



UNIVERSITAS BINA SARANA INFORMATIKA

Gedung Rektorat Jl. Kramat Raya No. 98, Senen. Jakarta Pusat 10450
Telp. (021) 23231170 Fax (021) 21236158 e-mail : rektorat@bsi.ac.id

SURAT TUGAS

Nomor : 632/3.01/UBSI/III/2025

Yang bertanda tangan di bawah ini, Rektor Universitas Bina Sarana Informatika, menugaskan kepada:

Nama : Mushliha, S.Pd., M.Si
NIP : 202203035

Sebagai Penulis Buku dengan judul “Statistika Terapan untuk Ekonomi dan Bisnis”, dengan masa penugasan:

Masa Penugasan : 3 Maret 2026 s/d 31 Agustus 2026

Demikian penugasan ini agar dapat dijalankan sebagaimana mestinya. Atas perhatian dan kerjasamanya kami mengucapkan terima kasih.



Jakarta, 2 Maret 2026
Rektor,

Prof. Dr. Ir. Mochamad Wahyudi, M.Kom, MM, M.Pd, IPU, ASEAN Eng.

Tembusan :
1. Divisi SDM
2. Wakil Rektor I & II

UNIVERSITAS

PSDKU

■ BOGOR ■ KARAWANG ■ PURWOKERTO ■ TASIKMALAYA ■ SURAKARTA
■ PONTIANAK ■ TEGAL ■ SUKABUMI ■ YOGYAKARTA





Statistika Terapan untuk Ekonomi dan Bisnis

Isti Kamila, S.Pd., M.Si., Nurul Maqfirah Rauf, S.Pd., M.Mat.,
Mushliha, S.Pd., M.Si., Indrawan, S.Pd., M.Si., Dulhadi, ST.,
Wiwit Pura Nurmayanti, S.Si., M.Si., Minharwan, S.E., S.H.,
Indah Gumala Andirasdini, M.Si., dan Rani, S.Pd., M.M.

Statistika Terapan untuk Ekonomi dan Bisnis

Isti Kamila, S.Pd., M.Si.

Nurul Maqfirah Rauf, S.Pd., M.Mat.

Mushliha, S.Pd., M.Si.

Indrawan, S.Pd., M.Si.

Dulhadi, ST.

Wiwit Pura Nurmayanti, S.Si., M.Si.

Minharwan, S.E., S.H.

Indah Gumala Andirasdini, M.Si.

Rani, S.Pd., M.M.

PT BUKULOKA LITERASI BANGSA

Anggota IKAPI: No. 645/DKI/2024



Statistika Terapan untuk Ekonomi dan Bisnis

Penulis : Isti Kamila, S.Pd., M.Si., Nurul Maqfirah Rauf, S.Pd., M.Mat., Mushliha, S.Pd., M.Si., Indrawan, S.Pd., M.Si., Dulhadi, ST., Wiwit Pura Nurmayanti, S.Si., M.Si., Minharwan, S.E., S.H., Indah Gumala Andirasdini, M.Si., dan Rani, S.Pd., M.M.

ISBN : 978-634-277-180-8 (PDF)

Penyunting Naskah : Putri Tiah Hadi Kusuma, S.Ikom.

Tata Letak : Putri Tiah Hadi Kusuma, S.Ikom.

Desain Sampul : Novikean Keysah Sanisri

Penerbit

Penerbit PT Bukuloka Literasi Bangsa

Distributor: PT Yapindo

Kompleks Business Park Kebon Jeruk Blok I No. 21, Jl. Meruya Ilir Raya No. 88, Desa/Kelurahan Meruya Utara, Kecamatan Kembangan, Kota Adm. Jakarta Barat, Provinsi DKI Jakarta, Kode Pos: 11620

Email: penerbit.blb@gmail.com

Whatsapp: 0878-3483-2315

Website: bukuloka.com

© Hak cipta dilindungi oleh undang-undang

Berlaku selama 50 (lima puluh) tahun sejak ciptaan tersebut pertama kali dilakukan pengumuman.

Dilarang mengutip atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari penerbit. Ketentuan Pidana Sanksi Pelanggaran Pasal 2 UU Nomor 19 Tahun 2002 Tentang Hak Cipta.

Barang siapa dengan sengaja dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam Pasal 2 ayat (1) atau Pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling singkat 1 (satu) bulan dan/atau denda paling sedikit Rp1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (Tujuh) tahun dan/atau denda paling banyak Rp5.000.000.000,00 (lima miliar rupiah).

Barang siapa dengan sengaja menyerahkan, menyiarkan, memamerkan, mengedarkan atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran Hak Cipta atau Hak Terkait sebagaimana dimaksud pada ayat (1) dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya, sehingga buku ajar berjudul **Statistika Terapan untuk Ekonomi dan Bisnis** dapat disusun dan dihadirkan kepada pembaca. Buku ini dirancang dengan bahasa yang lugas dan alur pembahasan yang mudah diikuti, agar konsep statistik dapat dipahami tanpa kesan rumit.

Materi dalam buku ini disusun untuk membantu pembaca memahami dan menggunakan statistik dalam berbagai aktivitas ekonomi dan bisnis sehari-hari. Contoh dan pembahasan disajikan secara praktis, relevan dengan kondisi nyata, serta dapat diterapkan oleh pembaca dari berbagai latar belakang.

Buku ini dijual untuk masyarakat umum dan ditujukan bagi siapa pun yang ingin meningkatkan pemahaman statistik sebagai alat bantu pengambilan keputusan. Harapannya, buku ini dapat menjadi bacaan yang bermanfaat, menambah wawasan, dan memberi nilai guna bagi pembaca secara luas.

Jakarta, Januari 2026

Tim Penyusun

DAFTAR ISI

KATA PENGANTAR.....	iii
DAFTAR ISI.....	iv
Bab 1: Konsep Dan Ruang Lingkup Statistika	1
1.1 Definisi Statistika	1
1.2 Proses Statistika.....	3
1.3 Jenis Data dan Pengukuran.....	4
1.4 Populasi dan Sampel.....	7
1.5 Teknik Sampling	8
1.6 Cakupan Statistika	10
1.7 Latihan Soal.....	13
Bab 2: Pengumpulan dan Penyajian Data.....	14
2.1 Konsep dan Sumber Data	14
2.2 Metode Pengumpulan Data	16
2.3 Jenis-jenis Data.....	18
2.4 Penyajian Data dalam Tabel dan Grafik.....	20
2.5 Distribusi Frekuensi.....	22
2.6 Latihan Soal.....	24
Bab 3: Ukuran Pemusatan dan Penyebaran Data	25
3.1 Pengantar	25
3.2 Ukuran Pemusatan.....	28
3.3 Ukuran Letak.....	34
3.4 Ukuran Penyebaran	37
3.5 Contoh Soal	40
3.6 Interpretasi dan Aplikasi.....	46
3.7 Latihan Soal.....	49
Bab 4: Analisis Data Kuantitatif dan Kualitatif	50

4.1 Konsep Analisis Data	50
4.2 Analisis Data Kuantitatif	52
4.3 Analisis Data Kualitatif	54
4.4 Perbandingan Analisis Kuantitatif dan Kualitatif	55
4.5 Latihan Soal	57
Bab 5: Probabilitas dan Distribusi	58
5.1 Konsep Dasar Probabilitas	58
5.2 Aturan dan Hukum Probabilitas	61
5.3 Variabel Acak dan Distribusi Probabilitas	62
5.4 Distribusi Probabilitas Diskrit	66
5.5 Distribusi Probabilitas Kontinyu	70
5.6 Latihan Soal	73
Bab 6. Statistika Inferensial	75
6.1 Konsep Dasar Statistika Inferensial	75
6.2 Estimasi Parameter	78
6.3 Pengujian Hipotesis	83
6.4 Uji Statistik Parametrik	86
6.5 Uji Statistik Nonparametrik	91
6.6 Latihan Soal	100
Bab 7: Analisis Korelasi dan Regresi	101
7.1 Pengantar Korelasi dan Regresi	101
7.2 Analisis Korelasi	104
7.3 Analisis Regresi Linier	105
7.4 Penilaian dan Aplikasi Model	108
7.5 Latihan Soal	111
Bab 8: Analisis Deret Waktu (Time Series)	112
8.1 Konsep Dasar Deret Waktu	112
8.2 Komponen Deret Waktu (Trend, Musiman, Siklus, dan Irregular)	114

8.3 Model Peramalan (Autoregressive, Moving Average, Exponential Smoothing, ARIMA).....	119
8.4 Penilaian Model Peramalan.....	125
8.5 Latihan Soal.....	129
Bab 9: Analisis Data dengan Software Statistik.....	130
9.1 Pendahuluan	130
9.2 Jenis-Jenis Software Statistik	132
9.3 Prosedur Umum Analisis Data	134
9.4 Contoh Kasus Penggunaan Software Statistik.....	135
9.5 Tantangan dan Solusi dalam Analisis Data Statistik	137
9.6 Latihan Soal.....	139
Bab 10: Statistika dalam Pengambilan Keputusan Bisnis	140
10.1 Pengertian dan Cakupan Statistika	140
10.2 Pengumpulan dan Penyajian Data	143
10.3 Analisis Deskriptif.....	146
10.4 Analisis Inferensial.....	151
10.5 Aplikasi Statistika dalam Pengambilan Keputusan	153
10.6 Latihan Soal.....	156
Profil Penulis	157
Daftar Pustaka.....	167

Bab 1: Konsep Dan Ruang Lingkup Statistika

1.1 Definisi Statistika

Statistika merupakan disiplin ilmu yang mempelajari prinsip serta prosedur dalam mengumpulkan, mengolah, menganalisis, menyajikan, dan menafsirkan data secara sistematis. Melalui proses tersebut, statistika memungkinkan data yang semula bersifat mentah diubah menjadi informasi yang bermakna, sehingga dapat digunakan untuk memahami fenomena serta mendukung pengambilan keputusan yang objektif dan berbasis bukti (Moore et al., 2015). Dalam konteks ilmiah maupun praktik profesional, peran statistika menjadi semakin penting karena menyediakan landasan metodologis untuk menilai pola, kecenderungan, dan ketidakpastian dalam berbagai jenis data.

