Konsentrasi data yang paling banyak digunakan adalah rata-rata (rata-rata). Melipat setiap nilai dalam pengumpulan data dan membaginya menjadi total data, Anda dapat menentukan rata-rata. Meskipun rata-rata mudah dihitung, rata-rata tersebut tidak selalu mewakili pusat data secara akurat karena pengaruh hasil ekstrem dan outlier. Misalnya, nilai rata-rata mungkin tidak secara akurat mewakili nilai-nilai lain dalam kumpulan yang memiliki nilai sangat tinggi atau rendah. Di beberapa bidang ilmu pengetahuan, termasuk biologi, ekonomi, dan ilmu sosial, mean sering kali digunakan untuk menghitung rata-rata berbagai peristiwa atau hasil.

b. Median

Nilai median setelah data diurutkan diwakili oleh metrik yang berpusat pada data, yaitu median. Jika banyaknya data ganjil maka nilai pusatnya disebut median. Median adalah rata-rata dari dua mean jika jumlah titik datanya genap. Median kurang sensitif terhadap outlier dibandingkan mean, sehingga sering digunakan ketika data mungkin berisi nilai ekstrem. Misalnya, untuk memberikan gambaran distribusi pendapatan yang lebih akurat, analisis pendapatan sering kali menggunakan median.

c. Modus

Nilai yang paling sering muncul dalam kumpulan data ditunjukkan oleh mode, yang merupakan metrik pusat data. Suatu kumpulan data mungkin berisi mode nol, mode bimodal atau multimodal, atau hanya satu mode (unimodal). Saat menganalisis data kategorikal, mode ini membantu mengidentifikasi kategori yang paling umum. Misalnya, mode survei kepuasan pelanggan mungkin menunjukkan produk atau layanan mana yang paling sering disebutkan. .

d. Perbandingan Mean, median, Modus

Ketiga ukuran pemusatan ini, mean, median, dan modus, memiliki peran yang berbeda dalam analisis data. Mean memberikan gambaran umum secara kuantitatif, namun sensitif terhadap outliers. Jika datanya asimetris atau terdapat outlier, median dapat memberikan gambaran yang lebih akurat. Kondisi tersebut juga berguna ketika kita ingin mengetahui nilai yang paling sering muncul dalam suatu data. Pilihan penggunaan terpusat tergantung pada jenis data dan tujuan analisis.

e. Peran Ukuran Pemusatan dalam Analisis Data

Pengukuran konsentrasi data memberikan pedoman penting untuk memahami dan menganalisis pengumpulan data. Melalui skala terpusat ini, para peneliti dan analis data dapat memahami foto umum tentang bagaimana data menyebar di sekitar median. Hal ini sangat penting dalam berbagai bidang penelitian dan bisnis untuk mengambil

keputusan yang lebih tepat berdasarkan informasi statistik terkini.

f. Kelebihan dan Kekurangan Ukuran Pemusatan

Pengukuran pemusatan berguna dalam banyak situasi, namun setiap pengukuran memiliki kekuatan dan kelemahannya sendiri. Misalnya, nilai rata-rata memberikan ringkasan data, namun sangat dipengaruhi oleh outlier. Sebaliknya, median lebih kuat terhadap outlier, namun tidak memberikan informasi tentang distribusi data. Mode tersebut mungkin tidak selalu memberikan informasi penting jika datanya kontinu atau jika nilai yang sering hilang.

g. Contoh Kasus

Misalnya, dalam survei pengeluaran rumah tangga, nilai rata-rata mungkin memberikan gambaran kasar tentang pengeluaran rata-rata, namun jika sejumlah kecil rumah tangga memiliki pengeluaran yang sangat tinggi, nilai median mungkin lebih akurat tentang pengeluaran sejumlah besar rumah tangga. Di sisi lain, moda dapat menunjukkan nilai pembelanjaan yang paling sering dilakukan, sehingga memberikan informasi tambahan mengenai perilaku konsumsi.

Berikutnya adalah contoh studi kasus mengenai mean, median dan modus dengan menggunakan angka.

Studi kasus : pengujian nilai ujian siswa

Sejumlah siswa mengikuti ujian matematika dan berikut adalah nilai yang mereka peroleh :

85, 90, 75, 85, 95, 80, 85, 70, 90, 100

1. Mean (rata rata)

Untuk menghitung mean, kita harus jumlahkan semua nilai dan bagi dengan jumlah siswa.

$$\text{Mean} = \frac{(85+90+75+85+95+80+85+70+90+100)}{10}$$

$$\text{Mean} = \frac{85+90+75+85+95+80+85+70+90+100}{10} = 92,5$$

2. Median (nilai tengah)

Untuk mencari median, kita urutkan nilai dari yang terkecil ke yang terbesar :

70, 75, 80, 85, 85, 90, 90, 95, 100

Karena jumlah data genap (10 data), median adalah rata-rata dari dua nilai tengah (nilai ke-5 dan ke-6).

$$\text{Median} = \frac{85 + 85}{2} = \frac{170}{2} = 85$$

3. Modus (Nilai yang Paling Sering Muncul)

Modus adalah nilai yang paling sering muncul dalam data. Dari daftar nilai, kita lihat :

85 muncul 3 kali

•70, 75, 80, 90, 95, dan 100 masing-masing muncul 1 kali.

Jadi, modus dari data ini adalah 85.

Ringkasan :

Mean: 92.5

Median: 85

Modus: 85

Dengan analisis ini, kita dapat memahami sebaran nilai ujian siswa dan menarik kesimpulan yang lebih dalam mengenai hasil belajar mereka.

Aplikasi Ukuran Pemusatan di Berbagai Bidang

Banyak industri menggunakan ukuran pemusatan data. Dalam ilmu sosial, median dan median sering digunakan untuk

memahami persebaran pendapatan atau preferensi dalam survei; dalam bidang ekonomi, mean biasanya digunakan untuk mengukur pendapatan rata-rata suatu populasi. Biologi menggunakan ukuran pemusatan ini untuk menganalisis statistik data eksperimen.

C. Cara Penghitungan Pemusatan Data

Kami biasanya menggunakan ukuran seperti mean, median, dan mode untuk menghitung sentralitas data. Sentralisasi data dapat dilakukan pada data yang terkluster dan tidak terkluster. Berikut adalah cara masing-masing:

1. Data yang Tidak Berkelompok

Data tidak berkelompok adalah data yang dicatat dalam bentuk individual

Berikut adalah cara menghitung pemusatan :

Mean (Rata-rata)

$$Mean = \frac{\sum xi}{n}$$

Keterangan :

\bar{x} = mean

$\sum x$ = jumlah nilai

n = banyak data

dimana xi adalah nilai data n adalah jumlah data.

Contoh :

Hitung mean dari data berikut : 2,3,4,5,6,1,7

Pembahasan

Jumlah seluruh data = $\sum x = 2+3+4+5+6+1+7 = 28$

Jumlah banyak data = n = 7

Mean = $\bar{x} = 28:7 = 4$

Jadi, mean dari data tersebut adalah 4

Median

$$\text{Median} = X \frac{n+1}{2} \text{ untuk data ganjil}$$

Keterangan :

n = banyak data

Contoh :

Diberikan data pengunjung pasar setiap hari Senin-Jumat.

Hari	Pengunjung
Senin	50
Selasa	30
Rabu	40
Kamis	20
Jumat	65

Berapakah median dari data pengunjung pasar tersebut?

Pembahasan

Langkah awal mengurutkan data dari terkecil menjadi terbesar yaitu :

Karena terdapat 5 data maka median = $X \frac{n+1}{2} = \frac{x_6}{2} = x_3$

Median terletak di data ke 3, yaitu 40

Modus

Dalam analisis data tunggal, tujuan kita adalah untuk menentukan keadaan yang paling umum dengan cara mencari nilai yang sering muncul. Proses ini melibatkan identifikasi nilai yang memiliki frekuensi tertinggi dalam dataset tersebut.

Contoh :

Berikut adalah data tinggi badan siswa kelas XII di SMA Tadika
150, 155, 157, 152, 154, 154, 152, 157, 157, 157, 160

Pembahasan

Langkah awal mencari data frekuensi terbanyak.

Nilai	Frekuensi
150	1
152	2
154	2
155	1
157	4
160	1

Dari tabel dapat dilihat bahwa frekuensi terbanyak yaitu 4, berada pada nilai 157. Jadi, modus tinggi badan siswa kelas XII di SMA Tadika adalah 157.

a. Data yang Berkelompok

Data yang berkelompok yaitu data yang telah diorganisir ke dalam interval atau kategori. Berikut adalah cara menghitung pemusatan :

Mean (Rata-rata)

$$\bar{X} = \frac{\sum_{i=1}^n f_i X_i}{\sum_{i=1}^n f_i}$$

Keterangan :

f_i = frekuensi kelas ke-i

X_i = nilai tengah kelas ke-i

Atau jika diketahui panjang kelas dalam satu tabel distribusi frekuensi sama dapat menggunakan :

$$\bar{X} = X_0 + p \left(\frac{\sum f_i c_i}{\sum f_i} \right)$$

Keterangan :

f_i = frekuensi kelas ke-i

X_0 = nilai tengah kelas dengan kode nol

p = panjang kelas

C_i = kode kelas ke-i (kode ditentukan berdasarkan frekuensi kelas)

Contoh 1

Data hasil ujian 60 orang mahasiswa:

NILAI UJIAN	f_i	NILAI TENGAH (X_i)	$f_i X_i$
1-15	4	8	32
16-30	4	23	92
31-45	7	38	266
46-60	13	53	689
61-75	24	68	1.632
76-90	8	83	664
Jumlah	60	-	3.375

Jadi, rata-rata hitung dari nilai ujian 60 mahasiswa adalah:

$$\bar{X} = \frac{\sum_{i=1}^n f_i X_i}{\sum_{i=1}^n f_i} = \frac{3.375}{60} = 56,25$$

Contoh:

Data hasil ujian 60 orang mahasiswa :

NILAI UJIAN	f_i	NILAI TENGAH (X_i)	c_i	$f_i c_i$
1-15	4	8	-4	-16
16-30	4	23	-3	-12
31-45	7	38	-2	-14
46-60	13	53	-1	-13
61-75	24	68	0	0
76-90	8	83	+1	8
Jumlah	60	-		-47

Jadi, rata-rata hitung dari nilai ujian 60 orang mahasiswa adalah :

$$\bar{X} = X_0 + p \left(\frac{\sum f_i c_i}{\sum f_i} \right) = 68 + 15 \left(\frac{-47}{60} \right) = 56,25$$

Median

$$Me = b + p \left(\frac{\frac{n}{2} - F}{f} \right)$$

Keterangan :

b = ujung bawah kelas $\left(\frac{1}{2} \sum f_i \right)$ median

p = panjang kelas

$n = \sum f_i$

F = frekuensi kumulatif sebelum kelas median

f = frekuensi kelas median

Contoh :

Data hasil ujian 60 orang mahasiswa :

NILAI UJIAN	RUANG KELAS	f_i	F
10 - 24	9,5 - 24,5	4	4
25 - 39	24,5 - 39,5	4	8
40 - 54	39,5 - 54,5	7	15
55 - 69	54,5 - 69,5	13	28
70 - 84	69,5 - 84,5	24	52
85 - 99	84,5 - 99,5	8	60
Jumlah		60	

Letak median dikelas 70-84

Berdasarkan tabel, diperoleh :

$$b = 69,5$$

$$p = 15$$

$$n = 60$$

$$F = 28$$

$$f = 24$$

Maka dari kelas median diperoleh :

$$\begin{aligned} Me &= 69,5 + 15 \left(\frac{\frac{60}{2} - 28}{24} \right) \\ &= 69,5 + 15 \left(\frac{30 - 28}{24} \right) \\ &= 70,75 \end{aligned}$$

Modus

$$Mo = b + p \left(\frac{b_1}{b_1 + b_2} \right)$$

Dengan :

b = ujung bawah kelas Modal (f terbesar)

b_1 = frekuensi kelas modal - frekuensi kelas sebelumnya

b_2 = frekuensi kelas modal - frekuensi kelas sesudahnya

p = panjang kelas

Contoh :

NILAI UJIAN	BATAS KELAS	f_i
10 - 24	9,5 - 24,5	4
25 - 39	24,5 - 39,5	4
40 - 54	39,5 - 54,5	7
55 - 69	54,5 - 69,5	13
70 - 84	69,5 - 84,5	24
85 - 99	84,5 - 99,5	8
Jumlah		60

Berdasarkan tabel, diperoleh :

$$b = 69,5$$

$$p = 15$$

$$b_1 = 24 - 13$$

$$b_2 = 24 - 8$$

Maka nilai ujian yang paling banyak diperoleh oleh mahasiswa adalah :

$$M_o = 69,5 + 15 \left(\frac{11}{11+16} \right)$$

$$= 69,5 + 15 \left(\frac{11}{27} \right)$$

$$= 75,61$$

Ukuran Letak Data

Dalam statistik, selain mengukur konsentrasi data, terdapat juga pengukuran lokasi data. Ukuran lokasi meliputi kuartil, desil, dan persentil, serta median, rata-rata, dan modus. Untuk menentukan ukuran lokasi ini, data harus disusun dari yang terkecil hingga terbesar.

Kuartil

Nilai triwulanan adalah nilai penyortiran data dalam empat bagian yang sama. Kuartil pertama (Q1), kuartil kedua (Q2) dan kuartil ketiga (Q3) adalah tiga jenis tipe triwulanan (kuartil) yang diwakili oleh Q.

1. Kuartil Untuk Data Tunggal (Tidak Berkelompok)

$$\text{Letak } Q_1 = \frac{i(n+1)}{4}$$

Keterangan :

Q_1 = Kuartil ke - i

N = Banyaknya data

Contoh soal Kuartil Data Tunggal

Tentukan Q_1 , Q_2 dan Q_3 dari data:

7,3,8,5,9,4,8,3,10,2,7,6,8,7,2,6,9.

Jawab:

Data Terurut: 2,2,3,3,4,5,6,6,7,7,7,8,8,8,9,9,10.

$n = 17$

$$\text{Letak } Q_1 = \frac{1(17+1)}{4} = \frac{18}{4} = 4,5$$

$$= 3 + 0,5(4-3)$$

$$= 3,5$$

$$\text{Letak } Q_2 = \frac{2(17+1)}{4} = \frac{36}{4} = 9$$

$$Q_2 = X_9 = 7$$

$$\text{Letak } Q_3 = \frac{3(17+1)}{4} = \frac{54}{4} = 13,5$$

$$Q_3 = X_{13} + 0,5(X_{14} + X_{13})$$

$$= 8 + 0,5(8-8)$$

$$= 8$$

2. Kuartil Untuk Data Bergolong (Berkelompok)

Berikut adalah cara untuk mencari lokasi kuartil untuk data yang dikelompokkan

$$Q_i = T_b + p \left(\frac{\frac{i \cdot n}{4} - F}{f} \right)$$

Keterangan :

Q1= kuartil ke-i

Tb = tepi bawah kelas kuartil

p = panjang kelas

n = banyak data

F = frekuensi kumulatif sebelum kelas kuartil

f = frekuensi kelas kuartil

Contoh :

Tentukan Q dari data berikut :

DATA	f
11 - 20	2
21 - 30	7
31 - 40	4
41 - 50	6
51 - 60	5
61 - 70	6

Jawab :

$$\begin{aligned} Q_i &= T_b + P \left(\frac{\frac{i \cdot n}{4} - F}{f} \right) \\ &= 20,5 + 10 \left(\frac{7,5 - 2}{7} \right) \\ &= 20,5 + 7,86 = 28,36 \end{aligned}$$

Desil

Desil adalah nilai-nilai yang membagi data menjadi sepuluh bagian yang sama. Desil biasanya dilambangkan dengan huruf D. Secara umum, ada 9 desil, yaitu $D_1, D_2, D_3, \dots, D_9$, yang membagi data menjadi sepuluh bagian yang berbeda.

1. Desil Untuk Data yang Tidak Berkelompok

$$\text{Letak } D_i = \frac{i(n+1)}{10}$$

Keterangan :

D = desil ke-i

n = banyaknya data

Contoh :

Tentukan desil ke-8 dari data: 6,3,8,9,5,9,9,7,5,7,4,5,8,3,7,6.

Jawab :

n = 16

data terurut = 3,3,4,5,5,5,6,6,7,7,7,8,8,9,9,9.

$$\begin{aligned} \text{letak } D_8 &= \frac{8(16+1)}{10} = 13,6 \\ D_8 &= X_{13} + 0,6 (X_{14} - X_{13}) \\ &= 8 + 0,6 (9-8) = 8,6 \end{aligned}$$

Jadi, nilai Desil ke-8 adalah 8,6

2. Desil Untuk Data Bergolong (berkelompok)

Menentukan letak desil untuk data berkelompok :

$$D_i = T_b + p \left(\frac{\frac{i \cdot n}{10} - F}{f} \right)$$

Keterangan:

D_i = desil ke-i

T_b = tepi bawah kelas kuartil

p = panjang kelas

n = banyak data

F = frekuensi kumulatif sebelum kelas kuartil

f = frekuensi kelas kuartil

Contoh :

Tentukan nilai D6 dari data berikut :

DATA	f
11-13	5
14-16	6
17-19	3
20-22	5
23-25	7
26-28	4

$$\begin{aligned}D_i &= \frac{i}{10} \times n \\&= \frac{6}{10} \times 30 \\&= 18\end{aligned}$$

Jawab :

DATA	f	f _k
11 - 13	5	5
14 - 16	6	11
17 - 19	3	14
20 - 22	5	19
23 - 25	7	26
26 - 28	4	30

$$\begin{aligned}D_6 &= Tb + p \left(\frac{\frac{6n}{10} - F}{f} \right) \\&= 19,5 + 3 \left(\frac{18-14}{5} \right) \\&= 19,5 + 2,4 = 21,9\end{aligned}$$

Jadi, nilai Desil ke-6 adalah 21,9

Persentil

Persentil adalah nilai yang membagi sekelompok data menjadi seratus bagian yang sama. Dalam notasinya, persentil biasanya dilambangkan dengan huruf P. Terdapat 99 jenis persentil, yaitu P1, P2, P3, hingga P99.

1. Persentil Untuk Data yang Tidak Berkelompok

$$\text{Letak } P_i = \frac{i(n+1)}{100}$$

Keterangan :

Pi = persentil ke-i

n = banyaknya data

Contoh :

Tentukan persentil ke-65 dari data: 6,5,8,7,9,4,5,8,4,7,8,5,8,4,5

Jawab:

n = 15

data terurut : 4,4,4,5,5,5,5,6,7,7,8,8,8,8,9

$$\text{Letak } P_{65} = \frac{65(15+1)}{100} = 10,4$$

$$\begin{aligned} P_{65} &= X_{10} + 0,4 (X_{11} - X_{10}) \\ &= 7 + 0,4 (8 - 7) = 7,4 \end{aligned}$$

Jadi, nilai persentil ke-65 adalah 7,4

2. Persentil untuk data Bergolong (Berkelompok)

Menentukan letak persentil untuk data berkelompok

$$\text{Letak } P_i = T_b + p \left(\frac{\frac{n}{100} - F}{f} \right)$$

Keterangan :

Pi = persentil ke-i

Tb = tepi bawah kelas persentil

p = panjang kelas

n = 1 banyak data

F = frekuensi kumulatif sebelum kelas persentil

f = frekuensi kelas persentil

Contoh :

Tentukan P30 dari data berikut :

DATA	f
10 - 14	4
15 - 19	6
20 - 24	5
25 - 29	7
30 - 34	3
35 - 39	5

$$\begin{aligned}P_i &= \frac{i}{100} \times n \\&= \frac{30}{100} \times 30 \\&= 9\end{aligned}$$

DATA	f	f _k
10 - 14	4	4
15 - 19	6	10
20 - 24	5	15
25 - 29	7	22
30 - 34	3	25
35 - 39	5	30

$$\begin{aligned}P_{30} &= Tb + p \left(\frac{\frac{30n}{100} - F}{f} \right) = 14,5 + 5 \left(\frac{9-5}{6} \right) \\&= 14,5 + 3,33 = 17,83\end{aligned}$$

Jadi, nilai P30 adalah 17,83

Rata-rata kuadrat

a. Data yang Tidak Berkelompok

Rata-rata kuadrat untuk data tidak terkelompok dihitung menggunakan persamaan tertentu.

$$MK = \sqrt{\frac{\sum_{i=1}^N X_i^2}{N}} = \sqrt{\frac{X_1^2 + X_2^2 + X_3^2 + \dots + X_n^2}{N}}$$

Keterangan :

MK = Nilai rata-rata kuadrat

$\sum X_i^2$ = Jumlah dari seluruh X_i^2

N= Jumlah data (pengamatan)

Xi (dalam ribuan)	X_i^2
500	250.000
400	160.000
550	302.500
600	360.000
350	122.500
	$\sum X_i^2 = 1.195.000$

Contoh data yang tak terkelompok mengenai perolehan keuntungan dari hasil penjualan barang jualan di lima toko yang terletak di sepanjang jalan Bandung.

X 1 = Rp 500.000

X 2 = Rp 400.000

X 3 = Rp 550.000

X 4 = Rp 600.000

X 5 = Rp 350.000

b. Data yang Berkelompok

Untuk data terkelompok, rata-rata kuadrat dihitung menggunakan persamaan.

$$MK = \sqrt{\frac{\sum_{i=1}^N X_i^2}{N}} = \sqrt{\frac{1.195.000}{5}} = \sqrt{239.000} = \text{Rp } 488,876,-$$

Untuk data terkelompok dengan data yang sama seperti pada ilustrasi, di dalamnya terdapat perhitungan

rata-rata (mean), median, dan modus.:

$$MK = \sqrt{\frac{\sum_{i=1}^N fM^2}{N}}$$

Keterangan :

MK=1

f = Frekuensi pada masing-masing kelas

M = nilai mid point pada masing-masing kelas

\sum frekuensi	M (C_m)	M^2	fM^2
1	52,5	2.756,25	2.756,25
2	58,5	3.422,25	6.844,50
17	64,5	4.160,25	70.724,25
13	70,5	4.970,25	64.613,25
24	76,5	5.852,25	14.454
9	82,5	6.806,25	61.256,25
7	88,5	7.832,25	54.825,75
7	94,5	8.930,25	62.511,75
80			$\sum fM^2 = 463.986$

Cara menghitung MK :

- 1) Temukan nilai titik tengah untuk setiap kelas
- 2) Kuadratkan titik tengah untuk setiap kelas (M^2)
- 3) Frekuensi dengan M harus dikalikan untuk setiap kelas (fM^2)
- 4) Semua nilai fM^2 (fM^2). (jumlahkan)
- 5) Bagi total (fM^2) dengan N. f. Hitung nilai MK dengan mengambil akar dari hasil pada langkah e

Rata-rata harmonik

1. Data yang Tidak Berkelompok

Rata-rata harmonik adalah tingkat normal dari penyebaran data. Jika hasil pengamatan data menunjukkan nilai-nilai $X_1, X_2, X_3, \dots, X_n$, maka rata-rata harmoniknya adalah.

$$M_H = \frac{N}{\sum_{i=1}^N \frac{1}{(X)_i}}$$

Keterangan :

M_H = rata-rata harmonik

N = jumlah data

Dengan menggunakan contoh data yang tidak terkelompok X_i
(Rp 000,-) : 500,400,550,650,350

X_i (dalam Rp000,-)	$1/X_i$
500	$1/500 \quad 2,00. 10^{-3}$
400	$1/400 \quad 2,50. 10^{-3}$
550	$1/550 \quad 1,81. 10^{-3}$
600	$1/600 \quad 1,66. 10^{-3}$
350	$1/350 \quad 2,85. 10^{-3}$
$\sum_{i=1}^N \frac{1}{X_i} = 10,82x 10^{-3} = 0,01081$	

$$M_H = \frac{N}{\sum_{i=1}^N \left(\frac{1}{X_i}\right)} = \frac{5}{0,01081} = Rp 461,173$$

2. Data yang Berkelompok

Untuk data terkelompok, rata-rata terhitung harus dihitung menggunakan persamaan.

$$M_H = \frac{N}{\sum_{i=1}^n \left(\frac{f}{M}\right)}$$

Contoh perhitungan :

TOKO	X_i (penjualan)	Log X_i
A	Rp 100.000,-	5,00
B	Rp 80.000,-	4,90
C	Rp 120.000,-	5,079
D	Rp 125.000,-	5,096
E	Rp 75.000,-	4,875
		= 24,95

$$\log M_G = \frac{1}{N} \sum_{i=1}^N \log X_i = 1/5 \times 24,95$$

Cara menghitungnya :

- Tentukan nilai midpoint (M) untuk setiap kelas.
- Bagi nilai frekuensi dengan M untuk setiap kelas.
- Hasil dari langkah sebelumnya. (jumlahkan)
- Untuk menghitung MH, bagi N dengan hasil dari langkah sebelumnya.

Rata-rata ukur

1. Data yang Tidak Berkelompok

Normal dari data yang tidak dikelompokkan dihitung dengan menggunakan persamaan.

$$M_G = \text{anti log} \frac{1}{N} \sum_{i=1}^N \log X_i$$

Menggunakan ilustrasi data tidak terkelompokkan X_i sebesar Rp. 100.000,-, Rp. 80.000,-, Rp. 120.000,-, Rp. 125.000,-, dan Rp. 75.000,-,-

$$\log M_G = \frac{1}{N} \sum_{i=1}^N \log X_i = 1/5 \times 24,95$$

$$M_G = \text{anti log } 4,99 = \text{Rp } 97.800,$$

Cara menghitungnya :

- Hitunglah log dari masing-masing X
- Jumlahkan setiap nilai log X ($\sum_{i=1}^N \log X_i$)
- Kalikan hasil dari b dengan $1/N$ sehingga akan diperoleh log M_G
- Hitunglah anti log hasil dari c untuk memperoleh M_G

2. Data yang Berkelompok

Untuk menghitung estimasi normal dari data terkelompok, persamaan yang digunakan adalah...

$$\log M_G = \frac{\sum_{i=1}^N f \log M_i}{N}$$

f = frekuensi masing-masing kelas

log Mi = nilai midpoint pada masing - masing kelas

N = Jumlah kasus

Contoh :

Kelas interval	∑ frekuensi	M (C _m)	log M	f log M
50-55	1	52,5	1,720	1,720
56-61	2	58,5	1,767	3,534
62-67	17	64,5	1,809	30,762
68-73	13	70,5	1,848	24,026
74-79	24	76,5	1,883	45,207
80-85	9	82,5	1,916	17,248
86-91	7	88,5	1,946	13,628
92-97	7	94,5	1,975	13,828
—————				
Total	80		$\sum_{i=1}^N f \log M = 149,953$	

$$\log M_G = \frac{\sum_{i=1}^N f \log M}{N} = \frac{149,953}{80} = 1,874$$

$$M_G = \text{anti log } 1,875 = 74,8$$

Cara menghitungnya :

- Hitung nilai titik tengah untuk setiap kelas
- Hitung logaritma dari setiap titik tengah tersebut
- Kalikan frekuensi dengan titik tengah pada setiap kelas
- Jumlahkan semua hasil dari langkah c
- Untuk hasil bagi dari langkah d dengan N
- Hitung anti log hasil e

Estimasi Kuartil, Desil, dan Persentil Data Berkelompok
Rumus menghitung kuartil data berkelompok :

$$Q_i = b + p \left(\frac{\frac{i}{4}n - f_k}{f_Q} \right)$$

$$D_i = b + p \left(\frac{\frac{i}{10}n - f_k}{f_D} \right)$$

$$P_i = b + p \left(\frac{\frac{i}{100}n - f_k}{f_P} \right)$$

Keterangan :

Q_i : kuartil ke-i,

D_i : desil ke-i,

P_i : persentil ke-i

b: batas bawah kelas kuartil, desil, atau persentil

p : panjang kelas

n : banyak data

f_k : frekuensi kumulatif sebelum kelas kuartil, desil, atau persentil.

f_Q : frekuensi kelas kuartil.

f_D : frekuensi kelas desil.

f_P : frekuensi kelas persentil.

D. Tantangan dalam Analisis Ukuran Pemusatan untuk Data yang Tidak Berkelompok dan Berkelompok

1. Tantangan Pada Data yang Tidak Berkelompok

a. Keberagaman Informasi

Tindakan sentralisasi yang tidak representatif mungkin terjadi karena banyaknya variabel atau keragaman data yang tidak dikelompokkan. Misalnya, jika informasi tentang tinggi badan siswa sangat bervariasi, rata-ratanya mungkin memberikan kesan yang salah tentang tinggi rata-rata.

b. Efek Outlier

Nilai yang berbeda dari nilai lainnya dapat memengaruhi hasil analisis. Outlier sering kali membuat mean tidak akurat; median, di sisi lain, cenderung lebih stabil, tetapi tetap memerlukan interpretasi yang hati-hati untuk menghindari kesalahan.

c. Penyampaian Informasi

Salah satu tantangannya adalah menyajikan data yang tidak dikelompokkan dalam bentuk yang mudah dipahami. Penting untuk menggunakan bagan atau tabel yang sesuai untuk menggambarkan distribusi data secara efektif.

d. Perspektif Median

Terlepas dari kenyataan bahwa median sering dianggap sebagai ukuran keterpusatan yang lebih baik dalam beberapa situasi.

2. Tantangan Pada Data yang Berkelompok

a. Keputusan Interval

Untuk data yang terkluster, memilih interval yang tepat merupakan masalah besar. Interval yang terlalu lebar dapat menyembunyikan variasi-variasi penting, sedangkan interval yang terlalu sempit dapat menyebabkan fluktuasi yang tidak stabil.

b. Penghapusan Informasi Detail

Keakuratan dan keterwakilan ukuran sentralitas yang dihitung dari data yang dikelompokkan mungkin terpengaruh karena informasi rinci tentang nilai individu hilang ketika data dikelompokkan.

c. Pengukuran Ukuran Pemusatan

Metode penilaian diperlukan untuk menghitung konsentrasi data grup. Misalnya, nilai rata-rata setiap interval digunakan untuk menghitung pentingnya data. Ini dapat mempengaruhi hasil akhirnya.

d. Pendistribusian Data

Analisis data kelompok harus memahami distribusi data. Jika distribusi data abnormal, konsentrasi (misalnya

nilai rata-rata) mungkin tidak menunjukkan pusat aktual.

3. Metodologi dalam Mengatasi Tantangan

a. Metode Grup yang Tepat

Untuk mengurangi kehilangan informasi dalam data berkelompok pemilihan Teknik pengelompokan yang tepat memerlukan pemahaman mendalam tentang sifat data.

b. Analisis Eksklusif

Pada data yang tidak dikelompokkan, analisis outlier dapat membantu memahami dampak outlier terhadap ukuran sentralitas. Teknologi ini dapat meningkatkan akurasi analisis.

Ringkasan Materi

1. Ukuran Pemusatan Data

a. Ukuran pemusatan data berfungsi untuk menggambarkan titik tengah atau kecenderungan sentral dari suatu distribusi data.

b. Ukuran pemusatan yang biasa digunakan meliputi mean (rata-rata), median, dan modus.

c. Perhitungan ukuran pemusatan data dilakukan berbeda pada data yang berkelompok dan tidak berkelompok.

d. Perhitungan untuk Data Tidak Berkelompok

1) Mean : Jumlah semua nilai dibagi dengan banyaknya data.

2) Median : Nilai tengah dari data yang sudah diurutkan.

3) Modus : Nilai yang paling sering muncul dalam kumpulan data.

e. Perhitungan untuk Data Berkelompok

1) Mean : Menggunakan nilai tengah setiap kelas dan frekuensi masing-masing kelas.

2) Median : Berdasarkan kelas median, yaitu kelas yang memiliki frekuensi kumulatif terdekat dengan 50%.

3) Modus : Kelas dengan frekuensi tertinggi dianggap sebagai kelas modal.

f. Tantangan dalam Analisis

Data Tidak Berkelompok

Efek dari outlier bisa mempengaruhi mean, dan keberagaman informasi bisa menyebabkan ukuran pemusatan tidak representatif.

Data Berkelompok

Kehilangan informasi detail akibat pengelompokan, serta penentuan interval yang tepat, dapat menjadi tantangan dalam analisis.

Kritik

1. Contoh yang Kurang Variatif

Contoh yang diberikan dalam materi lebih banyak berfokus pada data kuantitatif. Akan lebih baik jika juga disertakan contoh dari data kualitatif atau kategorikal, terutama terkait penggunaan modus untuk memperkaya pemahaman pembaca.

2. Tantangan Analisis Kurang Diperluas

Tantangan dalam menganalisis data yang tidak dikelompokkan dan dikelompokkan hanya dibahas secara singkat. Kritik yang lebih mendalam tentang cara memilih interval kelas yang sesuai untuk data yang dikelompokkan atau cara menangani distribusi data asimetris akan memberikan wawasan yang lebih baik kepada pembaca.

3. Minimnya Saran Solusi untuk Mengatasi Tantangan:

Meskipun tantangan dalam analisis data sudah disebutkan, materi ini tidak memberikan saran yang jelas tentang bagaimana mengatasi tantangan tersebut, misalnya dengan teknik statistik yang lebih canggih atau pendekatan lain yang relevan. Memberikan rekomendasi teknik atau metode yang dapat digunakan dalam menghadapi tantangan akan menambah nilai pada materi.

Saran :

1. Memilih Ukuran Pemusatan yang Tepat

Saat menghadapi data dengan *outlier* atau distribusi tidak normal, median mungkin lebih representatif dibandingkan mean.

Untuk data berkelompok, pertimbangkan pengelompokan yang tepat agar tidak ada informasi penting yang hilang.

2. Perhatikan Outlier

Dalam analisis data yang tidak berkelompok, identifikasi dan analisis outlier sangat penting untuk memastikan hasil yang lebih akurat.

3. Pentingnya Kesesuaian dalam Pengelompokan Data

Untuk data yang berkelompok, pastikan interval yang digunakan sesuai dengan tujuan analisis agar hasilnya lebih representatif.

Kesimpulan

Memahami perbedaan dalam penghitungan ukuran pemusatan antara data berkelompok dan tidak berkelompok sangat penting untuk memperoleh hasil analisis yang akurat. Penggunaan mean, median, dan modus harus disesuaikan dengan karakteristik data yang sedang dianalisis. Untuk data yang memiliki distribusi asimetris atau outlier, median sering kali menjadi ukuran yang lebih tepat, sementara mean dapat memberikan gambaran umum yang baik pada data dengan distribusi yang lebih seragam.

Pemilihan ukuran pemusatan data, baik untuk data berkelompok maupun tidak berkelompok, sangat bergantung pada jenis data, tujuan analisis, dan konteks penggunaannya. Data tidak berkelompok memberikan hasil yang lebih akurat dan detail karena menganalisis setiap nilai individu secara langsung. Metode ini cocok untuk dataset kecil hingga sedang, di mana setiap data memiliki peranan penting dalam analisis.

Sementara itu, analisis data berkelompok memungkinkan penyajian yang lebih sederhana dan terstruktur, khususnya untuk dataset yang sangat besar. Dengan mengelompokkan data ke dalam kelas-kelas interval, ukuran pemusatan seperti mean, median, dan modus tetap dapat memberikan gambaran umum tentang distribusi data, meskipun kehilangan informasi spesifik pada tingkat individu.

Kedua pendekatan ini saling melengkapi dan penting dalam statistik, karena memberikan fleksibilitas dalam analisis berdasarkan situasi. Pemahaman mendalam tentang masing-masing metode tidak hanya membantu dalam memilih pendekatan yang tepat, tetapi juga memastikan bahwa interpretasi data dilakukan secara akurat, relevan, dan sesuai dengan tujuan penelitian atau pengambilan keputusan.

Soal Latihan Dan Jawaban

Soal 1 (Data Tidak Berkelompok)

Sebuah toko pakaian mencatat pendapatan harian (dalam ribuan rupiah) dari 7 hari berikut:

250, 300, 275, 250, 400, 350, 250

1. Tentukan:

- a. Mean
- b. Median
- c. Modus

Jawaban:

- a. Mean = $(250 + 300 + 275 + 250 + 400 + 350 + 250) / 7 = 2.075 / 7 = 296,43$
- b. Median → Urutkan: 250, 250, 250, 275, 300, 350, 400 → 275
- c. Modus = 250 (muncul 3 kali)

Soal 2 (Data Berkelompok)

Tabel berikut menunjukkan hasil ujian matematika dari 50 siswa:

Table

Copy

Nilai	Frekuensi
40-49	5
50-59	10
60-69	15
70-79	12
80-89	8

Tentukan:

- Rata-rata(mean)kelompok
- Modus kelas

Jawaban:

- Mean

Gunakan nilai tengah (X_i) dan frekuensi (f):

Table

Copy

Kelas	X_i	f	$f \times X_i$
40-49	44,5	5	222,5
50-59	54,5	10	545

Kelas	X_i	f	$f \times X_i$
60-69	64,5	15	967,5
70-79	74,5	12	894
80-89	84,5	8	676
Total		50	3.305

$$\text{Mean} = 3.305 / 50 = 66,1$$

b. Modus → Kelas dengan frekuensi tertinggi = 60-69

Soal 3 (Aplikasi Bisnis)

Sebuah perusahaan ingin mengetahui gaji rata-rata karyawan dari sampel 5 orang dengan gaji sebagai berikut (dalam jutaan rupiah): 5, 6, 5, 8, 20

Tentukan:

- a. Mean
- b. Apakah mean ini representative? Jelaskan
- c. Ukuran pemusatan mana yang lebih tepat digunakan?

Jawaban:

- a. Mean = $(5 + 6 + 5 + 8 + 20) / 5 = 44 / 5 = 8,8$ juta
- b. Tidak representatif, karena ada outlier (20 juta) yang menyebabkan mean terlalu tinggi.
- c. Median lebih tepat → Urutkan: 5, 5, 6, 8, 20 → Median = 6 juta

DAFTAR PUSTAKA

- Gozali, I., & A. Fauzi, A. (2024). *Statistik Bisnis I: Ukuran Pemusatan Data Berkelompok dan Tidak Berkelompok*. Fakultas Ekonomi dan Bisnis, Universitas Bhayangkara Jakarta Raya.
- Sudjana, M. (2002). *Metode Statistika*. Bandung: Tarsito.
- Trihendradi, C. (2015). *Statistika Deskriptif & Inferensial: Menggunakan Microsoft Excel dan SPSS*. Yogyakarta: Andi.
- Walpole, R. E., & Myers, R. H. (2012). *Probability & Statistics for Engineers and Scientists*. Boston: Pearson.
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.
- Gujarati, D. N. (2004). *Basic Econometrics*. New York: McGraw-Hill.
- Arikunto, S. (2010). *Prosedur Penelitian: Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta.
- Mendenhall, W., Beaver, R. J., & Beaver, B. M. (2012). *Introduction to Probability and Statistics*. Boston: Cengage Learning.
- Anderson, D. R., Sweeney, D. J., & Williams, T. A. (2011). *Statistics for Business and Economics*. Boston: Cengage Learning.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Upper Saddle River, NJ: Pearson.
- Levine, D. M., Stephan, D. F., & Szabat, K. A. (2016). *Statistics for Managers Using Microsoft Excel*. Boston: Pearson.
- Agresti, A. (2018). *Statistical Methods for the Social Sciences*. Boston: Pearson.
- Montgomery, D. C., & Runger, G. C. (2014). *Applied Statistics and Probability for Engineers*. New York: Wiley.
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2017). *Introduction to the Practice of Statistics*. New York: W.H. Freeman and Company.
- Black, K. (2013). *Business Statistics: For Contemporary Decision Making*. Hoboken, NJ: Wiley.
- Newbold, P., Carlson, W. L., & Thorne, B. (2013). *Statistics for Business and Economics*. Boston: Pearson.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. New York: McGraw-Hill.

- Agresti, A., & Franklin, C. (2013). *Statistics: The Art and Science of Learning from Data*. Boston: Pearson.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. New Jersey: Pearson Prentice Hall.
- Newbold, P., Carlson, W. L., & Thorne, B. (2013). *Statistics for Business and Economics*. Boston: Pearson.
- Black, K. (2019). *Business Statistics: For Contemporary Decision Making*. New York: John Wiley & Sons.
- Sudjana. (1991). In *Statistika*. Bandung: Tarsito.
- Suharyadi, & Purwanto S. K. (2007). *Statistika: Untuk Ekonomi dan Keuangan Modern, Edisi 2*. Jakarta: Penerbit Salemba Empat..
- Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. London: Sage Publications.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall.
- Bluman, A. G. (2018). *Elementary Statistics: A Step-by-Step Approach*. New York: McGraw-Hill Education.
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2012). *Introduction to the Practice of Statistics*. New York: W.H. Freeman.
- Weiss, N. A. (2015). *Introductory Statistics*. Boston: Pearson.
- Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics*. New York: W.W. Norton & Company.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
- Rice, J. A. (2006). *Mathematical Statistics and Data Analysis*. Belmont, CA: Duxbury Press.
- Kenney, J. F., & Keeping, E. S. (1962). *Mathematics of Statistics*. Princeton: Van Nostrand.
- Hogg, R. V., McKean, J., & Craig, A. T. (2013). *Introduction to Mathematical Statistics*. Boston: Pearson.

BAB 4

UKURAN SIMPANGAN DAN VARIASI UNTUK DATA YANG DIKELOMPOKKAN DAN TIDAK DIKELOMPOKKAN

Dalam statistik deskriptif, ukuran deviasi dan variabilitas adalah konsep penting untuk memahami penyebaran atau distribusi suatu kumpulan data. Ukuran-ukuran ini memberikan informasi tentang sejauh mana nilai-nilai dalam sebuah dataset menyimpang dari titik pusat seperti rata-rata atau median. Dua jenis data yang umumnya dianalisis adalah data tak terkelompok dan data terkelompok.

Untuk data tak terkelompok, beberapa ukuran deviasi dan variabilitas yang sering digunakan meliputi rentang (range), deviasi rata-rata, variansi, simpangan baku, dan koefisien variasi. Sementara itu, untuk data terkelompok, perhitungan ukuran-ukuran ini memerlukan penyesuaian karena data telah dikelompokkan ke dalam interval atau kelas-kelas tertentu.

Memahami ukuran deviasi dan variabilitas ini sangat penting dalam berbagai bidang, termasuk penelitian ilmiah, analisis bisnis, dan pengambilan keputusan berbasis data. Dengan mengetahui penyebaran data, peneliti dan analis dapat membuat kesimpulan yang lebih akurat tentang karakteristik populasi dan membandingkan berbagai kelompok data.

Bab ini akan membahas konsep-konsep dan dasar teori terkait dengan ukuran deviasi dan variabilitas, dengan fokus pada perbedaan antara data terkelompok dan data tak terkelompok.

Rentang (Range)

Dalam statistik, rentang mengacu pada selisih antara nilai tertinggi dan nilai terendah dalam suatu kumpulan data. Rentang memberikan ukuran dasar mengenai sebaran atau penyebaran data. Untuk menghitung rentang, kita cukup mengurangi nilai terendah dari nilai tertinggi.

Rentang adalah konsep statistik yang mendasar yang membantu kita memahami sebaran atau variabilitas data dalam suatu kumpulan data. Rentang dalam statistik memberikan wawasan yang

berguna tentang sejauh mana nilai-nilai dalam kumpulan data berbeda satu sama lain. Pengukuran rentang ini menunjukkan selisih antara nilai tertinggi dan nilai terendah dalam kumpulan data tersebut.

Data yang tidak dikelompokkan

Rumus untuk menemukan rentang data yang tidak dikelompokkan atau distribusi data diskrit diberikan sebagai:

Rentang = Nilai tertinggi dari kumpulan data - Nilai terendah dari kumpulan data

Kita dapat memahaminya dari contoh yang disebutkan di bawah ini:

Carilah rentang data: 21,6,17,18,12,8,4,13

Jawab:

Di berikan, 21,6,17,18,12,8,4,13

Nilai tertinggi = 21

Nilai terendah = 4

Rentang = nilai tertinggi - nilai terendah = $21 - 4 = 17$

Data yang dikelompokkan

Dalam hal distribusi frekuensi kontinu atau data terkelompok, rentang didefinisikan sebagai selisih antara batas atas dari interval data terkelompok terbesar dan batas bawah dari interval data terkelompok terkecil. Ini merupakan ukuran penyebaran data yang paling sederhana. Rentang memberikan gambaran umum mengenai distribusi seluruh data dalam suatu kelompok. Oleh karena itu, rumus untuk menghitung rentang data terkelompok adalah sebagai berikut:
Rentang = Batas kelas atas interval tertinggi - Batas kelas bawah interval terendah

Pada data berkelompok yang kumpulan datanya disusun dalam Interval Kelas, Rentangnya ditemukan dengan cara mengurangkan batas bawah interval kelas pertama dan batas atas interval kelas terakhir. Kita dapat memahaminya dari contoh yang disebutkan di bawah ini:

Interval Kelas	Frekuensi
0-10	12
10-20	10
20-30	15
30-40	13
40-50	11

Rentang = Batas Atas Interval Kelas Terakhir - Batas Bawah Interval Kelas Pertama = 50-0 = 50

Variasi Dan Simpangan Baku

Variansi sampel yang disimbolkan dengan s^2 didefinisikan sebagai jumlah kuadrat deviasi terhadap mean dibagi dengan $n-1$

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

Simpangan baku sampel yang disimbolkan dengan s didefinisikan sebagai akar positif dari variansi sampel

$$s = \sqrt{s^2}$$

Contoh

x_i	$(x_i - \bar{x})^2$
56	350,06
59	246,80
65	94,28
76	1,66
85	105,88
89	204,20
93	18,29

$$\bar{x} = 74,71$$

$$\sum_{i=1}^n (x_i - \bar{x})^2 = 1021,17$$

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

$$= \frac{1021,17}{7-1}$$

$$= 170,195$$

$$s = \sqrt{s^2} = \sqrt{170,195} = 13,07$$

Variasi Yang Dikelompokkan

Variansi terkelompok mengacu pada variabilitas data yang dibagi ke dalam kelompok atau kategori tertentu. Variansi ini sering digunakan ketika data memiliki karakteristik yang dapat dikelompokkan secara alami atau berdasarkan kriteria tertentu. Misalnya, dalam data terkelompok, informasi dibagi ke dalam interval atau kategori, dan variansi terkelompok mengukur sejauh mana nilai-nilai dalam setiap kelompok tersebut tersebar atau bervariasi dari rata-rata kelompok tersebut.

Contoh penggunaan:

1. Membandingkan prestasi akademik siswa berdasarkan kelas atau tingkat pendidikan.
2. Menganalisis penjualan produk berdasarkan kategori atau wilayah geografis.

Variasi Yang Tidak Dikelompokkan

Variansi tak terkelompok mengacu pada variabilitas data yang tidak dibagi menjadi kelompok-kelompok tertentu. Data ini biasanya dianggap sebagai satu kesatuan tanpa pembagian ke dalam kategori atau interval. Dalam analisis statistik, variansi tak terkelompok mengukur sejauh mana nilai-nilai dalam dataset tersebar atau bervariasi dari rata-rata keseluruhan data.

Contoh penggunaan:

1. Menganalisis distribusi tinggi badan dalam suatu populasi.
2. Menghitung variasi dalam waktu tempuh perjalanan harian.

Simpangan Baku Yang Tidak Dikelompokkan

Untuk menghitung simpangan baku data tunggal, ikuti langkah-langkah berikut:

1. Perhitungan rata-rata melibatkan penjumlahan nilai data dan pembagian dengan jumlah observasi.
2. Perhitungan penyimpangan melibatkan pengurangan nilai data dari nilai rata-rata.
3. Kemudian, perhitungan varian melibatkan pengkuadratan penyimpangan dan pencarian rata-rata.
4. Simpangan baku adalah akarkuadrat dari nilai varian tersebut.

Simpangan Baku Yang Dikelompokkan

Untuk menghitung simpangan baku data kelompok, data harus dikelompokkan ke dalam interval kelas yang jelas dan terstruktur. Berikut cara menghitung simpangan baku dari data kelompok, yakni:

1. Tahap awal analisis statistik adalah menghitung nilai rata-rata data kelompok.
2. Selanjutnya, untuk menghitung penyimpangan kuadrat, kurangi nilai tengah setiap kelas frekuensi dengan nilai Mean, kemudian kuadratkan hasilnya.
3. Hitung hasil kuadrat penyimpangan dikali frekuensi untuk setiap kelas interval.
4. Tahap akhir adalah menjumlahkan hasil perhitungan dan membaginya dengan total jumlah data.

Simpangan Rata-Rata

Simpangan rata-rata merupakan ukuran penyebaran data yang menunjukkan seberapa jauh setiap data menyimpang dari nilai rata-ratanya. Dengan demikian, kita dapat memahami tingkat konsistensi data tersebut. Semakin kecil simpangan rata-rata, semakin seragam dan konsisten data tersebut. Perhatikan contoh berikut:

Data 1: 23, 22, 22, 21, 25 → rata-rata = 23

Data 2: 32, 24, 18, 19, 20 → rata-rata = 23

Kedua data di atas memiliki rata-rata yang sama, yaitu 23. Menurutmu, data mana yang lebih terpusat atau konsisten? Tentu data 1, karena data 1 menunjukkan konsistensi yang lebih tinggi dari perbedaan antara nilai data dan rata-rata relatif kecil

Simpangan Rata-Rata Yang Tidak Dikelompokkan

Data tunggal adalah data yang tidak disusun secara interval. Secara matematis, rumus simpangan rata-rata data tunggal adalah:
Dengan:

$$\partial_{\bar{x}} = \frac{1}{n} \sum_{i=1}^n |X_i - \bar{X}|$$

$\partial_{\bar{x}}$ = simpangan rata-rata data yang tidak dikelompokkan
 n = jumlah keseluruhan data
 i = nomor data
 X_i = nilai data nomor ke- i
 $\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \text{mean keseluruhan data}$

Simpangan Rata-Rata Yang Dikelompokkan

Data berkelompok adalah data yang disajikan dalam bentuk tabel dengan interval kelas tertentu. Secara matematis, rumus simpangan rata-rata data berkelompok adalah sebagai berikut:

$$\partial_{\bar{x}} = \frac{1}{n} \sum_{i=1}^n |Cm_i - \bar{X}| f_i$$

$\partial_{\bar{x}}$ = simpangan rata-rata data yang telah dikelompokkan
 n = jumlah kelas
 i = nomor kelas
 Cm_i = nilai tengah kelas (*class mark*) dari kelas i $Cm_i = B$
 $\bar{X} = \frac{\sum_{i=1}^n Cm_i \cdot f_i}{\sum_{i=1}^n f_i} = \text{mean data dikelompokkan}$
 f_i = frekuensi kelas nomor i

1. Simpangan rata-rata memberikan gambaran lebih lengkap tentang variasi data karena memperhitungkan setiap nilai, tidak hanya nilai tertinggi dan terendah.
2. Simpangan rata-rata merupakan indikator variasi data berdasarkan pengukuran simpangan absolut.

Dispersi Relatif

Dispersi relatif adalah ukuran penyebaran data yang memungkinkan perbandingan antara distribusi data yang berbeda. Dispersi relatif menunjukkan seberapa besar penyebaran data

dibandingkan dengan nilai rata-rata atau median. Dispersi relative juga mencakup beberapa aspek utama, yaitu:

1. Tujuan
 - a. Membandingkan penyebaran data antar distribusi.
 - b. Menghindari kesalahpahaman perbandingan karena perbedaan skala.
 - c. Memberikan gambaran penyebaran yang lebih akurat.

2. Jenis Dispersi Relatif
 - a. Koefisien Variasi (CV): Rasio standar deviasi terhadap mean.
 - b. Rasio Standar Deviasi: Standar deviasi dibagi dengan mean atau median.
 - c. Indeks Dispersi Relatif: Perbandingan antara standar deviasi dan nilai tertentu.

3. Kelebihan
 - a. Memungkinkan perbandingan antar distribusi.
 - b. Menghindari kesalahpahaman perbandingan.
 - c. Memberikan gambaran penyebaran yang lebih akurat.
 - d. Membantu dalam pengambilan keputusan statistik.

4. Kekurangan
 - a. Sensitif terhadap nilai ekstrim.
 - b. Tidak cocok untuk data yang tidak normal.
 - c. Memerlukan perhitungan yang rumi..

Pengukuran dispersi relatif terdiri dari:

- a. Koefisien variasi
- b. Koefisien variasi kuartil

Dispersi Relatif Data Tidak Dikelompokkan

Dispersi relatif untuk data tidak dikelompokkan adalah ukuran penyebaran relatif yang dihitung dari data mentah yang belum dikelompokkan ke dalam kelas-kelas interval.

1. Koefisien Variasi (CV)
 - a. Formula: $CV = (\text{Standar Deviasi} / \text{Mean}) \times 100\%$

- b. Mengukur variabilitas relatif terhadap nilai rata-rata
- c. Semakin besar nilai CV, semakin heterogen data tersebut

2. Koefisien Kuartil

Formula: $KK = (Q3 - Q1) / (Q3 + Q1)$

Q1 = Kuartil pertama

Q3 = Kuartil ketiga

Mengukur dispersi relatif berdasarkan nilai kuartil

Dispersi Relatif Data Dikelompokkan

Dispersi relatif untuk data yang telah dikelompokkan dalam kelas-kelas interval memerlukan perhitungan yang sedikit berbeda.

1. Koefisien Variasi (CV) Data Berkelompok

Formula: $CV = (\text{Standar Deviasi Kelompok} / \text{Mean Kelompok}) \times 100\%$

Standar Deviasi Kelompok = $\sqrt{[\sum(fi(xi - \bar{x})^2) / (n-1)]}$

Mean Kelompok = $\sum(fi \times xi) / \sum fi$

2. Koefisien Kuartil Data Berkelompok

Formula: $KK = (Q3 - Q1) / (Q3 + Q1)$

Q1 dan Q3 dihitung menggunakan rumus interpolasi kelas

Perbedaan Utama

1. Metode Perhitungan:

Data tidak dikelompokkan: menggunakan data mentah

Data dikelompokkan: menggunakan nilai tengah kelas dan frekuensi

2. Tingkat Ketelitian:

Data tidak dikelompokkan: lebih akurat

Data dikelompokkan: terdapat sedikit kehilangan informasi karena pengelompokan

3. Efisiensi:

Data tidak dikelompokkan: lebih rumit untuk data besar

Data dikelompokkan: lebih efisien untuk data besar

Aplikasi Dalam Analisis Data

Ukuran simpangan dan variasi memiliki berbagai aplikasi penting dalam analisis data:

1. Pemahaman Distribusi Data: Ukuran-ukuran ini membantu dalam memahami bagaimana data tersebar di sekitar ukuran pemusatan.
2. Deteksi Outlier: Nilai-nilai yang berada jauh di luar rentang normal (biasanya didefinisikan sebagai ± 3 standar deviasi dari mean) dapat diidentifikasi sebagai outlier potensial.
3. Analisis Risiko: Dalam keuangan, standar deviasi sering digunakan sebagai ukuran risiko investasi.
4. Kontrol Kualitas: Dalam manufaktur, ukuran variabilitas digunakan untuk memantau dan mengendalikan kualitas produk.
5. Analisis Komparatif: Koefisien variasi memungkinkan perbandingan variabilitas antara dataset dengan skala atau satuan yang berbeda.
6. Inferensi Statistik: Banyak uji statistik dan interval kepercayaan bergantung pada asumsi tentang variabilitas populasi.
7. Pemodelan Statistik: Ukuran variabilitas sering digunakan dalam pemodelan regresi dan analisis varians (ANOVA).

Perbandingan Metode

1. Pemilihan ukuran simpangan atau variasi yang tepat tergantung pada karakteristik data dan tujuan analisis:
2. Rentang: Cocok untuk gambaran cepat tentang sebaran data, tetapi sangat sensitif terhadap outlier.
3. Simpangan Rata-rata: Lebih stabil daripada rentang, tetapi kurang umum digunakan karena sifat matematis yang kurang menguntungkan.
4. Variansi dan Standar Deviasi: Paling umum digunakan karena sifat matematis yang baik dan kegunaannya dalam analisis lanjutan.
5. Koefisien Variasi: Berguna untuk membandingkan variabilitas antara dataset dengan skala yang berbeda.
6. Simpangan Kuartil: Kurang sensitif terhadap outlier dibandingkan dengan ukuran berbasis mean.
7. Kurtosis: Berguna untuk memahami bentuk distribusi data, terutama dalam kaitannya dengan distribusi normal.

Simpangan Baku

Simpangan baku merupakan ukuran statistik yang menunjukkan tingkat penyebaran data. Simpangan baku mengilustrasikan seberapa jauh titik-titik data berpenjar dari nilai tengah.

Perhitungan simpangan baku melibatkan pengkuadratan selisih antara nilai data dan rata-rata, diikuti dengan penjumlahan dan pengambilan akar kuadrat. Simpangan baku merupakan indikator statistik yang penting untuk memahami ketidakpastian dan dispersi data.

Fungsi Simpangan Baku

Fungsi utama simpangan baku dalam statistik adalah sebagai berikut:

1. Mengukur Sebaran Data

Simpangan baku memungkinkan Anda menentukan seberapa jauh titik data Anda dari rata-rata.

2. Perbandingan Data

Membandingkan deviasi standar membantu Anda menentukan data mana yang lebih seragam dan homogen.

3. Mengidentifikasi Outlier

Simpangan baku membantu Anda mengidentifikasi pengecualian atau data anomali. Data yang menyimpang secara signifikan dari rata-rata cenderung memiliki pengaruh besar pada nilai deviasi standar.

4. Keputusan Statistik

Simpangan baku memainkan peran kunci dalam statistik inferensial dengan membantu memperkirakan karakteristik populasi melalui analisis sampel.

5. Memprediksi Suatu Hal

Dalam analisis prakiraan, deviasi standar membantu memperkirakan derajat ketidakstabilan atau variabilitas dalam suatu prakiraan. Semakin besar simpangan baku, semakin besar pula ketidakpastian dalam perkiraan tersebut. Simpangan baku menyediakan wawasan statistik untuk analisis dan pengambilan keputusan strategis.

Apa Itu Simpangan Baku?

Seperti yang telah disebutkan sebelumnya, simpangan baku erat kaitannya dengan nilai rata-rata. Simpangan baku, atau yang juga bisa disebut sebagai deviasi standar, digunakan untuk mengukur sejauh mana setiap data mendekati rata-rata. Lalu, apa sebenarnya simpangan baku itu?

Berdasarkan definisinya, simpangan baku adalah ukuran statistik yang digunakan untuk mengetahui sejauh mana sebaran data dalam suatu sampel mendekati rata-rata. Simpangan baku membantu menentukan apakah sampel data mencerminkan karakteristik populasi.

KARAKTERISTIK UTAMA

1. Properti Matematis:
 - a. Selalu bernilai non-negatif
 - b. Menggunakan satuan yang sama dengan data asli
 - c. Sensitif terhadap nilai ekstrim (outlier)
 - d. Merupakan akar kuadrat dari variansi

2. Interpretasi Statistik:
 - a. Nilai kecil menunjukkan data berkelompok dekat rata-rata
 - b. Nilai besar menunjukkan data tersebar jauh dari rata-rata
 - c. Berguna untuk membandingkan sebaran dua atau lebih kelompok data
 - d. Dasar untuk interval kepercayaan dan uji hipotesis

APLIKASI DALAM BERBAGAI BIDANG

1. Pendidikan:
 - a. Mengukur variasi nilai siswa
 - b. Mengevaluasi konsistensi hasil belajar
 - c. Membandingkan prestasi antar kelas atau sekolah
 - d. Standarisasi penilaian

2. Ekonomi dan Bisnis:
 - a. Mengukur risiko investasi
 - b. Analisis volatilitas harga saham
 - c. Kontrol kualitas produksi

- d. Prediksi pasar dan tren ekonomi
- 3. Penelitian Ilmiah:
 - a. Validasi hasil eksperimen
 - b. Analisis kesalahan pengukuran
 - c. Kontrol kualitas data
 - d. Pembuatan model statistik
- 4. Industri:
 - a. Quality control
 - b. Process capability analysis
 - c. Optimization studies
 - d. Performance measurement

Rumus Simpangan Baku

Simpangan baku adalah ukuran penyebaran data yang menunjukkan seberapa jauh data berada dari rata-rata. Ada dua rumus umum yang digunakan untuk menghitung deviasi standar, tergantung pada jenis data yang tersedia: untuk data individual (sampel) dan untuk data kelompok (populasi).

Rumus Simpangan Baku Data Tunggal

Untuk data individual, Anda dapat menghitung deviasi standar menggunakan rumus berikut:

$$S = \sqrt{\sum (x_i - \bar{x})^2 / n}$$

Keterangan:

S= simpangan baku

x_i = nilai x ke- i

\bar{x} = nilai rata-rata data

n = jumlah data

Contoh soal:

1. Analisis Nilai Ujian Matematika Kelas XII

Konteks: Seorang guru matematika ingin menganalisis sebaran nilai ujian akhir semester dari 15 siswa kelas XII untuk mengevaluasi tingkat pemahaman materi.

Data nilai: 75, 82, 68, 90, 85, 78, 92, 70, 88, 83, 79, 87, 76, 84, 81

Langkah penyelesaian:

- a. Menghitung rata-rata (\bar{x})

$$\Sigma x = 1218$$

$$n = 15$$

$$\bar{x} = 1218/15 = 81.2$$

- b. Menghitung deviasi ($x - \bar{x}$)

$$75 - 81.2 = -6.2$$

$$82 - 81.2 = 0.8$$

$$68 - 81.2 = -13.2$$

$$90 - 81.2 = 8.8$$

$$85 - 81.2 = 3.8$$

$$78 - 81.2 = -3.2$$

$$92 - 81.2 = 10.8$$

$$70 - 81.2 = -11.2$$

$$88 - 81.2 = 6.8$$

$$83 - 81.2 = 1.8$$

$$79 - 81.2 = -2.2$$

$$87 - 81.2 = 5.8$$

$$76 - 81.2 = -5.2$$

$$84 - 81.2 = 2.8$$

$$81 - 81.2 = -0.2$$

- c. Menghitung kuadrat deviasi ($(x - \bar{x})^2$)

$$38.44, 0.64, 174.24, 77.44, 14.44, 10.24, 116.64, 125.44, 46.24, 3.24, 4.84, 33.64, 27.04, 7.84, 0.04$$

- d. Menjumlahkan kuadrat deviasi

$$\Sigma(x - \bar{x})^2 = 680.4$$

- e. Menghitung simpangan baku

$$s = \sqrt{[\Sigma(x - \bar{x})^2 / (n-1)]}$$

$$s = \sqrt{(680.4/14)}$$

$$s = \sqrt{48.6}$$

$$s = 6.97$$

Interpretasi: Nilai simpangan baku 6.97 menunjukkan bahwa rata-rata penyimpangan nilai siswa dari nilai rata-rata kelas adalah sekitar 7 poin. Ini mengindikasikan variasi yang cukup moderat dalam pemahaman siswa.

2. Pengukuran Kadar Glukosa Darah Pasien Diabetes

Konteks: Seorang dokter melakukan pemantauan kadar glukosa darah pada pasien diabetes selama 12 hari berturut-turut untuk menilai stabilitas gula darah.

Data (mg/dL): 142, 156, 138, 165, 149, 157, 145, 152, 160, 147, 151, 155

Langkah penyelesaian:

- a. Menghitung rata-rata (\bar{x})

$$\Sigma x = 1817$$

$$n = 12$$

$$\bar{x} = 1817/12 = 151.42$$

- b. Menghitung deviasi ($x - \bar{x}$)

$$142 - 151.42 = -9.42$$

$$156 - 151.42 = 4.58$$

$$138 - 151.42 = -13.42$$

$$165 - 151.42 = 13.58$$

$$149 - 151.42 = -2.42$$

$$157 - 151.42 = 5.58$$

$$145 - 151.42 = -6.42$$

$$152 - 151.42 = 0.58$$

$$160 - 151.42 = 8.58$$

$$147 - 151.42 = -4.42$$

$$151 - 151.42 = -0.42$$

$$155 - 151.42 = 3.58$$

- c. Menghitung kuadrat deviasi ($(x - \bar{x})^2$)

88.74, 20.98, 180.10, 184.42, 5.86, 31.14, 41.22, 0.34, 73.62, 19.54, 0.18, 12.82

- d. Menjumlahkan kuadrat deviasi

$$\Sigma(x - \bar{x})^2 = 658.96$$

- e. Menghitung simpangan baku

$$s = \sqrt{[\Sigma(x - \bar{x})^2 / (n-1)]}$$

$$s = \sqrt{(658.96 / 11)}$$

$$s = \sqrt{59.91}$$

$$s = 7.74$$

Interpretasi: Simpangan baku 7.74 mg/dL menunjukkan variabilitas yang relatif kecil dalam kadar glukosa darah pasien, mengindikasikan kontrol diabetes yang cukup baik.

3. Pengukuran Waktu Pelayanan Kasir Supermarket

Konteks: Manajer supermarket menganalisis efisiensi pelayanan kasir dengan mengukur waktu pelayanan (dalam detik) untuk 20 pelanggan.

Data (detik): 125, 138, 142, 130, 135, 128, 145, 132, 140, 133, 127, 136, 141, 129, 134, 137, 131, 139, 143, 126

Langkah penyelesaian:

a. Menghitung rata-rata (\bar{x})

$$\Sigma x = 2691$$

$$n = 20$$

$$\bar{x} = 2691/20 = 134.55$$

b. Menghitung deviasi ($x - \bar{x}$)

$$125 - 134.55 = -9.55$$

$$138 - 134.55 = 3.45$$

$$142 - 134.55 = 7.45$$

$$130 - 134.55 = -4.55$$

$$135 - 134.55 = 0.45$$

$$128 - 134.55 = -6.55$$

$$145 - 134.55 = 10.45$$

$$132 - 134.55 = -2.55$$

$$140 - 134.55 = 5.45$$

$$133 - 134.55 = -1.55$$

$$127 - 134.55 = -7.55$$

$$136 - 134.55 = 1.45$$

$$141 - 134.55 = 6.45$$

$$129 - 134.55 = -5.55$$

$$134 - 134.55 = -0.55$$

$$137 - 134.55 = 2.45$$

$$131 - 134.55 = -3.55$$

$$139 - 134.55 = 4.45$$

$$143 - 134.55 = 8.45$$

$$126 - 134.55 = -8.55$$

- c. Menghitung kuadrat deviasi $(x - \bar{x})^2$
91.20, 11.90, 55.50, 20.70, 0.20, 42.90, 109.20, 6.50, 29.70, 2.40,
57.00, 2.10, 41.60, 30.80, 0.30, 6.00, 12.60, 19.80, 71.40, 73.10

- d. Menjumlahkan kuadrat deviasi

$$\Sigma(x - \bar{x})^2 = 684.90$$

- e. Menghitung simpangan baku

$$s = \sqrt{[\Sigma(x - \bar{x})^2 / (n-1)]}$$

$$s = \sqrt{(684.90 / 19)}$$

$$s = \sqrt{36.05}$$

$$s = 6.00$$

Interpretasi: Simpangan baku 6 detik menunjukkan konsistensi yang cukup baik dalam waktu pelayanan kasir, dengan variasi rata-rata 6 detik dari waktu rata-rata pelayanan.

Rumus Simpangan Baku Data Kelompok

Untuk data yang dikelompokkan, Anda harus menggunakan rumus yang sedikit berbeda. Rumus untuk menghitung simpangan baku data kelompok adalah:

$$S = \sqrt{\sum f_i (x_i - \bar{x})^2 / n}$$

S = simpangan baku

f = frekuensi kelompok

x_i = nilai x ke- i

\bar{x} = nilai rata-rata data

n = jumlah data

Di sebuah kelas terdapat 40 siswa dengan tinggi badan sebagai berikut:

Kelas Rentang Tinggi Badan Frekuensi

CONTOH SOAL

Perhatikan tabel data berkelompok populasi suatu kelas berikut!

Nilai	Frekuensi
50 – 54	4
55 – 59	6
60 – 64	8
65 – 69	10
70 – 74	8
75 – 79	4

Pembahasan

Nilai	f_i	f_k	x_i	$f_i \cdot x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
50 – 54	4	4	52	208	-13	169	676
55 – 59	6	10	57	342	-8	64	384
60 – 64	8	18	62	496	-3	9	72
65 – 69	10	28	67	670	2	4	40
70 – 74	8	36	72	576	7	49	392
75 – 79	4	40	77	308	12	144	576
Jumlah				2.600			2140

Pada contoh soal simpangan baku ini, kita perlu mencari nilai rata-ratanya terlebih dahulu. Untuk mempermudah perhitungan, kita gunakan bantuan tabel.

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i}$$

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{2.600}{40} = 65$$

$$\sigma = \sqrt{\frac{\sum f_i (x_i - \mu)^2}{n}}$$

$$\sigma = \sqrt{\frac{2140}{40}} = \sqrt{53,5}$$

Berdasarkan tabel distribusi frekuensi di atas, dapat kita cari rata-ratanya dengan rumus berikut:

Berdasarkan tabel di atas, simpangan bakunya adalah:

Pengertian Varians

Variansi umumnya digunakan dalam statistik untuk mengukur tingkat kesamaan atau kedekatan dalam suatu kelompok. Variansi adalah nilai statistik yang sering digunakan untuk menentukan sejauh mana distribusi data dalam sebuah sampel dan sejauh mana titik-titik data individu mendekati rata-rata atau nilai tengah dari sampel itu sendiri.

Fungsi Varians

Fluktuasi umumnya digunakan oleh para analis atau orang yang terlibat dalam dunia statistik untuk mengetahui apakah sampel data yang diambil mewakili seluruh populasi. Mencari data yang tepat untuk suatu populasi sangatlah sulit. Oleh karena itu, untuk mempermudah pencarian, dipilihlah sampel data yang mewakili seluruh populasi. Rumus variansi data kelompok digunakan dalam statistik untuk menentukan kedekatan distribusi data dalam sampel tersebut. Variansi adalah nilai statistik yang digunakan untuk menentukan bagaimana data tersebar relatif terhadap rata-rata atau nilai tengah dari suatu kumpulan data. Dikutip dari buku Informatika Statistik oleh Drs. Syafril, M.Pd (2019), statistik adalah ilmu yang mempelajari metode yang digunakan dalam mengumpulkan,

menyajikan, menganalisis, dan menginterpretasi data kuantitatif (dalam bentuk angka), sehingga kesimpulan dan makna dapat diambil dari data tersebut.

Rumus Varians

Data Kelompok

Rumus variansi data kelompok dalam statistik adalah nilai yang digunakan untuk berbagai tujuan dalam sekelompok data yang disajikan. Dengan rumus variansi, kita dapat melihat bagaimana data tersebar dalam populasi atau sampel. Secara sederhana, variansi adalah ukuran statistik yang menggambarkan sejauh mana penyebaran data dari nilai rata-rata. Simbol yang digunakan untuk menyatakan variansi adalah σ^2 untuk data populasi.

Variansi dapat digunakan untuk mengukur sejauh mana data tersebar di sekitar nilai rata-rata. Tidak hanya itu, variansi juga memberikan pemahaman tentang bagaimana nilai-nilai individu dalam sampel atau populasi berbeda satu sama lain.

Contoh dan Pembahasan Soal

Cara Mencari Nilai Varians Data Kelompok

Tentukan Varians dari tabel berikut :

1. 150 – 154 cm 5
2. 155 – 159 cm 10
3. 160 – 164 cm 15
4. 165 – 169 cm 8
5. 170 – 174 cm 2

1. Hitunglah simpangan baku dari data tersebut.

Penyelesaian :

2. Hitung nilai tengah dari setiap kelompok.

$$\text{Kelas 1: } (150 + 154) / 2 = 152$$

$$\text{Kelas 2: } (155 + 159) / 2 = 157$$

$$\text{Kelas 3: } (160 + 164) / 2 = 162$$

$$\text{Kelas 4: } (165 + 169) / 2 = 167$$

$$\text{Kelas 5: } (170 + 174) / 2 = 172$$

3. Hitung nilai rata-rata dari nilai tengah kelompok.

$$\text{Rata - rata} = ((152 * 5) + (157 * 10) + (162 * 15) + (167 * 8) + (172 * 2))/40$$

$$\text{Rata-rata} = 162.1$$

4. Hitung penyimpangan setiap nilai tengah kelompok dari rata-ratanya.

Penyimpangan = nilai tengah kelompok - rata-rata

$$\text{Kelas 1: } 152 - 162.1 = -10.1$$

$$\text{Kelas 2: } 157 - 162.1 = -5.1$$

$$\text{Kelas 3: } 162 - 162.1 = -0.1$$

$$\text{Kelas 4: } 167 - 162.1 = 4.9$$

$$\text{Kelas 5: } 172 - 162.1 = 9.9$$

5. Hitung Varians.

Varians = (frekuensi * nilai penyimpangan²) / jumlah data

$$\text{Varians} = ((5 * (-10.1)^2) + (10 * (-5.1)^2) + (15 * (-0.1)^2) + (8 * 4.9^2) + (2 * 9.9^2))/40$$

$$\text{Varians} = 98.49 / 40$$

$$\text{Varians} = 2.46$$

6. Hitung Simpangan Baku.

Simpangan baku = akar kuadrat dari varians

Simpangan baku = akar kuadrat dari 2.46

Simpangan baku = 1.57

Jadi, simpangan baku dari data tersebut adalah 1.57

CONTOH SOAL

Tentukan varians dari tabel berikut!