Secara umum, statistika memiliki dua fungsi utama. Fungsi pertama adalah statistika deskriptif, yaitu rangkaian metode yang digunakan untuk merangkum dan menggambarkan karakteristik data melalui ukuran pemusatan, ukuran penyebaran, tabel, grafik, dan bentuk penyajian lainnya. Fungsi ini membantu peneliti memperoleh gambaran awal mengenai struktur dan sifat dasar data, termasuk pola maupun penyimpangannya. Fungsi kedua adalah statistika inferensial, yang mencakup teknik untuk menarik kesimpulan

mengenai populasi berdasarkan informasi dari sampel, seperti estimasi parameter dan pengujian hipotesis (Triola, 2018; Montgomery & Runger, 2018). Pendekatan ini memungkinkan generalisasi ilmiah tanpa harus mengobservasi seluruh populasi.

Selain menyediakan perangkat analitis, statistika juga membentuk pola pikir ilmiah yang kritis dan terukur. Melalui statistika, peneliti diajarkan untuk mengevaluasi kualitas data, mempertimbangkan variabilitas, serta mengidentifikasi tingkat ketidakpastian dalam setiap kesimpulan yang diambil. Dengan demikian, statistika tidak hanya berperan sebagai seperangkat alat hitung, tetapi juga sebagai kerangka konseptual yang membantu memahami fenomena secara lebih komprehensif dan rasional (Field, 2018).

Seiring perkembangan teknologi dan meningkatnya volume data pada berbagai sektor, statistika menjadi instrumen yang esensial dalam bidang-bidang seperti ekonomi, kesehatan, pendidikan, teknik, biologi, dan kebijakan publik. Kemampuannya dalam mengolah informasi secara sistematis menjadikan statistika dasar penting dalam analisis data modern, termasuk dalam konteks big data dan kecerdasan buatan.

Dengan demikian, statistika dapat dipandang sebagai ilmu yang menyediakan fondasi bagi pemahaman data dan pengambilan keputusan yang dapat dipertanggungjawabkan secara ilmiah, sekaligus tetap relevan dalam berbagai perkembangan penelitian dan aplikasi kontemporer.

1.2 Proses Statistika

Proses statistika merujuk pada rangkaian langkah sistematis yang digunakan untuk mengubah data mentah menjadi informasi bermakna yang dapat digunakan dalam penarikan kesimpulan dan pengambilan keputusan ilmiah. Setiap tahapan dalam proses ini saling berkaitan dan menentukan kualitas hasil akhir analisis. Dengan memahami proses statistika, peneliti dapat merancang penelitian secara lebih terstruktur dan memastikan bahwa data yang diperoleh relevan serta dapat diinterpretasikan dengan tepat (Montgomery & Runger, 2018).

Secara umum, proses statistika terdiri atas empat tahap utama:

1. Pengumpulan Data

Tahap awal ini mencakup perencanaan dan pelaksanaan kegiatan untuk memperoleh data secara sistematis melalui survei, eksperimen, observasi, atau data sekunder. Kualitas data sangat menentukan validitas analisis selanjutnya.

2. Pengolahan Data

Data yang diperoleh disusun dan dibersihkan dengan melakukan pemeriksaan kelengkapan, konsistensi, dan penanganan data hilang. Pengolahan data bertujuan mempersiapkan data agar dapat dianalisis dengan tepat.

3. Analisis Data

Analisis dilakukan dengan pendekatan deskriptif atau inferensial. Analisis deskriptif menggambarkan karakteristik data, sedangkan analisis inferensial membuat generalisasi atau prediksi mengenai populasi (Field, 2018).

4. Interpretasi dan Pelaporan

Hasil analisis harus disampaikan dalam bahasa yang jelas, naratif maupun visual. Tahap ini memastikan temuan dapat dipahami oleh berbagai pemangku kepentingan dan digunakan dalam pengambilan keputusan berbasis bukti.

1.3 Jenis Data dan Pengukuran

Dalam statistika, pemahaman mengenai jenis data dan skala pengukuran merupakan langkah penting sebelum melakukan analisis lebih lanjut. Setiap data memiliki karakteristik tertentu yang menentukan bagaimana data tersebut dapat diolah, dianalisis, dan ditafsirkan. Kesalahan dalam mengidentifikasi jenis data atau skala pengukurannya dapat berakibat pada penggunaan teknik analisis yang tidak tepat dan menghasilkan kesimpulan yang bias.

1.3.1 Jenis Data

Data merupakan elemen fundamental dalam penelitian, yang digunakan untuk menggambarkan fenomena, menguji hipotesis, dan membuat keputusan berbasis bukti. Secara umum, data dapat dikategorikan menjadi dua jenis utama, yaitu data kuantitatif dan data kualitatif. Data kuantitatif bersifat numerik dan dapat diukur

serta dianalisis secara matematis, sehingga memudahkan perhitungan statistik dan pengujian hipotesis. Data kuantitatif selanjutnya dapat dibagi menjadi dua bentuk: data diskrit, yang hanya memiliki nilai tertentu dan biasanya berupa bilangan bulat; serta data kontinu, yang dapat mengambil nilai dalam rentang tertentu dan sering digunakan dalam pengukuran yang membutuhkan presisi tinggi, misalnya tinggi badan atau suhu.

Sebaliknya, data kualitatif bersifat kategorik dan biasanya menggambarkan atribut, karakteristik, atau kualitas dari objek yang diamati, seperti jenis kelamin, status pekerjaan, atau preferensi konsumen. Analisis data kualitatif memerlukan pendekatan berbeda, misalnya pengkodean, pengelompokan, atau analisis konten, yang menekankan pada pemahaman makna dan pola di balik data tersebut (Creswell & Creswell, 2018). Pemahaman yang tepat terhadap jenis data sangat penting, karena jenis data menentukan metode pengumpulan, teknik analisis, serta interpretasi hasil penelitian.

1.3.2 Skala Pengukuran

Skala pengukuran merupakan konsep penting dalam statistika karena menentukan bagaimana suatu variabel dicatat, diinterpretasikan, dan dianalisis. Pemilihan skala pengukuran yang tepat sangat berpengaruh terhadap jenis analisis statistik yang dapat digunakan, serta keakuratan hasil penelitian. Literatur statistika modern menjelaskan empat skala pengukuran utama yang digunakan secara luas, yaitu nominal, ordinal, interval, dan rasio (Triola, 2018; Field, 2018).

Skala nominal digunakan untuk mengelompokkan objek atau individu ke dalam kategori tanpa memberikan urutan tertentu. Skala ini hanya menunjukkan perbedaan antar kategori dan tidak memiliki makna kuantitatif, misalnya jenis kelamin, fakultas, atau jenis pekerjaan.

Skala ordinal tidak hanya mengelompokkan data, tetapi juga memberikan urutan atau peringkat antar kategori. Meskipun demikian, jarak antar peringkat tidak dapat diukur secara pasti. Contohnya adalah tingkat kepuasan, peringkat kelas, atau preferensi.

Skala interval memungkinkan pengukuran jarak yang sama antar nilai, namun tidak memiliki titik nol absolut. Titik nol pada skala interval bersifat arbitrer, sehingga rasio antar nilai tidak dapat ditafsirkan secara proporsional. Contoh umum skala interval adalah suhu dalam derajat Celsius dan Fahrenheit.

Sedangkan skala rasio memiliki seluruh karakteristik skala interval, tetapi dilengkapi dengan nol absolut yang menunjukkan ketiadaan nilai secara nyata. Hal ini memungkinkan interpretasi perbandingan proporsional, seperti dua kali lebih besar atau dua kali lebih berat. Contoh variabel dengan skala rasio adalah tinggi badan, berat badan, usia, dan pendapatan (Gravetter & Wallnau, 2021).

Memahami perbedaan antara empat skala tersebut sangat penting karena setiap skala memerlukan teknik statistik yang berbeda dalam proses analisis. Kesalahan dalam mengidentifikasi skala pengukuran dapat menghasilkan penggunaan metode yang tidak sesuai dan berpotensi menyebabkan kesimpulan yang bias. Oleh karena itu, peneliti perlu memastikan bahwa skala yang

digunakan konsisten dengan tujuan penelitian dan sifat variabel yang diukur (Creswell & Creswell, 2018).

1.4 Populasi dan Sampel

Dalam penelitian statistik, konsep populasi dan sampel memegang peranan yang sangat fundamental. Keduanya menjadi dasar dalam perancangan penelitian, penentuan metode pengumpulan data, serta penarikan kesimpulan ilmiah. Populasi mencerminkan keseluruhan elemen yang menjadi sasaran kajian, sementara sampel merupakan bagian dari populasi yang dipilih untuk mewakili karakteristiknya. Pemahaman yang tepat mengenai kedua konsep ini penting karena kualitas sampel yang diambil sangat memengaruhi validitas generalisasi yang akan dibuat terhadap populasi secara keseluruhan. Dengan mengenali hubungan antara populasi dan sampel, peneliti dapat merancang studi yang lebih efisien, akurat, dan sesuai dengan tujuan analisis statistik.