<u>Nilai</u>	<u>Frekuensi</u>
63-67	3
68-72	2
73-77	7
78-82	3
83-87	4
88-92	1
Total	20

Untuk mencari nilai varians, maka langkah yang perlu dilakukan adalah:

1. Tentukan nilai tengah dari setiap kelompok yang ada. Selanjutnya hitung jumlah nilai kelompok dengan cara nilai tengah dikali frekuensi.

<u>Nilai</u>	<u>Frekuensi</u>	x_i	$f_i \cdot x_i$
63-67	3	65	195
68-72	2	70	140
73-77	7	75	525
78-82	3	80	240
83-87	4	85	340
88-92	1	90	90
Total	20		1.530

2. Hitung nilai rata-rata dengan cara jumlah nilai kelompok dibagi dengan total data tersebut.

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{195 + 140 + 525 + 240 + 340 + 90}{20} = \frac{1530}{20} = 76,5$$

3. Hitung simpangan setiap kelompok dengan cara mengkalikan frekuensi dengan kuadrat nilai tengah yang dikurang rata-rata data.

<u>Nilai</u>	<u>Frekuensi</u>	x_i	$f_i \cdot x_i$
63-67	3	65	195
68-72	2	70	140
73-77	7	75	525
78-82	3	80	240
83-87	4	85	340
88-92	1	90	90
Total	20		1.530

4. Jumlah simpangan setiap kelompok dibagi dengan total data disebut varians.

$$s^2 = \frac{\sum f_i (x_i - \bar{x})^2}{n} = \frac{1065}{20} = 53,25$$

Jadi, varians kelompok tersebut adalah 53,25

Data Yang Tidak Dikelompokkan

Varian mengukur rata-rata kuadrat deviasi setiap data dari rata-ratanya. Untuk data tidak dikelompokkan, rumusnya:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Dimana:

σ : Varian populasi

x_i : nilai data ke - 1

\bar{x} : Rata-rata data ($\bar{x} = \frac{\sum x_i}{n}$)

n: Jumlah data

Contoh soal

seorang guru matematika bernama Ibu Sarah ingin menganalisis sebaran nilai ujian semester matematika di kelas XI IPA 1. Analisis ini bertujuan untuk mengevaluasi tingkat pemahaman siswa dan efektivitas metode pengajaran yang diterapkan. Ibu Sarah mengumpulkan nilai dari 15 siswa untuk dianalisis menggunakan metode variasi data tidak dikelompokkan.

Data Nilai:

75, 82, 68, 90, 85, 78, 92, 70, 88, 83, 95, 65, 87, 73, 80

Analisis Statistik Lengkap:

a. Menghitung Rata-rata (Mean)

$$\begin{aligned}\text{Mean} &= \sum X \div n \\ &= (75 + 82 + 68 + 90 + 85 + 78 + 92 + 70 + 88 + 83 + 95 + 65 + 87 + \\ &73 + 80) \div 15 \\ &= 1211 \div 15 \\ &= 80.73\end{aligned}$$

b. Menghitung Simpangan (Deviasi)

Simpangan = $X_i - \text{Mean}$

- Untuk nilai 75: $75 - 80.73 = -5.73$
- Untuk nilai 82: $82 - 80.73 = 1.27$
- Untuk nilai 68: $68 - 80.73 = -12.73$
- Untuk nilai 90: $90 - 80.73 = 9.27$
- Untuk nilai 85: $85 - 80.73 = 4.27$
- Untuk nilai 78: $78 - 80.73 = -2.73$
- Untuk nilai 92: $92 - 80.73 = 11.27$
- Untuk nilai 70: $70 - 80.73 = -10.73$
- Untuk nilai 88: $88 - 80.73 = 7.27$
- Untuk nilai 83: $83 - 80.73 = 2.27$
- Untuk nilai 95: $95 - 80.73 = 14.27$
- Untuk nilai 65: $65 - 80.73 = -15.73$

- Untuk nilai 87: $87 - 80.73 = 6.27$
- Untuk nilai 73: $73 - 80.73 = -7.73$
- Untuk nilai 80: $80 - 80.73 = -0.73$

c. Menghitung Simpangan Kuadrat
 Simpangan Kuadrat = $(X_i - \text{Mean})^2$

- $(-5.73)^2 = 32.83$
- $(1.27)^2 = 1.61$
- $(-12.73)^2 = 162.05$
- $(9.27)^2 = 85.93$
- $(4.27)^2 = 18.23$
- $(-2.73)^2 = 7.45$
- $(11.27)^2 = 127.01$
- $(-10.73)^2 = 115.13$
- $(7.27)^2 = 52.85$
- $(2.27)^2 = 5.15$
- $(14.27)^2 = 203.63$
- $(-15.73)^2 = 247.43$
- $(6.27)^2 = 39.31$
- $(-7.73)^2 = 59.75$
- $(-0.73)^2 = 0.53$

d. Menghitung Jumlah Simpangan Kuadrat

$$\sum(X_i - \text{Mean})^2 = 1158.89$$

e. Menghitung Variasi

$$\begin{aligned} \text{Varians} &= \sum(X_i - \text{Mean})^2 \div n \\ &= 1158.89 \div 15 \\ &= 77.26 \end{aligned}$$

f. Menghitung Standar Deviasi

$$\begin{aligned} \text{Standar Deviasi} &= \sqrt{\text{Varians}} \\ &= \sqrt{77.26} \\ &= 8.79 \end{aligned}$$

5. Interpretasi Varians

- a. Varians kecil: Data cenderung dekat dengan rata-rata, menunjukkan konsistensi yang tinggi.

- b. Varians besar: Data tersebar jauh dari rata-rata, menunjukkan adanya variasi yang besar dalam dataset.
6. Hubungan Varians dengan Standar Deviasi
- Standar deviasi adalah akar kuadrat dari varians dan sering digunakan untuk interpretasi karena nilainya berada pada satuan yang sama dengan data aslinya.
- Standar Deviasi = $\sqrt{\text{Varians}}$
7. Penggunaan Varians dalam Statistik
- a. Analisis data: Untuk mengukur seberapa bervariasi data dalam satu populasi atau sampel.
 - b. Statistik inferensial: Sebagai dasar untuk menghitung standar deviasi, ANOVA, dan uji hipotesis.
 - c. Model probabilitas: Untuk menggambarkan distribusi peluang suatu variabel acak.
 - d. Manajemen risiko: Untuk mengukur volatilitas dalam data keuangan, seperti harga saham.
8. Contoh Aplikasi dalam Kehidupan Nyata
- a. Bisnis: Varians dalam penjualan digunakan untuk memahami fluktuasi dalam pendapatan.
 - b. Pendidikan: Varians nilai siswa digunakan untuk mengevaluasi keberhasilan program pembelajaran.
 - c. Keuangan: Varians harga saham digunakan untuk mengukur risiko investasi.
9. Keterbatasan Varians
- Nilainya dalam satuan kuadrat membuatnya sulit diinterpretasikan dibandingkan standar deviasi.
- Rentan terhadap nilai ekstrem (outliers) yang dapat meningkatkan varians secara signifikan.

BAB 5 PENYAJIAN DATA

PENYAJIAN DATA DALAM STATISTIK: KONSEP, METODE, DAN APLIKASINYA DALAM EKONOMI BISNIS

1. Pendahuluan

Dalam era informasi saat ini, data menjadi aset strategis yang sangat berharga. Setiap detik, miliaran byte data dihasilkan dari berbagai sumber – transaksi penjualan, interaksi pelanggan, sensor IoT, media sosial, laporan keuangan, dan lain-lain. Namun, data mentah saja tidak memiliki nilai jika tidak diolah, dianalisis, dan disajikan secara tepat. Di sinilah peran penyajian data dalam statistik menjadi krusial.

Penyajian data adalah proses mengorganisasi, merangkum, dan menampilkan informasi numerik atau kategorikal dalam bentuk yang mudah dipahami, baik melalui tabel, grafik, diagram, atau visualisasi interaktif. Tujuan utamanya adalah mengkomunikasikan pola, tren, hubungan, dan anomali dalam data kepada audiens yang relevan – baik itu manajer, investor, analis, atau pembuat kebijakan – sehingga mereka dapat mengambil keputusan yang berbasis bukti (*evidence-based decision making*). Dalam konteks ekonomi dan bisnis, penyajian data bukan hanya alat komunikasi, tetapi juga instrumen strategis. Sebuah laporan penjualan yang disajikan dalam bentuk grafik garis dapat mengungkap tren pertumbuhan bulanan. Diagram batang perbandingan laba antar divisi bisa membantu alokasi anggaran. Heatmap perilaku pelanggan di situs web *e-commerce* dapat mengoptimalkan desain antarmuka. Semua ini menunjukkan bahwa kualitas penyajian data secara langsung memengaruhi kualitas keputusan bisnis.

Dokumen ini akan membahas secara mendalam:

- a. Konsep dasar penyajian data dalam statistik
- b. Jenis-jenis data dan implikasinya terhadap metode penyajian
- c. Metode penyajian data: tabel dan grafik
- d. Prinsip-prinsip penyajian data yang efektif

- e. Alat dan teknologi modern untuk visualisasi data
- f. Studi kasus nyata dalam ekonomi bisnis
- g. Peran penyajian data dalam pengambilan keputusan strategis dan operasional

Tujuannya adalah memberikan pemahaman holistik tentang bagaimana penyajian data bukan hanya soal estetika, tetapi **inti dari literasi data dan kecerdasan bisnis**.

2. Konsep Dasar Penyajian Data dalam Statistik

a. Definisi dan Tujuan Penyajian Data

Penyajian data (data presentation) dalam statistik merujuk pada teknik dan metode yang digunakan untuk menyampaikan informasi statistik kepada pembaca atau audiens dengan cara yang jelas, ringkas, akurat, dan menarik. Ini merupakan tahap penting dalam siklus analisis data, yang biasanya mengikuti tahapan: pengumpulan → pembersihan → analisis → penyajian → interpretasi → pengambilan keputusan. Tujuan utama penyajian data meliputi:

- 1) Menyederhanakan kompleksitas: Data mentah sering kali voluminous dan tidak terstruktur. Penyajian data mereduksi kompleksitas tersebut menjadi bentuk yang lebih mudah dicerna.
- 2) Mengungkap pola dan tren: Visualisasi dapat menunjukkan hubungan yang tidak terlihat dalam tabel angka.
- 3) Memfasilitasi perbandingan: Misalnya, membandingkan kinerja dua produk dalam satu grafik.
- 4) Mendukung komunikasi: Menyampaikan temuan kepada pihak yang tidak memiliki latar belakang statistik.
- 5) Mendorong tindakan: Penyajian yang persuasif dapat memicu respons strategis, seperti perubahan harga, ekspansi pasar, atau inovasi produk.

b. Perbedaan antara Penyajian Data dan Analisis Data

Meskipun saling terkait, **penyajian data** dan **analisis data** adalah dua hal yang berbeda:

- 1) **Analisis data** berfokus pada *apa yang terjadi* dan *mengapa* – menggunakan teknik statistik inferensial, regresi, uji hipotesis, dll.
- 2) **Penyajian data** berfokus pada *bagaimana menyampaikan hasil analisis tersebut* agar bermakna bagi audiens.

Namun, keduanya harus selaras. Penyajian yang buruk dapat menyesatkan, meskipun analisisnya benar. Sebaliknya, penyajian yang cerdas dapat memperkuat insight dari analisis.

3. Jenis-Jenis Data dan Implikasinya terhadap Penyajian

Sebelum memilih metode penyajian, penting memahami jenis data yang dimiliki, karena setiap jenis data memiliki representasi visual yang optimal.

a. Data Kuantitatif vs. Kualitatif

Data Kuantitatif: Data numerik yang dapat diukur (misalnya: pendapatan, jumlah unit terjual, suhu).

Diskrit: Nilai terpisah (jumlah karyawan).

Kontinu: Nilai dalam rentang (berat badan, waktu).

Data Kualitatif (Kategorikal): Data non-numerik yang menggambarkan kategori atau atribut (misalnya: jenis kelamin, merek favorit, status pekerjaan).

Nominal: Tidak ada urutan (warna: merah, biru, hijau).

Ordinal: Ada urutan (tingkat kepuasan: sangat puas, puas, netral, tidak puas).

b. Implikasi terhadap Pemilihan Visualisasi

Data kuantitatif cocok untuk:

Histogram (distribusi frekuensi)

Grafik garis (tren waktu)

Scatter plot (hubungan antar variabel)

Data kualitatif cocok untuk:

Diagram batang (frekuensi kategori)

Pie chart (proporsi, meski sering dikritik)

Treemap (hierarki kategori)

Kombinasi keduanya (misalnya: rata-rata penjualan per wilayah) memerlukan visualisasi campuran seperti **bar chart dengan warna berbeda per kategori**.

4. Metode Penyajian Data: Tabel dan Grafik

a. Penyajian Data dalam Bentuk Tabel

Tabel adalah bentuk paling dasar dan universal dalam penyajian data. Kelebihannya:

Presisi tinggi (menampilkan angka eksak)

Cocok untuk data detail dan referensi

Mudah dibuat dan dipahami

Jenis Tabel Statistik:

- 1) Tabel Satu Arah: Menampilkan satu variabel (misalnya: jumlah penjualan per bulan).
- 2) Tabel Dua Arah (Cross-tabulation): Menampilkan hubungan antara dua variabel (misalnya: penjualan per produk per wilayah).
- 3) Tabel Distribusi Frekuensi: Menunjukkan seberapa sering nilai tertentu muncul.

Prinsip Desain Tabel yang Baik:

Judul jelas dan deskriptif

Label kolom dan baris informatif

Gunakan alignment yang tepat (angka kanan, teks kiri)

Hindari garis berlebihan (minimalis)

Sertakan total dan persentase jika relevan

b. Penyajian Data dalam Bentuk Grafik/Visualisasi

Grafik mengubah angka menjadi gambar, memanfaatkan kemampuan otak manusia dalam memproses visual. Berikut jenis-jenis utama:

1) Diagram Batang (Bar Chart)

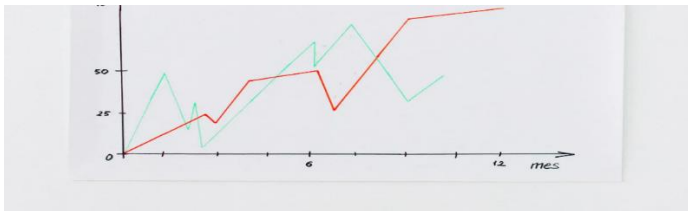


Digunakan untuk membandingkan kategori.

Vertikal (column chart) atau horizontal.

Contoh bisnis: Penjualan per produk, laba per cabang.

2) Grafik Garis (Line Chart)

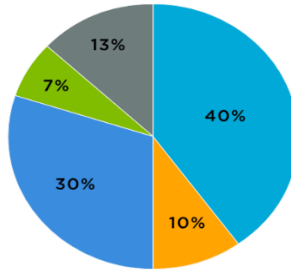


Ideal untuk data time-series.

Menunjukkan tren, fluktuasi, dan proyeksi.

Contoh bisnis: Pertumbuhan pendapatan bulanan, indeks saham.

3) Pie Chart

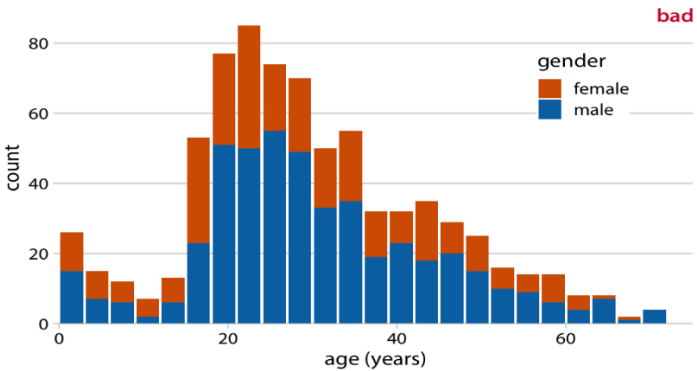


Menunjukkan proporsi bagian terhadap keseluruhan.

Kritik: Sulit membandingkan irisan yang ukurannya mirip.

Alternatif: Donut chart, stacked bar chart.

4) Histogram

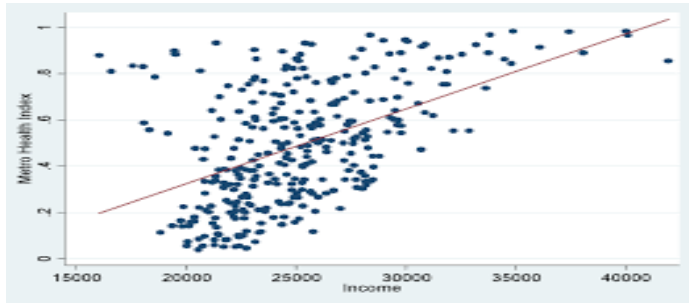


Menampilkan distribusi frekuensi data kuantitatif.

Mirip bar chart, tetapi batang saling menempel (karena data kontinu).

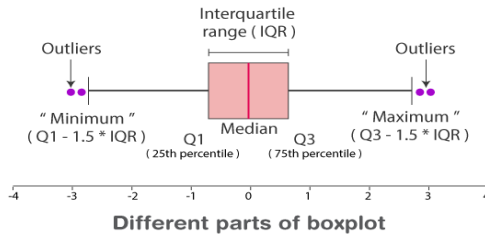
Contoh bisnis: Distribusi usia pelanggan, durasi kunjungan website.

5) Scatter Plot



Menunjukkan hubungan antara dua variabel kuantitatif.
Dapat mengungkap korelasi positif/negatif.
Contoh bisnis: Hubungan antara iklan dan penjualan.

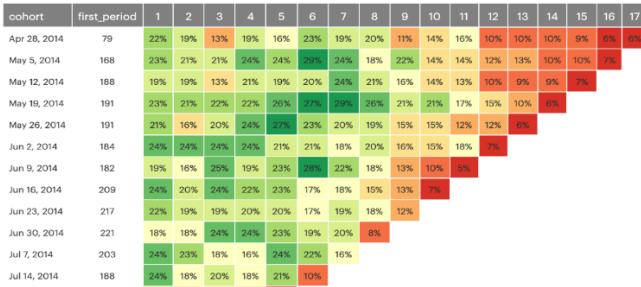
6) Box Plot (Box-and-Whisker Plot)



© Byjus.com

Menampilkan distribusi data melalui kuartil, median, dan outlier.
Berguna untuk membandingkan variasi antar kelompok.
Contoh bisnis: Perbandingan gaji antar departemen.

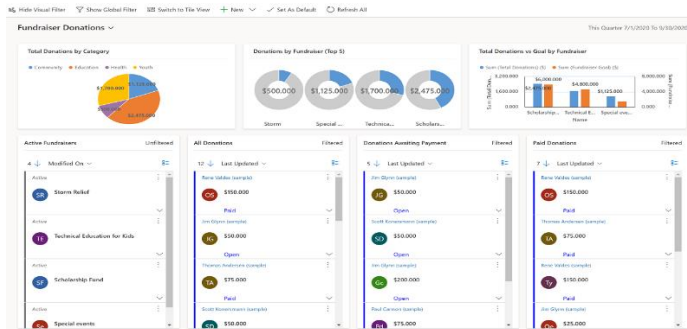
7) Heatmap



Menggunakan warna untuk merepresentasikan nilai dalam matriks.

Contoh bisnis: Matriks korelasi, aktivitas pengguna per jam/hari.

8) Dashboard Interaktif



Kombinasi beberapa visualisasi dalam satu tampilan.

Memungkinkan filter, drill-down, dan real-time update.

Alat: Power BI, Tableau, Google Data Studio.

5. Prinsip-Prinsip Penyajian Data yang Efektif

Edward Tufte, pakar visualisasi data, menekankan bahwa "**graphical excellence**" adalah seni mengkomunikasikan data secara kompleks dengan kesederhanaan, presisi, dan efisiensi. Berikut prinsip utamanya:

a. Kejelasan (Clarity)

Hindari elemen yang tidak perlu (*chart junk*): dekorasi berlebihan, 3D efek, bayangan. Gunakan label langsung, bukan legenda jika memungkinkan.

b. Akurasi (Accuracy)

Skala sumbu harus proporsional.

Jangan memanipulasi sumbu Y untuk memperbesar/mengecilkan perbedaan.

c. Relevansi (Relevance)

Sajikan hanya data yang mendukung tujuan analisis.

Hindari informasi yang membingungkan atau tidak relevan.

d. Konsistensi (Consistency)

Gunakan skema warna, font, dan format yang konsisten di seluruh laporan.

Warna harus memiliki makna (misalnya: merah = negatif, hijau = positif).

e. Audiens-Sentris

Sesuaikan kompleksitas dengan latar belakang audiens.

Eksekutif butuh ringkasan eksekutif; analis butuh detail teknis.

6. Alat dan Teknologi Modern untuk Visualisasi Data

Perkembangan teknologi telah merevolusi cara kita menyajikan data:

Microsoft Excel: Masih populer untuk laporan sederhana.

Tableau: Kuat untuk visualisasi interaktif dan dashboard.

Power BI: Terintegrasi dengan ekosistem Microsoft, ideal untuk bisnis.

Python (Matplotlib, Seaborn, Plotly): Fleksibel untuk analisis lanjutan.

R (ggplot2): Populer di kalangan akademisi dan data scientist.

Google Data Studio: Gratis, berbasis cloud, cocok untuk laporan marketing.

Alat-alat ini memungkinkan:

Real-time data updating

Drill-down interaktif

Integrasi multi-sumber data

Kolaborasi tim

7. Studi Kasus: Penyajian Data dalam Ekonomi Bisnis

a. Studi Kasus 1: Ritel Modern - Optimasi Persediaan dengan Visualisasi Penjualan

Latar Belakang:

Sebuah jaringan supermarket nasional mengalami masalah *overstock* dan *stockout* secara bersamaan di berbagai cabang. Manajemen ingin mengoptimalkan sistem persediaan berdasarkan pola permintaan riil.

Data yang Dikumpulkan:

Penjualan harian per produk (SKU)

Stok akhir hari

Waktu kedatangan pasokan

Musim/promosi

Penyajian Data yang Digunakan:

- 1) **Heatmap Penjualan per Hari & Produk:** Menunjukkan hari-hari dengan penjualan tinggi untuk tiap kategori (misalnya: susu laris di Senin pagi).
- 2) **Line Chart Tren Bulanan:** Mengidentifikasi musiman (misalnya: peningkatan penjualan mie instan saat musim hujan).
- 3) **Scatter Plot Hubungan Antara Harga Diskon dan Volume Penjualan:** Mengukur elastisitas harga.

Keputusan Bisnis yang Diambil:

Menyesuaikan jadwal pengiriman stok berdasarkan heatmap.

Meningkatkan stok produk musiman 2 minggu sebelum puncak permintaan.

Mengoptimalkan diskon berdasarkan respons elastisitas.

Hasil:

Pengurangan biaya penyimpanan 18%

Penurunan stockout sebesar 32%

Peningkatan margin kotor 5%

Pelajaran: Visualisasi data memungkinkan manajemen melihat pola mikro yang tidak terlihat dalam laporan agregat bulanan.

b. Studi Kasus 2: Startup Fintech - Analisis Risiko Kredit dengan Dashboard

Latar Belakang:

Sebuah startup fintech menawarkan pinjaman mikro berbasis aplikasi. Mereka perlu menilai risiko kredit calon peminjam secara cepat dan akurat.

Data yang Digunakan:

Riwayat transaksi bank

Skor kredit alternatif (dari media sosial, pembayaran tagihan)

Data demografis

Riwayat pinjaman sebelumnya

Penyajian Data:

Dashboard Kredit Risk Score: Gabungan indikator warna (merah/kuning/hijau) berdasarkan skor risiko.

Bar Chart Distribusi Default Rate per Segmen Usia.

ROC Curve Visualization: Untuk mengevaluasi performa model prediksi.

Keputusan Bisnis:

Menolak aplikasi dengan skor risiko merah.

Menawarkan suku bunga diferensial berdasarkan segmen risiko.

Melatih ulang model prediksi setiap bulan berdasarkan data baru.

Hasil:

Tingkat gagal bayar turun dari 12% menjadi 6% dalam 6 bulan.

Efisiensi proses underwriting meningkat 40%.

Pelajaran: Penyajian data real-time memungkinkan keputusan kredit yang cepat dan akurat, krusial dalam bisnis berbasis volume tinggi.

c. Studi Kasus 3: Manufaktur - Pengendalian Kualitas dengan Control Chart

Latar Belakang: Pabrik otomotif mengalami peningkatan keluhan pelanggan terkait cacat produksi.

Data:

Jumlah cacat per shift

Jenis cacat (goresan, salah rakit, dll.)

Mesin yang digunakan

Penyajian Data:

Control Chart (Shewhart Chart): Menampilkan batas kendali atas/bawah (UCL/LCL).

Pareto Chart: Menunjukkan 20% jenis cacat yang menyebabkan 80% masalah.

Keputusan:

Menghentikan mesin yang melewati UCL.

Fokus perbaikan pada 3 jenis cacat utama (prinsip Pareto).

Hasil:

Cacat produksi turun 55% dalam 3 bulan.

Penghematan biaya recall dan garansi.

Pelajaran: Visualisasi statistik kendali kualitas memungkinkan deteksi dini masalah proses.

8. Peran Penyajian Data dalam Pengambilan Keputusan Bisnis

a. Keputusan Strategis vs. Operasional

Strategis (jangka panjang): Ekspansi pasar, investasi R&D, merger.

→ Butuh visualisasi tren makro, benchmarking industri, skenario simulasi.

Operasional (harian/mingguan): Penjadwalan staf, penetapan harga, manajemen inventori.

→ Butuh dashboard real-time, alert otomatis, drill-down detail.

b. Data-Driven Culture

Organisasi yang sukses membangun **budaya berbasis data**, di mana:

Setiap keputusan didukung oleh data.

Visualisasi data tersedia di semua level.

Literasi data menjadi bagian dari pelatihan karyawan.

Contoh: Amazon menggunakan "six-page memo" yang selalu didukung oleh data visual untuk rapat eksekutif—bukan PowerPoint.

c. Menghindari Bias Kognitif

Penyajian data yang baik juga membantu mengurangi bias manusia:

Confirmation bias: Data visual yang objektif menantang asumsi.

Anchoring: Grafik tren menunjukkan perubahan, bukan hanya angka awal.

Overconfidence: Interval kepercayaan dalam grafik mengingatkan ketidakpastian.

9. Tantangan dan Etika dalam Penyajian Data

a. Manipulasi Visual (Chart Junk & Distorsi)

Contoh umum:

Memotong sumbu Y untuk memperbesar perbedaan kecil.

Menggunakan pie chart 3D yang menyesatkan proporsi.

Memilih skala waktu yang bias (misalnya: hanya menampilkan 3 bulan terakhir saat tren sedang naik).

Solusi: Transparansi, dokumentasi metodologi, dan peer review.

b. Privasi dan Keamanan Data

Dalam bisnis, data sering mengandung informasi sensitif.

Penyajian harus:

Menghindari identifikasi individu (anonymization).

Mematuhi regulasi (GDPR, UU PDP).

Menggunakan akses berbasis peran (role-based access).

Tugas dalam menyampaikan informasi adalah untuk menggambarkan perubahan suatu kondisi atau membuat perbandingan pada waktu tertentu.

Tabel

Tabel merupakan kumpulan angka-angka yang disusun menurut kategori-kategori sehingga memudahkan untuk pembuatan analisis data (Supranto, 20). Terdapat tiga jenis tabel, yaitu:

- 1) Tabel satu arah ialah tabel yang memuat keterangan mengenai satu hal atau satukarakteristik saja.

Tabel 2.1 Banyaknya produksi kayu menurut jenis 1984/1985

Jenis	Banyaknya (x1000m ³)
(1)	(2)
Kayu Rimba	14.951
Kayu jati	1.006
Jumlah	15.957

- 2) Tabel dua arah ialah tabel yang menunjukkan dua hal atau dua karakteristik.

Tabel 2.2 Jumlah mahasiswa universitas persada jakarta, menurut fakultas agama, 1986

Fakultas	Katolik	Bukan Katolik	Jumlah
Fak. Ekonomi	266	292	558
Fak.Teknik	150	162	312
Fak.Hukum	55	65	120
Fak.Kedokteran	273	186	459
Jumlah	924	861	1.785

- 3) Tabel tiga arah ialah tabel yang menunjukkan tiga kategori atau tiga karakteristik.

Tabel 2.3 Jumlah kendaraan perusahaan x menurut umur, merek & jenis, pada tahun 1986

Golongan										Jumlah
	Sedan	Bis	Dll	Sedan	Bis	Dll	Sedan	Bis	Dll	
1	2	3	4	5	6	7	8	9	10	11
< 1th	5	2	1	4	1		3	1	1	18
1th < 2th										
2th < 3th										
3th < 4 th										
4th < 5 th										
≥ 5 th										
Jumlah										

Ukuran Bentuk

Ukuran bentuk adalah suatu ukuran yang digunakan untuk menentukan bentuk kelengkungan data yang telah diperoleh atau diproses. Berdasarkan ukuran bentuk, data dibagi menjadi dua, yaitu sebagai berikut.

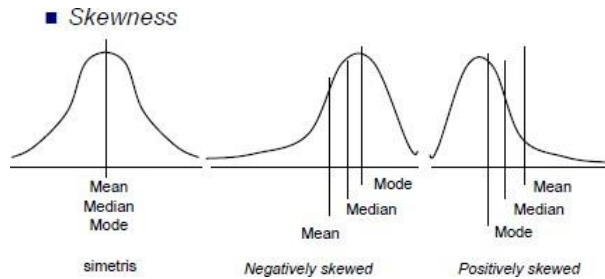
a) Skewness

Skewness atau yang sering disebut sebagai derajat kemiringan, adalah angka yang dapat menunjukkan apakah bentuk kelengkungan distribusi frekuensi condong. Distribusi yang memiliki ekor lengkungan lebih panjang ke kanan disebut *skewed to the right* atau skewness positif. Sebaliknya, jika ekornya lebih panjang ke kiri, disebut *skewed to the left* atau skewness negatif. Dalam distribusi yang miring, nilai rata-rata (mean)

cenderung berada di sisi ekor yang lebih panjang, dengan modus (mode) berada pada frekuensi tertinggi. Skewness dapat dihitung dengan rumus berikut :

$$\text{Skewness} = \frac{\sum \left(\frac{x_i - \bar{x}}{s} \right)^3}{(n-1)(n-2)}$$

x_i = data ke - $i\bar{x}$ = x rata - ratas = standar deviasi (simpangan baku)



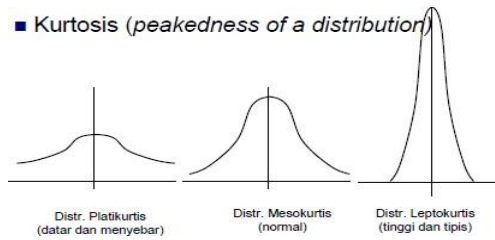
b) Kurtosis

Kertosis adalah ukuran statistik yang menggambarkan bentuk distribusi data, khususnya tingkat "tajamnya" atau "rata-ratanya" distribusi. Kertosis membantu memahami apakah distribusi data lebih "tajam" atau "rata" dibandingkan dengan distribusi normal. Distribusi normal memiliki kurtosis = 3, sementara distribusi yang leptokurtic memiliki kurtosis > 3, dan platikurtik < 3. Berikut rumus kurtosis:

$$\text{Kurtosis} = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

n = banyaknya data/sampel
 \bar{x} = x rata - rata

x_i = data ke - i
 s = standar devias



Perhitungan Pada Data

Statistika membagi data menjadi dua jenis, yaitu: data tunggal dan data kelompok, yang memudahkan analisis. Mengetahui contoh-contoh data tunggal akan membuat siswa lebih terampil dalam melakukan perhitungan data dalam statistika.

Perhitungan data dalam statistika meliputi perhitungan rata-rata (mean), median, modus, rentang, kuartil, deviasi kuartil, deviasi rata-rata, varians, dan deviasi standar. Baik data tunggal maupun data kelompok memiliki metode perhitungan masing-masing.

Pengolahan Data Tunggal

Menurut Koeshartati Saptorini (2009), data tunggal didefinisikan sebagai kumpulan data dengan jumlah yang terbatas sehingga tidak menunjukkan pola pengulangan yang signifikan.

Data individual juga disebut data yang tidak dikelompokkan. Contoh data pribadi yang kami temukan setiap hari meliputi: Misalnya: data umur, data tinggi badan, data berat badan, dan lain-lain. Contoh data individual statistik meliputi pengukuran kuantitatif seperti ketinggian gedung tertinggi dan suhu saat ini

Pengolahan dan analisis data individual memungkinkan pemahaman mendalam tentang karakteristik populasi. Berikut adalah beberapa contoh titik data individual dalam statistik:

- Nilai ulangan matematika Ani adalah 7,8, 6, 10, 9,10
- Nilai ulangan harian matematika kelas VIII C SMP Bumi Pertiwi: 8, 7, 6, 8, 9, 9, 8, 7, 6, 9, 7, 8, 8, 7, 9, 8, 8, 6, 8, 7
- Berikut adalah data berat badan (dalam kg) siswa laki-laki kelas VIII.2 SMPN 6 Parepare: 47 57 53 50 45 48 52 49 55 57 46 57
- Umur siswa kelas XI A adalah sebagai berikut 30kg, 55kg, 49kg, 48kg, 35kg, 60kg, 65kg, 54kg, 48kg, 67kg, 57kg
- Jumlah buku yang dimiliki siswa di kelas: 5, 3, 2, 4, 1, 3
- Nilai matematika siswa di ujian: 80, 90, 85, 70, 95, 80
- Jumlah siswa per kelas di sebuah sekolah: 30, 25, 28, 32, 26, 30

Contoh Soal

Nilai matematika yang diperoleh siswa kelas VIII C SMP Bumi Pertiwi disajikan dalam data berikut:

8,7,6,8,9,9,8,7,6,9,7,8,8,7,9,8,8,6,8,7

Berdasarkan data di atas, carilah mean, median, dan modus!

Jawab:

$$\text{Mean} = \frac{X_1 + X_2 + X_3 + X_4 + X_5 + X \dots n}{n}$$

$$\text{Mean} = \frac{8 + 7 + 6 + 8 + 9 + 9 + 8 + 7 + 6 + 9 + 7 + 8 + 8 + 7 + 9 + 8 + 8 + 6 + 8 + 7}{20}$$

$$\text{Mean} = 7,65$$

$$\text{Median} = \frac{1}{2}(X_{(n/2)} + X_{(N/2+1)})$$

$$\text{Me} = \frac{1}{2} (X_{(20/2)} + X_{(20/2+1)})$$

$$\text{Me} = \frac{1}{2} (X_{10} + X_{11})$$

$$\text{Me} = \frac{1}{2} (8 + 8)$$

$$\text{Me} = 8$$

Modus = Pada data yang disajikan, nilai 6 muncul 3 kali, nilai 7 muncul 5 kali, nilai 8 muncul 8 kali, dan nilai 9 muncul 4 kali. Jadi, modus pada soal di atas adalah 8.

BAB 6

KOEFSISIEN KORELASI DAN BERBAGAI BENTUK KURVA DALAM STATISTIK

Analisis korelasi adalah metode analisis temuan yang diperoleh dari sampel yang digunakan untuk mengetahui kekuatan hubungan antar Kumpulan data. Tujuan korelasi sendiri adalah untuk mencari hubungan korelasi antar variabel, tingkat kekuatan hubungannya, memperoleh kepastian hubungan tersebut. Korelasi positif dan negative adalah dua jenis korelasi. Hubungan statistika antara dua variabel yang bergerak searah disebut korelasi positif. Variabel-variabel lain biasanya meningkat seiring dengan peningkatan suatu variabel dan menurun seiring dengan penurunan variabel lainnya. Beberapa contoh dari korelasi positif adalah sebagai berikut:

1. Jika titik-titik korelasi dihubungkan satu sama lain, maka akan dihasilkan garis lurus miring ke kanan yang menunjukkan korelasi positif tertinggi.
2. Korelasi positif yang tinggi atau kuat terjadi jika satu atau lebih titik pancaran pada peta korelasi mulai menyimpang dari garis lurus, terpancar, atau miring ke kanan di sekitar garis lurus.
3. Bila titik-titik pada peta korelasi semakin memburuk atau menyebar di sekitar garis lurus dengan kemiringan ke kanan, hal ini menunjukkan korelasi positif rendah atau sedang.