1.4.1 Populasi

Populasi dalam suatu penelitian merujuk pada keseluruhan objek atau individu yang menjadi subjek penelitian. Populasi bisa berupa manusia, organisasi, peristiwa, atau unit analisis lain yang sesuai dengan tujuan penelitian. Peneliti mendefinisikan populasi dengan kriteria tertentu agar mencakup semua elemen relevan yang terkait dengan pertanyaan penelitian. Pemahaman yang jelas terhadap populasi penting agar hasil penelitian dapat digeneralisasikan secara tepat ke keseluruhan populasi.

1.4.2 Sampel

Sampel adalah bagian dari populasi yang dipilih untuk dianalisis dalam penelitian sehingga dapat mewakili karakteristik populasi secara keseluruhan. Karena seringkali tidak memungkinkan secara praktis untuk mengamati seluruh populasi, baik karena keterbatasan waktu, biaya, maupun akses, sampel digunakan sebagai representasi. Agar hasil penelitian valid dan dapat digeneralisasikan, pemilihan sampel harus dilakukan dengan metode yang tepat sehingga karakteristik populasi tercermin dengan baik dalam sampel yang digunakan.

1.5 Teknik Sampling

Teknik sampling merupakan prosedur yang digunakan untuk memilih sebagian anggota populasi agar dapat mewakili karakteristik populasi tersebut secara akurat. Penggunaan sampel menjadi penting ketika peneliti tidak memungkinkan untuk mengobservasi seluruh populasi karena keterbatasan waktu, biaya, atau akses. Pemilihan teknik sampling yang tepat sangat berpengaruh terhadap validitas hasil penelitian, baik dari segi representativitas maupun kemampuan generalisasi temuan (Creswell & Creswell, 2018).

Secara umum, teknik sampling dibagi menjadi dua kelompok besar, yaitu probability sampling dan non-probability sampling.

1.5.1 Probability Sampling

Probability sampling adalah metode pengambilan sampel di mana setiap anggota populasi memiliki peluang yang sama atau peluang yang dapat diketahui untuk terpilih sebagai sampel. Teknik ini dianggap lebih kuat secara metodologis karena mampu menghasilkan sampel yang lebih representatif sehingga kesimpulan penelitian dapat digeneralisasikan dengan lebih baik (Taherdoost, 2016). Beberapa teknik yang termasuk dalam kategori ini antara lain:

1. *Simple Random Sampling*, yaitu pemilihan sampel secara acak sehingga setiap elemen memiliki peluang yang sama untuk terpilih.
2. *Stratified Sampling*, yaitu pembagian populasi ke dalam kelompok homogen (strata), kemudian pengambilan sampel dilakukan dari setiap strata.
3. *Cluster Sampling*, yaitu pemilihan sampel dari kelompok atau klaster yang ada dalam populasi, terutama ketika populasi tersebar luas secara geografis.

Teknik-teknik ini banyak digunakan dalam penelitian kuantitatif yang memerlukan ketepatan estimasi dan generalisasi hasil.

1.5.2 Non-Probability Sampling

Non-probability sampling merupakan teknik pengambilan sampel di mana tidak semua anggota populasi memiliki peluang yang sama untuk terpilih. Metode ini sering digunakan dalam penelitian eksploratif, penelitian kualitatif, atau kondisi ketika daftar populasi tidak tersedia secara lengkap. Meskipun tidak sekuat

probability sampling dari segi generalisasi, teknik ini tetap relevan karena fleksibel, efisien, dan sesuai untuk konteks penelitian tertentu (Etikan, Musa, & Alkassim, 2016). Beberapa teknik yang termasuk dalam kategori ini adalah:

1. *Convenience Sampling*, yaitu pemilihan sampel berdasarkan kemudahan akses dan ketersediaan responden.
2. *Purposive Sampling*, yaitu pemilihan sampel berdasarkan pertimbangan atau kriteria tertentu yang relevan dengan tujuan penelitian.
3. *Quota Sampling*, yaitu pemilihan sampel hingga jumlah tertentu yang mewakili karakteristik tertentu dalam populasi.

Penggunaan teknik non-probability sampling perlu disertai pertimbangan metodologis yang matang, termasuk potensi bias dan keterbatasan dalam generalisasi hasil.

1.6 Cakupan Statistika

Ruang lingkup statistika mencakup berbagai konsep, metode, dan pendekatan yang digunakan untuk memahami data serta menarik kesimpulan yang dapat dipertanggungjawabkan. Secara umum, statistika terbagi ke dalam dua cabang utama, yaitu statistika deskriptif dan statistika inferensial, yang masing-masing memiliki tujuan dan teknik analisis yang berbeda. Pemahaman mengenai kedua ruang lingkup ini penting karena keduanya saling melengkapi dalam proses analisis data. Statistika deskriptif memberikan gambaran awal mengenai karakteristik data, sedangkan statistika

inferensial memungkinkan peneliti membuat generalisasi atau prediksi berdasarkan informasi dari sampel. Dengan memahami ruang lingkup ini secara menyeluruh, peneliti dapat memilih pendekatan yang sesuai untuk menjawab pertanyaan penelitian secara ilmiah dan akurat.

1.6.1 Statistika Deskriptif

Statistika deskriptif merupakan cabang ilmu statistika yang berfokus pada pengumpulan, pengolahan, dan penyajian data sehingga informasi yang terkandung di dalamnya dapat dengan mudah dipahami. Proses ini meliputi pengorganisasian data dalam bentuk tabel, grafik, dan diagram, serta perhitungan ukuran pemusatan dan ukuran penyebaran data, seperti *mean*, median, modus, varians, dan standar deviasi (Triola, 2018). Penerapan statistika deskriptif sangat penting dalam berbagai bidang, termasuk penelitian sosial, ekonomi, dan kesehatan, karena membantu peneliti memahami karakteristik data sebelum melakukan analisis lebih lanjut. Dengan memvisualisasikan data melalui diagram batang, diagram lingkaran, atau histogram, peneliti dapat mengidentifikasi pola, tren, dan anomali yang mungkin tidak terlihat dalam bentuk data mentah. Selain itu, ukuran penyebaran memberikan gambaran mengenai variabilitas data, sehingga keputusan yang diambil berdasarkan data tersebut dapat lebih akurat dan terpercaya.

1.6.2 Statistika Inferensial

Statistika inferensial adalah cabang statistika yang digunakan untuk membuat kesimpulan, prediksi, atau generalisasi tentang populasi berdasarkan sampel yang diambil. Pendekatan ini melibatkan teknik estimasi parameter, seperti estimasi titik dan interval, serta pengujian hipotesis untuk menilai signifikansi perbedaan atau hubungan antarvariabel (Montgomery & Runger, 2018). Statistika inferensial memungkinkan peneliti untuk mengukur ketidakpastian dalam pengambilan keputusan, serta memperkirakan perilaku populasi tanpa harus mengumpulkan data dari seluruh populasi. Dalam praktiknya, statistika inferensial diterapkan pada penelitian eksperimental maupun survei, di mana sampel dipilih secara representatif. Dengan metode ini, hasil analisis dapat digeneralisasikan untuk populasi yang lebih luas, sehingga penelitian menjadi lebih efisien dan ekonomis. Penerapan statistika inferensial juga penting dalam pengembangan kebijakan, evaluasi program, dan penelitian ilmiah, karena memberikan dasar ilmiah untuk membuat keputusan berdasarkan bukti.

Penggabungan statistika deskriptif dan inferensial dalam penelitian memberikan pemahaman yang komprehensif tentang data. Statistika deskriptif menyediakan gambaran awal, sedangkan statistika inferensial memungkinkan penarikan kesimpulan yang lebih luas, sehingga kedua pendekatan saling melengkapi dalam proses analisis data yang sistematis dan ilmiah.

1.7 Latihan Soal

1. Jelaskan perbedaan antara data kuantitatif dan kualitatif beserta contohnya.
2. Apa yang dimaksud dengan populasi dan sampel? Berikan contohnya.
3. Uraikan empat skala pengukuran dalam statistika dan fungsinya.
4. Jelaskan perbedaan antara statistika deskriptif dan inferensial dengan contoh.
5. Apa tujuan penggunaan teknik sampling dan sebutkan beberapa jenisnya beserta kegunaannya.

Bab 2: Pengumpulan dan Penyajian Data

2.1 Konsep dan Sumber Data

Konsep data dalam bidang ekonomi dan bisnis merupakan fondasi penting bagi proses analisis, perencanaan, dan pengambilan keputusan yang berbasis bukti. Data didefinisikan sebagai sekumpulan fakta atau informasi kuantitatif maupun kualitatif yang menggambarkan suatu fenomena ekonomi. Dalam praktiknya, data ekonomi dapat diperoleh dari berbagai sumber, baik yang bersifat primer maupun sekunder. Sumber data primer meliputi hasil survei lapangan, wawancara, observasi langsung, serta eksperimen yang dilakukan oleh peneliti atau lembaga tertentu untuk tujuan spesifik. Sementara itu, sumber data sekunder mencakup data yang telah dikumpulkan dan dipublikasikan oleh pihak lain, seperti lembaga pemerintah, organisasi internasional, perusahaan swasta, atau lembaga penelitian independen. Contoh umum sumber data sekunder meliputi laporan keuangan perusahaan, publikasi statistik resmi, arsip transaksi perbankan, serta basis data ekonomi digital yang disediakan oleh lembaga seperti World Bank, OECD, dan Badan Pusat Statistik (BPS).

Pemahaman terhadap karakteristik dan kredibilitas sumber data menjadi aspek krusial dalam penelitian ekonomi dan bisnis.

Data yang tidak akurat atau bias dapat mengarah pada interpretasi keliru dan keputusan yang tidak efektif. Oleh karena itu, kualitas data harus dinilai berdasarkan prinsip validitas, reliabilitas, konsistensi, serta relevansi terhadap masalah yang sedang dikaji. Dalam konteks analisis ekonomi terapan, pengelolaan data yang baik juga mencakup proses pembersihan (*data cleaning*), pengkodean, serta transformasi agar siap digunakan dalam model statistik atau ekonometrika. Kemajuan teknologi digital semakin memperluas akses dan kecepatan pengumpulan data melalui otomatisasi transaksi, sistem *big data*, dan integrasi data lintas sektor, yang membuka peluang analisis lebih mendalam terhadap perilaku konsumen, dinamika pasar, dan kinerja makroekonomi (Varian, 2021).

Selain aspek teknis, sumber data juga memiliki implikasi etis dan kebijakan. Transparansi metodologi pengumpulan data, perlindungan privasi, serta kepatuhan terhadap regulasi menjadi pertimbangan penting dalam era ekonomi digital yang sarat pertukaran informasi. Para peneliti dan praktisi bisnis diharapkan mampu mengombinasikan berbagai sumber data secara bijak untuk menghasilkan analisis yang komprehensif dan berdaya guna bagi pengambilan keputusan strategis. Dengan demikian, pemahaman mendalam mengenai konsep dan sumber data tidak hanya berfungsi sebagai keterampilan metodologis, tetapi juga sebagai landasan epistemologis bagi pembentukan pengetahuan ekonomi yang akurat dan relevan dalam mendukung pembangunan berkelanjutan (Mayer-Schönberger & Cukier, 2021).

2.2 Metode Pengumpulan Data

Metode pengumpulan data merupakan tahapan fundamental dalam penelitian karena menentukan kualitas informasi yang digunakan untuk menjawab rumusan masalah. Keberhasilan penelitian sangat dipengaruhi oleh ketepatan metode dalam menangkap fenomena yang diteliti serta kesesuaiannya dengan konteks lapangan. Pemilihan metode umumnya mempertimbangkan tujuan penelitian, desain metodologis, ketersediaan waktu dan biaya, serta karakteristik populasi sasaran. Dalam praktiknya, kombinasi beberapa metode sering digunakan untuk meningkatkan validitas temuan, terutama pada penelitian sosial dan kesehatan masyarakat yang membutuhkan data objektif sekaligus pemahaman mendalam mengenai perilaku atau pengalaman individu (Creswell, 2021).

Metode yang sering digunakan meliputi wawancara, observasi, survei kuesioner, dan pengumpulan data administratif. Masing-masing memiliki keunggulan, keterbatasan, serta prosedur operasional yang berbeda sehingga pemilihannya harus disesuaikan dengan jenis data yang dibutuhkan.

2.2.1 Wawancara dan observasi

Wawancara merupakan metode pengumpulan data yang memungkinkan peneliti memperoleh informasi langsung melalui percakapan terstruktur, semi-terstruktur, atau tidak terstruktur. Metode ini bermanfaat untuk menggali persepsi, pengalaman, dan motivasi partisipan secara mendalam. Selain itu, wawancara

memberikan fleksibilitas bagi peneliti untuk menyesuaikan pertanyaan berdasarkan dinamika interaksi.

Observasi digunakan ketika peneliti ingin memahami perilaku atau situasi dalam konteks alamiah. Observasi dapat bersifat partisipatif maupun nonpartisipatif, dengan fokus pada tindakan, interaksi, atau proses yang tidak selalu terungkap melalui wawancara. Kombinasi wawancara dan observasi sering meningkatkan ketepatan interpretasi data karena kedua metode saling melengkapi dalam menangkap dimensi verbal dan nonverbal dari fenomena sosial (Marshall, 2020).

2.2.2 Survei kuesioner

Survei kuesioner merupakan metode yang efektif untuk mengumpulkan data dari populasi besar dalam waktu relatif singkat. Instrumen ini biasanya berisi pertanyaan tertutup atau terbuka yang disusun secara sistematis untuk mengukur variabel tertentu. Keunggulan utama survei adalah efisiensi, kemampuan menghasilkan data kuantitatif terstandar, serta kemudahan dalam analisis statistik. Namun, kualitas data sangat bergantung pada kejelasan item pertanyaan, tingkat respons, dan kejujuran responden.

Survei cocok digunakan ketika penelitian membutuhkan generalisasi temuan, terutama dalam studi perilaku, opini publik, atau karakteristik demografis. Penggunaan survei daring semakin meningkat karena biaya rendah dan jangkauan luas.

2.2.3 Data administratif

Pengumpulan data administratif dilakukan melalui dokumen organisasi, catatan klinis, arsip pemerintahan, atau basis data institusi. Metode ini bermanfaat untuk mendapatkan informasi objektif yang sudah terdokumentasi, seperti laporan keuangan, rekam medis, atau data kependudukan. Keunggulannya adalah reliabilitas yang tinggi dan ketersediaan data historis dalam rentang waktu panjang.

Namun, penggunaan data administratif perlu mempertimbangkan aspek privasi, etika, serta kesesuaian format data dengan kebutuhan penelitian. Peneliti sering mengombinasikan data administratif dengan survei atau wawancara guna memperoleh gambaran yang lebih komprehensif.

2.3 Jenis-jenis Data

2.3.1 Klasifikasi Umum Data: Kualitatif dan Kuantitatif

Secara umum, data diklasifikasikan menjadi dua kelompok utama, yaitu data kualitatif dan kuantitatif. Data kualitatif menggambarkan karakteristik atau atribut yang tidak dapat diukur secara numerik, seperti warna, jenis kelamin, atau preferensi konsumen. Data ini bersifat deskriptif dan biasanya diperoleh melalui observasi, wawancara, atau kuesioner terbuka. Sebaliknya, data kuantitatif melibatkan angka dan ukuran, mencerminkan besaran yang dapat dihitung atau dibandingkan. Contohnya meliputi jumlah produk terjual, pendapatan bulanan, atau skor kepuasan

pelanggan. Pemahaman terhadap perbedaan ini penting karena akan memengaruhi metode pengumpulan, analisis, dan interpretasi data (Frankfort-Nachmias & Leon-Guerrero, 2020).

2.3.2 Empat Skala Pengukuran Data: Nominal, Ordinal, Interval, dan Rasio

Skala pengukuran data terbagi menjadi empat jenis utama: nominal, ordinal, interval, dan rasio. Data nominal merupakan data kategorikal yang tidak memiliki urutan atau hierarki, seperti warna, jenis kelamin, atau nama merek. Sementara itu, data ordinal memiliki urutan atau peringkat, namun jarak antar kategori tidak terukur secara pasti; contohnya adalah tingkat kepuasan pelanggan (misalnya: tidak puas, netral, puas). Data interval memiliki jarak antar nilai yang tetap dan terukur, tetapi tidak memiliki nol mutlak; sebagai contoh, suhu dalam derajat Celsius atau Fahrenheit termasuk dalam kategori ini. Adapun data rasio memiliki seluruh karakteristik skala interval, namun dilengkapi dengan nol absolut, yang memungkinkan perbandingan proporsional—contohnya termasuk tinggi badan, berat, atau pendapatan. Pemahaman terhadap perbedaan masing-masing skala ini penting karena akan menentukan teknik statistik yang dapat digunakan dalam proses analisis data, seperti pemilihan uji parametrik atau nonparametrik sesuai dengan jenis skala yang dimiliki (Gravetter & Wallnau, 2021).

2.3.3 Implikasi Jenis Data dalam Analisis Bisnis

Dalam praktik bisnis, pengelompokan data yang tepat menjadi fondasi analisis statistik dan pengambilan keputusan. Misalnya, data kuantitatif rasio digunakan untuk menghitung *return on investment*, sedangkan data ordinal berguna dalam survei kepuasan pelanggan. Data kualitatif nominal, meskipun tidak dapat dihitung secara langsung, tetap penting dalam segmentasi pasar dan analisis preferensi konsumen.

Oleh karena itu, pemahaman terhadap jenis dan skala data sangat krusial bagi analis, peneliti, maupun manajer dalam memilih alat analisis yang relevan dan menghasilkan kesimpulan yang valid.

2.4 Penyajian Data dalam Tabel dan Grafik

Penyajian data merupakan tahap penting dalam proses analisis statistik yang bertujuan untuk menyampaikan informasi secara sistematis dan mudah dipahami. Melalui penyajian yang efektif, data mentah yang kompleks dapat diubah menjadi bentuk yang lebih ringkas dan komunikatif, sehingga mempermudah interpretasi serta pengambilan keputusan. Dalam konteks pembelajaran dan riset, kemampuan menyajikan data dengan benar merupakan keterampilan dasar yang mendukung pemahaman terhadap pola, tren, dan perbandingan antar kelompok data.

2.4.1 Penyajian data dalam bentuk tabel

Tabel merupakan bentuk penyajian data yang paling dasar dan umum digunakan, terutama dalam tahap eksplorasi awal. Tabel frekuensi, sebagai salah satu contohnya, menyajikan jumlah kemunculan suatu nilai atau kategori dalam kumpulan data. Tabel ini dapat digunakan untuk data kuantitatif maupun kualitatif, dan sering menjadi dasar untuk penyajian grafik selanjutnya.