Korelasi yang berjalan berlawanan arah antara dua variabel atau lebih dan jika variabel independen naik sedang kan variabel dependen turun atau sebaliknya, hal ini disebut sebut korelasi negative. Beberapa contoh korelasi negative adalah sebagai berikut:

1. Korelasi negative yang minimal atau sempurna
2. Apabila titik-titik pada peta korelasi dihubungkan satu sama lain, maka akan terbentuk garis lurus memanjang ke kiri. Jika beberapa titik pada peta korelasi tersebar, pada garis lurus dengan kemiringan ke kiri, atau mulai menjauhi garis lurus, maka terdapat korelasi negative tinggi dan kuat.
3. Korelasi negative lemah atau rendah menjadi semakin parah atau memanjang dan menjauhi dari garis lurus yang condong ke kiri.

a. Pengujian Pada Korelasi

Para peneliti sering menggunakan metode berikut untuk menguji korelasi:

1) Pearson Product Moment

Rumus korelasi product moment Karl Pearson dapat digunakan untuk memperoleh korelasi. Jika terdapat skala-skala pengukuran variabel-variabel yang berkorelasi. Menggunakan rumus untuk menentukan korelasi dalam regresi linear adalah sebagai berikut:

$$r_{xy} = \frac{N \sum XY - \sum X \times \sum Y}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}}$$

Keterangan: Type equation here.

r_{xy} : Korelasi antara variabel X dan variabel Y

N : Jumlah dari data

X : Variabel Independen

Y : Variabel Dependen

2) Rank Spearman

Koefisien korelasi spearman menentukan derajat hubungan ketika variabel independent dan dependen berada pada skala ordinal. Berikut rumus korelasi spearman adalah

$$\rho = 1 - \frac{6 \sum d_i^2}{N^3} - N$$

Keterangan:

d : beda antara 2 pengamatan berpasangan

N : total pengamatan

ρ : koefisien korelasi spearman

b. Kriteria Korelasi

Statistika yang menunjukkan derajat hubungan antara dua variabel, atau lebih tepatnya satu variabel terikat, disebut koefisien korelasi. Hasil koefisien korelasi berganda secara konsisten positif. Korelasi antara dua variabel yang diwakili oleh koefisien korelasi yang dimulai dari -1 hingga korelasi.

Kriteria berikut yang digunakan untuk menilai bagaimana dua variabel berhubungan satu sama lain, yaitu:

$r = 0$ artinya tidak ada hubungan korelasi

$0 < r \leq 0,25$ artinya hubungan korelasi yang sangat lemah

$0,25 < r \leq 0,5$ artinya korelasi sedang

$0,5 < r \leq 0,75$ artinya korelasi kuat

$0,75 < r \leq 0,99$ artinya korelasi sangat kuat

$r = 1$ artinya korelasi sempurna

c. Asumsi Pada Korelasi

Analisis korelasi dilakukan terhadap data yang terdistribusikan dengan baik. Misalnya data yang simetris sempurna atau berdistribusi normal.

H_0 : Distribusi normal populasi

H_1 : Populasi tidak berperilaku normal sesuai dengan kriteria pengujiannya.

Jika nilai α 0,05 dan nilai sig. 0,05 maka terima H_0 salah satu paling efektif untuk mengupulkan data. Dengan kata lain, nilai sig juga dikenal sebagai taraf keberatian, adalah kemampuan untuk menentukan nilai sampel yang sama tingginya dengan nilai yang secara pasti ditetapkan sebagai H_0 . Hipotesis nol akan lebih sulit diterima bila nilai sig. nya lebih kecil. Sulit untuk mempertahankan hipotesis yang kecil untuk didukung. Jelas bahwa sig bergantung pada beberapa faktor, seperti ukuran sampel, jaraknya nilai H_0 dan nilai H_1 yang sesungguhnya, rentang sampel. Biasanya, ukuran sampel yang lebih besar menghasilkan sig yang lebih kecil. Demikian pula, sig yang lebih kecil dihasilkan oleh H_0 yang lebih besar dibandingkan H_1 yang umumnya lebih mudah dipisahkan.

Karakteristik Korelasi

Menurut Sukardi (2004:166), Penelitian korelasi mempunyai tiga ciri penting bagi penelitian yang enggan menggunakannya. Ciri-ciri tiga ini adalah sebagai berikut.

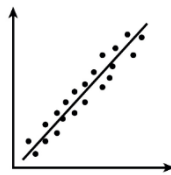
1. Analisis korelasi dikatakan akurat jika variabelnya kompleks dan peneliti tidak mampu memanipulasi dan mengendalikannya seperti dalam eksperimen.

2. Memungkinkan variabel tersebut diintensifikan dalam lingkungan (setting) nyata
3. Memungkinkan peneliti memperoleh asosiasi derajat yang signifikan.

Uji Korelasi

Salah satu cara untuk memikirkan korelasi adalah suatu hubungan. Menemukan pola dan derajat hubungan antara dua variabel atau lebih merupakan tujuan analisis korelasi. Ada beberapa cara untuk mengkategorikan arah hubungan antara dua variabel sebagai berikut:

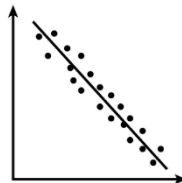
1. Direct Correlation (positive correlation), ketika satu variabel berubah, variabel lain sering kali mengikuti dan bergerak ke arah yang sama.



Positive Correlation

Gambar 1. 1 Korelasi Positif

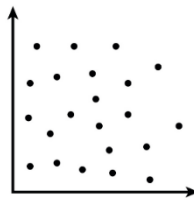
2. Inverse Correlation (negative correlation), biasanya, ketika satu variabel berubah, variabel lain akan mengikuti dan bergerak ke arah yang berlawanan.



Negative Correlation

Gambar 1. 2 Korelasi Negatif

3. Nihil Correlation, ada hubungan yang tidak teratur antara kedua variabel.



No Correlation

Gambar 1. 3 Korelasi Nol

Huruf (r) sendiri sering digunakan sebagai koefisien korelasi. Koefisien korelasi adalah nilai numerik yang berkisar antara 0 hingga +1 atau 0 hingga -1. Hubungan yang kuat ditunjukkan dengan korelasi yang mendekati +1 atau -1, sedangkan hubungan yang lemah ditunjukkan dengan korelasi yang mendekati 0. Tidak ada hubungan sama sekali jika korelasi kedua variabel bernilai 0. Kedua mempunyai hubungan sempurna jika korelasinya + atau -1. Arah hubungan antara variabel ditunjukkan dengan notasi positif (+) dan notasi negative (-). Notasi positif (+) menunjukkan bahwa kedua variabel mempunyai hubungan yang positif, artinya jika salah satu variabel meningkat maka variabel lainnya akan mengikuti. Pada kedua variabel dalam notasi negative maka berbanding terbalik, kenaikan salah satu variabel justru mengakibatkan penurunan variabel lainnya.

Pengukuran skala

Skala pengukuran yang digunakan dalam penelitian menentukan metode statistik yang digunakan untuk menguji data. Karena pengukuran adalah penerapan norma-norma tertentu pada situasi tertentu, maka hukum atau peraturan yang berlaku pada suatu situasi mungkin mempunyai dampak yang berbeda pada situasi lain, sehingga menghasilkan jenis pengukuran yang berbeda. Skala pengukuran adalah hukum yang menentukan bagaimana angka digunakan sebagai simbol. Data dapat dibagi menjadi empat

kelompok menurut jenis skala pengukurannya: interval, rasio, ordinal, dan nominal (Iqbal Hasan, 2002:34). Menurut Daniel (1989: 18–21), ada tiga kategori utama pengukuran skala yang dapat diidentifikasi: interval, rasio, ordinal, dan nominal. Keempat skala pengukuran tersebut dijelaskan sebagai berikut.

1. Skala nominal

Skala nominal merupakan skala yang paling dasar dari keempat jenis skala pengukuran karena hanya digunakan untuk mengklasifikasikan benda, individu, atau sifat tertentu menurut nama atau kategori. Skala ini tidak menunjukkan peringkat atau ukuran; sebaliknya, ia menggunakan angka atau simbol sebagai penanda. Misalnya dalam proses produksi, angka 0 menunjukkan produk tidak cacat, sedangkan angka 1 menunjukkan kumpulan barang cacat. Dalam hal ini angka 0 dan 1 digunakan sebagai simbol untuk membedakan barang yang cacat dan yang tidak.

2. Skala Ordinal

Skala ordinal merupakan nama lain dari skala ordinal. Karena memungkinkan pengukuran berdasarkan perbandingan relatif antara benda atau kejadian, skala ini lebih maju dibandingkan skala nominal. Skala ordinal memungkinkan seseorang untuk membandingkan sesuatu atau kejadian berdasarkan seberapa banyak kesamaan yang mereka miliki. Setiap benda dapat disusun menggunakan skala ini berdasarkan letaknya pada skala tersebut. Misalnya, bergantung pada berapa lama waktu yang dibutuhkan untuk melewati garis finis dalam suatu kompetisi lari, peserta mungkin menduduki peringkat pertama, kedua, dan ketiga.

3. Skala Interval

Jika benda-benda atau peristiwa yang digunakan dapat dibandingkan satu sama lain dan kemudian dengan yang lain, dan jika perbedaan antara satu dan lainnya mempunyai ciri-ciri (yaitu tetap pengukurannya Tunggal), maka interval skala dapat diterapkan. Tidak ada titik nilai nol mutlak untuk skala interval. Contoh skala interval adalah perubahan suhu antara Fahrenheit dan Celcius. Nol baik pada termometer yang bertuliskan Fahrenheit atau Celsius tidak menunjukkan bahwa tidak ada suhu;

melainkan mengacu pada apa pun yang diukur dengan menggunakan instrumen yang disebutkan di atas.

4. Skala Rasio

Apabila pengukuran-pengukuran yang dilakukan mempunyai interval dan skala jarak, maka skala yang dimaksud disebut skala rasio. Salah satu contohnya adalah pengukuran berat badan. Menurut skala rasio, dapat dikatakan bahwa orang yang memiliki berat badan 90 kg atau lebih secara teratur umumnya lebih kuat dibandingkan orang yang memiliki berat badan 45 kg atau kurang. Skala rasio mempunyai nilai mutlak nol; Artinya, jika tanaman tumbuh dari minggu sampai 2 nol, maka tidak ada pertumbuhan. Skala rasio merupakan skala yang pengukuran tingkatnya paling tinggi.

Analisis Regresi Linear Sederhana

Bentuk dan model hubungan antara dua variabel atau lebih, serta variasi beberapa variabel yang berkaitan dengan variabel tertentu dalam suatu fenomena, semuanya dipelajari melalui analisis regresi yang juga digunakan sebagai memperkirakan sejauh mana pengaruh suatu variabel atau lebih variabel independent berhubungan dengan variabel dependen. Regresi linear sederhana didasarkan pada persyaratan bahwa variabel dependen dan variabel independent mempunyai teori hubungan sebab akibat atau hubungan didasarkan pada temuan penelitian sebelumnya, atau dapat dijelaskan secara logis. Dengan satu variabel independent dan satu variabel dependen, rumus persamaan regresi adalah sebagai berikut:

$$Y = a + bX$$

Keterangan:

Y = Variabel Dependen

X = Variabel Independen

a = Konstanta

b = Koefisien arah

Peran Statistika Dalam Berbagai Bidang

Penggunaan statistika sangat luas dan mencakup hampir setiap aspek kehidupan manusia, termasuk ekonomi, bisnis, kesehatan, sosiologi, kontruksi, dan analisis perilaku manusia. Bisnis kesehatan, dan pertumbuhan, inflasi, dan pengembangan bisnis semuanya dapat dijelaskan dalam ekonomi sebaliknya; statistika digunakan untuk membuat analisis keuangan, meningkatkan perencanaan pemasaran, dan menentukan nilai pasar. Statistika digunakan dalam bidang kesehatan untuk menentukan prevalensi penyakit tertentu serta tingkat kematian dan kesakitannya. Pertumbuhan dalam untuk menentukan tingkat kemiskinan pertumbuhan ekonomi, keberhasilan pemerintah sebaliknya dapat digunakan untuk memahami perilaku manusia dan sebagai ukuran kemajuan sosial di bidang manusia di permukaan, statistika sangat berguna untuk membantu menciptakan keputusan yang lebih akurat dan efisien di berbagai bidang.

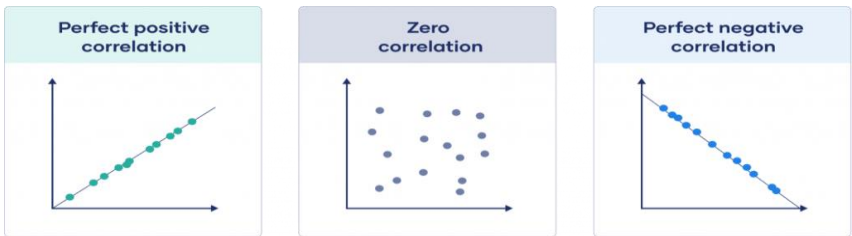
Pengertian Koefisien Korelasi

Korelasi (*correlation*) adalah adanya saling keterhubungan atau hubungan timbal balik. Dalam statistika sebagai ilmu, istilah korelasi mengandung dua variabel atau lebih. Hubungan antara dua variabel dikenal sebagai *bivariate correlation*, sedangkan hubungan antar lebih disebut juga *multivariate correlation*. Tujuan korelasi adalah Untuk mencari apakah hubungan antar variabel, bagaimana Tingkat keeratan hubungannya, dan untuk memperoleh kepastian (signifikan).

Konsep dasar korelasi adalah tinggi rendahnya data suatu variabel, akan diikuti oleh tinggi rendahnya data variabel yang lain. Dalam hal kekuatan hubungan, nilai koefisien korelasi bervariasi antara +1 dan -1. Nilai ± 1 menunjukkan Tingkat hubungan yang sempurna antar kedua variabel. Ketika koefisien korelasi turun di bawah nol, hubungan variabel akan melemah mendekati 0, hubungan antara kedua variabel akan lemah. Tujuan menganalisis koefisien korelasi adalah untuk memahami kekuatan dan arah hubungan antara dua variabel.

Tabel 3. 1 Koefisien Korelasi

1	0	-1
Korelasi positif sempurna	Korelasi nol	Korelasi negative sempurna
Ketika satu variabel berubah, variabel lain berubah ke arah yang sama.	Tidak ada hubungan antara variabel	Ketika satu variabel berubah, variabel lain berubah ke arah yang berlawanan



Gambar 3. 1 Koefisien Korelasi

JENIS - JENIS KOEFISIEN KORELASI

Adapun berbagai banyak jenis korelasi lainnya, yang paling umum meliputi korelasi pearson, korelasi spearman, korelasi point-biserial, korelasi kendall, korelasi parsial, korelasi linear ganda, korelasi determinasi, dan korelasi lambda serta yang lainnya.

1. Korelasi Pearson

Pearson coefficient of correlation atau korelasi pearson dikembangkan oleh Karl Pearson pada tahun 1900. Korelasi pearson, sering dikenal sebagai korelasi product moment adalah ukuran yang digunakan untuk memperkirakan derajat korelasi antara data linear yang interval. Fungsi dari korelasi ini adalah untuk menentukan hubungan dan kontribusi pada variabel independen dan variabel dependen.

Tinggi rendah atau kuat lemahnya, dapat dipahami dengan melihat beberapa indeks korelasi, yang disimbolkan dengan ρ

(baca rho, untuk populasi) atau r (untuk sampel). Jika hubungan antara Y dan X adalah regresi linier dalam kasus ini I dapat ditentukan r^2 dan disimpulkan sebagai berikut:

$$r^2 = \frac{\sum(Y_i - \underline{Y})^2 - \sum(Y_i - \hat{Y}_i)^2}{\sum(Y_i - \underline{Y}_i)^2}$$

Dengan demikian, koefisien korelasi (r) akan kira-kira antara -1 sampai dengan $+1$. Korelasi $r = 1$, baik negatif atau positif, itu menunjukkan bahwa ada hubungan yang sempurna. Koefisien r positif sempurna akan terjadi jika tinggi rendahnya setiap data pada suatu variabel memiliki tinggi rendahnya setiap data pada variabel yang lain. Koefisien r negatif sempurna akan terjadi jika ada keterbalikkan, rendahnya data dalam sebuah variabel memiliki data yang tinggi pada variabel yang lain.

Untuk menghitung koefisien korelasi r , rumus yang digunakan adalah yang langsung dihitung dari data mentah, sedangkan rumus yang pada prinsipnya sama. Rumus untuk mencari nilai koefisien korelasi r antara variabel X dan variabel Y, yang diturunkan dari rumus r^2 diatas dan diambil akarnya, sebagai berikut

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n(\sum X^2) - (\sum X)^2][n(\sum Y^2) - (\sum Y)^2]}}$$

Keterangan:

- r : Nilai koefisien korelasi
- $\sum x$: Jumlah pengamatan variabel x
- $\sum y$: Jumlah pengamatan variabel y
- $(\sum x^2)$: Jumlah kuadrat dari pengamatan variabel x
- $(\{\sum x\})^2$: Jumlah kuadrat dari jumlah pengamatan variabel x
- $(\sum y^2)$: Jumlah kuadrat dari pengamatan variabel y
- $(\{\sum y\})^2$: Jumlah kuadrat dari jumlah pengamatan variabel y
- n : Jumlah pasangan pengamatan y dan x

Contoh:

Misalkan seorang peneliti ingin mengetahui hubungan antara Kepemimpinan dan Kinerja. Hasil perolehan data dari 12 orang yang dijadikan responden sebagai berikut.:

Nama Karyawan	Kepemimpinan (X)	Kinerja (Y)
A	60	450
B	70	475
C	75	450
D	65	470
E	70	475
F	60	455
G	80	475
H	75	470
I	85	485
J	90	480
K	70	475
L	85	480

Tabel 3. 2 Data Kepemimpinan dan Kinerja

No.Resp.	X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$
1	60	450	3.600	202.500	27.000
2	70	475	4.900	225.625	33.250
3	75	450	5.625	202.500	33.750
4	65	470	4.225	220.900	30.550
5	70	475	4.900	225.625	33.250
6	60	455	3.600	207.025	27.300
7	80	475	6.400	225.625	38.000
8	75	470	5.625	220.900	35.250
9	85	485	7.225	235.225	41.225
10	90	480	8.100	230.400	43.200
11	70	475	4.900	225.625	33.250
12	85	480	7.225	230.400	40.800
Jumlah	855	5.640	66.325	2.652.350	416.825

Tabel Koefisien Korelasi r

$$r = \frac{12(416.825) - (855)(5.640)}{\sqrt{[12(66.325) - (855)^2][12(2.652.350) - (15.640)^2]}} = 0,684$$

Dengan demikian, koefisien yang menunjukkan hubungan antara Kepemimpinan dan Kinerja adalah 0,684. Secara sederhana untuk dapat mengetahui kuat lemahnya keamatan hubungan antara variabel X dan variabel Y, dapat dibandingkan berdasarkan tabel nilai koefisien korelasi dari *Guilford* (cara lain dari nilai kritis *r Product Moment*) sebagai berikut:

Tabel 3. 4 Tingkat Keamatan Hubungan Variabel X dan Variabel Y

Nilai Koefisien Korelasi	Keterangan
0,00 - 0,20	Hubungan lemah (dianggap tidak ada)
0,20 - 0,40	Hubungan rendah
0,40 - 0,70	Hubungan sedang
0,70 - 0,90	Hubungan kuat/tinggi
0,90 - 1,00	Hubungan sangat kuat/sangat tinggi

Dari contoh diatas, dengan r sebesar 0,684 terletak antara nilai 0,40 - 0,70, yang dapat disimpulkan bahwa Tingkat keamatan hubungan variabel kepemimpinan dengan variabel kinerja adalah sedang.

2. Korelasi Spearman

Korelasi Spearman merupakan suatu metode uji korelasi yang digunakan terhadap dua variabel apabila datanya ordinal, atau berupa interval rasio dan rasio, namun tidak meniadakan syarat analisis parametrik, seperti normalitas data dan linearitas hubungan (SPSS, 2001). Indeks korelasi Spearman dapat dihitung dengan menggunakan rumus berikut.

Angka indeks korelasi spearman dapat dihitung dengan rumus berikut.

$$r_s = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$$

Keterangan:

r_s : Korelasi spearman

n : Banyaknya sampel

$\sum D_i^2$: jumlah kuadrat dari selisih variabel X dengan variabel Y

Menghitung nilai koefisien korelasi spearman dengan menggunakan rumus diatas, apabila data pada variabel mempunyai nilai yang sama kurang 20%. Jika nilai data sama lebih dari 20%, maka rumus korelasi spearman sebagai berikut.

$$r_s = \frac{\sum x^2 + \sum y^2 - \sum D^2}{2\sqrt{\sum x^2 \sum y^2}}$$

$$\sum x^2 = \frac{n(n^2 - 1)}{12} - \sum \frac{t(t^2 - 1)}{12}$$

$$\sum y^2 = \frac{n(n^2 - 1)}{12} - \sum \frac{t(t^2 - 1)}{12}$$

D : selisih dari nilai variabel X dengan nilai variabel Y.

t : banyaknya nilai yang kembar pada suatu variabel.

Contoh :

Misalkan seseorang melakukan penelitian untuk menghimpun data empiris, agar dapat menjawab pertanyaan: apakah ada hubungan antara kompensasi yang diterima seorang pegawai dengan kinerja. Peneliti tersebut, kemudian mengukur kompensasi dan kinerja dengan skala "Likert", untuk 15 orang karyawan (n). data yang diperoleh adalah sebagai berikut:

Kompensasi

$(X_i) = 94, 83, 62, 58, 94, 72, 72, 69, 62, 80, 75, 70, 68, 85, 90$

Kinerja

$(Y_i) = 51, 63, 65, 97, 53, 68, 70, 81, 90, 53, 62, 63, 65, 58, 55$

Untuk menghitung koefisien korelasi spearman data disusun dalam tabel sebagai berikut:

Tabel 3. 5 Perhitungan Koefisien Korelasi Spearman

X	Y	R_X	R_Y	D_i $(R_x - R_y)$	D_i^2
94	51	14,5	1	13,5	182,25
83	63	11	7,5	3,5	12,5
62	65	2,5	10	-7,5	56,25
58	97	1	15	-14	196
94	53	14,5	2,5	12	144
72	65	7,5	10	-2,5	6,25
72	70	7,5	12	4,5	20,5
69	81	5	13	-8	64
62	90	2,5	14	-11,5	132,25
80	53	10	2,5	7,5	56,25
75	62	9	6	3	9
70	63	6	7,5	-1,5	2,25
68	65	4	10	6	36
85	58	12	5	7	49
90	55	13	4	9	81
					$\Sigma = 1.047$

Keterangan:

Kolom 1 : Diisi data variabel X

Kolom 2 : Diisi data variabel Y

Kolom 3 : Diisi peringkat data variabel X, jika ada data yang sama, peringkatnya dijumlahkan kemudian dibagi sebanyak data yang sama, contoh; data 62 ada 2 buah yang menempati peringkat ke-2 dan ke-3, maka peringkat data tersebut masing masing adalah 2,5.

Kolom 4 : Diisi data variabel Y

Kolom 5 : Diisi hasil pengurangan data variabel X dan variabel Y

Kolom 6 : Diisi kuadrat hasil pengurangan variabel X dan variabel Y.

Dalam tabel 5, pada variabel X, terdapat tiga variabel yang kembar, yaitu variabel 2,5 ada dua unsur ($t = 2$), variabel ke-7,5 ada dua unsur, dan unsur ke-14,5 ada dua unsur. Banyaknya variabel kembar tersebut ada 6 buah ($2 + 2 + 2$); sehingga persentasenya adalah $\frac{6}{15} \times 100\% = 40\%$. Pada variabel Y, ada 7 buah ($2 + 2 + 3$); sehingga persentasenya adalah $\frac{7}{15} \times 100\% = 46,67\%$.

Dengan demikian, karena jumlah persentase variabel yang kembar, baik pada variabel X ataupun variabel Y lebih dari 20%, maka untuk menghitung koefisien korelasiya digunakan pada rumus sebagai berikut.

Tabel 3. 6 Kembar variabel X

Peringkat Kembar Ke-	Jumlah Kembar (t)	$\frac{t(t^2 - 1)}{12}$
2,5	2	0,5
7,5	2	0,5
14,5	2	0,5
		$\sum \frac{t(t^2 - 1)}{12}$ = 1,5

$$\sum x^2 = \frac{15(15^2 - 1)}{12} - 1,5 = 278,5$$

Tabel 3. 7 Kembar variabel Y

Peringkat Kembar ke-	Jumlah Kembar (t)	$\frac{t(t^2 - 1)}{12}$
2,5	2	0,5
7,5	2	0,5
10	3	2
		$\sum_{=3} \frac{t(t^2 - 1)}{12}$

$$\sum y^2 = \frac{15(15^2 - 1)}{12} - 3 = 277$$

Dari tabel 5, diperoleh $\sum D^2 = 1.047$ dan $n = 15$, sehingga koefisien korelasi adalah:

$$r_s = \frac{\sum x^2 + \sum y^2 - \sum D^2}{2\sqrt{\sum x^2 \sum y^2}} = \frac{(278,5 + 277) - 1.047}{2\sqrt{(278,5)(277)}} = 0.7355$$

3. Korelasi Point-Biserial

Korelasi point biserial dipergunakan untuk mencari hubungan antara data berskala interval dan berskala nominal yang bersifat dikotomis. Data variabel yang bersifat dikotomis antara lain, jenis kelamin pria dan Wanita, jawaban benar dan salah, dan lain-lain. Data berskala nominal biasanya dikategorikan dengan simbol-simbol tertentu, misalnya untuk jawaban benar = 1, dan jawaban salah = 0. Rumus korelasi point biserial sebagai berikut:

$$r_b = \left(\frac{X_p - X_q}{s} \right) \sqrt{pq}$$

Keterangan :

r_b : Koefisien korelasi point biserial

X_p : Rata-rata hitung data interval yang berkategori dikotomi 1

\bar{X}_q : Rata-rata hitung data interval yang berkategori dikotomi 0

s : Standar deviasi dari keseluruhan data

p : Proporsi kasus berkategori dikotomi 1

q : Proporsi kasus berkategori dikotomi 0

Contoh:

Hasil tes mata kuliah Statistika dari 20 mahasiswa suatu PTS, yang terdiri atas mahasiswa laki-laki dan Perempuan. Pertanyaannya apakah ada korelasi hasil ujian Statistika berdasarkan statusnya sebagai laki-laki atau Perempuan. Mahasiswa laki-laki dikategorikan dengan 1, sedangkan mahasiswa dikategorikan dengan 0.

Tabel 3. 8 Data Nilai 20 Mahasiswa Suatu PTS

No.	Nama	Data	Kategori	Data kategori 1 (p)	Data Kategori 0 (q)
1	A	60	1	60	
2	B	70	0		70
3	C	75	1	75	
4	D	65	0		65
5	E	70	1	70	
6	F	60	1	60	
7	G	80	0		80
8	H	75	1	75	
9	I	85	1	85	
10	J	90	0		90
11	K	70	1	70	
12	L	85	1	85	
13	M	58	1	58	
14	N	42	0		42
15	O	50	0		50
16	P	52	1	52	
17	Q	50	1	50	
18	R	68	1	68	
19	S	46	0		46
20	T	58	1	58	
				$n_1 = 13$ $\bar{X}_p = 66,62$ $p = 0,65$	$n_0 = 7$ $\bar{X}_q = 63,29$ $q = 0,35$

Data tersebut dimasukkan ke dalam rumus sebagai berikut:

$$r_b = \left(\frac{X_p - X_q}{s} \right) \sqrt{pq} = \left(\frac{66,62 - 63,29}{13,74} \right) \sqrt{(0,65)(0,35)} = 0,116$$

Artinya, tidak terdapat hubungan nilai statistika antara mahasiswa laki-laki dan Perempuan, atau dengan kata lain da perbedaan nilai yang disebabkan oleh perbedaan kelamin.

4. Korelasi Kendall

Korelasi kendall adalah ukuran korelasi yang mengharuskan kedua variabel diukur sekurangnya dalam skala ordinal. Rumus yang dapat digunakan Korelasi Kendall sebagai menghitung koefisien korelasi variabel X dan Y, yang kedua-duanya memiliki Tingkat pengukuran ordinal. Korelasi Kendal digunakan sebagai mencari hubungan dan menguji hipotesis antara dua variabel atau lebih dan sering diterapkan secara bergantian dengan korelasi spearman.

Rumus korelasi Kendal (r_k) adalah sebagai berikut:

$$r_k = \frac{2(P - Q)}{n(n - 1)}$$

Keterangan:

P : Jumlah angka peringkat yang lebih tinggi

Q : Jumlah angka peringkat yang lebih rendah

n : Banyaknya data

Rumus yang dapat digunakan untuk menentukan koefisien korelasi, jika tidak lebih dari satu variabel yang sama (kembar), sebagai berikut:

$$r_{k,c} = \frac{2(P - Q)}{n(n - 1)} \left\{ \frac{\left[\frac{n(n-1)}{2} - \frac{z_{11}(z_{11}-1) + \dots + z_{1n}(z_{1n}-1)}{2} \right] \left[\frac{n(n-1)}{2} - \frac{z_{21}(z_{21}-1) + \dots + z_{2n}(z_{2n}-1)}{2} \right]}{\left[\frac{n(n-1)}{2} - \frac{z_{11}(z_{11}-1) + \dots + z_{1n}(z_{1n}-1)}{2} \right] \left[\frac{n(n-1)}{2} - \frac{z_{21}(z_{21}-1) + \dots + z_{2n}(z_{2n}-1)}{2} \right]} \right\}$$

Keterangan:

z_1 : Peringkat-peringkat kembar data pertama

z_2 : Peringkat-peringkat kembar data kedua

Contoh:

Hasil observasi 15 orang karyawan sebuah Perusahaan, telah diperoleh data kompensasi (X) dan kinerja (Y).

Tabel 3. 9 Data Kompensasi dan Kinerja

X	Y	R_x	R_y	D_i ($R_x - R_y$)	D_i^2
94	51	14,5	1	13,5	182,25
83	63	11	7,5	3,5	12,5
62	65	2,5	10	-7,5	56,25
58	97	1	15	-14	196
94	53	14,5	2,5	12	144
72	65	7,5	10	-2,5	6,25
72	70	7,5	12	4,5	20,5
69	81	5	13	-8	64
62	90	2,5	14	-11,5	132,25
80	53	10	2,5	7,5	56,25
75	62	9	6	3	9
70	63	6	7,5	-1,5	2,25
68	65	4	10	6	36
85	58	12	5	7	49
90	55	13	4	9	81

Tabel 3. 10 Peringkat Variabel Kompensasi dan Variabel kinerja

Nama Karyawan	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Kompensasi (X)	14,5	11	2,5	1	14,5	7,5	7,5	5	2,5	10	9	6	4	12	13
Kinerja (Y)	1	7,5	10	15	2,5	10	12	13	14	2,5	6	7,5	10	5	4

Tabel 3. 11 Peringkat Hasil Pengamatan Kompensasi (X) Dan Kinerja (Y)

No.	Nama Karyawan	Peringkat Kompensasi	Peringkat Kinerja	Jumlah Peringkat Yang Lebih Tinggi	Jumlah Peringkat Lebih Rendah
1	D	1	15	0	14
2	C	2,5	10	3	8
3	I	2,5	14	0	12
4	M	4	10	2	8
5	H	5	13	0	10
6	L	6	7,5	2	6
7	F	7,5	10	1	7
8	G	7,5	12	0	7
9	K	9	6	1	5
10	J	10	2,5	3	1
11	B	11	7,5	0	4
12	N	12	5	0	3
13	O	13	4	0	2
14	A	14,5	1	1	0
15	E	14,5	2,5	0	0
n = 15				$\sum vP = 13$	$\sum vQ = 87$

Karena variabel X ataupun variabel Y mempunyai lebih dari satu peringkat yang sama, maka dapat dicari faktor korelasi dalam mencari koefisien korelasi, sebagai berikut:

a. $z_1 = \frac{2(2-1)+2(2-1)+2(2-1)}{2} = 3$
 b. $z_2 = \frac{2(2-1)+2(2-1)+3(3-1)}{2} = 5$

Maka, dapat dimasukkan ke rumus korelasi kendall, sebagai berikut:

$$r_k = \frac{13 - 87}{\sqrt{\left[\frac{15(15 - 1)}{2} - 3\right]\left[\frac{15(15 - 1)}{2} - 5\right]}} = -0,733$$

Koefisien korelasi tersebut, menunjukkan bahwa hubungan antara kompensasi dengan kinerja pegawai adalah **-0,733**.

5. Korelasi Parsial

Korelasi Parsial (*Partial Correlation*) adalah suatu nilai yang memberikan kuatnya hubungan dua atau lebih variabel independen $\{(X_j)_i\}$ dengan variabel dependen (Y), yang salah satu atau beberapa variabelnya dianggap konstan. Beberapa rumus untuk menghitung korelasi parsial sebagai berikut:

- a. Untuk tiga variabel terdiri dari Y sebagai variabel dependen, X_1 dan X_2 variabel independent.
- 1) Koefisien korelasi parsial antara Y dan X_1 dengan menganggap X_2 tetap, dinyatakan dalam rumus:

$$r_{y1-2} = \frac{r_{y1} - (r_{y2})(r_{12})}{\sqrt{(1 - r_{y2}^2)(1 - r_{12}^2)}}$$

- 2) Koefisien korelasi parsial antara variabel Y dan X_2 dengan menganggap variabel X_1 tetap, rumusnya:

$$r_{y2-1} = \frac{r_{y2} - (r_{y1})(r_{12})}{\sqrt{(1 - r_{y1}^2)(1 - r_{12}^2)}}$$

Keterangan:

r_{y1-2} : Korelasi Parsial

r_{y1} : Korelasi antara variabel Y dengan variabel X_1 .

r_{y2} : Korelasi antara variabel Y dengan variabel X_2 .

r_{12} : Korelasi antara variabel X_1 dan X_2 .

Contoh :

Dalam contoh ini, diambil hasil perhitungan koefisien korelasi antar variabel yang disajikan sebagai berikut:

Tabel 3. 12 Data Hasil Angket Kepemimpinan, Kompensasi, Kepuasan Kerja, dan Motivasi Berprestasi

Responden	Kepemimpinan (X_1)	Kompensasi (X_2)	Kepuasan Kerja (X_3)	Motivasi Berprestasi (Y_1)
1	60	56	134	54
2	70	54	130	52
3	75	66	136	62
4	65	62	140	56
5	70	52	120	50
6	60	56	124	48
7	80	56	112	50
8	75	60	110	46
9	85	62	132	52
10	90	62	132	56

11	70	64	136	54
12	85	52	126	44
13	58	68	126	56
14	42	60	102	38
15	50	50	128	52
16	52	54	128	54
17	50	48	120	44
18	68	66	140	60
19	46	40	112	38
20	58	48	124	54
Jumlah	$\sum vX_1 = 1.309$ $\sum vX_{1^2} = 89.261$ $\sum vX_{1X_2} = 75.134$ $\sum vX_{1X_3} = 165.332$ $\sum vX_{1Y} = 67.392$	$\sum vX_2 = 1.136$ $\sum vX_{2^2} = 65.520$ $\sum vX_{2X_3} = 143.292$ $\sum vX_{2Y} = 58.472$	$\sum vX_3 = 2.512$ $\sum vX_{3^2} = 317.500$ $\sum vX_{3Y} = 129.156$	$\sum vY = 1.020$ $\sum vY^2 = 52.808$

Berdasarkan data koefisien korelasi tersebut telah diperoleh: $r_{y1} = 0,377$, $r_{y2} = 0,605$, $r_{12} = 0,414$, seanjutnya dapat dihitung korelasi parsial sebagai berikut:

- Kepemimpinan $\{(X)_1\}$ dan motivasi berprestasi (Y) , dengan beranggapan kompensasi (X_2) tetap.

$$r_{y1-2} = \frac{r_{y1} - (r_{y2})(r_{12})}{\sqrt{(1 - r_{y2}^2)(1 - r_{12}^2)}} = \frac{0,377 - (0,605)(0,414)}{\sqrt{(1 - 0,605^2)(1 - 0,414^2)}} = 0,175$$

- Kompensasi (X_2) , dan motivasi berprestasi (Y) , dengan beranggapan Kepemimpinan $\{(X)_1\}$ tetap

$$r_{y2-1} = \frac{r_{y2} - (r_{y1})(r_{12})}{\sqrt{(1 - r_{y1}^2)(1 - r_{12}^2)}} = \frac{0,605 - (0,377)(0,414)}{\sqrt{(1 - 0,377^2)(1 - 0,414^2)}} = 0,532$$

6. Korelasi Linear Ganda

Analisis korelasi berganda merupakan perluasan dari analisis korelasi sederhana. Dalam korelasi ini bertujuan untuk mengetahui bagaimana derajat hubungan antara beberapa variabel independent secara bersama - sama dengan variabel dependen. Ada beberapa rumus korelasi ganda sebagai berikut:

- Korelasi ganda antara dua variabel independen dengan variabel dependen, yaitu:

$$R_{y12} = \sqrt{\frac{r_{y1}^2 + r_{y2}^2 - 2(r_{y1})(r_{y2})(r_{12})}{1 - r_{12}^2}}$$

R_{y12} = Korelasi ganda antara variabel dependen dengan dua variabel independen

- b. Koefisien korelasi ganda dengan dua variabel independent, yaitu:

$$R_{y12} = \sqrt{1 - [(1 - r_{y1}^2)(1 - r_{y2}^2)]}$$

r_{y2-1} = korelasi parsial antara variabel dependen (Y) dan variabel independen ($\{X\}_2$) dengan menganggap variabel ($\{X\}_1$)

- c. Korelasi ganda dengan tiga variabel, yaitu:

$$R_{y123} = \sqrt{1 - [(1 - r_{y1}^2)(1 - r_{y2-1}^2)(1 - r_{y3-12}^2)]}$$

R_{y2-1} = korelasi parsial antara variabel dependen (Y) dan variabel independen ($\{X\}_2$) dengan menganggap variabel (X_1) tetap

r_{y3-12} = korelasi parsial antara variabel dependen (Y) dan variabel independen X_3 dengan menganggap variabel (X_3), dengan menganggap variabel (X_1) dan (X_2) tetap.