Tabel juga memungkinkan penyajian data silang (*cross-tabulation*) antara dua atau lebih variabel. Kelebihan utama tabel adalah kemampuannya menyampaikan informasi numerik secara akurat dan terstruktur. Namun, untuk menyampaikan pesan visual secara cepat, grafik sering kali lebih efektif.

2.4.2 Penyajian data dalam bentuk grafik

Grafik digunakan untuk menyampaikan hubungan, distribusi, atau perubahan data secara visual. Beberapa jenis grafik yang umum digunakan meliputi:

- a. Diagram batang (*bar chart*): Cocok untuk membandingkan frekuensi atau nilai antar kategori diskrit. Tinggi batang mewakili nilai variabel, sehingga perbedaan antar kategori mudah diamati.
- b. Diagram garis (*line chart*): Efektif untuk menunjukkan tren atau perubahan nilai dari waktu ke waktu. Diagram ini umum digunakan dalam data deret waktu (*time series*).
- c. Diagram lingkaran (*pie chart*): Menyajikan proporsi tiap kategori terhadap total keseluruhan, dengan sudut sektor mencerminkan persentase relatif.

- d. Histogram: Digunakan untuk menyajikan distribusi frekuensi data numerik dalam interval kelas, berbeda dari diagram batang karena sumbu horizontalnya bersifat kontinu.

Pemilihan jenis grafik sangat bergantung pada sifat data dan tujuan analisis. Penggunaan grafik yang tidak tepat dapat menyesatkan interpretasi, sehingga prinsip kejelasan dan proporsionalitas visual perlu dijaga (Zhou et al., 2020).

2.4.3 Konteks aplikasi dan keterampilan literasi data

Dalam dunia akademik dan profesional, keterampilan menyajikan data secara visual dianggap bagian dari literasi data. Kemampuan ini tidak hanya mendukung komunikasi hasil analisis, tetapi juga memengaruhi pemahaman audiens terhadap kompleksitas informasi. Oleh karena itu, pemilihan bentuk penyajian harus mempertimbangkan audiens sasaran serta konteks penggunaannya (Gould & Sun, 2021).

2.5 Distribusi Frekuensi

2.5.1 Pengertian dan Tujuan

Distribusi frekuensi merupakan teknik statistik yang digunakan untuk menyajikan data kuantitatif dalam bentuk kelompok-kelompok nilai atau *interval kelas*. Setiap kelas menunjukkan rentang nilai tertentu, sementara frekuensinya menunjukkan jumlah kemunculan data dalam rentang tersebut. Tujuan utama distribusi frekuensi adalah menyederhanakan

kumpulan data yang besar dan kompleks menjadi tampilan yang lebih mudah dibaca dan dianalisis. Dengan menggunakan tabel atau grafik distribusi frekuensi, seperti *histogram* atau *poligon frekuensi*, peneliti dapat dengan cepat mengidentifikasi pola, kecenderungan, atau anomali dalam data (Weiss & Weiss, 2020). Pendekatan ini juga memungkinkan perbandingan yang lebih efisien antar kelompok data dan mendukung proses pengambilan keputusan berbasis data.

2.5.2 Manfaat dalam Analisis Skala Besar

Distribusi frekuensi sangat berguna dalam konteks analisis data berskala besar, seperti dalam bidang bisnis, ekonomi, dan ilmu sosial. Dalam pengelolaan data penjualan, misalnya, distribusi frekuensi dapat menunjukkan berapa banyak transaksi terjadi dalam kisaran harga tertentu, atau berapa banyak pelanggan yang melakukan pembelian dalam jumlah tertentu. Teknik ini membantu perusahaan mengidentifikasi segmen pasar, mengevaluasi efektivitas strategi promosi, dan merancang kebijakan harga yang lebih akurat. Selain itu, dalam bidang riset pasar, distribusi frekuensi dapat digunakan untuk mengolah hasil survei dalam skala besar dan menemukan distribusi opini konsumen terhadap suatu produk atau layanan. Keandalan metode ini dalam menyaring informasi penting dari data mentah menjadikannya alat penting dalam eksplorasi data awal (*exploratory data analysis*) (Umar, 2021).

2.5.3 Penyusunan dan Interpretasi

Untuk menyusun distribusi frekuensi, langkah awal adalah menentukan jumlah kelas dan lebar kelas yang sesuai berdasarkan sebaran data. Selanjutnya, data diklasifikasikan ke dalam kelas-kelas

tersebut dan dihitung frekuensinya. Hasilnya kemudian disajikan dalam bentuk tabel distribusi frekuensi. Interpretasi dilakukan dengan memperhatikan distribusi nilai, apakah bersifat simetris, miring ke kiri atau kanan, atau memiliki nilai pencilan. Distribusi yang normal atau simetris menunjukkan data tersebar merata di sekitar nilai tengah, sedangkan distribusi yang miring menandakan adanya ketidakseimbangan persebaran nilai. Pemahaman terhadap bentuk distribusi ini sangat penting dalam memilih metode analisis statistik lanjutan, seperti uji parametrik atau nonparametrik, serta dalam menyusun laporan data yang akurat dan informatif.

2.6 Latihan Soal

1. Jelaskan perbedaan antara data primer dan data sekunder dalam konteks bisnis.
2. Mengapa pemilihan metode pengumpulan data penting dalam penelitian ekonomi?
3. Berikan contoh data nominal, ordinal, interval, dan rasio dalam bisnis.
4. Jelaskan kelebihan diagram batang dibandingkan tabel frekuensi.
5. Bagaimana distribusi frekuensi membantu pengambilan keputusan bisnis?

Bab 3: Ukuran Pemusatan dan Penyebaran Data

3.1 Pengantar

Dalam statistika deskriptif, ukuran pemusatan dan ukuran penyebaran merupakan dua komponen utama yang digunakan untuk merangkum dan menyajikan data secara lebih sistematis sehingga pola, kecenderungan, serta karakteristik suatu kumpulan data dapat dipahami dengan lebih jelas. Ketika data dikumpulkan dalam bentuk mentah, nilai-nilai tersebut sering kali tidak memberikan gambaran yang langsung dapat ditafsirkan. Oleh karena itu, diperlukan metode peringkasan yang membantu menganalisis struktur data secara lebih ringkas namun tetap informatif. Ukuran pemusatan berfungsi menunjukkan nilai representatif dari suatu data set, sedangkan ukuran penyebaran menggambarkan derajat variasi yang terdapat dalam data tersebut. Kombinasi keduanya memungkinkan peneliti atau analis menyusun interpretasi yang lebih komprehensif mengenai kondisi maupun dinamika data yang diobservasi.

Ukuran pemusatan meliputi *mean*, median, dan mode. *Mean* atau nilai rata-rata aritmetika memberikan gambaran umum mengenai kecenderungan nilai-nilai dalam data set, terutama ketika data tidak mengandung pencilan ekstrem. Median, sebagai nilai tengah dari data yang telah diurutkan, umumnya lebih stabil terhadap

pencilan dan sangat berguna untuk distribusi yang tidak simetris. Sementara itu, mode menggambarkan nilai yang paling sering muncul sehingga relevan untuk data kategorikal maupun data kuantitatif tertentu. Pemilihan ukuran pemusatan yang tepat harus mempertimbangkan karakteristik distribusi data, skala pengukuran, serta tujuan analisis yang dilakukan. Dalam berbagai konteks analisis modern, ukuran pemusatan juga menjadi dasar dalam metode analitis lebih lanjut, karena banyak teknik statistik dan *machine learning* memerlukan pemahaman awal mengenai distribusi nilai data (James et al., 2021).

Di sisi lain, ukuran penyebaran memberikan informasi tentang seberapa jauh nilai-nilai individual dalam data set menyimpang dari ukuran pemusatannya. Simpangan baku, varians, rentang, dan rentang antarkuartil merupakan ukuran penyebaran yang umum digunakan. Simpangan baku menggambarkan deviasi rata-rata nilai data terhadap *mean*, sehingga semakin besar simpangan baku, semakin tinggi tingkat variasi yang terkandung. Varians, yang merupakan kuadrat dari simpangan baku, juga memberikan ukuran yang menggambarkan keragaman data secara matematis. Rentang menunjukkan perbedaan antara nilai minimum dan maksimum, sedangkan rentang antarkuartil menawarkan ukuran yang lebih tahan terhadap pencilan karena hanya mempertimbangkan 50% nilai tengah dalam distribusi data. Dengan menggunakan ukuran penyebaran, analis dapat menilai stabilitas, pola ketidakpastian, serta tingkat konsistensi suatu fenomena.

Korelasi antara ukuran pemusatan dan penyebaran merupakan hal yang sangat penting dalam interpretasi data. Ukuran pemusatan tanpa memahami penyebarannya dapat menghasilkan kesimpulan yang menyesatkan, sebab dua data set dapat memiliki mean yang sama tetapi penyebarannya sangat berbeda. Dalam kasus tersebut, tingkat risiko atau ketidakstabilan kedua data set jelas tidak serupa. Oleh karena itu, memahami bagaimana nilai-nilai tersebar di sekitar pusat distribusi merupakan langkah penting dalam penarikan kesimpulan yang bertanggung jawab. Informasi ini dapat digunakan dalam berbagai konteks, seperti evaluasi performa sistem, pengendalian kualitas produksi, atau penilaian reliabilitas proses bisnis.