Contoh:

Dilakukan suatu penelitian yang bertujuan untuk mempelajari tentang “pengaruh pendapatan keluarga per hari ($\{X\}_1$) dan jumlah anggota keluarga ($\{X\}_2$) terhadap pengeluaran konsumsi keluarga per hari (Y)”. Penelitian tersebut menggunakan sampel sebanyak 10 keluarga. Hasil pengumpulan data diperoleh data sebagai berikut:

Tabel 3. 13 Pengaruh pendapatan keluarga per hari

Responden	X_1 (Ratusan Ribu)	X_2 (Orang)	Y (Ratusan ribu)
1	100	7	23
2	20	3	7
3	40	2	15
4	60	4	17
5	80	6	23
6	70	5	22
7	40	3	10
8	60	3	14
9	70	4	20
10	60	3	19

Berdasarkan data diatas, maka carilah koefisien korelasi berganda dan parsial (jika jumlah anggota keluarga dianggap konstan).

7. Pengujian keberatian Korelasi Linear ganda dan Parsial

Pengujian keberatian korelasi linear ganda dan parsial, terlebih dahulu dirumuskan dalam hipotesis sebagai berikut:

$H_0 : \rho = 0$, tidak terdapat hubungan yang berarti antara variabel dependen dan variabel independent

$H_1 : \rho \neq 0$, terdapat hubungan yang berarti antara variabel dependen dan variabel independent

Pengujian korelasi parsial dapat menggunakan rumus sebagai berikut:

$$t_{hit} = r \frac{\sqrt{n-k-1}}{\sqrt{1-r^2}}$$

Tolak H_0 jika nilai t_{hit} lebih besar dari nilai t_{tab} , atau lebih kecil dari $(-t_{tab})$ dengan $dk = n - k - 1$ dan $\frac{1}{2}\alpha(t_{0,975}(dk))$

Pengujian korelasi linear ganda dapat dilakukan dengan menggunakan rumus sebagai berikut:

$$t_{hit} = \frac{\frac{R_{y.x_1x_2x_3...x_k}}{k}}{\frac{1 - R_{y.x_1x_2x_3...x_k}^2}{n - k - 1}}$$

Keterangan:

F_{hit} = Nilai uji F yang akan dibandingkan dengan nilai F_{tabel}

R = Koralasi ganda

k = Banyaknya variabel independent

n = Banyaknya data

8. Korelasi determinasi

Korelasi determinasi merupakan nilai diterapkan untuk menghitung besarnya kontribusi variabel independen (X) terhadap naik atau turunnya variabel dependen (Y). Korelasi determinasi biasanya dilambangkan dengan r^2 dan dinyatakan dalam persentase (%). Definisi r^2 sendiri adalah rasio variabilitas nilai yang dibuat dalam model dengan varibilitas nilai data asli. Secara

umum nilai r^2 digunakan sebagai informasi mengenai kecocokan suatu model. Rumus yang dapat digunakan dalam menghitung korelasi determinasi sebagai berikut:

$$KD = R = r^2 \times 100\%$$

Keterangan:

$KD = R$ = Nilai Korelasi Determinasi

r^2 = Nilai koefisien penentu (antara 0 dan +1)

Contoh: Jika diketahui data X adalah kenaikan gaji dan Y adalah kenaikan harga dibawah ini sebagai berikut

Tabel 3. 14 Kenaikan gaji dan Kenaikan harga

X	14	18	16	20	22	26	12	20
Y	18	12	10	16	8	20	10	14

Hitunglah nilai r dalam data diatas dan nilai dari korelasi determinasi!

Tabel 3. 15 Data hasil kenaikan gaji dan kenaikan harga

X	Y	XY	x^2	y^2
14	18	252	196	324
18	12	216	324	144
16	10	160	256	100
20	16	320	400	256
22	8	176	484	64
26	20	520	676	400
12	10	120	144	100
20	14	280	400	196
$\sum = 148$	$\sum = 108$	$\sum = 2044$	$\sum = 2880$	$\sum = 1584$

Penyelesaian:

a. Nilai r (korelasi pearson)

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$$

$$= \frac{8(2044) - 2044}{\sqrt{[8(2880) - (148)^2][8(1584) - (108)^2]}}$$

$$= \frac{16352 - 2044}{\sqrt{(23040 - 21904)(12672 - 11664)}}$$

$$\frac{14308}{(33.70)(31,74)} = \frac{14308}{1069.638}$$

$$r = 13,376$$

b. Korelasi determinasi

$$r^2 = (\{13,376\})^2 \times 100\%$$

$$= 178.917 \times 100\%$$

$$= 17891,7\%$$

9. Korelasi PHI

Korelasi PHI merupakan analisis korelasi yang digunakan ketika data yg dikorelsikan adalah data yang benar-benar dikotomik (tepisah atau dipisahkan) dengan istilah lain yaitu variabel diskrit murni. Variabel diskrit murni adalah variabel yang memiliki nilai dapat dihitung dan dalam bentuk bilangan bulat atau integer.

Tabel 3. 16 Uji PHI

Kategori peubah II	Kategori peubah I		Total
	1	2	
1	a	b	a + b
2	b	d	c + d
Total	a + c	b + d	n

Rumus korelasi PHI dapat diartikan sebagai berikut:

$$\phi = \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}}$$

Nilai PHI sendiri mempunyai rentang diantara -1 dan 1. Nilai PHI didapatkan dari perhitungan *chi square*. Rumus dalam hubungan dengan *chi square* di bawah ini:

$$\phi^2 = \frac{\chi^2}{n} \text{ atau } \chi^2 = n\phi^2$$

Keterangan:

ϕ : Korelasi PHI

Contoh:

Seorang peneliti ingin mengetahui apakah ada korelasi antara jenis kelamin dengan pilihan jurusan di perguruan tinggi. Jenis Kelamin dapat dibedakan menjadi laki-laki dan Perempuan. Sedangkan, pilihan jurusan dibedakan ke dalam matematika dan bahasa.

Tabel 3. 17 Hasil Pilihan Jurusan di Perguruan Tinggi

Jenis Kelamin	Pilihan Jurusan		Total
	Matematika	Bahasa	
Laki-laki	40	15	55
Perempuan	25	25	50
Jumlah	65	40	105

Penyelesaian:

$$\begin{aligned} \phi &= \frac{ad - bc}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} \\ \phi &= \frac{(40 \times 25) - (15 \times 25)}{\sqrt{(40 + 15)(25 + 25)(40 + 25)(15 + 25)}} \\ \phi &= \frac{1000 - 375}{\sqrt{(55)(50)(65)(40)}} \\ \phi &= \frac{625}{2673,948} \\ \phi &= 0,233736 \end{aligned}$$

10. Korelasi Lambda (λ)

Rumus korelasi lambda (λ), digunakan dalam analisis korelasi sederhana untuk variabel nominal. Korelasi lambda diringkas sebagai:

- Simetris lambda, tidak mempersonalisasikan variabel yang ditetapkan sebagai bebas.

$$\lambda = \frac{f_b + f_k - (F_b + F_k)}{2n - (F_b + F_k)}$$

Keterangan:

λ : Korelasi *Lambda*

f_b : Frekuensi baris

- f_k : Frekuensi kolom
- F_b : Frekuensi Majinal baris
- F_k : Frekuensi kolom
- n : Jumlah data

- b. Lambda asimetris, mengubah variabel tepat menjadi variabel bebas.

$$\lambda = \frac{f_1 - f_d}{n - F_d}$$

Keterangan:

- f_1 : Frekuensi variabel independent
- F_d : frekuensi variabel dependen
- n : Jumlah data

Contoh:

Berikut ini diberikan data tentang tingkat partisipasi mahasiswa dalam kegiatan politik dengan jenis media yang paling sering diikutinya. Tentukan nilai korelasi lambda

Tabel 3. 18 Jumlah tingkat partisipasi politik

Jenis Media	Tingkat Partisipasi Politik			Jumlah
	Tinggi	Menengah	Rendah	
Media Cetak	32	26	11	69
Media Elektronik	10	14	47	71
Jumlah	42	40	58	140

Penyelesaian:

$$\lambda = \frac{f_b + f_k - (F_d + F_d)}{2n - (F_d + F_k)}$$

$$\lambda = \frac{79 + 105 - (71 + 58)}{2(140) - (71 + 58)}$$

$$\lambda = \frac{184 - 129}{280 - 129}$$

$$\lambda = \frac{55}{151}$$

$$\lambda = 0,3642$$

c. Korelasi Kontingensi

Korelasi Kontingensi digunakan untuk menghitung hubungan antar variabel bila datanya berbentuk nominal. Teknik dari *chi kuadrat* sendiri digunakan sebagai menguji hipotesis komparatif k sampel independent. Korelasi kontingensi digunakan apabila kedua variabel berskala nominal. Banyaknya baris dan kolom dalam tabel kontingensi menentukan nilai maksimum dari C yang tidak pernah >1 . Jika banyaknya baris dan kolom dari tabel kontingensi sama dengan k , maka nilai maksimum diberikan rumus sebagai berikut:

$$C = \sqrt{\frac{(k-1)}{k}}$$

Rumus yang digunakan dalam korelasi kontingasi adalah

$$C = \sqrt{\frac{x^2}{N + x^2}}$$
$$x^2 = \sum \frac{(f_o - f_t)^2}{f_t}$$

Keterangan:

C : Korelasi Kontingansi

N : Jumlah banyaknya data

x^2 : Hasil dari perhitungan *chi square*

f_o : Frekuensi Observasi

f_t : Frekuensi Teoretis

Contoh:

Suatu penelitian bertujuan ingin mengetahui apakah ada hubungan antara kegairahan olahraga dengan semangat kerja. Hitunglah nilai C pada Data yang diambil dari penyebaran kuesioner ditabelkan dalam data sebagai berikut:

Tabel 3. 19 Kegairahan Olahraga dengan Semangat Kerja

Kegairahan olahraga dengan semangat kerja	Besar	Sedang	Rendah	Jumlah
Tinggi	18	12	10	40
Sedang	34	43	33	110
Rendah	10	10	30	50
Jumlah	62	65	73	200

Penyelesaian:

Tabel 3. 20 Tabel Perhitungan

Sel	f_o	f_t	$(f_o - f_t)$	$(f_o - f_t)^2$	$\left(\frac{f_o - f_t}{f_t}\right)^2$
1	18	$\frac{62 \times 40}{200} = 12,40$	5,60	31,3600	2,5290
2	12	$\frac{65 \times 40}{200} = 13$	-1	1	0,0770
3	10	$\frac{73 \times 40}{200} = 14,60$	-4,60	21,1600	1,4490
4	34	$\frac{62 \times 110}{200} = 34,10$	-0,10	0,0100	0,0003
5	43	$\frac{65 \times 110}{200} = 35,75$	7,25	52,5625	1,4703
6	33	$\frac{73 \times 110}{200} = 40,15$	-7,15	51,1225	1,2733
7	10	$\frac{62 \times 50}{200} = 15,50$	-5,50	30,2500	1,9516
8	10	$\frac{65 \times 50}{200} = 16,25$	-6,25	39,0625	2,4038
9	30	$\frac{73 \times 50}{200} = 18,25$	11,75	138,0625	7,5650
Jumlah	200	-	0,00	-	18,7134

$$C = \sqrt{\frac{X^2}{N + X^2}}$$

$$C = \sqrt{\frac{18,7134}{200 + 18,7134}}$$

$$C = \sqrt{\frac{18,7134}{218,7134}} = 0,29255$$

11. Korelasi Biserial

Korelasi Biserial digunakan untuk data interval menggunakan data dikotomi buatan. Rumus yang digunakan dalam korelasi biserial adalah sebagai berikut:

$$r = \left(\frac{X_p - X_q}{S_x} \right) \frac{pq}{h} \quad q = 1 - p \quad S_x = \sqrt{\frac{\sum_{i=1}^n vX_i^2}{n} - \left(\sum_{i=1}^n vX_i \right)^2}$$

Keterangan:

r : Korelasi Biserial

X_p : Rata - rata data interval I

X_q : Rata - rata data interval II

S_x : Simpangan baku seluruh data x

p : Proporsi kelompok data interval I menurut data y

q : Proporsi kelompok data interval 0 menurut data x

h : Tinggi ordinat p dan q

Contoh:

Tabel 3. 21 Hasil Indeks Prestasi Mahasiswa

Mahasiswa	Indeks Prestasi			Ujian	
	X	x	X^2	Lulus (p)	Gagal (q)
A	2,50	-0,63	0,3969		0
B	3,20	0,07	0,0049	1	
C	3,35	0,22	0,4840	1	
D	3,25	0,12	0,0144		0
E	2,75	-0,38	0,1444	1	
F	2,80	-0,33	0,1089		0
G	3,40	-0,27	0,0729		0
H	2,50	-0,63	0,3969		0
I	3,75	0,62	0,3844	1	
J	3,80	0,67	0,4489	1	
Jumlah	31,30	0	2,0210	16,85	14,45

Penyelesaian:

$$\sum_{i=1}^5 vX_p = 3,20 + 3,35 + 2,75 + 3,75 + 3,80 = 16,85$$

$$\sum_{i=1}^5 vX_q = 2,50 + 3,25 + 2,80 + 3,40 + 2,50 = 14,45$$

$$\underline{X}_p = \frac{16,85}{5} = 3,37 \rightarrow p = 0,5$$

$$\underline{X}_q = \frac{14,45}{5} \rightarrow q = 0,5$$

$$S_x = \sqrt{\frac{\sum_{i=1}^n X_i^2}{n} - \left(\frac{\sum_{i=1}^n X_i}{n}\right)^2} = \sqrt{\frac{2,021}{10} - \left(\frac{0}{10}\right)^2} = 0,45 \rightarrow h = 0,3989$$

$$r = \left(\frac{X_p - X_q}{S_x}\right) \frac{pq}{h} = \left(\frac{3,37 - 2,89}{0,45}\right) \frac{(0,5)(0,5)}{0,3989} = 0,6685$$

BAB 7 REGENERASI LINIER GANDA DAN REGENERASI NON LINIER

Regresi Non-linear

Analisis regresi adalah suatu teknik dalam statistik yang di gunakan untuk memahami pola hubungan antara variable independen dan variabel dependen (hosman dan lemeshow,2000) berdasarkan pola hubungan yang ada, analisis regresi dapat di bedakan menjadi analisis regresi linier dan analisis regresi non-linier, menurut hasan (1999), sebuah model dikategorikan sebagai model regresi non-linier jika terdapat variable yang memiliki pangkat. Contoh dari model regresi non-linier meliputi model parabola, kuadratik, hiperbola, dan lainnya. Montgomery dan peck (1992) menyebutkan bahwa model regresi non-linier juga dapat di lihat dari parameter yang digunakan. Salah satu contoh model regresi non-linier dalam hal parameter adalah model regresi logistik, sebagaimana di ungkapkan oleh Montgomery dan peck (1992). dapat dituliskan sebagai:

$$y_i = f(x_i, \theta) + \epsilon_i, i = 1, 2, \dots, n. (2.1) \text{ dengan, } y_i$$

Regresi linier

Analisis regresi linier adalah metode statistik yang digunakan untuk menganalisis hubungan antara satu variabel dependen dan satu atau lebih variabel independen. Tujuan utama dari regresi linier adalah untuk memprediksi nilai variabel dependen berdasarkan nilai variabel independen yang ada. Model regresi linier mengasumsikan bahwa hubungan antara variabel-variabel tersebut dapat digambarkan dengan garis lurus. Dapat dituliskan sebagai:

$$y_1 = F(X_i, \theta) + \epsilon_i, i = 1, 2, \dots, n.$$

Dengan,

y_i = variabel terikat ke- i

x_i = variabel bebas ke- i

θ_0 = parameter yang tidak diketahui

E_i = error, Dimana $E \sim N(0, \sigma^2)$

Di bawah ini adalah contoh model regresi non linear dalam parameter:

$Y_i = e^{-\theta_0 x_i} + e_i$. Karena $df/d\theta_0 = -x_i e^{-\theta_0 x_i}$ merupakan fungsi dalam θ_0 maka model di atas adalah model non linear dalam parameter.

pengukuran data

Data dalam penelitian dapat dibagi menjadi beberapa kategori. Proses pengelompokan ini harus didasarkan pada pengukuran yang tepat. Menurut Siegel (1994), terdapat empat jenis pengukuran:

1. Skala Nominal: Ini adalah jenis pengukuran yang digunakan untuk mengklasifikasikan objek tanpa adanya urutan atau tingkatan.
2. Skala Interval: Jenis pengukuran ini menunjukkan tingkatan, seperti preferensi, tingkat kesulitan, dan sebagainya.
3. Skala Ordinal: Pengukuran ini memiliki semua karakteristik skala ordinal, di mana jarak antara dua angka pada skala ini dapat diketahui ukurannya.
4. Skala Rasio: Ini adalah pengukuran semua ciri dari skala rasio, serta memiliki titik nol sejati yang berfungsi sebagai titik awal.

Regresi logistik multi nominal

Pemodelan regresi logistik untuk variabel respon nominal dan ordinal:

Pendekatan ini digunakan untuk menggambarkan hubungan antara beberapa variabel bebas dan satu variabel respon yang bersifat multi nominal. Data nominal adalah data yang menggunakan angka sebagai label tanpa menunjukkan tingkatan, sedangkan data ordinal menunjukkan tingkatan pada variabel tersebut. Jika ada k kategori pada variabel bebas, maka akan ada k model yang terbentuk.

1. Model Regresi Logistik

Menurut Agresti (1990), model regresi logistik multinomial dapat diterapkan ketika variabel dependen dinyatakan dalam bentuk vektor \mathbf{x}_i . Jika kategori respon memiliki urutan (respon ordinal), maka model yang digunakan adalah regresi logistik ordinal. Sebagai contoh, variabel kontinu z dapat

dibagi menjadi beberapa kategori ordinal dengan menggunakan titik potong tertentu. Beberapa model yang dapat diterapkan untuk regresi logistik ordinal termasuk model logit.

2. Pengertian regresi

Regresi berperan dalam memperkirakan nilai suatu variabel yang tidak diketahui berdasarkan satu atau beberapa variabel yang ada. Analisis regresi didefinisikan sebagai studi tentang hubungan antara satu variabel yang disebut variabel terjelaskan (*explained variable*), yang juga dikenal sebagai variabel yang dipengaruhi, terdapat juga variabel lain yang bersifat independen atau variabel bebas. Metode regresi linier dan non-linier. Namun, jika variabel bebas bersifat diskrit, maka analisis linier dapat digunakan. Namun, jika variabel bebas bersifat diskrit, analisis linier mungkin tidak cocok digunakan karena beberapa alasan : hubungan antara dua variabel atau lebih. Dalam regresi linier, variabel tidak bebas harus memiliki sifat kontinu.

- a. Dalam regresi linier, variabel tidak bebas harus mampu menerima nilai negatif.
- b. Variabel diskrit berada dalam kategori tertentu dan biasanya disebut sebagai variabel nominal atau variabel kategori. persamaan garis yang paling mewakili hubungan antara dua variabel beberapa asumsi. Statistik yang diperlukan dalam melakukan analisis regresi adalah : distribusi normal atau setidaknya mendekati

1) Model Regresi Linier Sederhana

Model ini dapat digunakan untuk memperkirakan hubungan antara dua variabel dan membuat asumsi mengenai bentuk hubungan yang dapat dinyatakan dalam fungsi tertentu. Setelah hubungan diperkirakan, beberapa asumsi dapat diperiksa.

Diantara beberapa hal, bisa juga di cek kedalam asumsi tersebut setelah hubungan diperkirakan.

Regresi linier sederhana memiliki beberapa fungsi, diantaranya sebagai berikut:

- a) Menguji hubungan / kolerasi / pengaruh satu variabel bebas terhadap satu variabel terikat.
- b) Melakukan perediksi atau estimasi variabel terikat berdasarkan variabel bebasnya.
- c) Data yang dianalisis juga berupa data yang berskala interval / rasio.

Fungsi linier, selain mudah dipahami, juga dapat digunakan sebagai pendekatan untuk menggambarkan hubungan yang tidak bersifat linier.

Analisis Regresi Linier Berganda

Analisis yang melibatkan lebih dari satu variabel bebas dikenal sebagai analisis regresi linier berganda. Teknik ini digunakan untuk menentukan apakah ada pengaruh yang signifikan dari dua atau lebih variabel bebas terhadap variabel yang terkait (Y).

Hubungan Linier Antara Dua Variabel

Analisis hubungan linier antara dua variabel, X dan Y, bertujuan untuk mengukur dampak kuantitatif dari perubahan satu variabel terhadap yang lain. Kebijakan pemerintah atau swasta seringkali ditujukan untuk menciptakan perubahan, seperti peningkatan produksi melalui penambahan pupuk. Jika X dan Y berkorelasi, perubahan pada X akan mempengaruhi Y, dan hubungan ini dapat dinyatakan dalam bentuk fungsi matematis. Kekuatan hubungan diukur dengan koefisien korelasi, sedangkan pengaruh X terhadap Y diukur dengan koefisien regresi.

Regenerasi dalam Variabel Terikat Data Kuantitatif

Data kuantitatif yang berfungsi sebagai variabel independen sering kali disebut sebagai variabel dummy. Salah satu contoh umum adalah ketika data kuantitatif diterapkan pada variabel dependen. Misalnya, kemampuan seseorang untuk memiliki kendaraan di kota dipengaruhi oleh faktor-faktor seperti jarak dan pendapatan. Model yang melibatkan variabel atau kategori terikat dapat dibagi menjadi dua jenis, yaitu:

1. Model regresi probabilitas linier (Linear Probability Model - LPM)
2. Model regresi logistik biner (Binary Logistic Regression Model)

1. Pengertian Persamaan Regresi

Persamaan regresi merupakan suatu rumus matematik yang memungkinkan kita untuk memprediksi nilai dari suatu variabel yang tidak bebas berdasarkan nilai satu atau lebih variabel yang bebas (Walpole, 1995, p340).

2. Metode Regresi

Secara umum, regresi adalah teknik yang digunakan untuk memperkirakan nilai harapan yang bersyarat. Regresi dikategorikan sebagai linear jika terdapat hubungan linier antara variabel bebas dan variabel tidak bebas. Sebaliknya, jika hubungan antara kedua jenis variabel tersebut tidak linier, maka regresi tersebut disebut regresi non-linier. Hubungan antara variabel bebas dan tidak bebas dapat dianggap linier jika diagram pencar data menunjukkan pola yang mendekati garis lurus.

3. Regresi Linier Sederhana

Regresi linier sederhana adalah model regresi di mana variabel bebasnya berwujud skalar.

4. Regresi Linier Berganda

Regresi linier berganda adalah jenis regresi yang menggambarkan hubungan antara satu variabel tidak bebas Y dengan beberapa variabel bebas lainnya, yaitu X_1 , X_2 , dan seterusnya.

Pengukuran Data

Pengukuran Data

Hasil penelitian dapat dikelompokkan ke dalam beberapa kategori. Dalam proses pengelompokan, penting untuk menggunakan pengukuran yang tepat. Siegel (1994) mengemukakan bahwa terdapat empat jenis skala pengukuran:

1. Skala Nominal: Merupakan pengukuran yang digunakan hanya untuk mengklasifikasikan objek-objek tertentu.
2. Skala Ordinal: Jenis pengukuran ini menunjukkan tingkatan, seperti preferensi, tingkat kesulitan, atau kriteria lainnya.
3. Skala Interval: Pengukuran ini mencakup semua karakteristik skala ordinal, dengan tambahan bahwa jarak antara dua angka pada skala ini dapat diukur.

5. Skala Rasio: Tipe pengukuran ini memiliki semua ciri dari skala interval dan juga memiliki titik nol sejati sebagai titik awal pengukuran.

Regresi logistik multinomial

Regresi Logistik Multinomial

Regresi logistik multinomial merupakan metode pemodelan yang efektif untuk data nominal maupun ordinal. Metode ini digunakan untuk menjelaskan hubungan antara beberapa variabel independen dengan satu variabel respon yang bersifat multinomial (polytomous) (Adisanto, 2010).

Data berskala nominal adalah data di mana angka yang diberikan kepada objek berfungsi sebagai label tanpa menunjukkan adanya tingkatan. Di sisi lain, data ordinal mencerminkan adanya tingkatan pada variabel dependen. Jika terdapat K kategori pada variabel independen, maka jumlah model logistik yang dapat dibentuk adalah $K-1$.

1. Pengujian Parameter

Pengujian Parameter Model

Menurut Hosmer dan Lemeshow (1989), pengujian parameter dalam model bertujuan untuk menilai pengaruh variabel independen terhadap model tersebut. Terdapat dua jenis pengujian yang dilakukan:

Pengujian Parameter dengan Uji Simultan atau Uji G

Uji statistik G digunakan untuk mengevaluasi pengaruh keseluruhan variabel independen dalam model. Dalam konteks pengujian hipotesis, jika hipotesis nol (H_0) dianggap benar, maka statistik uji G akan mengikuti distribusi Chi-Square dengan derajat bebas K , di mana K adalah jumlah prediktor dalam model. Dengan demikian, kriteria untuk menolak H_0 adalah ketika nilai G lebih besar dari X^2 .

Pengujian Parameter dengan Uji Wald (Uji Parsial)

Pengujian terhadap masing-masing variabel dilakukan secara terpisah menggunakan statistik uji Wald (Hosmer dan Lemeshow, 1989). Hipotesis yang diuji adalah sebagai berikut.

Koefisien Determinasi

Koefisien determinasi (R-Square) adalah ukuran yang menunjukkan sejauh mana variasi kadar gula darah pada pasien Diabetes Mellitus dapat dijelaskan oleh model regresi yang telah dikembangkan. Koefisien ini mencerminkan kemampuan variabel dalam menjelaskan variasi tersebut.

Pengujian koefisien determinasi bertujuan untuk menilai seberapa besar pengaruh variabel-variabel bebas terhadap nilai variabel terikat. Menurut Rizki (2016), sebuah model dianggap baik jika koefisien Nagelkerke lebih dari 70%, yang berarti bahwa variabel bebas dalam model tersebut mempengaruhi 70% dari variabel terikat. Pengujian Parameter Model

Menurut Hosmer dan Lemeshow (1989), pengujian parameter dalam model bertujuan untuk menilai dampak variabel independen terhadap model tersebut. Ada dua jenis pengujian yang dilakukan:

Pengujian Parameter dengan Uji Simultan atau Uji G

Uji statistik G digunakan untuk mengevaluasi pengaruh keseluruhan variabel independen dalam model. Dalam pengujian hipotesis, jika hipotesis nol (H_0) dianggap benar, maka statistik uji G akan mengikuti distribusi Chi-Square dengan derajat bebas K , di mana K adalah jumlah prediktor dalam model. Oleh karena itu, kriteria untuk menolak H_0 adalah jika nilai G melebihi X^2 .

Pengujian Parameter dengan Uji Wald (Uji Parsial)

Pengujian terhadap masing-masing variabel dilakukan secara terpisah menggunakan statistik uji Wald (Hosmer dan Lemeshow, 1989). Hipotesis yang diuji adalah sebagai berikut.

Koefisien Determinasi

Koefisien determinasi (R-Square) adalah ukuran yang menunjukkan sejauh mana variasi kadar gula darah pada pasien Diabetes Mellitus dapat dijelaskan oleh model regresi yang telah dikembangkan. Koefisien ini mencerminkan seberapa baik variabel dapat menjelaskan variasi yang terjadi.

Odd Ratio

Rasio kecenderungan, menurut Hosmer dan Lemeshow (1989), adalah ukuran yang digunakan untuk memperkirakan pengaruh variabel bebas terhadap variabel terikat. Odd Ratio (OR) membantu dalam memahami hubungan antara kedua jenis variabel tersebut. Jika $OR = 1$, berarti tidak ada hubungan; $OR > 1$ menunjukkan peluang sukses lebih tinggi dibandingkan nilai pembanding, sedangkan $OR < 1$ menunjukkan peluang sukses lebih rendah. Contoh penerapannya terlihat pada model regresi logistik multinomial dengan variabel bebas yang memiliki tiga kategori dan dua variabel terikat, di mana salah satu variabel terikat berskala kategori dan lainnya berskala continue.

1. Jenis-Jenis Persamaan Linear

Terdapat beberapa jenis persamaan linear, yaitu:

Persamaan linear dengan satu variabel

Persamaan linear dengan dua variabel

Persamaan linear dengan tiga variabel

2. Ciri-Ciri Persamaan Linear

Persamaan linear memiliki pangkat satu.

Tidak terdapat perkalian antara variabel.

Terdapat dua ruas yang dihubungkan oleh tanda sama dengan (=).

Operasi penjumlahan, pengurangan, perkalian, dan pembagian pada kedua ruas tidak mempengaruhi nilai persamaan.

3. Unsur-Unsur dalam Persamaan Linear

Dalam rumus persamaan linear, terdapat beberapa unsur penting yang perlu diperhatikan, antara lain:

4. Contoh Unsur dalam Persamaan Linear

Variabel: Dalam konteks persamaan linear, variabel berfungsi sebagai pengganti atau penambah suatu angka yang nilainya belum diketahui secara pasti.

Koefisien: Koefisien adalah angka yang menunjukkan jumlah dari variabel sejenis dalam persamaan tersebut

Konstanta

Nilai tetap yang tidak berubah bersamaan dengan perubahan variabel lainnya.

Suku

Persamaan Linear: Bagian-bagian utama sebuah persamaan linear, mencakup koefisien, variabel, dan konstanta.

Model Analisis Data

1. Menggunakan Model Eksponensial Negatif

Dalam beberapa situasi, model linear tidak cukup efektif untuk mendeskripsikan data. Salah satu alternatif adalah menggunakan model eksponensial negatif, seperti misalnya untuk menghitung peluruhan zat radioaktif. Menurut model ini, $x + B$ merupakan berat awal suatu zat pada waktu $T = 0$, X adalah berat akhir bagian zat yang tidak rusak pada waktu T , dan parameter lainnya terkait dengan kecepatan peluruhan dan hubungan dengan waktu paruh zat. Penggunaan model nonlinear diperlukan untuk deskripsi yang tepat.

2. Fungsi Konstansi Elastisitas Substitusi (CES)

Fungsi CES dikembangkan oleh Arrow, Chenery, Minhas, dan Solow pada tahun 1961 [2]. Elastis substitusi mengukur kemampuan sebuah perusahaan untuk mensubstitusi input satu dengan input lainnya agar produksi tetap pada level yang sama. Model produksi CES didefinisikan sebagai berikut:

$$Y = f(K^{\alpha}, L^{\beta})$$

dimana K adalah modal, L adalah tenaga kerja, dan α serta β adalah parameter efisiensi, distribusi, substitusi, dan skala balik. Namun, seperti tampak pada persamaan (3), model produksi CES tidak dapat distrukturasi menjadi bentuk linear.

3. Metode Untuk Mengestimasi Model Regresi Nonlinier

Metode Coba dan Ralat (Trial and Error)

Langkah-langkah:

- Menguji beberapa nilai untuk parameter B dalam persamaan regresi.
- Menghitung jumlah kesalahan kuadrat (galat kuadrat).
- Model yang dianggap terbaik adalah yang memiliki nilai galat kuadrat terkecil.

Kelemahan:

- a. Memerlukan pengetahuan sebelumnya tentang data yang akan dimodelkan agar dapat memperkirakan nilai B dengan akurat.
- b. Tidak ada jaminan bahwa parameter yang dipilih akan menghasilkan jumlah kesalahan kuadrat (SSE) yang paling rendah.
- c. Jika nilai SSE tidak mencapai minimum, proses pencarian solusi bisa memakan waktu yang cukup lama.

4. Metode Optimalisasi Langsung (Direct Optimizion)

Langkah-langkah :

- a. -menurunkan persamaan SSE dengan suatu nilai parameter atau koefisien yang tidak diketahui
- b. -menyamakan persamaan tersebut dengan nol
- c. -menyelesaikan secara simultan

Kelemahan:

- a. Tidak dapat diselesaikan dengan mudah (secara eksplisit)
- b. Tetap memerlukan Teknik coba dan ralat untuk menentukan nilai parameter
- c. Proses sangat lambat untuk mencapai nilai estimasi yang konvergen

Pedoman Memilih Antara Regresi Linear dan Nonlinear

Sebagai langkah awal, disarankan untuk menggunakan regresi linier guna mengevaluasi apakah data dapat diwakili oleh jenis kurva tertentu. Jika regresi linier tidak memberikan kecocokan yang memadai, maka regresi nonlinier mungkin menjadi pilihan yang lebih tepat.

Regresi linier lebih mudah diterapkan dan lebih sederhana dalam hal interpretasi, serta memberikan statistik yang lebih banyak untuk menilai model. Meskipun regresi linier dapat digunakan untuk memodelkan kurva, ada batasan pada bentuk kurva yang dapat disesuaikan. Dalam beberapa kasus, regresi linier mungkin tidak mampu menyesuaikan kurva tertentu dalam data.

Di sisi lain, regresi nonlinier mampu menyesuaikan berbagai jenis kurva, tetapi sering kali memerlukan usaha lebih untuk menemukan model yang paling sesuai dan untuk memahami peran variabel independen. Selain itu, nilai R-squared tidak berlaku untuk regresi nonlinier, dan perhitungan nilai p untuk estimasi parameter tidak dapat dilakukan.

Apa Itu Regresi Linier? Dan Bagaimana Model Regresi Linier

Regresi linier merupakan teknik analisis data yang digunakan untuk memprediksi nilai yang tidak diketahui dengan memanfaatkan nilai-nilai lain yang sudah diketahui dan saling terkait. Secara matematis, teknik ini memodelkan variabel yang tergantung (tidak diketahui) dan variabel independen (dikenal) dalam bentuk persamaan linier. Sebagai contoh, jika kita memiliki data mengenai pengeluaran dan pendapatan tahun lalu, teknik regresi linier akan menganalisis data tersebut dan menemukan bahwa pengeluaran adalah setengah dari pendapatan. Dengan demikian, informasi ini dapat digunakan untuk prediksi di masa depan.

Mengapa regresi linier penting ?

Model nonlinier tepat digunakan ketika model nonlinier lebih berguna daripada model linier

Model regresi linier tergolong sederhana dan menyediakan rumus matematika yang mudah dipahami untuk menghasilkan prediksi. Teknik statistik ini telah ada sejak lama dan dapat dengan mudah diterapkan menggunakan perangkat lunak serta dalam proses komputasi. Banyak perusahaan memanfaatkannya untuk mengubah data mentah menjadi informasi yang dapat diandalkan dan diimplementasikan. Para peneliti di berbagai disiplin ilmu, termasuk biologi, ilmu perilaku, lingkungan, dan ilmu sosial, menggunakan regresi linier untuk melakukan analisis data awal serta meramalkan tren di masa depan. Selain itu, banyak metode dalam ilmu data, seperti pembelajaran mesin dan kecerdasan buatan, juga memanfaatkan regresi linier untuk menyelesaikan masalah yang kompleks

Memodelkan probabilitas respons tersebut sebagai fungsi dari tingkat suatu zat di lingkungan, misalnya insektisida.

- a. Probabilitas itu berada dalam rentang antara 0 dan 1, dan batasan ini dapat dicapai.
- b. Model regresi linier tidak memiliki batasan tersebut.

Oleh karena itu, setiap model linier yang menggambarkan $P = P\{\text{Kutu busuk mati}\}$ terkait dengan paparan insektisida menggunakan fungsi linear dari konsentrasi insektisida hanya akan efektif dalam rentang konsentrasi tertentu. Di luar rentang ini, model tersebut dapat memprediksi probabilitas negatif atau lebih dari 1, yang keduanya tidak mungkin terjadi dan bukan merupakan prediksi yang berguna. Sebagai alternatif yang lebih modern, kita bisa menggunakan model linear umum, yang merupakan model nonlinier tetapi dapat dilinearisasi. Salah satu jenis model linear umum yang akan dibahas di sini adalah regresi logistik (meskipun regresi probit juga bisa dipertimbangkan). Dalam kedua kasus ini, kita berasumsi bahwa ada suatu fungsi (disebut fungsi tautan) dari probabilitas yang merupakan fungsi linear dari variabel penjelas (regressor).

Metode Kuadrat Terkecil

Metode Kuadrat Terkecil

Metode kuadrat terkecil adalah teknik statistik yang penting untuk menentukan garis regresi atau garis terbaik yang sesuai dengan pola data tertentu. Metode ini dijelaskan melalui rumus yang melibatkan parameter tertentu. Penggunaan metode ini sangat umum dalam analisis dan evaluasi regresi. Dalam konteks analisis regresi, metode ini sering dianggap sebagai pendekatan standar untuk memperkirakan sistem persamaan yang memiliki lebih banyak persamaan dibandingkan jumlah variabel yang tidak diketahui.

Secara fundamental, metode kuadrat terkecil bertujuan untuk meminimalkan jumlah kuadrat dari deviasi atau kesalahan dalam hasil setiap persamaan. Rumus untuk menghitung jumlah kuadrat kesalahan ini berguna untuk mengidentifikasi variasi dalam data yang telah diamati.

Metode kuadrat terkecil juga banyak diterapkan dalam pencocokan data. Pencocokan yang paling baik diharapkan dapat mengurangi total kesalahan kuadrat atau residual, yang merupakan

selisih antara nilai yang diamati atau eksperimen dan nilai yang diprediksi oleh model.

Ada dua kategori utama dalam masalah kuadrat terkecil, yaitu:

1. Kuadrat Terkecil Biasa (Linier)
2. Kuadrat Terkecil Nonlinier

Kategorisasi ini tergantung pada apakah residual bersifat linier atau non-linier. Masalah linier sering muncul dalam analisis regresi statistik, sedangkan masalah non-linier biasanya digunakan dalam metode penyempurnaan iteratif, di mana model didekati dengan model linier pada setiap langkah iterasi.

Metode Newton-Raphson

Metode Newton-Raphson adalah teknik yang digunakan untuk menemukan pendekatan atau estimasi terhadap akar dari fungsi real. Rumus untuk metode Newton-Raphson adalah sebagai berikut:

Parafrazed Algoritma Metode Newton-Raphson

Berikut adalah urutan langkah-langkah untuk menggunakan metode Newton-Raphson:

1. Identifikasikan Persamaan Nonlinear: Tentukan persamaan yang ingin Anda pecahkan.
3. Atur Batas Epsilon: Tetapkan batas toleransi kesalahan (epsilon) untuk iterasi.
4. Mulailah dengan Titik Awal: Pilih titik awal untuk proses iterasi.
5. Perhitungkan Derivatif: Hitung nilai derivatif dari fungsi pada titik awal.
6. Iterasi Selanjutnya: Gunakan formula $x_{i+1} = x_i - f'(x_i)/f''(x_i)$ untuk mendapatkan nilai berikutnya.
7. Cek Konvergensi: Jika perubahan nilai x_i lebih kecil dari batas toleransi, maka iterasi berhasil.
8. Ubah Titik Awal: Ubah nilai x_i menjadi nilai terkini dan ulangi langkah 4 sampai kondisi konvergensi tercapai.

Metode Newton-Raphson adalah sebuah metode iterasi yang efektif untuk menyelesaikan persamaan nonlinear dengan konvergensi kuadratik. Prosesnya melibatkan perhitungan derivat dan iterasi hingga mencapai titik solusi yang akurat. Dengan menggunakan batas toleransi (epsilon) dan iterasi maksimum, Anda dapat memastikan bahwa proses iterasi stabil dan efisien.

Parafrazed Pentingnya Menggunakan Data untuk Memprediksi Tren Masa Depan

Menggunakan data dalam bisnis sangatlah penting. Pengumpulan data luas dan akurat memungkinkan organisasi untuk melihat tren dan peluang baru. Dengan menganalisis data secara terus-menerus, organisasi dapat menemukan cara-cara baru untuk melakukan perbaikan.

Pengaruh Variabel Bebas Terhadap Variabel Tidak bebas

Setelah mengerti definisi variabel bebas dan variabel tidak bebas, saya juga ingin tahu apa saja perbedaan antara variabel bebas dan variabel tidak bebas. Secara umum, variabel bebas dan variabel tidak bebas adalah dua komponen penting dalam suatu penelitian. Perbedaan utama antara keduanya adalah hubungan yang saling mempengaruhi.

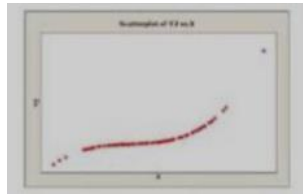
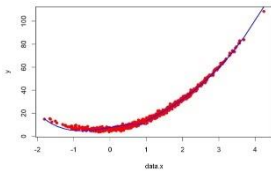
Perbedaan yang tampak adalah jika variabel tersebut adalah variabel bebas, maka variabel tersebut dapat berubah sendiri tanpa dipengaruhi oleh variabel lain. Di sisi lain, jika variabel tersebut adalah variabel tidak bebas, maka ia akan ikut berubah atau dipengaruhi oleh variabel lain. Sebaliknya, variabel bebas biasanya tidak dipengaruhi oleh nilai variabel tidak bebas. Sehingga, variabel bebas adalah variabel yang mandiri dalam suatu eksperimen, sedangkan variabel tidak bebas tidak dapat hadir tanpa adanya variabel bebas. Contohnya, misalkan judul penelitian “Pengaruh Pendidikan dan Pendapatan” dan “Hubungan Antara Pendidikan dengan Pendapatan.

Contoh kasusnya adalah judul penelitian yang awalnya fokus pada pendidikan. Judul penelitian tersebut ingin mengeksplorasi apakah pengaruh pendidikan seseorang berdampak pada pendapatannya. Oleh karena itu, pendidikan dalam judul tersebut

dianggap sebagai variabel bebas (X), sedangkan pendapatan dianggap sebagai variabel terikat (Y). Dengan demikian, contoh variabel bebas dalam judul tersebut adalah pendidikan (variabel X), sedangkan contoh variabel tidak bebasnya adalah pendapatan (variabel Y).

Contoh Gambar Regresi Non Linier

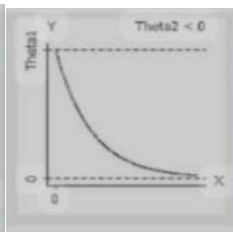
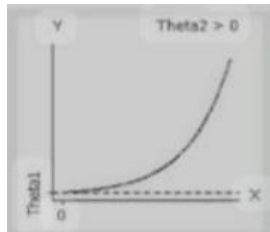
Polinomial

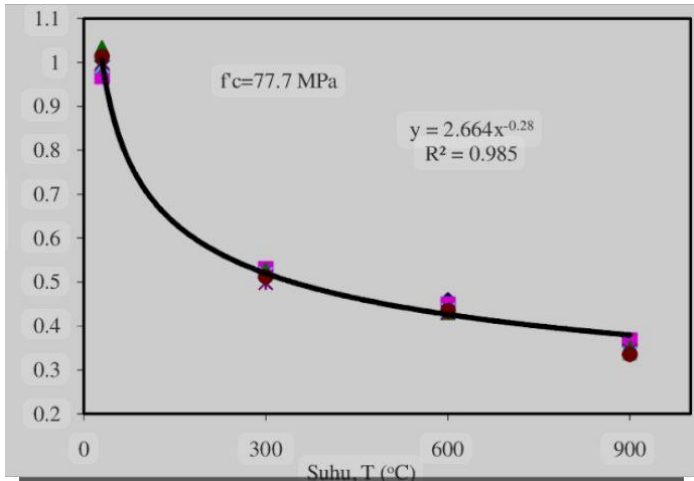


$$Y = \theta_0 + \theta_1 X + \theta_2 X^2 + \theta_3 X^3$$

$$Y = \theta_0 + \theta_1 X + \theta_2 X^2$$

Ekspensial $\rightarrow Y = \theta_1 e^{\theta_2 X}$





Tabel Data kecepatan rata-rata (y) untuk berbagai jarak (x) lari dari 5 pelari berusia di atas 70 tahun:

No.	Jarak (x)	Kecepatan (y)
1	1.6	4.7
2	3	4.2
3	5	4.1
4	10	3.9
5	42	3.8

Catatan: Data ini diambil dari majalah *American Scientist*, Mei-Juni 1981.

PROSES REGRESI NONLINIER (POWER FUNCTION)

Karena hubungan antara jarak dan kecepatan tidak linear, digunakan model regresi nonlinier bentuk power function:

$$y = c \cdot x^d$$

Untuk mempermudah perhitungan, dilakukan transformasi logaritmik natural (\ln) pada kedua sisi

Ini mengubah model menjadi regresi linier dalam bentuk:

$$Y' = a + bX' \quad Y' = a + bX'$$

TABEL PERHITUNGAN TRANSFORMASI LOGARITMIK

Berikut tabel perhitungan nilai-nilai logaritma dan kuadratnya:

No.	x	y	ln x	ln y	(ln x) ²	(ln y) ²	ln x · ln y
1	1.6	4.7	0.47	1.55	0.22	2.40	0.73
2	3	4.2	1.10	1.44	1.21	2.07	1.58
3	5	4.1	1.61	1.41	2.59	1.99	2.27
4	10	3.9	2.30	1.36	5.29	1.85	3.13
5	42	3.8	3.74	1.34	13.99	1.80	5.01
Total	61.6	20.7	9.22	7.1	23.3	10.11	12.72

MENGHITUNG KOEFISIEN REGRESI LINIER (a dan b)

Rumus regresi linier sederhana:

$$b = \frac{n \sum(X'Y') - \sum X' \sum Y'}{n \sum(X')^2 - (\sum X')^2}$$

Substitusi nilai:

$$\begin{aligned}
 n &= 5 \quad n = 5 \\
 \sum X' &= 9.22 \quad \sum X' = 9.22 \\
 \sum Y' &= 7.1 \quad \sum Y' = 7.1 \\
 \sum X'Y' &= 12.72 \quad \sum X'Y' = 12.72 \\
 \sum (X')^2 &= 23.3 \quad \sum (X')^2 = 23.3 \\
 b &= \frac{5 \cdot 12.72 - 9.22 \cdot 7.1}{5 \cdot 23.3 - (9.22)^2} \\
 &= \frac{63.6 - 65.462}{116.5 - 85.0084} = \frac{-1.862}{31.4916} \\
 &\approx -0.06 \\
 X' &= 9.225 = 1.844, Y' = 7.15 = 1.42X' = 59.22 = 1.844, Y' = 57.1 \\
 &= 1.42
 \end{aligned}$$

data berikut menyatakan kecepatan rata-rata untuk berbagai jarak dari 5 pelari yang semuanya di atas 70 tahun, tertua di American scientist (mei-juni 1981).

X	Jarak (S)		1,6	3	5	10	42
Y	Kecepatan (v)		4,7	4,2	4,1	3,9	3,8
No	S	V_1	$\ln S_1$	$\ln V_1$	$(\ln \{S\})^2$	$(\ln \{V\})^2$	$\ln S_1 \cdot \ln V_1$
1	1,6	4,7	0,47	1,55	0,22	2,4	0,73
2	3	4,2	1,1	1,44	1,21	2,07	1,58
3	5	4,1	1,61	1,41	2,59	1,99	2,27
4	10	3,9	2,3	1,36	5,29	1,85	3,13
5	42	3,8	3,74	1,34	13,99	1,8	5,01
Tota 1	61, 6	20, 7	9,22	7,1	23,3	10,11	12,72

Prediksikan kecepatan rata-rata seorang pelari berusia diatas 70 tahun yang jarak larinya 10 km dengan menggunakan regresi tak linier $B = Cx^d$

Solusi :

Persamaan ini jika diturunkan kedalam bentuk logaritma natural maka didapatkan

$$\ln B = \ln c + d \ln s$$

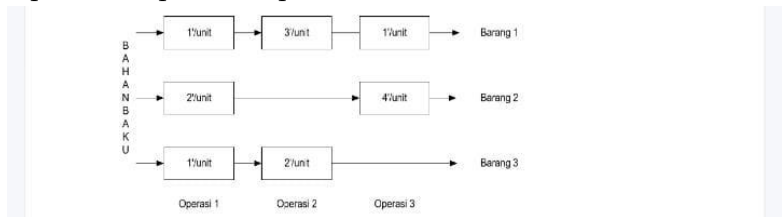
Koefisien c dan d bisa kita cari secara manual dengan diperoleh c = 4,60 dan d = -0,06.

Table

SAMPEL	JUMLAH MAHASISWA DALAM 1 KELAS (X)	PEMAHAMAN MATERI (Y)
1	33	80
2	35	82
3	34	85
4	28	87
5	25	95
6	22	94
7	20	94
8	26	88
9	27	90
10	25	92

STUDI KASUS

1. Sebuah Perusahaan memproduksi tiga macam barang, pembuatan barang-barang tersebut di lakukan melalui tiga proses produksi seperti pada gambar dibawah ini. Waktu pengerjaan setiap barang dapat dilihat pada setiap kotak.



Oleh karna mesin terebut juga dipakai untuk pembuatan barang lain, maks waktu produksi yang tersedia dari proses terbatas sebesar 430,460, dan 420 menit untuk setiap prosesnya. Studi pasar memperlihatkan keuntungan setiap macam barang berturut-turut sebesar \$3, \$2, dan \$5 per unit. Tentukan Tingkat produksi yang optimal

Penyelesaian :

Model Program Linear Variabel

x_j = jumlah produksi barang j ; $j = 1, 2, 3$

fungsi tujuan

Maks. $Z = 3x_1 + 2x_2 + 5x_3$

Batasan kendala

Proses produksi 1: $1x_1 + 2x_2 + 1x_3 < 430$

Proses produksi 2: $3x_1 + 0x_2 + 2x_3 < 460$

Proses produksi 3: $1x_1 + 4x_2 + 0x_3 < 420$ $x_1 ; x_2 ; x_3 > 0$

2. Contoh kasus

PT . Adil Sejahtera memproduksi dua jenis mainan yang terbuat dari kayu , berupa boneka dan kereta. Boneka dijual dengan harga Rp 15.000,-/ lusin yang setiap lusinnya memerlukan biaya material sebesar Rp 10.000,- serta biaya tenaga kerja sebesar Rp 14.000,- kereta api yang 1jam pemolesan dan 1jam pekerjaan kayu.

Meskipun pada setiap minggunya perusahaan ini dapat memenuhi seluruh material yang diperlukan ,jam kerja yang tersedia hanya 100jam untuk pemolesan dan 80 jam untuk pekerjaan kayu .

Dari pengamatan pasar selama ini dapat dikatakan bahwa kebutuhan dari 40 lusin yang terjual setiap minggunya. Bagaimana kah formulasi dari persoalan diatas untuk mengetahui berapa lusin jenis mainan masing-masing yang harus dibuat setiap minggu nya agar diperoleh keuntungan yang maksimum pendapatan Perusahaan untuk sebesar Rp 27.000 dan kereta api Rp 21.000, karakteristik-karakteristik yang di gunakan dalam persoalan program linier:

Variabel Keputusan adalah variabel yang menguraikan secara lengkap Keputusan-keputusan yang akan di buat. Dalam persoalan ini, variabel Keputusan yang akan dibuat. Dalam persoalan ini, variabel Keputusan akan menentukan berapa banyak boneka dan kereta api masing-masing yang harus dibuat setiap minggunya. Misalkan X_1 = banyaknya yang dibuat setiap minggunya.

X_2 =banyaknya kereta api yang di buat setiap minggu

Fungsi tujuan merupakan fungsi dari variabel Keputusan yang akan di maksimumkan (untuk pendapatan atau keuntungan) atau diminimumkan (untuk ongkos).

Pada persoalan ini akan dimaksimumkan (pendapatan/minggu)-(ongkos material/minggu)- ongkos tenaga kerja minggu).

Pendapatan dan ongkos-ongkos ini dapat diekspresikan dengan menggunakan variabel keputusan x_1 dan x_2 sebagai berikut:

Pendapatan/minggu pendapatan/minggu dari boneka
pendapatan/minggu dari kereta api

$$27x_1 + 21x_2$$

Ongkos material/minggu $10x_1 + 9x_2$

Ongkos tenaga kerja/minggu $14x_1 + 11x_2$

- Kasus : seorang mahasiswa akan meneliti apakah terdapat pengaruh promosi terhadap volume penjualan pada perusahaan-perusahaan di kabupaten malang untuk kepentingan penelitian tersebut diambil 10 perusahaan sejenis yang telah melakukan kegiatan promosi
- Alternatif judul penelitian : pengaruh biaya promosi terhadap volume penjualan Perusahaan
- Rumusan masalah : apakah terdapat pengaruh antar biaya promosi terhadap penjualan Perusahaan?
- Hipotesis : terhadap pengaruh antara biaya promosi terhadap penjualan perusahaan?

no	Penjualan	promosi
1	73600	203000
2	70150	18400
3	96600	39100
4	80500	26450
5	101200	31050

6	105800	36800
7	82800	20700
8	88550	25300
9	86250	23000
10	72450	20700

HASIL ANALISIS

Model	Unstandardized Coefficients		Standardized Coefficients Beta	t	sig
	B	/std.error			
1 (constant)	45800,119		8618,582		,001
PROMOSI	1,512		,316	,861	,001

Maka persamaan regresi linear sederhana tersebut adalah :

. konstanta sebesar 45800,119 , tanpa kegiatan promosi (X) atau biaya promosi sama dengan nol (0) maka volume penjualan (Y) nilainya sebesar 45800,119

. koefisien regresi variabel biaya promosi (X) sebesar 1,512 ; jika biaya promosi mengalami kenaikan sebesar Rp 1 mak , volume penjualan (Y) akan mengalami peningkatan sebesar Rp 1,512. Sebaliknya jika biaya promosi mengalami penurunan Rp 1 , maka volume penjualan (Y) akan mengalami penurunan sebesar Rp1,512

Modal	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std. Error			
1 (constant)	-672,770	1041,144	,564		,539
HARGA	,844	,222	,965	3,794	,007
PENDAPATAN	,437	,067		5,490	,000

Uji t dimaksud untuk mengetahui apakah dalam model regresi variabel independent secara parsial berpengaruh signifikan terhadap variabel dependen (Y).

Hipotesis 1 ; nilai uji t X1 diperoleh sebesar 3,794 dengan Tingkat signifikansi sebesar 0,007 yang bermakna bahwa hasil berpengaruh signifikan karna $0,007 < 0,05$ ($\alpha=5\%$) dan secara otomatis H_a diterima dan H_o ditolak.

Nilai signifikansi sebesar 0,007 memberikan makna bahwa Tingkat kepercayaan yang di berikan sebesar 99,3% ; Tingkat kepercayaan ini lebih tinggi di banding dengan Tingkat kepercayaan yang diisyaratkan yaitu 99% ($\alpha=5\%$).

Hipotesis 2 ; nilai uji t X2 diperoleh sebesar 6,490 dengan Tingkat signifikansi sebesar 0,000 yang bermakna bahwa hasil berpengaruh signifikan karna $0,000 < 0,05$ ($\alpha=5\%$) dan secara otomatis H_a di terima dan H_o ditolak.

Nilai signifikansi sebesar 0,000 memberikan makna bahwa Tingkat kepercayaan yang di berikan sebesar 100%; Tingkat kepercayaan ini lebih tinggi di banding dengan Tingkat kepercayaan yang di isyaratkan yaitu 99% ($\alpha=5\%$).

modal	Sum of squares	Df	Mean square	F	Sig
1 regression	20044128	2	10022064,14	22,343	,001a
Residual	3139872	7	448553,103		
total	23184000	9			

a.predictors ; (constant), PENDAPATAN, HARGA

b. Dependent variable : KONSUMSI

Uji F dimaksudkan untuk mengetahui apakah dalam model regresi variabel independen secara simultan.

Berdasarkan tabel tersebut diperoleh nilai F sebesar 22,343 dengan Tingkat signifikansi sebesar 0,001. Yang bermakna bahwa hasil berpengaruh signifikan karena $0,001 < 0,05$ ($\alpha=5\%$).

Nilai signifikansi sebesar 0,001 meberikan makna bahwa Tingkat kepercayaan yang diberikan sebesar 99,9%; Tingkat kepercayaan ini lebih tinggi dibandingkan dengan Tingkat kepercayaan yang diisyaratkan yaitu 99% ($\alpha=5\%$).

MODEL	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,930a	,865	,826	669,74107

Analisis Korelasi Ganda dipergunakan untuk mengetahui apakah korelasi antara dua atau lebih variable independent terhadap variable dependen (Y) secara serentak.

Berdasarkan output pada tabek Model summary diperoleh nilai koefisien korelasi Ganda (R) sebesar 0,930, hal tersebut menunjukkan bahwa terdapat hubungan yang sangat kuat antara variable harga dan pendapatan terhadap konsumsi produksi.

Nilai koefisien of Determinasi menunjukkan besarnya prosentase sumbangan pengaruh variable independen terhadap variable dependen (Y) secara serentak.

Berdasarkan output pada tabek model summary diperoleh nilai koefisien of deteminasi (r^2) sebesar 0,865 atau 86,5% yang lain bermakna bahwa prosentase sumbangan pengaruh varible independent sebesar 86,5% sedangkan sisanya 13,5% dipengaruhi oleh variable lain

Sedangkan nilai Adjusted R Square merupakan nilai R Squire yang telah di standarisasi

Standart Error of The Estimate menunjukkan besarnya nilai kesalahan model regresi dalam memprediksi nilai Y

DAFTAR PUSTAKA

- Achmad, F. (2017). *Regresi linier berganda dan non linear*. Scribd.
<https://id.scribd.com/presentation/438237543/36104-regresi-Linier-Berganda-Dan-Non-Linear>
- Amazon Web Services. (n.d.). *Apa itu regresi linear?*
<https://aws.amazon.com/id/what-is/linear-regression/>
- azizah, n. (2021). *Regresi linier sederhana*. Universitas Gunadarma.
- Gozali, I., & Fauzi, A. (2024). *Statistik bisnis I: Ukuran pemusatan data berkelompok dan tidak berkelompok*. Fakultas Ekonomi dan Bisnis, Universitas Bhayangkara Jakarta Raya.
- Gusniar, Anggi, L., & Endang, L. (2017). Analisis regresi logistik multinomial pada faktor-faktor yang mempengaruhi kadar gula darah penderita diabetes mellitus. *eprints.uny.ac.id*.
<https://eprints.uny.ac.id/>
- ispriyanti, d. (2012). *Analisis regresi linier*.
- Kenton, W. (2023, December 26). What is nonlinear regression? Comparison to linear regression. *Investopedia*.
https://translate.google.com/translate?u=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FNonlinear_regression&hl=id&sl=en&tl=id&client=srp&prev=search&source=sh%2F%2Fsrp%2Fwr%2Fm1%2F4&kgs=ea83864b019d0e5e
- nafi'iyah, n. (2016). *Perbandingan regresi linier*.
- ratna, m. (2012). *Regresi linier multi nominal*.
- Regresi non linear. (2017, January 31). *Repository Universitas Komputer Indonesia*.
<https://repository.unikom.ac.id/33244/1/regresi%20non%20linier.pdf>
- SAPUTRA, R., YUNIARTI, D., & WAHYUNINGSIH, S. (n.d.). *Analisis regresi eksponensial berganda*. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman.
<https://fmipa.unmul.ac.id>
- sartono, b. (2021). *Regresi Newton-Raphson*.
- siregar, t. m. (2023). *Model regresi non linier*.
- Spada UNS. (n.d.). *Non linear regression*.
<https://spada.uns.ac.id/mod/resource/view.php?id=181303>

Studocu. (n.d.). *Analisis regresi non linear berganda*.
<https://www.studocu.com/id/document/universitas-negeri-makassar/education-statistics/analisis-regresi-non-linear-berganda/47401374>

Ulum, H. M. (n.d.). *Statistika*. Universitas Negeri Malang.

BAB 8

ANALISIS DATA BERKALA METODE UNTUK MEMPEROLEH DATA BERKALA DAN BPS

Analisis data berkala adalah proses sistematis yang melibatkan pengumpulan, pengolahan, dan interpretasi data yang diperoleh secara teratur dalam interval waktu tertentu. Data berkala ini dapat mencakup informasi tentang berbagai aspek kehidupan masyarakat, seperti demografi, ekonomi, kesehatan, dan pendidikan. Analisis ini bertujuan untuk menemukan pola, tren, dan hubungan antara variabel yang berkaitan, dan juga untuk memberikan informasi yang bisa digunakan dalam membuat keputusan dan perencanaan kebijakan. Analisis data berkala merupakan alat yang sangat berharga dalam memahami dinamika sosial, ekonomi, dan politik. Melalui analisis yang sistematis dan terstruktur, kita dapat memperoleh wawasan yang mendalam, yang mendasari perencanaan dan pengambilan keputusan yang lebih baik. Dengan memanfaatkan data berkala dari lembaga seperti BPS, para pemangku kepentingan dapat merumuskan kebijakan yang lebih baik dan cepat tanggap terhadap kebutuhan masyarakat.

Karakteristik Data Berkala

1. **Pengumpulan Rutin:** Data berkala dihasilkan melalui pengumpulan informasi yang dilakukan secara rutin, misalnya setiap bulan, kuartal, atau tahunan. Contoh data berkala termasuk data pengangguran, inflasi, dan statistik kesehatan masyarakat.
2. **Interval Waktu yang Jelas:** Setiap periode pengumpulan data ditentukan dengan jelas, sehingga memungkinkan analisis perubahan dari waktu ke waktu. Misalnya, survei pengeluaran rumah tangga yang dilakukan setiap tahun.
3. **Variabel yang Konsisten:** Data berkala sering kali melibatkan variabel yang sama untuk setiap periode, sehingga memudahkan perbandingan. Misalnya, dalam survei kepuasan masyarakat, pertanyaan yang diajukan dapat konsisten dari tahun ke tahun.

4. Keterbandingan: Salah satu keunggulan analisis data berkala adalah kemampuannya untuk membandingkan data dari berbagai periode, yang membantu dalam memahami tren jangka panjang dan fluktuasi.

Manfaat Analisis Data Berkala

1. Identifikasi Tren dan Pola: Analisis ini memungkinkan peneliti untuk mengidentifikasi pola yang berulang atau perubahan signifikan dalam variabel yang dianalisis, seperti tren pertumbuhan ekonomi atau perubahan tingkat kemiskinan.
2. Dasar Pengambilan Keputusan: Dengan informasi yang diperoleh dari analisis data berkala, pengambil keputusan dapat merumuskan kebijakan yang berbasis bukti, meningkatkan efektivitas program, dan merespons kebutuhan masyarakat secara lebih tepat.
3. Evaluasi Program dan Kebijakan: Data berkala memungkinkan evaluasi terhadap efektivitas program atau kebijakan yang telah diterapkan. Misalnya, analisis data dapat membantu mengevaluasi dampak suatu program pengentasan kemiskinan.
4. Prediksi Masa Depan: Dengan menganalisis pola yang ada, analisis data berkala dapat digunakan untuk membuat prediksi tentang perkembangan di masa depan, membantu perencanaan jangka panjang yang lebih baik.
5. Keterlibatan Masyarakat: Data yang transparan dan mudah diakses dapat meningkatkan keterlibatan masyarakat dalam proses pengambilan keputusan, sehingga menciptakan kebijakan yang lebih inklusif.

Analisis data berkala adalah alat yang sangat berharga dalam memahami dinamika sosial, ekonomi, dan politik. Melalui analisis yang sistematis dan terstruktur, kita dapat memperoleh wawasan yang mendalam, yang mendasari perencanaan dan pengambilan keputusan yang lebih baik. Dengan memanfaatkan data berkala dari lembaga seperti BPS, para pemangku kepentingan dapat merumuskan lebih baik dan cepat tanggap terhadap kebutuhan masyarakat.

Konsep Tentang Analisis Data Berkala

Analisis data berkala adalah proses sistematis yang melibatkan pengumpulan, pengolahan, dan interpretasi data yang diperoleh secara teratur dalam interval waktu tertentu. Data berkala ini dapat mencakup informasi tentang berbagai aspek kehidupan masyarakat, seperti demografi, ekonomi, kesehatan, dan pendidikan. Analisis ini bertujuan Untuk menemukan pola, tren, dan hubungan antara variabel yang berkaitan, serta untuk menghasilkan wawasan yang dapat digunakan dalam pengambilan keputusan dan perencanaan kebijakan.

Karakteristik Data Berkala

1. Pengumpulan Rutin: Data berkala dihasilkan melalui pengumpulan informasi yang dilakukan secara rutin, misalnya setiap bulan, kuartal, atau tahunan. Contoh data berkala termasuk data pengangguran, inflasi, dan statistik kesehatan masyarakat.
2. Interval Waktu yang Jelas: Setiap periode pengumpulan data ditentukan dengan jelas, sehingga memungkinkan analisis perubahan dari waktu ke waktu. Misalnya, survei pengeluaran rumah tangga yang dilakukan setiap tahun.
3. Variabel yang Konsisten: Data berkala sering kali melibatkan variabel yang sama untuk setiap periode, sehingga memudahkan perbandingan. Misalnya, dalam survei kepuasan masyarakat, pertanyaan yang diajukan dapat konsisten dari tahun ke tahun.
4. Keterbandingan: Salah satu keunggulan analisis data berkala adalah kemampuannya untuk membandingkan data dari berbagai periode, yang membantu dalam memahami tren jangka panjang dan fluktuasi.

Proses Analisis Data Berkala

1. Pengumpulan Data: Data dikumpulkan dengan berbagai cara, seperti survei, sensus, pengukuran administratif, atau observasi. Lembaga seperti BPS sering kali menggunakan metode yang telah terstandarisasi untuk memastikan keakuratan dan keterbandingan.

2. Pengolahan Data: Setelah data dikumpulkan, langkah berikutnya adalah mengolah data tersebut, yang meliputi pembersihan, pengkodean, dan pengorganisasian informasi agar siap untuk dianalisis. Proses ini juga mencakup penanganan data hilang atau tidak valid.
3. Analisis Data: Di sini, berbagai teknik analisis statistik diterapkan untuk menggali informasi dari data. Metode yang umum digunakan termasuk:
4. Interpretasi dan Penyajian Hasil: Setelah analisis, hasilnya diinterpretasikan dan disajikan dalam bentuk yang dapat dimengerti, seperti grafik, tabel, atau laporan. Penyajian yang baik sangat penting agar informasi yang dihasilkan dapat digunakan oleh pemangku kepentingan.
Statistik Deskriptif: Untuk memberikan gambaran umum tentang data, seperti rata-rata, median, dan distribusi.
Analisis Regresi: Untuk mengeksplorasi hubungan antar variabel.
Analisis Tren: Untuk mengidentifikasi pola dari waktu ke waktu.
Analisis Time Series: Untuk memprediksi nilai di masa depan berdasarkan data yang telah dikumpulkan.

Manfaat Analisis Data Berkala

1. Identifikasi Tren dan Pola: Analisis ini memungkinkan peneliti untuk mengidentifikasi pola yang berulang atau perubahan signifikan dalam variabel yang dianalisis, seperti tren pertumbuhan ekonomi atau perubahan tingkat kemiskinan.
2. Dasar Pengambilan Keputusan: Dengan informasi yang diperoleh dari analisis data berkala, pengambil keputusan dapat merumuskan kebijakan yang berbasis bukti, meningkatkan efektivitas program, dan merespons kebutuhan masyarakat secara lebih tepat.
3. Evaluasi Program dan Kebijakan: Data berkala memungkinkan evaluasi terhadap efektivitas program atau kebijakan yang telah diterapkan. Misalnya, analisis data dapat membantu mengevaluasi dampak suatu program pengentasan kemiskinan.

4. **Prediksi Masa Depan:** Dengan menganalisis pola yang ada, analisis data berkala dapat digunakan untuk membuat prediksi tentang perkembangan di masa depan, membantu perencanaan jangka panjang yang lebih baik.
5. **Keterlibatan Masyarakat:** Data yang transparan dan mudah diakses dapat meningkatkan keterlibatan masyarakat dalam proses pengambilan keputusan, sehingga menciptakan kebijakan yang lebih inklusif.

Kesimpulan

Analisis data berkala adalah alat yang sangat berharga dalam memahami dinamika sosial, ekonomi, dan politik. Melalui analisis yang sistematis dan terstruktur, kita dapat memperoleh wawasan yang mendalam, yang mendasari perencanaan dan membuat keputusan yang lebih baik.. Dengan menggunakan data berkala dari lembaga seperti BPS, para pemangku kepentingan dapat merumuskan kebijakan yang lebih efektif dan responsif terhadap kebutuhan masyarakat.

Macam Macam Analisis Data

Terdapat beberapa jenis analisis data sering dipakai di berbagai bidang, Setiap jenis memiliki cara dan tujuan yang berbeda. Berikut ini adalah beberapa macam analisis data yang paling umum:

1. Analisis Deskriptif (Descriptive Analysis)

Analisis deskriptif adalah cara untuk menganalisis data penelitian yang bertujuan untuk melihat hasil penelitian berdasarkan satu sampel. Analisis ini dilakukan dengan menguji hipotesis deskriptif.

Teknik statistik yang digunakan untuk menguji hipotesis deskriptif harus sesuai dengan jenis data atau variabelnya, yang bisa berupa skala nominal, ordinal, atau interval/rasio. Ada 2 cara untuk menguji data nominal yaitu:

- a. Uji binomial adalah jenis distribusi yang terdiri dari dua kategori (dua peristiwa yang biasanya saling melengkapi). Jadi, jika dalam suatu populasi dengan jumlah n ada satu kategori yang disebut x , maka kategori lainnya adalah $n - x$. Probabilitas

untuk memperoleh nilai x dirumuskan : $P(x) = \binom{n}{x} p^x q^{n-x}$

- b. Uji chi-square satu sampel digunakan ketika populasi terdiri dari dua kategori atau lebih dan ukuran sampelnya besar. $\chi^2 = \sum \frac{(f_0 - f_h)^2}{f_h}$

Keterangan : χ^2 : kai kuadrat f_0 : frekuensi yang diamati f_h : frekuensi yang diinginkan

Analisis Inferensial (Inferential Analysis)

Statistika Inferensial (Inferential Statistics) membahas cara menganalisis data dan menarik kesimpulan. Ini berkaitan dengan pengambilan keputusan, seperti memperkirakan parameter dan menguji hipotesis. Statistika inferensial sering disebut sebagai Statistik Induktif. Metode statistik inferensia adalah cara yang digunakan untuk menganalisis sebagian data dan membuat perkiraan atau kesimpulan tentang keseluruhan data.

Contoh analisis data inferensial :

Uji Normalitas

Uji normalitas dengan menggunakan metode Liliefors, dengan ketentuan jika $L_{hitung} < L_{tabel}$ maka data normal. Nilai L_{tabel} diperoleh dari table uji Liliefors. Sedangkan L_{hitung} adalah harga terbesar dari $|F() - S(Z_i)|$, dimana dihitung dengan rumus angka normal baku:

$Z_i = \frac{X_i - \bar{x}}{s}$ \bar{x} = Rata - Rata s = simpangan baku

Analisis Prediktif (Predictive Analysis)

Tujuan: Untuk meramalkan hasil di masa depan berdasarkan data yang sudah ada.

Penjelasan: Analisis ini menggunakan data yang ada untuk memprediksi tren atau perilaku masa depan. Ini sering kali menggunakan algoritma pembelajaran mesin atau statistik.