Statistika deskriptif, melalui ukuran pemusatan dan penyebaran, membantu menyederhanakan data kompleks menjadi bentuk yang lebih mudah dibaca dan dianalisis. Selain memberikan gambaran umum, ukuran-ukuran ini juga menjadi fondasi bagi penerapan metode statistik yang lebih mendalam, baik dalam analisis prediktif, pengujian hipotesis, maupun pemodelan. Dengan meningkatnya pemanfaatan teknologi analitik dan big data, pemahaman mendalam terhadap ukuran-ukuran deskriptif menjadi semakin penting agar analisis yang dilakukan tetap valid, akurat, dan bermakna. Pendekatan sistematis dalam menggunakan ukuran pemusatan dan penyebaran juga mendukung proses pengambilan keputusan berbasis data yang lebih rasional dan dapat dipertanggungjawabkan (Wilcox, 2020).

3.2 Ukuran Pemusatan

3.2.1 Konsep Umum Ukuran Pemusatan

Ukuran pemusatan (*measures of central tendency*) merupakan parameter statistik yang digunakan untuk menggambarkan nilai representatif dari suatu kumpulan data. Dalam analisis statistik deskriptif, ukuran pemusatan berfungsi merangkum data numerik menjadi satu nilai inti yang mewakili keseluruhan distribusi. Keberadaannya penting karena memungkinkan analisis memahami pola umum, pergeseran distribusi, karakteristik data, dan potensi penyimpangan sebelum dilakukan analisis lebih lanjut. Ukuran pemusatan juga menjadi dasar untuk menghitung ukuran variabilitas, melakukan perbandingan antar kelompok, serta membangun pemodelan statistik. Dalam banyak konteks penelitian modern, penggunaan ukuran pemusatan direkomendasikan sebagai langkah awal untuk mengidentifikasi struktur distribusi dan mendeteksi ketidakwajaran data (Pratama & Lukman, 2021).

3.2.2 Rata-Rata (*Mean*)

Mean atau rata-rata merupakan ukuran pemusatan yang paling umum digunakan karena rumusnya yang sederhana dan kemampuannya dalam memberikan gambaran matematis yang komprehensif mengenai pusat data. Rata-rata dihitung dengan menjumlahkan seluruh nilai dalam satu set data lalu membaginya dengan jumlah observasi. *Mean* sangat efektif ketika data memiliki distribusi simetris dan tidak menunjukkan adanya *outlier*. Nilai ini

mencerminkan kecenderungan umum data dan sering menjadi indikator awal dalam menentukan kondisi normal suatu populasi.

Namun, *mean* memiliki kelemahan utama, yaitu sangat sensitif terhadap nilai ekstrem. Satu atau dua nilai yang sangat besar atau sangat kecil dapat menggeser rata-rata secara signifikan sehingga tidak lagi merepresentasikan data secara akurat. Oleh karena itu, dalam konteks penelitian empiris, *mean* sering dilengkapi dengan median atau ukuran dispersi seperti simpangan baku untuk memberikan pemahaman lebih menyeluruh terkait penyebaran data dan potensi distorsi distribusi. Sensitivitas mean terhadap *outlier* menjadi pertimbangan penting dalam pemilihan ukuran pemusatan yang tepat untuk analisis data tertentu (Rahardjo, 2022).

Formula untuk menentukan rata-rata data tunggal adalah sebagai berikut.

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

Keterangan:

\bar{x} : rata-rata

x_i : datum ke-*i*

n : banyak datum

Sedangkan rata-rata untuk data kelompok, formula menentukan rata-rata adalah sebagai berikut.

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

Keterangan:

\bar{x} : rata-rata

f_i : frekuensi pada kelas ke-i

x_i : nilai tengah pada kelas ke-i

3.2.3 Median

Median merupakan nilai tengah dari suatu data yang telah diurutkan berdasarkan besarnya. Median diperoleh dengan memilih nilai yang berada di posisi tengah jika jumlah data ganjil. Jika jumlah data genap, median dihitung dari rata-rata dua nilai tengah. Karena tidak dipengaruhi oleh *outlier*, median menjadi alternatif yang tepat ketika data memiliki distribusi tidak simetris (*skewed distribution*) atau banyak mengandung nilai ekstrem.

Dalam banyak aplikasi, median digunakan untuk menggambarkan kondisi tipikal yang lebih stabil dibandingkan mean. Sebagai contoh, dalam analisis pendapatan atau harga rumah, perbedaan ekstrem antara kelompok ekonomi membuat *mean* sering kali tidak mewakili kondisi nyata, sehingga median digunakan sebagai parameter yang lebih relevan. Penggunaan median juga umum dalam kebijakan publik, penelitian sosial, dan epidemiologi, terutama ketika tujuan analisis adalah memahami kecenderungan populasi tanpa distorsi yang dihasilkan nilai ekstrem.

Langkah pertama sebelum menentukan median adalah urutkan datum dari yang terkecil ke yang terbesar. Setelah itu

menentukan datum yang menjadi median dengan formula untuk menentukan median data tunggal adalah sebagai berikut.

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \text{ untuk } n \text{ genap.....(3)}$$

$$Me = x_{\frac{n+1}{2}} \text{ untuk } n \text{ ganjil.....(4)}$$

Keterangan:

Me : nilai median

$x_{\frac{n}{2}}$: datum ke- $\frac{n}{2}$

$x_{\frac{n}{2}+1}$: datum ke- $\frac{n}{2} + 1$

$x_{\frac{n+1}{2}}$: datum ke- $\frac{n+1}{2}$

n : banyak datum

Sedangkan nilai median untuk data kelompok, diperoleh dengan langkah-langkah sebagai berikut.

1. Memastikan kolom interval kelas, frekuensi, dan frekuensi kumulatif tersedia pada tabel.
2. Menentukan kelas median dengan melihat frekuensi kumulatif yang nilainya sama dengan atau frekuensi pertama yang lebih besar dari $\frac{\sum_{i=1}^n f_i}{2}$
3. Menentukan nilai median dari kelas median dengan formula sebagai berikut.

$$Me = Tb + \left(p \times \frac{\frac{\sum_{i=1}^n f_i}{2} - F_{Me}}{f_{Me}} \right) \text{(1)}$$

Keterangan:

Me : nilai median

Tb : Tepi bawah kelas median = Batas bawah kelas median $- 0,5$

p : panjang kelas = Batas atas kelas $-$ Batas bawah kelas $+ 1$

F_{Me} : Frekuensi kumulatif sebelum kelas median

f_{Me} : frekuensi kelas median

3.2.4 Modus

Modus adalah nilai dalam data yang paling sering muncul. Berbeda dengan *mean* dan median, modus tidak memerlukan pengurutan atau perhitungan matematis sehingga mudah diterapkan, terutama pada data kategorikal. Karena dapat digunakan pada skala nominal, modus menjadi satu-satunya ukuran pemusatan yang berlaku ketika data tidak berbentuk angka. Misalnya, dalam survei preferensi konsumen, modus dapat menggambarkan produk yang paling banyak dipilih tanpa mempertimbangkan besaran nilai.

Modus juga dapat diterapkan pada data numerik, khususnya ketika distribusi menunjukkan pola multimodal, yaitu memiliki beberapa puncak frekuensi. Dalam kondisi ini, penggunaan *mean* saja dapat menyesatkan karena tidak mencerminkan variasi pola dalam distribusi data. Modus memberikan informasi tambahan yang relevan mengenai kecenderungan frekuensi dan pola dominasi nilai tertentu dalam data set.

Nilai modus pada data tunggal ditentukan dengan datum terbanyak pada data tunggal. Berbeda halnya dengan menentukan modus pada data kelompok. Adapun langkah-langkah yang

dilakukan dalam menentukan modus data kelompok adalah sebagai berikut.

1. Menentukan kelas modus yang memiliki frekuensi paling besar.
2. Menentukan nilai modus

$$Mo = Tb + \left(p \times \frac{s_1}{s_1 + s_2} \right)$$

Keterangan :

Mo: nilai modus

Tb : Tepi bawah kelas modus = Batas bawah kelas modus – 0,5

p : panjang kelas = Batas atas kelas – Batas bawah kelas + 1

*s*₁: selisih frekuensi kelas modus dengan frekuensi sebelum kelas modus

*s*₂: selisih frekuensi kelas modus dengan frekuensi setelah kelas modus

3.2.5 Pemilihan Ukuran Pemusatan yang Tepat

Pemilihan ukuran pemusatan harus mempertimbangkan bentuk distribusi data, konteks analisis, serta keberadaan *outlier*. *Mean* tepat digunakan ketika data bersifat simetris dan tidak mengandung nilai ekstrem. Median menjadi pilihan lebih representatif ketika data mencerminkan distribusi miring atau ketika tujuan analisis adalah menggambarkan kondisi tipikal secara stabil. Sementara itu, modus cocok digunakan untuk data kategorikal atau ketika analisis menekankan frekuensi kemunculan nilai tertentu. Pemilihan ukuran pemusatan yang tidak tepat dapat menyebabkan bias interpretasi dan menghasilkan kesimpulan yang keliru, sehingga

pemahaman mengenai struktur distribusi menjadi faktor penting dalam tahap awal analisis data (Sulaiman, 2023).

3.3 Ukuran Letak

3.3.1 Konsep Dasar dan Fungsi Ukuran Letak

Ukuran letak merupakan indikator statistik yang menunjukkan posisi suatu nilai dalam distribusi data. Berbeda dengan ukuran pemusatan yang menggambarkan nilai tipikal, ukuran letak berfokus pada posisi relatif sebuah data terhadap keseluruhan kelompok. Dengan demikian, ukuran letak membantu menjelaskan bagaimana nilai tertentu berada dalam konteks distribusi, apakah berada pada kelompok bawah, menengah, atau atas. Konsep ini sangat penting dalam analisis yang memerlukan perbandingan antar individu atau kelompok, terutama ketika data memiliki variasi yang tinggi atau distribusi yang tidak simetris (Zhang & Ma, 2020).