Analisis prediktif adalah metode yang menggunakan data, model statistik, dan teknologi pembelajaran mesin untuk memprediksi hasil yang mungkin terjadi. Metode ini mencakup berbagai teknik dan teknologi statistik. Ada dua jenis model analitik prediktif yang sering digunakan, yaitu model klasifikasi dan regresi. Model klasifikasi membantu kita memprediksi hasil, seperti kapan

sebuah komponen mungkin akan gagal. Di sisi lain, model regresi digunakan untuk memprediksi angka, seperti berapa lama rata-rata sebelum kerusakan terjadi. Analisis prediktif semakin banyak dipakai oleh berbagai organisasi di banyak bidang, seperti rumah sakit, ritel, manufaktur, layanan keuangan, dan lainnya.

Analisis Diagnostik (Diagnostic Analysis)

Tujuan: Untuk memahami penyebab atau alasan di balik tren atau peristiwa tertentu dalam data.

Penjelasan: Analisis ini berfokus pada menjawab pertanyaan "mengapa" sesuatu terjadi. Hal ini melibatkan penggalan data untuk mencari faktor penyebab dari hasil tertentu.

Contoh:

Melakukan analisis korelasi untuk melihat hubungan antara variabel. Analisis regresi untuk menentukan variabel mana yang paling mempengaruhi hasil tertentu.

Menganalisis outlier atau kejadian ekstrem untuk memahami penyebab yang mendasari.

Kapan digunakan: Ketika ingin memahami alasan di balik suatu hasil atau peristiwa, seperti memahami mengapa penjualan menurun di suatu cabang toko.

Analisis Preskriptif (Prescriptive Analysis)

Tujuan: Untuk memberikan saran tindakan yang jelas berdasarkan hasil analisis.

Penjelasan: Analisis ini lebih berkembang karena Tidak hanya meramalkan hasil, tetapi juga memberikan saran atau Saran tentang langkah-langkah yang perlu diambil. Ini sering melibatkan simulasi dan pengoptimalan.

Contoh:

Simulasi Monte Carlo untuk mengeksplorasi berbagai skenario dan hasil potensial.

Penggunaan algoritma pengoptimalan untuk menemukan cara terbaik dalam meminimalkan biaya atau memaksimalkan keuntungan.

Rekomendasi produk dalam e-commerce, di mana sistem merekomendasikan produk berdasarkan riwayat pembelian pelanggan.

Kapan digunakan: Ketika ingin mendapatkan saran tindakan untuk mencapai hasil optimal, misalnya merancang strategi pemasaran terbaik berdasarkan pola perilaku konsumen.

Analisis Klaster (Cluster Analysis)

Tujuan: Analisis kluster adalah cara statistik yang dipakai untuk mengelompokkan sekumpulan objek atau data ke dalam kelompok (kluster) yang memiliki kesamaan di antara mereka. Tujuan dari analisis kluster adalah untuk memisahkan data ke dalam grup yang lebih homogen, sehingga objek dalam satu kluster lebih mirip satu sama lain dibandingkan dengan objek di kluster lain.

Analisis Jaringan (Network Analysis)

Tujuan: Untuk menganalisis hubungan antar elemen dalam suatu jaringan.

Penjelasan: Analisis jaringan memfokuskan pada hubungan antara node (entitas) dan bagaimana mereka terhubung. Ini berguna untuk menganalisis interaksi di dalam jaringan, seperti jejaring sosial atau jaringan komunikasi.

Contoh:

Analisis media sosial untuk memahami hubungan dan pengaruh antar pengguna.

Menganalisis alur informasi di dalam perusahaan untuk meningkatkan komunikasi dan efisiensi.

aringan transportasi untuk mengoptimalkan rute pengiriman atau perjalanan.

Kapan digunakan: Ketika perlu menganalisis hubungan antar entitas dalam sistem, seperti dalam media sosial atau jaringan transportasi.

Analisis Regresi (Regression Analysis)

Tujuan: Untuk memahami hubungan antara variabel yang dipengaruhi dan satu atau lebih variabel yang mempengaruhi.

Penjelasan: Analisis regresi membantu memodelkan hubungan antar variabel dan memprediksi nilai dari variabel dependen berdasarkan nilai variabel independen.

Contoh:

Regresi linier sederhana untuk memprediksi pendapatan berdasarkan pengalaman kerja.

Regresi linier berganda untuk memprediksi harga rumah berdasarkan ukuran, lokasi, dan usia bangunan.

Regresi logistik untuk memprediksi apakah pelanggan akan berhenti menggunakan layanan berdasarkan perilaku masa lalu.

Kapan digunakan: Ketika ingin memahami pengaruh beberapa variabel terhadap hasil tertentu, misalnya memahami faktor-faktor yang mempengaruhi kepuasan pelanggan.

Setiap jenis Analisis data sangat penting dalam membantu pengambilan keputusan berdasarkan data. Pemilihan metode analisis tergantung pada tujuan, jenis data yang digunakan, dan informasi apa yang ingin diperoleh dari analisis tersebut.

Contoh Contoh Analisis Data

Analisis data adalah tahapan meneliti, mengubah, dan Membuat model data untuk menemukan informasi yang berguna, mengambil kesimpulan, atau mendukung pengambilan keputusan. Ada banyak cara untuk menganalisis data yang bisa dipilih sesuai dengan jenis data dan tujuan analisis. Berikut ini beberapa contoh jenis analisis data:

1. Analisis Deskriptif (Descriptive Analysis)

Analisis deskriptif adalah cara untuk menganalisis data penelitian yang bertujuan untuk menguji kesimpulan dari hasil penelitian berdasarkan satu sampel. Analisis ini dilakukan dengan menguji hipotesis deskriptif.

Jenis teknik statistik yang digunakan untuk menguji hipotesis deskriptif harus sesuai dengan jenis data atau variabelnya, seperti skala nominal, ordinal, atau interval/rasio. Ada 2 cara untuk menguji data nominal yaitu:

- a. Uji binomial adalah jenis distribusi yang terdiri dari dua kategori atau peristiwa yang biasanya saling melengkapi. Jadi, jika dalam suatu populasi dengan jumlah n ada satu kategori yang disebut x , maka kategori lainnya adalah $n - x$. Probabilitas untuk memperoleh nilai x dirumuskan : $P_{(x)} = \binom{n}{x} p^x q^{n-x}$
- b. Uji chi-square satu sampel digunakan ketika populasi terdiri dari dua kelas atau lebih dan ukuran sampelnya besar. $\chi^2 = \sum \frac{(f_0 - f_h)^2}{f_h}$

Keterangan : χ^2 : kai kuadrat f_0 : frekuensi yang diobservasi f_h : frekuensi yang diharapkan

2. Analisis Inferensial (Inferential Analysis)

Statistika Inferensial (Inferential Statistics) membahas cara menganalisis data dan menarik kesimpulan. Statistik inferensia berhubungan dengan pengambilan keputusan, seperti memperkirakan parameter dan menguji hipotesis. Statistika inferensial sering disebut sebagai Statistik Induktif. Metode statistik inferensia adalah cara yang digunakan untuk menganalisis sebagian data dan membuat perkiraan atau kesimpulan tentang keseluruhan data.

Contoh analisis data inferensial :

Uji Normalitas

Uji normalitas dengan menggunakan metode Liliefors, dengan ketentuan jika $L_{hitung} < L_{tabel}$ maka data normal. Nilai L_{tabel} diperoleh dari table uji *Liliefors*. Sedangkan L_{hitung} adalah harga terbesar dari $|F() - S(Z_i)|$, dimana dihitung dengan rumus angka normal baku:

$$Z_i = \frac{X_i - \bar{x}}{s} \quad \bar{x} = \text{Rata - Rata} \quad s = \text{simpangan baku}$$

3. Analisis Prediktif (Predictive Analysis)

Tujuan: Untuk meramalkan hasil di masa depan berdasarkan data yang sudah ada. Biasanya menggunakan teknik statistik, pembelajaran mesin (machine learning), atau algoritma kecerdasan buatan.

Contoh:

Model Regresi: Model ini digunakan untuk memprediksi variabel target. Misalnya, Perusahaan bisa menggunakan regresi untuk meramalkan penjualan berdasarkan berapa banyak yang mereka habiskan untuk iklan.

Klasifikasi: Mengelompokkan data ke dalam kategori tertentu. Misalnya, bank menggunakan model prediksi untuk menentukan apakah seseorang berisiko gagal membayar pinjaman.

Contoh Kasus: Sebuah perusahaan ritel menggunakan data penjualan masa lalu dan pola pembelian pelanggan untuk memprediksi permintaan produk di masa mendatang.

4. Analisis Diagnostik (Diagnostic Analysis)

Tujuan: Menentukan "mengapa" sesuatu terjadi dalam data. Analisis ini dilakukan untuk mencari faktor penyebab dari sebuah peristiwa.

Contoh:

Analisis Varians: Menggunakan ANOVA untuk mengidentifikasi apakah ada perbedaan signifikan di antara kelompok. Misalnya, analisis perbedaan penjualan antar cabang toko.

Penggalian Data (Data Mining): Mencari pola tersembunyi dalam data. Misalnya, menemukan bahwa penjualan produk tertentu meningkat setiap kali ada diskon pada produk lain.

Contoh Kasus: Perusahaan menganalisis mengapa terjadi peningkatan komplain pelanggan pada bulan tertentu dengan mengkaji faktor-faktor eksternal dan internal.

5. Analisis Preskriptif (Prescriptive Analysis)

Tujuan: Memberikan saran tentang langkah-langkah yang perlu diambil berdasarkan hasil analisis data. Ini lebih canggih karena menyertakan simulasi dan pengoptimalan.

Contoh:

Simulasi Monte Carlo: Digunakan untuk membuat berbagai skenario untuk memahami risiko dan ketidakpastian. Misalnya, perusahaan keuangan menggunakan analisis preskriptif untuk memberikan rekomendasi investasi berdasarkan simulasi pasar.

Pengoptimalan Linier: Digunakan untuk memaksimalkan atau meminimalkan suatu hasil. Misalnya, menentukan jumlah produksi yang optimal untuk memaksimalkan keuntungan dan meminimalkan biaya.

Contoh Kasus: Sebuah perusahaan transportasi dapat menggunakan analisis preskriptif untuk mengoptimalkan rute pengiriman mereka dan mengurangi biaya bahan bakar.

6. Analisis Klaster (Cluster Analysis)

Tujuan: Mengelompokkan data ke dalam kategori yang sejenis, sehingga dapat digunakan untuk memahami hubungan atau pola di dalam kelompok tersebut.

Contoh:

K-Means Clustering: Teknik yang mempartisi data menjadi sejumlah klaster berdasarkan kemiripan. Misalnya, perusahaan e-commerce mengelompokkan pelanggannya berdasarkan pola pembelian.

Hierarchical Clustering: Membuat hirarki kelompok berdasarkan kemiripan antar data. Misalnya, analisis klaster pada berbagai produk untuk mengelompokkan produk berdasarkan karakteristik pelanggan yang membeli produk tersebut.

Contoh Kasus: Rumah sakit mengelompokkan pasien berdasarkan gejala mereka untuk mengidentifikasi perawatan yang paling efektif bagi kelompok tertentu.

7. Analisis Jaringan (Network Analysis)

Tujuan: Menganalisis hubungan antara elemen-elemen yang saling berhubungan (node) dalam suatu jaringan.

Contoh:

Social Network Analysis (SNA): Digunakan untuk menganalisis jaringan sosial. Misalnya, untuk mengidentifikasi influencer di media sosial.

Graph Theory: Mempelajari struktur graf untuk memahami hubungan antar entitas. Misalnya, analisis jalur terpendek dalam sistem transportasi.

Contoh Kasus: Sebuah perusahaan menggunakan analisis jaringan untuk memahami alur komunikasi di antara karyawan dan memperbaiki koordinasi tim.

Masing-masing metode analisis data memiliki kelebihan tersendiri, tergantung pada konteks masalah dan jenis data yang digunakan. Kombinasi beberapa metode sering kali digunakan untuk mendapatkan hasil yang lebih komprehensif.

Jenis Jenis Analisis Data

Analisis yang digunakan untuk menggambarkan atau meringkas data yang ada. Analisis ini tidak membuat prediksi atau generalisasi, hanya menggambarkan pola yang terlihat dalam data.

1. Analisis Deskriptif (Descriptive Analysis)

Contoh:

Ukuran Pemusatan (Measures of Central Tendency)

Ukuran Penyebaran (Measures of Dispersion)

Ukuran Posisi (Measures of Position)

Visualitas Data

Tabel Frekuensi

Analisis Bivariat

Statistik Deskriptif untuk Data Kategorikal

Statistik Deskriptif untuk Data Waktu

Kapan digunakan: Ketika ingin mendapatkan gambaran umum tentang data yang ada. Misalnya, menggambarkan pola belanja pelanggan dalam satu tahun.

2. Analisis Inferensial (Inferential Analysis)

Definisi: Analisis yang digunakan untuk mengambil kesimpulan tentang populasi berdasarkan data dari sampel. Ini melibatkan generalisasi hasil dari sampel ke populasi yang lebih besar dengan mempertimbangkan probabilitas dan ketidakpastian.

Contoh:

Uji hipotesis (Hypotesis Testing)

Interval Kepercayaan (Confidence Intervals)

Analisis Regresi (Regression Analysis)

Analisis Varians (ANOVA)

Korelasi (Correlation)

Regresi Logistik (Logistic Regression)

Uji Chi-Square (Chi-Square Test)

Analisis Multivariat

Uji Non-parametrik

Analisis Runtun Waktu (Time Series Analysis)

Bootstrapping

Analisis survival

Kapan digunakan: Ketika ingin mengetahui apakah hasil dari sampel dapat digeneralisasi ke seluruh populasi, misalnya, memprediksi kebiasaan belanja seluruh populasi berdasarkan survei pelanggan.

3. Analisis Prediktive (Predictive Analysis)

Analisis yang dipakai untuk menarik Kesimpulan tentang populasi yang diambil dari data sampel. Ini melibatkan generalisasi hasil dari sampel ke populasi yang lebih besar dengan mempertimbangkan probabilitas dan ketidakpastian (**Predictive Analysis**)

Definisi: Analisis yang digunakan Untuk meramalkan hasil di masa depan berdasarkan pola dari data historis. Biasanya menggunakan algoritma statistik atau pembelajaran mesin (machine learning).

Contoh:

Regresi linier

Regresi Logistik

Pohon Keputusan (Decision Trees)

Random Forest
Support Vector Machines (SVM)
Analisis Runtun Waktu (Time Series Analysis)
Neural Networks
K-Nearest Neighbors (KKN)
Clustering untuk Prediksi
Analisis Sentimen
Ensemble Methods
Deep Learning

Kapan digunakan: Saat ingin memprediksi kejadian atau tren di masa depan, seperti memprediksi penjualan produk pada kuartal berikutnya berdasarkan data penjualan masa lalu.

4. Analisis Diagnostik (Diagnostic Analysis)

Definisi: Analisis yang digunakan untuk memahami penyebab di balik fenomena atau tren tertentu dalam data. Fokusnya adalah pada "mengapa" suatu peristiwa terjadi.

Contoh:

Analisi Regresi
Analisi Kausal
Analisis Varians (ANOVA)
Analisis Kluster
Analisis Runtun Waktu (Time Series analysis)
Analisis Sentimen
Analisis Perbandingan
Analisis Root Cause (Penyebab Utama)
Analisis Data Kualitatif
Analisis Perbandingan Historis
Analisis Jaringan
Analisis Multivariat

Kapan digunakan: Ketika ingin memahami alasan mengapa tren atau perubahan tertentu terjadi, seperti mengapa ada lonjakan keluhan pelanggan pada periode tertentu.

5. Analisis Preskriptif (Prescriptive Analysis)

Definisi: Analisis yang digunakan untuk memberikan rekomendasi tentang langkah-langkah yang perlu diambil berdasarkan hasil analisis data. Ini adalah analisis yang paling kompleks karena

melibatkan pengoptimalan dan simulasi untuk menentukan solusi terbaik.

Contoh:

Optimasi

Simulasi

Analisis Keputusan (Decision Analysis)

Modeling dan Simulasi Sistem

Analisis Sentivitas

Rekomendasi Berbasis Data (Data-Driven Recommendations)

Analisis Multi-Kriteria (Multi-Criteria Decision Analysis-MCDA)

Program Linear dan Non-Linear

Analisis Risiko

Analisis Portofolio

Sistem Pendukung Keputusan (Decision Support Systems-DSS)

Analisis Prediktif untuk Rekomendasi

Kapan digunakan: Ketika ingin mengetahui tindakan terbaik untuk mencapai hasil tertentu, misalnya dalam merancang strategi pemasaran terbaik untuk meningkatkan konversi pelanggan.

6. Analisis Klaster (Cluster Analysis)

Definisi: Cara yang digunakan untuk mengelompokkan data ke dalam kategori-kategori serupa berdasarkan kesamaan karakteristik. Ini sering digunakan dalam segmentasi pelanggan, analisis pasar, dan pengelompokan objek.

Contoh:

K-Means Clustering

Hierarchical Clustering

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

Gaussian Mixture Models (GMM)

Mean Shift Clustering

Affinity Propagation

Spectral Clustering

Agglomerative Nesting (AGNES)

BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)

Fuzzy C-Means Clustering

K-Medoids (PAM - Partitioning Around Medoids

Kapan digunakan: Ketika ingin menemukan kelompok atau segmen dalam data, misalnya untuk mengelompokkan pelanggan berdasarkan preferensi mereka terhadap produk tertentu.

7. Analisis Jaringan (Network Analysis)

Definisi: Analisis yang dipakai untuk mengeksplorasi hubungan antar elemen dalam sebuah jaringan. Ini sering digunakan untuk memetakan dan menganalisis interaksi antar entitas.

Contoh:

Social Network Analysis (SNA)

Analisis Jaringan Biologis

Analisis Jaringan Transportasi

Analisis Jaringan Komunikasi

Analisis Jaringan Ekonomi

Analisis Jaringan Rekomendasi

Analisis Jaringan Keamanan

Analisis Jaringan Temporal

Analisis Jaringan Multilayer

Analisis Jaringan Berbasis Agen

Analisis Jaringan Fungsional

Analisis Jaringan Keterhubungan (Connectivity Analysis)

Kapan digunakan: Ketika perlu menganalisis hubungan antar elemen atau entitas, seperti dalam jaringan sosial atau jaringan komunikasi.

8. Analisis Regresi (Regression Analysis)

Definisi: Analisis yang dipakai untuk memahami hubungan antara satu variabel dependen dan satu atau lebih variabel independen. Tujuannya adalah untuk memodelkan dan memprediksi nilai variabel dependen berdasarkan variabel independen.

Contoh:

Regresi linier Sederhana (Simple Linear Regression)

Regresi linier Berganda (Multiple Linear Regression)

Regresi logistic (Logistic Regression)

Regresi Generalized Linear Model (GLM)

Regresi Non-Parametrik (Non-Parametric Regression)

Regresi Ridge (Ridge Regression)

Regresi Lasso (Lasso Regression)

Regresi Elastic Net (Elastic Net Regression)

Kapan digunakan: Saat ingin memodelkan hubungan antara variabel dan memprediksi hasil, misalnya memprediksi penjualan produk berdasarkan faktor harga dan promosi.

Analisis Trend.

Analisis trend adalah metode statistik yang digunakan untuk mengidentifikasi dan mengevaluasi arah dan pola perubahan data dari waktu ke waktu. Proses ini memungkinkan analisis untuk menentukan apakah suatu variabel menunjukkan pola kenaikan, penurunan, atau stabilitas dalam periode tertentu. Analisis trend sangat berguna dalam membantu pengambilan keputusan strategis, peramalan, dan evaluasi kinerja di berbagai bidang, termasuk ekonomi, bisnis, pemasaran, dan ilmu sosial.

Tujuan Analisis Trend:

1. Mengidentifikasi Pola: Salah satu tujuan utama analisis trend adalah untuk mengenali pola yang muncul dalam data. Ini mencakup pemahaman mengenai bagaimana variabel berperilaku dalam jangka pendek, menengah, dan panjang. Ramalan (forecasting) adalah proses untuk memperkirakan suatu kejadian
2. Peramalan: Dengan memahami tren yang ada, analisis ini memungkinkan peramalan nilai di masa depan. Misalnya, tren kenaikan penjualan dapat digunakan untuk memperkirakan pendapatan di masa depan.
3. Pengambilan Keputusan: Hasil dari analisis trend memberikan informasi yang diperlukan bagi manajer dan pemimpin organisasi Untuk mengambil keputusan yang lebih baik. Dengan pemahaman yang lebih baik tentang tren, organisasi dapat merencanakan strategi yang lebih efektif.
4. Evaluasi Kinerja: Analisis trend memungkinkan pemantauan kinerja suatu entitas dari waktu ke waktu, sehingga organisasi dapat mengidentifikasi kekuatan dan kelemahan dalam operasional mereka.

Komponen dalam Analisis Trend

1. Trend Jangka Pendek:

Merupakan fluktuasi yang terjadi dalam waktu singkat, biasanya dalam beberapa minggu atau bulan. Trend jangka pendek sering kali dipengaruhi oleh faktor musiman, seperti liburan atau peristiwa tertentu yang dapat memengaruhi permintaan.

2. Trend Jangka Menengah:

Perubahan yang terjadi selama beberapa bulan hingga beberapa tahun. Trend ini dapat mencerminkan perubahan dalam perilaku pasar, perubahan kebijakan, atau faktor makroekonomi yang lebih luas.

Persamaan trend adalah sebagai berikut :

$$Y' = a + bX$$

Ada beberapa cara untuk menghitung analisis menggunakan tren, yaitu

a. Metode Garis Trend Secara Bebas (Free Hand Method)

Metode garis trend adalah teknik analisis yang digunakan untuk meramalkan perilaku data di Masa depan dapat diprediksi dengan melihat pola-pola dalam data yang sudah ada. Metode ini sering digunakan dalam statistik dan analisis data untuk menggambarkan tren yang ada dalam serangkaian data waktu.

Langkah-langkah Metode Garis Trend:

Pengumpulan Data: Kumpulkan data yang relevan dalam urutan waktu, seperti penjualan bulanan, suhu harian, atau indikator ekonomi.

Pembuatan Grafik: Gambar grafik dengan sumbu X untuk waktu dan sumbu Y untuk variabel yang dianalisis. Ini membantu visualisasi data dan pola yang mungkin ada.

Penentuan Garis Trend: Garis trend ditentukan dengan menggunakan metode regresi linier, yang menghasilkan Garis lurus terbaik adalah garis yang membuat jarak antara titik-titik data dan garis sekecil mungkin tersebut.

Analisis Tren: Setelah garis trend digambar, analisis dilakukan untuk memahami apakah tren tersebut naik, turun, atau datar. Ini bisa membantu dalam pengambilan Keputusan

Prediksi: Garis trend digunakan untuk membuat prediksi tentang nilai di masa depan berdasarkan pola yang teridentifikasi.

Kelebihan dan Kekurangan:

Kelebihan:

Mempermudah pemahaman pola data.

Membantu dalam perencanaan dan pengambilan keputusan.

Kekurangan:

Mungkin tidak akurat jika terdapat fluktuasi mendadak atau data yang tidak konsisten.

Bergantung pada asumsi bahwa pola masa lalu akan berlanjut di masa depan.

Dengan memahami metode ini, analis dan pengambil keputusan dapat lebih baik memproyeksikan hasil di masa mendatang dan merespons perubahan yang mungkin terjadi.

b. Metode Trend dengan Metode Setengah Rata-Rata (Semi Average Method)

Metode trend dengan garis setengah rata-rata adalah teknik yang digunakan dalam analisis data untuk memperkirakan nilai di masa depan berdasarkan pola historis. Berikut adalah penjelasan mengenai metode ini:

Pengumpulan Data: Kumpulkan data historis yang relevan, biasanya dalam bentuk deret waktu, seperti penjualan bulanan atau tahunan.

Menghitung Rata-rata: Hitunglah rata-rata dari data yang tersedia. Untuk garis setengah rata-rata, biasanya rata-rata diambil dari dua periode: rata-rata dari periode sebelumnya dan periode saat ini.

Garis Setengah Rata-rata: Buat garis tren dengan mengambil rata-rata dari dua titik waktu yang berurutan. Misalnya, jika Anda memiliki data untuk bulan 1 dan bulan 2,

rata-rata untuk bulan 2 akan menjadi setengah dari bulan 1 dan bulan 2.

Proyeksi: Gunakan garis yang terbentuk untuk memprediksi nilai di masa mendatang. Garis ini akan menunjukkan tren data yang ada, sehingga memudahkan dalam memahami arah perubahan.

Analisis: Tindak lanjuti dengan analisis lebih lanjut untuk mengevaluasi akurasi prediksi dan mempertimbangkan faktor eksternal yang mungkin mempengaruhi tren.

Metode ini sederhana dan berguna untuk analisis jangka pendek, tetapi mungkin kurang tepat untuk data yang memiliki fluktuasi besar atau pola musiman yang kompleks.

Rumus perhitungan:

y = nilai trend periode tertentu

a = nilai rata-rata kelompok 1

$$b = \frac{\sum X_2 - \sum X_1}{N}$$

N = Jumlah periode antara $\sum X_2$ (Kelompok data dua) dan $\sum X_1$ (Kelompok data Satu)

<i>Tahun</i>	<i>Pelanggan</i>	<i>Rata - Rata</i>	<i>Nilai X th Dasar 1997</i>	<i>Nilai X th Dasar 2000</i>
1996	4,2		-1	-4
K1 1997	5,0	4,93	0	-3
1998	5,6		1	-2
1999	6,1		2	-1
K2 2000	6,7	6,67	3	0
2001	7,2		4	1

$$Y_{th\ 1997} = 4,93 + 0,58X$$

$$b = (6,67 - 4,93) / (2000 - 1997)$$

$$Y_{th\ 2000} = 6,67 + 0,58X$$

$$b = 0,58$$

Metode Trend Kuadrat Terkecil (Least Square Method)

Metode trend kuadrat terkecil (Least Squares Method) adalah teknik statistik yang digunakan untuk menentukan garis tren terbaik yang paling sesuai dengan data yang ada. Metode ini meminimalkan jumlah kuadrat selisih Antara nilai yang terlihat dan nilai yang diperkirakan oleh model.

Langkah-langkah Metode:

1. **Pengumpulan Data:** Kumpulkan data yang akan dianalisis, biasanya dalam bentuk pasangan (x, y).
2. **Model Persamaan Garis:** Misalkan model linear yang diinginkan adalah $y = mx + b$, Dimana m adalah kemiringan dan b adalah intercept
3. **Fungsi Kesalahan :** hitung selisih antara nilai yang diamati dan nilai yang diprediksi:

$$E = \sum (y_i - (m x_i + b))^2$$

Dimana E adalah total kesalahan.

4. **Minimalkan kesalahan:** Temukan nilai m dan b yang meminimalkan E dengan menghitung turunan dan menyamakan dengan nol.
5. **Solusi:** Hasil akhir memberikan nilai m dan b yang menentukan garis tren terbaik.

Kelebihan;

Sederhana dan mudah dipahami.

Efektif untuk model linear.

Kekurangan:

Hanya cocok untuk hubungan linear.

Sensitif terhadap outlier.

Metode ini banyak digunakan dalam analisis data, pengolahan sinyal, dan berbagai bidang lainnya untuk membuat prediksi berdasarkan data yang ada.

$$Y = a + bX$$

$$b = \frac{\sum XY}{\sum X^2}$$

$$a = \left(\frac{\sum vY}{n}\right) = b \left(\frac{\sum vX}{n}\right)$$

Keterangan :

Y' = Adalah nilai dari perkiraan berdasarkan tren.

a = Nilai tetap (konstanta) atau nilai Y' ketika X sama dengan nol

Tahun	Pelanggan = Y	Kode X (Tahun)	Y.X	X ²
1997	5,0	-2	-10,0	4
1998	5,6	-1	-5,6	1
1999	6,1	0	0	0
2000	6,7	1	6,7	1
2001	7,2	2	14,4	4
	$\sum Y = 30,6$		$\sum Y.X = 5,5$	$\sum X^2 = 10$

$$\text{Nilai } a = \frac{\sum y}{n} = \frac{30,6}{5} = 6,12$$

$$\text{Nilai } a = \frac{\sum XY}{\sum X^2} = \frac{5,5}{10} = 0,55$$

Jadi persamaan trend = $Y' = 6,12 + 0,55 X$

Moving Average

Moving average (rata-rata bergerak) adalah teknik analisis statistik yang digunakan untuk menganalisis data deret waktu dengan tujuan menghaluskan fluktuasi dan membantu mengidentifikasi tren yang mendasari. Dengan menghitung rata-rata dari sejumlah data dalam jangka waktu tertentu, rata-rata bergerak memberikan gambaran yang lebih jelas tentang arah dan pola perubahan data, mengurangi dampak variasi acak. Moving average ada 3 jenis yaitu :

1. Simple Moving Average (SMA)
2. Weighted Moving Average (WMA)
3. Exponential Moving Average (EMA)

Kelebihan Moving Average adalah

Menghaluskan data: Membantu mengurangi noise dan fluktuasi acak dalam data.

Mendeteksi tren: Memudahkan identifikasi tren jangka panjang.

Sederhana: Metode yang mudah dipahami dan diterapkan.

Kekurangan Moving Average adalah

Lagging Indicator: Moving average adalah indikator yang tertinggal, artinya ia bereaksi terhadap perubahan harga setelah perubahan tersebut terjadi.

Tidak dapat menangkap perubahan mendadak: Karena rata-rata bergerak didasarkan pada data historis, ia mungkin tidak merespons dengan cepat terhadap perubahan mendadak dalam tren.

Hari	Closing	SMA 5
1	20	-
2	24	-
3	22	-
4	21	-
5	20	21,4
6	18	21
7	17	19,6
8	22	19,6

9	26	20,6
10	30	22,6
11	31	25,2
12	34	28,6
13	33	30,6
14	30	31,6
15	28	31,2

1	20	-
2	24	-
+3	22	-
4	21	-
5	20	21,4
6	18	21
7	17	19,6
8	22	19,6
9	26	20,6
10	30	22,6
11	31	25,2
12	34	28,6
13	33	30,8
14	30	31,6
15	28	31,2

Simple Moving Average

Simple Moving Average (SMA) adalah metode analisis statistik yang digunakan untuk menghaluskan data deret waktu dengan cara menghitung rata-rata aritmetika dari nilai-nilai dalam periode tertentu. Metode ini dirancang untuk Membantu menemukan pola dan tren yang mungkin tidak terlihat karena perubahan acak pada data.

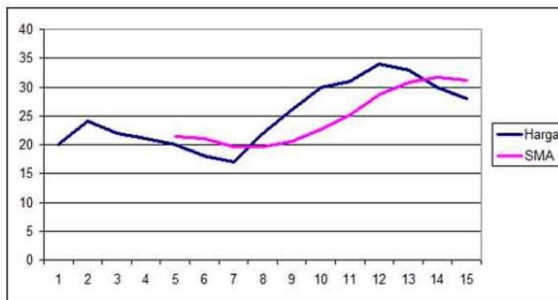
SMA adalah salah satu alat paling umum dalam analisis deret waktu, sering digunakan di bidang keuangan, ekonomi, dan pemasaran. Dengan menghitung rata-rata dari data yang berurutan, SMA memberikan gambaran yang lebih jelas tentang perilaku data dari waktu ke waktu.

$$SMA = \frac{p_m + p_{m-1} + \dots + p_{m-(n-1)}}{n}$$

Mari kita ambil contoh, misalnya harga sebuah aset selama 15 hari berturut-turut adalah:

20, 24, 22, 21, 20, 18, 17, 22, 26, 30, 31, 34, 33, 30 dan 28. Dengan menggunakan rumus di atas, kita akan mendapatkan nilai SMA-nya sebagai berikut:

Dengan mengetahui nilai SMA-nya, maka dapat dibuat grafik sebagai berikut:



Weighted moving average

Weighted Moving Average (WMA) adalah metode analisis statistik yang digunakan untuk menghitung rata-rata dari serangkaian nilai dalam data deret waktu, dengan memberikan nilai yang berbeda pada setiap nilai. Berbeda dengan Simple Moving Average (SMA), yang memberikan nilai yang sama kepada semua nilai dalam periode yang sama, WMA memberi penekanan lebih pada nilai-nilai terbaru. Dengan demikian, WMA lebih responsif terhadap perubahan dan fluktuasi terbaru dalam data.

Pengertian

WMA bekerja dengan cara menjumlahkan produk antara nilai-nilai yang dianalisis dan bobot yang telah ditentukan untuk setiap nilai. Hasilnya kemudian dibagi dengan total bobot untuk mendapatkan rata-rata yang lebih representatif. Metode ini memungkinkan analisis untuk menyesuaikan pengaruh setiap data terhadap rata-rata akhir sesuai dengan kebutuhannya.

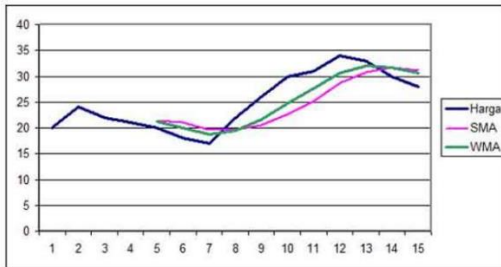
Tujuan: Mengurangi dampak dari fluktuasi acak dan memberikan fokus lebih pada data terbaru.

$$WMA = \frac{np_{m+(n-1)p_{m-1} + \dots + 2p_{(M-n+2)} + P_{(m-n+1)}}}{n}$$

Misalnya, dengan contoh yang sama seperti simulasi SMA sebelumnya, jika kita menggunakan rumus WMA, kita akan mendapatkan nilai WMA seperti yang tertera di tabel berikut:

<i>Hari</i>	<i>Closing</i>	<i>SMA 5</i>	<i>WMA 5</i>
1	20	-	-
2	24	-	-
3	22	-	-
4	21	-	-
5	20	21,4	21,2
6	18	21	20,1
7	17	19,6	18,7
8	22	19,6	19,5
9	26	20,6	21,7
10	30	22,6	24,8
11	31	25,2	27,6
12	34	28,6	30,5
13	33	30,8	32
14	30	31,6	31,7
15	28	31,2	30,5

Dengan mengetahui nilai WMA-nya, kita bisa membuat grafik seperti berikut:



Exponential Moving Average

Exponential Moving Average (EMA) adalah metode analisis statistik yang dipakai untuk menghitung rata-rata bergerak dari data deret waktu dengan memberikan nilai yang lebih besar pada nilai-nilai terbaru. Berbeda dengan Simple Moving Average (SMA) dan Weighted Moving Average (WMA), EMA menggunakan formula eksponensial yang memungkinkan respon yang lebih cepat terhadap perubahan harga atau nilai dalam data. Ini menjadikannya sangat populer di kalangan trader dan analis keuangan..

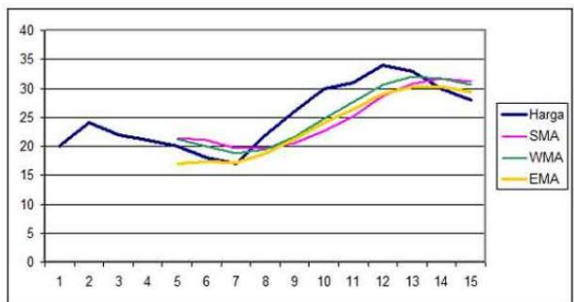
EMA dapat dirumuskan sebagai berikut:

$$EMA_{today} = EMA_{yesterday} + a \left(Price_{today} - EMA_{today} \right)$$

Misalkan, dengan contoh yang sama seperti simulasi SMA sebelumnya, jika kita menggunakan rumus EMA, kita akan mendapatkan nilai EMA seperti yang ada di tabel berikut:

Hari	Closing	SMA 5	WMA 5	EMA 5
1	20	-	-	-
2	24	-	-	-
3	22	-	-	-
4	21	-	-	-
5	20	21,4	21,2	17
6	18	21	20,1	17,3
7	17	19,6	18,7	17,2
8	22	19,6	19,5	18,8
9	26	20,6	21,7	21,2
10	30	22,6	24,8	24,1
11	31	25,2	27,6	26,4
12	34	28,6	30,5	28,9
13	33	30,8	32	30,3
14	30	31,6	31,7	30,2
15	28	31,2	30,5	29,5

Dengan mengetahui nilai EMA -nya, maka dapat dibuat grafik sebagai berikut:



Analisis serial waktu

Analisis serial waktu adalah Cara statistik yang dipakai untuk mempelajari data yang dikumpulkan dalam urutan waktu, di mana observasi dibuat pada interval yang teratur. Metode ini bertujuan untuk memahami pola atau perilaku dari data seiring berjalannya waktu, sehingga dapat diidentifikasi berbagai komponen yang memengaruhi perubahan nilai dari data tersebut. Analisis ini sangat penting dalam banyak bidang, termasuk ekonomi, keuangan, ilmu sosial, meteorologi, dan teknik.