Fungsi ukuran letak mencakup penyederhanaan interpretasi distribusi, identifikasi kelompok berdasarkan posisi nilai, dan penentuan batas-batas klasifikasi dalam evaluasi. Ukuran ini juga memungkinkan analisis yang lebih komprehensif dalam situasi di mana nilai ekstrem atau pencilan berpotensi mengganggu interpretasi statistik berbasis rata-rata. Dalam berbagai penelitian, ukuran letak digunakan sebagai dasar pengelompokan data untuk tujuan akademik, medis, maupun organisasi (Jebb, Tay, Wang, & Huang, 2020).

3.3.2 Kuartil, Desil, dan Persentil sebagai Alat Identifikasi Posisi Data

Kuartil membagi data menjadi empat bagian yang sama besar: kuartil pertama (Q1) menunjukkan 25% populasi terbawah, kuartil kedua (Q2) adalah median, dan kuartil ketiga (Q3) menunjukkan 75% data. Kuartil sering dipakai untuk memahami struktur penyebaran data dan mengidentifikasi kecenderungan distribusi. Selain itu, selisih antara Q1 dan Q3 (IQR) digunakan untuk mendeteksi pencilan, sehingga ukuran ini sangat bermanfaat dalam analisis statistik dasar maupun terapan (Zhang & Ma, 2020).

Desil membagi data menjadi sepuluh bagian, memberikan tingkat ketelitian lebih tinggi dibanding kuartil. Setiap desil menunjukkan interval 10% dari distribusi. Desil kerap digunakan dalam analisis sosial, ekonomi, dan evaluasi kinerja karena dapat menggambarkan kesenjangan atau variasi antar kelompok secara lebih rinci.

Persentil, yang membagi data menjadi seratus bagian, merupakan ukuran letak yang paling sensitif terhadap variasi dalam distribusi. Persentil umum digunakan dalam pendidikan, kesehatan, dan psikometri, misalnya dalam kurva pertumbuhan, penilaian kemampuan, atau evaluasi posisi akademik. Dengan persentil, posisi individu dapat dibandingkan secara lebih presisi terhadap populasi, sehingga analisis menjadi lebih akurat dan terstandarisasi (Jebb et al., 2020).

3.3.3 Penerapan Ukuran Letak dalam Berbagai Bidang Analisis

Dalam bidang pendidikan, ukuran letak membantu membandingkan pencapaian siswa terhadap kelompoknya. Penggunaan persentil memudahkan guru atau institusi pendidikan dalam menilai apakah seorang siswa berada pada kategori rendah, sedang, atau tinggi. Kuartil dan desil digunakan untuk membuat klasifikasi kinerja yang objektif, sehingga keputusan akademik dapat diambil berdasarkan pertimbangan yang terukur.

Pada bidang kesehatan, ukuran letak sangat penting dalam evaluasi pertumbuhan, status gizi, serta interpretasi indikator klinis. Persentil banyak digunakan dalam penilaian perkembangan anak untuk menentukan apakah pertumbuhan seseorang berada dalam batas normal atau memerlukan perhatian khusus. Hal ini memungkinkan tenaga medis memberikan intervensi yang relevan sesuai posisi individu terhadap standar populasi.

Dalam evaluasi kinerja organisasi, ukuran letak digunakan untuk menilai pencapaian pegawai atau unit kerja berdasarkan pembandingan dalam kelompok yang sama. Persentil sering dipakai dalam sistem *performance ranking*, sementara kuartil digunakan untuk menentukan kelompok kinerja. Dengan demikian, ukuran letak tidak hanya menjadi alat analisis statistik, tetapi juga instrumen manajerial yang mendukung pengambilan keputusan berbasis data dalam lingkungan kerja modern.

3.4 Ukuran Penyebaran

3.4.1 Konsep Dasar Ukuran Penyebaran

Ukuran penyebaran merupakan komponen penting dalam analisis statistik karena memberikan gambaran mengenai seberapa jauh data tersebar di sekitar nilai pemusatan. Tanpa informasi penyebaran, nilai rata-rata atau median sering kali dapat menyesatkan, terutama ketika data memiliki variasi tinggi atau mengandung outlier. Dalam praktik statistik modern, ukuran penyebaran digunakan untuk menilai konsistensi data, mendeteksi anomali, serta menganalisis stabilitas suatu fenomena. Pemahaman mengenai penyebaran data memungkinkan peneliti melakukan interpretasi yang lebih komprehensif, baik dalam konteks penelitian akademis maupun pengambilan keputusan praktis. Literatur statistik kontemporer menegaskan bahwa ukuran penyebaran merupakan pelengkap esensial dari ukuran pemusatan dalam menghasilkan analisis yang reliabel (Triola, 2020).

3.4.2 *Range* dan Sifat Dasarnya

Range atau jangkauan adalah ukuran penyebaran yang paling sederhana, dihitung sebagai selisih antara nilai maksimum dan nilai minimum dalam suatu kumpulan data. Ukuran ini memberikan informasi awal mengenai luasnya sebaran, namun memiliki kelemahan karena sangat sensitif terhadap *outlier*. Satu nilai ekstrem dapat membuat *range* terlihat besar, meskipun sebagian besar data sebenarnya berada dalam rentang yang lebih sempit. Meskipun demikian, *range* tetap berguna sebagai langkah awal eksplorasi data,

terutama pada tahap pemeriksaan distribusi umum dan identifikasi potensi anomali. Dalam praktik statistik, range jarang digunakan sebagai indikator tunggal, tetapi sering dikombinasikan dengan ukuran penyebaran lainnya untuk memperoleh pemahaman yang lebih seimbang.

3.4.3 Rentang Interkuartil (IQR) dan Ketahanannya terhadap *Outlier*

Rentang interkuartil (*Interquartile Range* atau IQR) dihitung dari selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1). IQR hanya memperhitungkan 50% data tengah, sehingga tidak dipengaruhi oleh nilai ekstrem pada bagian bawah maupun atas data. Ketahanannya terhadap *outlier* menjadikan IQR sebagai ukuran penyebaran yang sangat relevan dalam analisis data yang tidak simetris atau memiliki distribusi berat ekor. Dalam banyak konteks penelitian, IQR digunakan untuk menentukan variasi inti dari data dan menjadi dasar konstruksi *boxplot*. Selain itu, IQR juga digunakan dalam aturan deteksi *outlier*, di mana suatu titik data dianggap ekstrem apabila berada di bawah $Q1 - 1,5 \cdot IQR$ atau di atas $Q3 + 1,5 \cdot IQR$. Ukuran yang stabil terhadap gangguan nilai ekstrem ini menjadikan IQR salah satu ukuran penyebaran yang banyak direkomendasikan dalam penelitian statistik terapan.

3.4.4 Varians sebagai Ukuran Penyebaran Kuadrat

Varians merupakan ukuran penyebaran yang dihitung dari rata-rata kuadrat selisih antara tiap nilai data dan *mean*. Varians memberikan gambaran matematis mengenai distribusi data di sekitar mean, dengan menekankan nilai yang jauh dari *mean* melalui

penguadratan selisih. Oleh karena ukurannya berupa satuan kuadrat, interpretasi varians tidak selalu intuitif, tetapi varians memiliki keunggulan analitis dan digunakan secara luas dalam perhitungan model statistik seperti *regression*, ANOVA, dan *machine learning*. Varians yang tinggi menunjukkan bahwa data tersebar jauh dari mean, sedangkan varians yang rendah menandakan data lebih homogen.

Formula yang digunakan untuk menentukan varians populasi adalah sebagai berikut.

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

Varians sampel dapat ditentukan dengan formula sebagai berikut.

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

3.4.5 Simpangan Baku sebagai Ukuran Penyebaran yang Lebih Interpretatif

Simpangan baku (*standard deviation*) merupakan akar kuadrat varians, sehingga memiliki satuan yang sama dengan data aslinya. Karena alasan tersebut, simpangan baku lebih mudah diinterpretasikan dibandingkan varians. Simpangan baku menggambarkan jarak rata-rata data dari *mean*; semakin besar nilai simpangan baku, semakin besar pula variasi data. Dalam penelitian kuantitatif, simpangan baku menjadi ukuran penyebaran yang paling umum digunakan karena kemudahannya dalam interpretasi, relevansinya dalam berbagai metode statistik, serta perannya dalam

distribusi probabilitas, terutama distribusi normal. Penggunaan simpangan baku secara konsisten dalam riset ilmiah modern memperkuat perannya sebagai indikator utama dalam memahami variabilitas data (Field, 2022). Simpangan baku adalah akar dari nilai varians.

3.5 Contoh Soal

3.5.1 Contoh Soal Data Tunggal

Diketahui data sampel pendapatan lima kepala keluarga RT 01 sebagai berikut : 3.000.000, 3.500.000, 4.000.000, 4.000.000, 2.000.000.

Tentukan rata-rata, median, modus, dan simpangan baku pendapatan lima kepala keluarga RT 01 tersebut!

a. Rata-rata

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{3.000.000+3.500.000+4.000.000+4.000.000+2.000.000}{5}$$

$$\bar{x} = 3.300.000$$

b. Median

Sebelum menentukan median, datum harus diurutkan dari terkecil hingga terbesar terlebih dahulu.