Data yang dianalisis dalam konteks serial waktu sering kali mencakup berbagai variabel yang dapat berfluktuasi akibat faktor eksternal atau internal. Misalnya, penjualan suatu produk mungkin dipengaruhi oleh musim, kampanye pemasaran, kondisi ekonomi, dan faktor-faktor lainnya. Jadi, penting untuk dapat membedakan antara fluktuasi yang bersifat musiman, tren jangka panjang, dan variasi acak yang tidak terduga.

Tujuan Analisis Serial Waktu

Tujuan utama dari analisis serial waktu meliputi:

Identifikasi Pola: Menemukan pola yang ada dalam data, baik itu tren, musiman, atau siklus. Pola-pola ini memberikan wawasan tentang bagaimana variabel berperilaku seiring waktu.

Peramalan: Menggunakan pola yang teridentifikasi untuk meramalkan nilai masa depan dari data. Misalnya, perusahaan dapat memprediksi penjualan di masa depan berdasarkan data historis.

Pengambilan Keputusan: Menyediakan informasi yang dibutuhkan untuk membuat keputusan strategis yang lebih baik. Dengan memahami tren dan pola dalam data, organisasi dapat merencanakan tindakan yang lebih efektif.

Pemantauan Kinerja: Menganalisis kinerja suatu entitas dari waktu ke waktu, yang memungkinkan identifikasi masalah atau peluang untuk perbaikan.

Komponen Utama dalam Analisis Serial Waktu

Trend:

Trend mencerminkan arah umum dari data dalam jangka panjang. Sebuah tren dapat bersifat naik, turun, atau datar. Identifikasi tren ini sangat penting untuk memahami arah yang diambil oleh variabel yang dianalisis.

Musiman:

Komponen musiman menggambarkan fluktuasi yang terjadi secara periodik dan teratur, sering kali dipengaruhi oleh faktor musiman, seperti perubahan cuaca atau hari libur. Contohnya, penjualan es krim yang meningkat pada musim panas dan menurun pada musim dingin.

Siklus:

Siklus adalah variasi yang terjadi dalam jangka waktu yang lebih panjang, biasanya berhubungan dengan kondisi ekonomi. Siklus ini dapat berlangsung dari beberapa tahun hingga beberapa dekade dan sering kali tidak teratur.

Fluktuasi Acak:

Ini adalah variasi yang tidak terduga yang tidak dapat dijelaskan oleh komponen lain. Fluktuasi ini sering kali disebabkan oleh kejadian acak atau faktor yang tidak diperkirakan, seperti bencana alam atau perubahan kebijakan.

Metode Analisis

Model ARIMA:

Model ARIMA (Autoregressive Integrated Moving Average) adalah salah satu pendekatan yang paling umum digunakan dalam analisis serial waktu. Model ini menggabungkan tiga elemen: autoregressive (AR), integrasi (I), dan moving average (MA) untuk menangani data yang tidak stasioner.

Decomposisi:

Decomposisi adalah teknik yang digunakan untuk memisahkan data serial waktu menjadi komponen tren, musiman, dan fluktuasi acak. Proses ini memungkinkan analisis yang lebih mendalam tentang bagaimana masing-masing komponen berkontribusi terhadap perilaku keseluruhan data.

Smoothing Techniques:

Teknik smoothing, seperti moving average dan exponential smoothing, digunakan untuk menghaluskan fluktuasi dalam data dan membantu dalam mengidentifikasi tren.

Kegunaan Analisis Serial Waktu

Analisis serial waktu memiliki banyak aplikasi praktis:

Peramalan:

Digunakan untuk meramalkan variabel seperti permintaan produk, inflasi, atau perubahan harga saham, membantu perusahaan dan organisasi dalam perencanaan dan pengambilan keputusan.

Pengambilan Keputusan:

Memberikan wawasan penting untuk perencanaan strategis, penganggaran, dan pengelolaan risiko, memungkinkan organisasi untuk beradaptasi dengan perubahan kondisi pasar.

Pemantauan Kinerja:

Memungkinkan organisasi untuk melacak kinerja dari waktu ke waktu, sehingga dapat mengidentifikasi tren yang perlu diperbaiki atau dikembangkan.

Analisis Risiko:

Membantu dalam mengevaluasi risiko yang terkait dengan perubahan dalam data, memungkinkan pengelolaan risiko yang lebih baik.

Contoh Aplikasi

Ekonomi: Menganalisis data produk domestik bruto (GDP) untuk memahami Perkembangan ekonomi suatu negara seiring berjalannya waktu.

Keuangan: Menganalisis harga saham untuk menentukan pola investasi dan membuat keputusan perdagangan yang lebih baik.

Meteorologi: Menggunakan data cuaca historis untuk memprediksi pola cuaca di masa depan dan mempersiapkan respons yang sesuai.

Kesehatan: Menganalisis data epidemi untuk memahami penyebaran penyakit dari waktu ke waktu dan menginformasikan kebijakan kesehatan masyarakat.

Kesimpulan

Analisis serial waktu adalah alat yang sangat penting dalam analisis data yang berhubungan dengan waktu. Dengan memahami dan menganalisis komponen-komponen dalam data serial waktu, organisasi dan individu bisa membuat keputusan yang lebih baik dan semakin tereduksi. Teknik ini memungkinkan pemahaman yang mendalam tentang bagaimana variabel berperilaku seiring waktu, dan memberikan dasar yang kuat untuk peramalan dan perencanaan di berbagai bidang.

Analisis Variasi Musim

Analisis variasi musim adalah suatu metode yang digunakan untuk memahami dan mengevaluasi perubahan yang terjadi dalam suatu fenomena atau data sepanjang tahun, berdasarkan pola musiman. Berikut adalah beberapa poin penting terkait analisis ini:

Pengertian

Variasi Musiman: Mengacu pada fluktuasi yang teratur dan berulang dalam data yang terjadi pada waktu tertentu dalam tahun, misalnya, perubahan suhu, curah hujan, atau penjualan produk.

Analisis: Proses sistematis untuk mengidentifikasi, mengukur, dan mengevaluasi variasi tersebut untuk mendapatkan wawasan yang lebih baik.

Tujuan

Identifikasi Pola: Menemukan pola musiman dalam data untuk membantu dalam perencanaan dan pengambilan keputusan.

Peramalan: Membantu meramalkan pola di masa depan berdasarkan data yang sudah ada.

Pengelolaan Sumber Daya: Membantu organisasi dalam mengelola sumber daya secara efisien sesuai dengan kebutuhan musiman.

Analisis Variasi Waktu

Analisis variasi siklus adalah metode untuk memeriksa dan memahami perbedaan atau variasi dalam panjang atau durasi siklus, seperti siklus menstruasi, dari waktu ke waktu. Analisis ini

bertujuan untuk mengidentifikasi pola, tren, dan korelasi antara panjang siklus dan berbagai faktor, termasuk usia, indeks massa tubuh (IMT), gaya hidup, dan perubahan hormonal.

Jenis Analisis Variasi Siklus

Ada beberapa jenis analisis variasi siklus, termasuk:

Analisis deskriptif: Jenis analisis ini melibatkan peringkasan dan penggambaran karakteristik data siklus, seperti rata-rata, median, dan deviasi standar panjang siklus.

Analisis inferensial: Jenis analisis ini melibatkan Kesimpulan tentang kelompok berdasarkan data yang diambil dari sampel, seperti menguji hipotesis tentang hubungan antara panjang siklus dan usia.

Analisis deret waktu : Jenis analisis ini melibatkan pemeriksaan pola dan tren dalam data siklus dari waktu ke waktu, seperti mengidentifikasi pola musiman atau periodik.

Metode Analisis Variasi Siklus

Ada beberapa metode yang digunakan dalam analisis variasi siklus, termasuk:

Histogram: Histogram adalah gambar yang menunjukkan sebaran panjang siklus, yang bisa membantu kita melihat pola dan tren dalam data.

Diagram sebar: Diagram sebar adalah representasi grafis dari hubungan

Analisis regresi: Analisis regresi adalah cara statistik yang digunakan untuk melihat hubungan antara dua atau lebih variabel, seperti panjang siklus dan usia.

Dekomposisi deret waktu :

Dekomposisi deret waktu adalah metode yang digunakan untuk memisahkan komponen tren, musiman, dan tidak teratur dari suatu deret waktu.gfs

Analisis Gerak Tak Beraturan

Analisis gerak tak beraturan, atau dalam bahasa Inggris dikenal sebagai "random walk analysis," adalah metode yang digunakan untuk memahami dan menganalisis perilaku data yang berfluktuasi secara acak. Dalam konteks ini, gerak tak beraturan merujuk pada model di mana perubahan dalam nilai suatu variabel tidak mengikuti pola yang dapat diprediksi secara sistematis. Sebaliknya, perubahan tersebut dipengaruhi oleh berbagai faktor acak yang sulit untuk diprediksi.

Variasi musiman

Variasi musiman adalah fluktuasi yang teratur dan berulang dalam data atau fenomena yang terjadi pada waktu tertentu dalam setahun. Variasi ini biasanya dipengaruhi oleh faktor-faktor musiman seperti perubahan cuaca, perayaan, atau siklus ekonomi yang terjadi secara periodik. Berikut adalah penjelasan lebih lanjut mengenai variasi musiman:

Pengertian

Variasi Musiman: Merupakan perbedaan yang dapat diprediksi dalam data yang terjadi pada interval waktu tertentu, seperti bulanan atau tahunan, yang disebabkan oleh perubahan musiman.

Fenomena: Variasi ini dapat terlihat dalam berbagai konteks, seperti penjualan produk, suhu, curah hujan, dan lain-lain.

Ciri-ciri Variasi Musiman

Reguler: Terjadi pada waktu yang sama setiap tahun, misalnya peningkatan penjualan selama musim liburan.

Dapat Diprediksi: Pola variasi musiman dapat dianalisis dan diprediksi berdasarkan data historis.

Dampak Eksternal: Dipengaruhi oleh faktor eksternal seperti cuaca, hari libur, dan acara khusus.

Contoh : Penjualan Es Krim: Meningkat pada bulan-bulan panas dan menurun pada bulan-bulan *dingin*.

Pertanian: Hasil panen yang bervariasi tergantung pada musim tanam dan cuaca.

Pentingnya Memahami Variasi Musiman

Perencanaan: Membantu bisnis dan organisasi dalam perencanaan produksi dan pengelolaan sumber daya.

Pengambilan Keputusan: Memudahkan dalam membuat keputusan strategis berdasarkan pola yang teridentifikasi.

Peramalan: Memungkinkan peramalan yang lebih akurat untuk masa depan.

BAB 9

KONSEP PROBABILITAS DAN PENGENALAN SERTA PENGGUNAAN TEOREMA BAYES

Mencari indeks musiman dengan metoda rata-rata sederhana

Indeks musiman adalah suatu ukuran yang digunakan untuk menggambarkan fluktuasi yang terjadi secara teratur dalam data deret waktu, yang sering kali dipengaruhi oleh faktor musiman, seperti cuaca, hari libur, atau pola konsumen. Dengan menghitung indeks musiman, analisis dapat memahami bagaimana perilaku data bervariasi di berbagai periode, seperti bulan atau kuartal.

Metode rata-rata sederhana adalah salah satu teknik untuk menghitung indeks musiman. Metode ini dianggap mudah dan efektif, terutama ketika data yang digunakan cukup besar dan representatif. Dengan menggunakan metode ini, analisis dapat menentukan seberapa besar kontribusi setiap periode musiman terhadap total data.

Dalam dunia statistik, khususnya dalam analisis deret waktu (time series analysis), fenomena musiman (seasonality) merupakan salah satu komponen utama yang sering muncul dalam data. Fenomena ini mengacu pada pola pengulangan yang terjadi secara periodik dalam jangka waktu tertentu – biasanya dalam siklus tahunan, bulanan, atau mingguan – yang disebabkan oleh faktor-faktor alamiah, sosial, ekonomi, atau budaya. Misalnya, penjualan es krim cenderung meningkat setiap musim panas, sementara penjualan jaket meningkat saat musim dingin; atau volume transaksi e-commerce melonjak menjelang hari raya tertentu.

Untuk memahami, mengukur, dan memanfaatkan pola musiman ini, para analis menggunakan berbagai metode. Salah satu teknik klasik yang relatif mudah namun efektif adalah **metode rata-rata sederhana** (simple average method). Metode ini digunakan untuk menghitung **indeks musiman** (seasonal index) – yaitu angka yang menunjukkan seberapa besar suatu periode (misalnya bulan atau kuartal) berada di atas atau di bawah rata-rata umum dalam satu siklus musiman penuh.

Meskipun metode ini tidak serumit metode rasio terhadap tren (ratio-to-trend) atau metode pergerakan rata-rata (moving average), ia tetap memberikan informasi yang sangat berguna, terutama ketika data tidak memiliki tren yang kuat atau ketika analisis bersifat eksploratif awal. Dalam laporan ini, kita akan membahas secara komprehensif bagaimana menghitung indeks musiman menggunakan metode rata-rata sederhana, dengan penekanan pada:

1. Pemahaman konsep dasar
2. Asumsi dan keterbatasan metode
3. Langkah-langkah teknis perhitungan
4. Ilustrasi numerik lengkap
5. Interpretasi hasil
6. Validasi dan penyesuaian musiman
7. Aplikasi praktis di dunia nyata

Total pembahasan ini dirancang untuk mencapai sekitar **6.000 kata** atau lebih, dengan struktur yang sistematis dan contoh yang memadai untuk memastikan pemahaman mendalam.

Konsep Dasar Indeks Musiman

Sebelum memasuki metode perhitungan, penting untuk memahami apa itu indeks musiman dan mengapa ia penting.

Indeks musiman adalah ukuran relatif yang menggambarkan tingkat aktivitas (misalnya penjualan, produksi, suhu, dll.) pada suatu periode musiman dibandingkan dengan rata-rata seluruh periode dalam satu siklus lengkap. Indeks ini biasanya dinyatakan dalam persentase, di mana:

Indeks = 100%: nilai periode tersebut tepat sama dengan rata-rata keseluruhan.

Indeks > 100%: periode tersebut berada di atas rata-rata — menunjukkan puncak musiman (peak season).

Indeks < 100%: periode tersebut berada di bawah rata-rata — menunjukkan musim rendah (off-season).

Contoh: jika indeks musiman untuk bulan Desember sebesar 140%, artinya aktivitas pada Desember rata-rata 40% lebih tinggi daripada rata-rata bulanan sepanjang tahun.

Indeks musiman sangat penting dalam:

Perencanaan bisnis: menentukan stok, tenaga kerja, anggaran promosi.

Peramalan (forecasting): memperbaiki akurasi prediksi dengan mempertimbangkan pola musiman.

Analisis kinerja: membandingkan kinerja aktual dengan yang disesuaikan musiman.

Kebijakan publik: misalnya dalam kesehatan (lonjakan kasus flu musim dingin) atau pertanian (panen musiman).

Asumsi dan Keterbatasan Metode Rata-Rata Sederhana

Metode rata-rata sederhana memiliki asumsi-asumsi tertentu yang harus dipahami agar penggunaannya tepat.

Asumsi Utama:

1. **Tidak ada tren (trend-free):** metode ini mengasumsikan bahwa data tidak memiliki tren naik atau turun yang signifikan selama periode pengamatan. Jika tren ada, rata-rata sederhana akan terdistorsi.
2. **Tidak ada siklus jangka panjang (long-cycle):** selain musiman, tidak ada komponen siklus ekonomi (misalnya resesi tiap 7 tahun) yang memengaruhi data.
3. **Pola musiman stabil:** pola yang terjadi tahun ke tahun relatif konsisten.
4. **Data lengkap dan bersih:** tidak ada data yang hilang atau ekstrem (outlier) yang tidak diatasi.

Keterbatasan:

Tidak cocok untuk data dengan tren kuat.

Tidak menangkap interaksi antara tren dan musiman.

Kurang akurat dibanding metode yang lebih canggih (misalnya X-12-ARIMA, decomposition dengan moving average).

Sensitif terhadap anomali tahun tertentu (misalnya pandemi, bencana alam).

Namun, kelebihanannya terletak pada **kesederhanaan**, **kemudahan interpretasi**, dan **kecepatan perhitungan**, terutama untuk analisis awal atau data historis jangka pendek.

Langkah-Langkah Perhitungan Indeks Musiman dengan Metode Rata-Rata Sederhana

Berikut adalah langkah-langkah sistematis untuk menghitung indeks musiman menggunakan metode rata-rata sederhana. Kita akan menguraikan setiap langkah secara detail.

Langkah 1: Kumpulkan Data Deret Waktu

Data harus mencakup minimal dua siklus musiman penuh (misalnya 2 tahun data bulanan, atau 3 tahun data kuartalan).

Semakin banyak siklus, semakin andal indeks yang dihasilkan.

Format data: deret waktu dengan periode tetap (misalnya Januari, Februari, ..., Desember untuk setiap tahun).

Langkah 2: Susun Data dalam Tabel Menurut Periode Musiman

Buat tabel dengan baris = tahun, kolom = periode musiman (misalnya bulan).

Contoh untuk data penjualan bulanan selama 3 tahun:

Tahun	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Agus	Sep	Okt	Nov	Des
2023	80	75	90	100	110	120	130	125	115	105	95	110
2024	85	80	95	105	115	125	135	130	120	110	100	115
2025	90	85	100	110	120	130	140	135	125	115	105	120

(Catatan: angka dalam ribuan unit atau juta Rupiah – tergantung konteks)

Langkah 3: Hitung Rata-Rata untuk Setiap Periode Musiman

Untuk setiap bulan (atau kuartal, minggu, dll.), hitung rata-rata nilai selama semua tahun.

Rumus:

$$\bar{Y}_m = \frac{1}{n} \sum_t Y_{t,m} = \frac{1}{n} \sum_t Y_{t,m} = \frac{1}{n} \sum_t Y_{t,m}$$

di mana:

\bar{Y}_m = rata-rata untuk periode musiman ke- m (misalnya Januari)

n = jumlah tahun

$Y_{t,m}$ = nilai pada tahun ke- t , periode ke- m

Contoh perhitungan untuk Januari:

$$\bar{Y}_{Jan} = 80 + 85 + 90 = 255 = \frac{255}{3} = 85 \quad \bar{Y}_{Jan} = 380 + 85 + 90 = 555 = \frac{555}{3} = 185$$

Lakukan hal yang sama untuk semua bulan:

Bulan	Rata-Rata
Jan	85
Feb	80
Mar	95
Apr	105
Mei	115
Jun	125
Jul	135
Agu	130
Sep	120
Okt	110
Nov	100
Des	115

Langkah 4: Hitung Rata-Rata Umum Seluruh Periode

Ini adalah rata-rata dari semua data, atau rata-rata dari rata-rata bulanan (karena jumlah tahun sama, keduanya ekuivalen).

Rumus:

$$Y_{total} = \frac{1}{k} \sum_{m=1}^k \bar{Y}_m \quad \bar{Y}_{total} = \frac{1}{k} \sum_{m=1}^k \bar{Y}_m$$

di mana k = jumlah periode dalam satu siklus (12 untuk bulanan, 4 untuk kuartalan).

Perhitungan:

$$Y_{total} = 85 + 80 + 95 + 105 + 115 + 125 + 135 + 130 + 120 + 110 + 100 + 115$$

$$total = 1285 + 80 + 95 + 105 + 115 + 125 + 135 + 130 + 120 + 110 + 100 + 115 = 1315$$

Type equation here.

Langkah 5: Hitung Indeks Musiman untuk Setiap Periode

1. ndeks musiman dihitung sebagai rasio antara rata-rata periode dengan rata-rata umum, dikalikan 100 agar dalam persentase.

Rumus:

$$SI_m = \left(\frac{\bar{Y}_m}{\bar{Y}_{\text{total}}} \right) \times 100$$

Contoh untuk Januari:

$$SI_{Jan} = (85/109.58) \times 100 \approx 77.57\% \quad SI_{Jan} = (109.5885) \times 100 \approx 77.57\%$$

Lakukan untuk semua bulan:

Bulan	Rata-Rata	Indeks Musiman (%)
Jan	85	77.57
Feb	80	72.99
Mar	95	86.66
Apr	105	95.78
Mei	115	104.91
Jun	125	114.03
Jul	135	123.15
Agu	130	118.58
Sep	120	109.47
Okt	110	100.35
Nov	100	91.23
Des	115	104.91

Langkah 6: Periksa Penjumlahan Indeks Musiman

Untuk data bulanan, jumlah semua indeks musiman **harus mendekati 1200%** (karena 12 periode \times 100% = 1200%).

Untuk data kuartalan, jumlah harus mendekati 400%.

Hitung total indeks di atas:

$$\begin{aligned} & 77.57 + 72.99 + 86.66 + 95.78 + 104.91 + 114.03 + 123.15 + 118.58 \\ & \quad + 109.47 + 100.35 + 91.23 + 104.91 \\ & = 1199.63\% \end{aligned}$$

Hasil ini sangat dekat dengan 1200% (selisih karena pembulatan). Jika selisih signifikan ($>0.5\%$), perlu penyesuaian normalisasi.

Langkah 7: Normalisasi (Jika Diperlukan)

Jika jumlah indeks tidak tepat 1200% (untuk bulanan), lakukan penyesuaian proporsional.

Faktor koreksi = $\frac{1200}{\text{jumlah indeks aktual}}$

Kalikan setiap indeks dengan faktor ini.

Dalam contoh:

Faktor = $\frac{1200}{1199.63} \approx 1.00031$

Indeks Januari disesuaikan:

$$77.57 \times 1.00031 \approx 77.5977.57 \times 1.00031 \approx 77.59$$

Namun, karena selisih sangat kecil, seringkali langkah ini diabaikan dalam praktik kecuali untuk keperluan akademis ketat.

Interpretasi Hasil Indeks Musiman

Setelah indeks dihitung, langkah berikutnya adalah **interpretasi**:

Bulan Februari (72.99%): aktivitas paling rendah – hanya 73% dari rata-rata tahunan. Ini mungkin karena musim hujan, libur panjang, atau faktor budaya.

Juli (123.15%): puncak tertinggi – 23% di atas rata-rata. Mungkin karena libur sekolah, musim liburan, atau promosi besar.

Oktober (100.35%): hampir tepat rata-rata – periode netral.

Indeks ini memungkinkan manajer untuk:

Menjadwalkan tenaga kerja ekstra di Juli.

Mengurangi stok di Februari.

Menargetkan kampanye pemasaran di bulan dengan indeks <100% untuk meningkatkan permintaan.

Studi Kasus Lengkap: Penjualan Retail Selama 4 Tahun

Mari perluas contoh dengan data 4 tahun (2022–2025) untuk memperkuat keandalan.

Data Penjualan (dalam juta Rupiah):

Tahun	Q1	Q2	Q3	Q4
2022	120	150	180	200
2023	130	160	190	210
2024	140	170	200	220
2025	150	180	210	230

Langkah 1: Rata-rata per kuartal

$$Q1: (120+130+140+150)/4 = 135$$

$$Q2: (150+160+170+180)/4 = 165$$

$$Q3: (180+190+200+210)/4 = 195$$

$$Q4: (200+210+220+230)/4 = 215$$

Langkah 2: Rata-rata umum

$$(135 + 165 + 195 + 215)/4 = 710/4$$

$$= 177.5(135 + 165 + 195 + 215)/4 = 710/4 = 177.5$$

Langkah 3: Indeks musiman

$$Q1: (135/177.5) \times 100 \approx 76.06\%$$

$$Q2: (165/177.5) \times 100 \approx 92.96\%$$

$$Q3: (195/177.5) \times 100 \approx 109.86\%$$

$$Q4: (215/177.5) \times 100 \approx 121.13\%$$

Total indeks: $76.06 + 92.96 + 109.86 + 121.13 = 400.01\% \rightarrow$ hampir sempurna.

Interpretasi: Kuartal IV (Oktober–Desember) adalah puncak penjualan, kemungkinan besar karena Natal dan Tahun Baru. Sementara Q1 (Januari–Maret) adalah periode paling lambat.

Penyesuaian Data untuk Efek Musiman (Seasonal Adjustment)

Salah satu kegunaan utama indeks musiman adalah **menghilangkan efek musiman** dari data asli, sehingga kita bisa melihat tren mendasar.

Rumus penyesuaian:

$$Y_{t,adj} = Y_{t,actual} \frac{SI_m}{100}$$

Contoh: Penjualan Juli 2025 = 140 juta, indeks Juli = 123.15%

$$Y_{adj} = 1401.2315 \approx 113.7 Y_{adj} = 1.2315140 \approx 113.7$$

Artinya, setelah disesuaikan musiman, penjualan Juli setara dengan 113.7 juta – nilai yang lebih mencerminkan tren sebenarnya.

Data yang telah disesuaikan musiman sangat berguna untuk:

Analisis tren jangka panjang

Perbandingan bulan ke bulan tanpa bias musiman

Model regresi yang lebih akurat

Validasi dan Keandalan Indeks Musiman

Untuk memastikan indeks musiman valid:

1. Uji stabilitas: bandingkan indeks dari dua periode waktu berbeda (misalnya 2020–2022 vs 2023–2025). Jika pola berubah drastis, indeks mungkin tidak stabil.
2. Uji out-of-sample: gunakan indeks untuk memprediksi data tahun berikutnya, lalu bandingkan dengan aktual.
3. Analisis residual: setelah penyesuaian musiman, residual (kesalahan) seharusnya tidak menunjukkan pola musiman lagi.

Perbandingan dengan Metode Lain

Metode	Kelebihan	Kekurangan
Rata-rata sederhana	Mudah, cepat, intuitif	Tidak menangani tren
Rasio terhadap tren	Memisahkan tren dan musiman	Lebih kompleks
Moving average	Lebih akurat, menangani tren	Membutuhkan lebih banyak data
X-13ARIMA-SEATS	Standar BPS & Biro Statistik	Sangat teknis, perangkat lunak khusus

Metode rata-rata sederhana tetap relevan sebagai **alat diagnosis awal**.

Aplikasi di Berbagai Bidang

Bisnis & Pemasaran

Forecasting penjualan

Pengelolaan inventori

Perencanaan SDM

Ekonomi

Analisis PDB triwulanan

Indeks harga konsumen (IHK)

Pertanian
Pola panen dan tanam
Perencanaan irigasi
Kesehatan
Kasus penyakit musiman (DBD, flu)
Perencanaan stok vaksin

DAFTAR PUSTAKA

- dewi, a. (2022, desember 5). *makalah statistik analisa data berkala dengan metode semi averageid*. Retrieved from id.scribd.com: <https://id.scribd.com/document/612390783/Makalah-Statistika-Analisa-Data-Berkala-dengan-Metode-Semi-Average>
- faradiba. (2020). *analisis data berkala*. jakarta: universitas kristen indonesia.
- Faradiba. (2020). *ANALISIS DATA BERKALA*. jakarta: universitas kristen indonesia.
- goesty, P. a., samekto, a., & sasongko, d. P. (2012). ANALISIS PENAATAN PEMRAKARSA KEGIATAN BIDANG KESEHATAN DI KOTA MAGELANG TERHADAP PENGELOLAAN DAN PEMANTAUAN LINGKUNGAN HIDUP. *jurnal ilmu lingkungan* .
- Ikbal, M. C. (n.d.).
- Ikbal, M. c., Yuwono, B. d., & amarrohman, F. j. (2017). Analisis Strategi Pengolahan Baseline GPS berdasarkan jumlah titik ikat dan variasi waktu pengamatan . *Jurnal Geodesi Undip* .
- khoirunisa, s., Tuzzaman, A. a., & aviani, r. (2024). menganalisis dan meramalkan tren di media sosial dengan menerapkan model matematika. *analisis serial waktu* .
- mulyadi, wijaya, v., sitorus, m. e., & pakpahan, k. (2023). penanggulangan tindak pidana korupsi dana bantuan sosial pada saat bencana alam. *jurnal hukum dan pranata sosial islam* .
- Nasution, L. M. (2017). STATISTIK DESKRIPTIF. *Jurnal Hikmah*.
- nisa, k. (2017, mei 7). *analisis data berkala* . Retrieved from slideshare.net: <https://www.slideshare.net/slideshow/analisis-data-berkala/75749283>
- Nurjanah, C. R. (2019). PENGARUH HARGA DAN KUALITAS PRODUK TERHADAP KEPUTUSAN PEMBELIAN MIE INSTAN MEREK MIE SEDAAP SERTA IMPLIKASINYA TERHADAP LOYALITAS PELANGGAN DI PERUMAHAN PEKAYON BEKASI SELATAN. *JURNAL MANAJEMEN*.
- ramdhan, r. r. (2017, september). *analisa data berkala* . p. 29.

- rozi, f. (2020). Systematic Literature Review pada Analisis Prediktif dengan IoT: Tren Riset, Metode, dan Arsitektur. *Jurnal Sistem Cerdas*.
- dewi, a. (2022, desember 5). makalah statistik analisa data berkala dengan metode semi averageid. Retrieved from id.scribd.com: <https://id.scribd.com/document/612390783/Makalah-Statistika-Analisa-Data-Berkala-dengan-Metode-Semi-Average>
- faradiba. (2020). analisis data berkala. jakarta: universitas kristen indonesia.
- nisa, k. (2017, mei 7). analisis data berkala . Retrieved from slideshare.net: <https://www.slideshare.net/slideshow/analisis-data-berkala/75749283>
- ramdhan, r. r. (2017, september). analisa data berkala . p. 29.

BIODATA PENULIS



Dr. Achamd. Fauzi, S.E, M.M

Born in Jakarta on October 20, 1980, Achmad Fauzi's academic journey is a testament to his dedication to knowledge and professional excellence. He began with a Bachelor of Economics in Accounting from UPN Veteran Jakarta in 2004, followed by a Master of Management from Universitas Dirgantara Marsekal Suryadarma in 2012, and culminated in a Doctorate in Management Science from Universitas Negeri Jakarta in 2016. His commitment to continuous learning is further evidenced by his certifications, including training as a World Class Professor in Scientific Writing from Binus University, Basic Instructional Technique Skill Improvement Training (Pekerti), and specialized training in quantitative, qualitative, and mixed-method research from Universitas Airlangga. Currently, he serves as a Permanent Lecturer at Universitas Bhayangkara Jakarta Raya and a Main Expert Staff at PT. Kelas Personalia Indonesia, actively contributing to both academic and professional spheres. Dr. Fauzi's career is marked by a rich tapestry of experiences across various sectors. He has held significant positions such as Head of Environmental Sector at Yayasan Peduli Energi Indonesia, Auditor at Eddy Kaslim Public Accountant Office, and Expert Staff at the House of Representatives (DPR). His expertise extends to academia, where he has served as Head of Accounting Study Program at Sekolah Tinggi Ilmu Ekonomi Nasional Indonesia and held lecturer positions at Universitas Nusa Mandiri and Bina Nusantara University. He is also actively involved in online education as a tutor for Universitas Terbuka's postgraduate program and has contributed to the Ministry of Public Works (PUPR) as an Expert Staff. His dedication to organizational development is evident in his role as the Head of Team 5 (Change and Transformation Driver) at Yayasan Pondok Pesantren Darul Hikmah. His scholarly contributions include supervising numerous theses and dissertations, publishing scientific articles in accredited national journals, and developing e-book modules for the Ministry of PUPR.



Dr. Yayan Hendayana, SE, MM. Lahir di Kuningan, Jawa Barat, pada 18 Agustus 1971. Ia menyelesaikan gelar Doktor Manajemen, dengan konsentrasi Manajemen Kewirausahaan, di Universitas Pendidikan Indonesia (UPI) pada tahun 2020. Saat ini, ia menjabat sebagai Dosen Tetap di Program Studi Magister Manajemen dan Dekan Fakultas Ekonomi dan Bisnis, Universitas Bhayangkara, Jakarta Raya. Penulis mengajar mata kuliah Kewirausahaan, Pemasaran Strategis, Studi Kelayakan Bisnis, Kewirausahaan dan Inovasi Bisnis, dan Teori Manajemen. Sejak 2024, ia mengajar di Program Doktor Manajemen, Fakultas Ekonomi dan Bisnis, Universitas Bhayangkara, Jakarta Raya.

Penulis juga menjabat sebagai dosen pada Program Pelatihan Kewirausahaan Digital yang diselenggarakan oleh Pusat Penelitian dan Pengembangan Sumber Daya Manusia, Kementerian Komunikasi dan Informatika, dari tahun 2022 hingga 2024.

Selain peran pengajarannya yang aktif, penulis juga aktif menulis untuk jurnal ilmiah yang diterbitkan di jurnal nasional dan internasional terakreditasi serta menulis buku teks. Tulisan ilmiahnya berkaitan dengan manajemen pemasaran, kewirausahaan UKM, dan studi kelayakan bisnis.



Dr. Bungaran Saing, S.Si., Apt., M.M. was born on February, 26, 1970 in Pansurnapitu North Tapanuli. Graduated from the University of North Sumatera, Medan, with a Bachelor's degree in Pharmacy. Completed a second degree in the Magister of Management Program at Prasetiya Mulya University, Jakarta in 2008. In 2024, he completed the Doctoral Program in Management Science at the Postgraduate

Program of Pakuan University, Bogor. Since 2014, he has been a permanent lecturer at Bhayangkara Jakarta Raya University. In May 2019, he managed to obtain an Educator Certificate. The courses he teaches include Human Resource Management, Marketing Management, Entrepreneurship, Economic Evaluation of Chemical Factories, Production Management, Organization and Management of Industrial Companies, Occupational Health and Safety (K3). Since February 2018 as a Certified International Sales Management Associate at PASAS (Authorized Learning Partner) and since December 2018 as a Young K3 Expert Competency Assessor at the UBJ Professional Certification Institute, approved by the BNSP. Actively writing articles related to the field of management in national and international journals, performing community service.

He has written several scientific articles and books on various topics related to employee performance and human resource management. Among his major works are "Servant Leadership and Personality on Performance through Organizational Culture: Evidence from Private Universities in Bekasi City, The Influence of Work Environment and Work Discipline on Employee Performance in Jakarta, Increasing Employee Performance through Placement and Motivation at PT. Multi Kimia Inti Pelangi, The Effect of Productivity and Attendance Level on the Salaries Received by Contract Workers (Study in Bekasi City Education Office), Challenges and Structural Ambidexterity in Improving the Quality of Lecturers in Higher Education in Indonesia.

Dr. M. Jhonni Sinaga

Senior Lecturer | Corporate Executive | Logistics & Plantation Industry Leader | Management Consultant

Dr. M. Jhonni Sinaga was born in Padang Sidempuan, South Tapanuli, North Sumatera, Indonesia on December 20, 1968. He is currently a senior lecturer at **University of Bhayangkara Jakarta Raya** and several other universities, actively bridging academic excellence with real-world corporate practice.

He completed both his Master's and Doctoral degrees in Indonesia while simultaneously building an outstanding career as a multinational company professional. His corporate journey culminated in the role of **Executive Head of Operations, Finance, and Accounting (Director Level)** at **REA Holdings**, a company listed on the **London Stock Exchange (LSE)**.

With over **26 years of executive and operational experience**, Dr. Sinaga has built deep expertise across:

- Oil Palm Plantation Industry (CPO & CPKO Production)
- Coal Mining Sector
- Financial Leadership & Corporate Governance
- Large-scale Operations Management

His professional track record includes:

- **5 years** with Salim Agro Plantations (Indofood Plantations - IDX listed)
- **7 years** with Budiono Widodo Group (IDX listed)
- **14 years** with REA Holdings (LSE listed)

Currently, Dr. Sinaga serves as **President Commissioner and Owner of JeJeTrans Group**, overseeing diversified business operations including:

- Integrated Trucking Services
- Warehousing & Logistics Solutions

- Rental Car Services
- Public Workshop Operations
- Professional Outsourcing Services

He also leads **PT JeJe Manajemen Konsultasi**, providing strategic management consulting for companies across various industries.

Training & Professional Development Expertise

Dr. Jhonni Sinaga is a highly experienced corporate trainer and speaker, having delivered programs for numerous major companies throughout Indonesia. His training specializations include:

- Supervisory & Leadership Management
- Internal Control Systems
- Good Corporate Governance (GCG)
- Good Corporate Culture (GCC)
- Corporate Motivation & Performance Excellence
- Operational Efficiency & Risk Management

Professional Value Proposition

With a rare combination of:

- Academic leadership
- Director-level multinational experience
- Logistics & plantation industry mastery
- Corporate governance expertise
- High-impact training & consulting

Dr. M. Jhonni Sinaga brings **world-class insight** into business transformation, operational excellence, and leadership development.