2.000.000, 3.000.000, 3.500.000, 4.000.000, 4.000.000

Karena banyak datum adalah 5 (ganjil), maka nilai mediannya adalah $= \frac{x_{n+1}}{2} = \frac{x_{5+1}}{2} = \frac{x_6}{2} = x_3 = 3.500.000$

c. Modus

Datum yang paling sering muncul adalah 4.000.000 maka nilai modulusnya adalah 4.000.000.

d. Simpangan baku

Karena yang diketahui data sampel, maka formula simpangan baku yang digunakan adalah formula simpangan baku sampel.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$
$$= \sqrt{\frac{(2.000.000 - 3.300.000)^2 + (3.000.000 - 3.300.000)^2 + (3.500.000 - 3.300.000)^2 + 2 \times (4.000.000 - 3.300.000)^2}{5-1}}$$
$$s = 836.660,0265$$

3.5.2 Contoh Soal Data Kelompok

Berikut adalah data hasil penjualan harian 75 toko:

Penjualan (juta Rp)	Frekuensi
10-14	15
15-19	20
20-24	25
25-29	10
30-34	5

Hitunglah rata-rata, median, modus, IQR, dan simpangan baku dari data tersebut!

Jawaban:

Langkah pertama, melengkapi kolom pada tabel yang diperlukan untuk membantu perhitungan.

a. Rata-rata

Penjualan (juta Rp)	Frekuensi (f_i)	x_i	$f_i x_i$
10-14	15	$\frac{10 + 14}{2} = 12$	$15 \times 12 = 180$
15-19	20	17	340
20-24	25	22	550
25-29	10	27	270
30-34	5	32	160
Σ	75		1500

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} = \frac{1500}{75} = 20$$

b. Median

Penjualan (juta Rp)	Frekuensi (f_i)	Frekuensi kumulatif (f_k)
10-14	15	15
15-19	20	35
20-24	25	60
25-29	10	70
30-34	5	75
Σ	75	

$$\text{Kelas median : } f_k \text{ pertama} \geq \frac{\sum_{i=1}^n f_i}{2} = \frac{75}{2} = 37,5$$

Hal ini berarti kelas mediannya ada di kelas yang $f_k = 60$ yaitu kelas ke-3.

$$Tb = 20 - 0,5 = 19,5$$

$$p = 24 - 20 + 1 = 5$$

$$F_{Me} = 35$$

$$f_{Me} = 25$$

$$Me = Tb + \left(p \times \frac{\frac{\sum_{i=1}^n f_i}{2} - F_{Me}}{f_{Me}} \right)$$

$$Me = 19,5 + \left(5 \times \frac{37,5 - 35}{25} \right)$$

$$Me = 20$$

c. Modus

Kelas Modus adalah kelas dengan frekuensi terbesar.

Penjualan (juta Rp)	Frekuensi (f_i)	Frekuensi kumulatif (f_k)
10-14	15	15
15-19	20	35
20-24	25	60
25-29	10	70
30-34	5	75
Σ	75	

$$Tb = 20 - 0,5 = 19,5$$

$$p = 24 - 20 + 1 = 5$$

$$s_1 = 25 - 20 = 5$$

$$s_2 = 25 - 10 = 15$$

$$Mo = Tb + \left(p \times \frac{s_1}{s_1 + s_2} \right)$$

$$Mo = 19,5 + \left(5 \times \frac{5}{5 + 15} \right)$$

$$Mo = 20,75$$

d. Quartil pertama (Q1)

$$\text{Kelas } Q_1 : f_k \text{ pertama} \geq \frac{\sum_{i=1}^n f_i}{4} = \frac{75}{4} = 18,75.$$

Hal ini berarti kelas quartil pertamanya ada di kelas yang $f_k = 35$ yaitu kelas ke-2

Penjualan (juta Rp)	Frekuensi (f_i)	Frekuensi kumulatif (f_k)
10-14	15	15
15-19	20	35
20-24	25	60
25-29	10	70
30-34	5	75
Σ	75	

$$Tb = 15 - 0,5 = 14,5$$

$$p = 5$$

$$F_{Q_1} = 15$$

$$f_{Q_1} = 20$$

$$Q_1 = Tb + \left(p \times \frac{\frac{\sum_{i=1}^n f_i}{4} - F_{Q_1}}{f_{Q_1}} \right)$$

$$Q_1 = 14,5 + \left(5 \times \frac{18,75 - 15}{20} \right)$$

$$Q_1 = 15,44$$

e. Quartil atas (Q3)

$$\text{Kelas } Q_3 : f_k \text{ pertama} \geq \frac{3 \times \sum_{i=1}^n f_i}{4} = \frac{3 \times 75}{4} = 56,25$$

Hal ini berarti Quartil atasnya ada di kelas yang $f_k = 60$ yaitu kelas ke-3.

Penjualan (juta Rp)	Frekuensi (f_i)	Frekuensi kumulatif (f_k)
10-14	15	15
15-19	20	35
20-24	25	60
25-29	10	70
30-34	5	75
Σ	75	

$$Tb = 20 - 0,5 = 19,5$$

$$p = 5$$

$$F_{Q_1} = 35$$

$$f_{Q_1} = 25$$

$$Q_3 = Tb + \left(p \times \frac{\frac{3 \times \sum_{i=1}^n f_i}{4} - F_{Q_3}}{f_{Q_3}} \right)$$

$$Q_3 = 19,5 + \left(5 \times \frac{56,25 - 35}{25} \right)$$

$$Q_3 = 24,15$$

f. IQR

$$\text{IQR} = Q_3 - Q_1 = 24,15 - 15,44 = 8,71$$

g. Simpangan Baku

Penjualan (juta Rp)	Frekuensi (f_i)	x_i	$(x_i - \bar{x})^2$
10-14	15	12	$(12 - 20)^2 = 64$
15-19	20	17	$(17 - 20)^2 = 9$
20-24	25	22	$(22 - 20)^2 = 4$
25-29	10	27	$(27 - 20)^2 = 49$
30-34	5	32	$(32 - 20)^2 = 144$
Σ	75		270

Anggap penjualan yang diketahui adalah sampel, maka gunakan rumus simpangan baku sampel.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{270}{75-1}} = 1,91$$

3.6 Interpretasi dan Aplikasi

3.6.1 Pemaknaan Kombinasi Ukuran Pemusatan dan Penyebaran

Interpretasi statistik yang komprehensif tidak cukup hanya bergantung pada ukuran pemusatan seperti *mean*, *median*, atau *mode*. Informasi tersebut baru memberikan gambaran mengenai titik pusat distribusi, sementara karakter variasi data memerlukan ukuran penyebaran seperti simpangan baku, varians, dan rentang. Kombinasi kedua kelompok ukuran ini menghasilkan pemahaman yang lebih utuh mengenai pola data. Sebagai contoh, dua kelompok data dapat memiliki *mean* yang sama namun simpangan baku yang

berbeda, menandakan bahwa tingkat variasinya tidak seragam. Kelompok dengan simpangan baku lebih tinggi menunjukkan distribusi yang lebih tersebar, sedangkan simpangan baku rendah mengindikasikan konsistensi nilai yang lebih kuat. Perbedaan ini penting untuk menghindari interpretasi yang menyesatkan, terutama dalam situasi ketika keputusan riset atau kebijakan didasarkan pada karakteristik distribusi data (Rahman & Abdullah, 2022).

3.6.2 Peran Ukuran Pemusatan dan Penyebaran dalam Analisis Penelitian

Dalam konteks penelitian, ukuran pemusatan dan penyebaran berfungsi sebagai fondasi untuk memahami karakteristik sampel secara deskriptif. Ukuran-ukuran ini digunakan untuk menggambarkan profil data awal, memeriksa kelayakan sampel, dan mengidentifikasi pola atau anomali yang mungkin memengaruhi analisis lebih lanjut. Selain itu, informasi mengenai penyebaran membantu peneliti menilai homogenitas data serta memeriksa apakah variasi yang ada mencerminkan karakteristik populasi atau merupakan akibat kesalahan pengukuran.

Ukuran pemusatan dan penyebaran juga digunakan untuk menguji asumsi dasar sebelum melanjutkan ke tahap analisis inferensial, misalnya asumsi normalitas, kesetaraan varians, dan independensi pengamatan. Apabila ukuran penyebaran menunjukkan ketidakwajaran, seperti simpangan baku yang sangat tinggi atau rentang yang ekstrem, maka peneliti perlu mempertimbangkan teknik transformasi data atau memilih metode analisis yang tidak sensitif terhadap pelanggaran asumsi. Oleh

karena itu, pemahaman yang tepat mengenai ukuran-ukuran dasar ini berperan penting dalam menjaga integritas proses penelitian (Hossain & Williams, 2021).

3.6.3 Aplikasi dalam Pengambilan Keputusan dan Pemilihan Metode Analisis

Penerapan ukuran pemusatan dan penyebaran tidak hanya sebatas deskripsi awal, tetapi juga memengaruhi pengambilan keputusan pada tahap lanjutan. Dalam penelitian kuantitatif, informasi tersebut digunakan untuk menentukan alat analisis yang paling sesuai. Misalnya, ketika data memiliki variasi yang tinggi atau distribusi yang sangat tidak simetris, peneliti dapat memilih teknik analisis nonparametrik sebagai alternatif metode parametris. Ukuran penyebaran, khususnya simpangan baku dan varians, juga membantu dalam menilai reliabilitas pengukuran, memeriksa stabilitas model, serta mengevaluasi kemungkinan bias dalam sampel.

Dalam konteks penerapan praktis, kombinasi ukuran tersebut memberikan nilai tambah bagi pengambil keputusan di bidang ilmu sosial, ekonomi, pendidikan, maupun kesehatan. Dengan memahami tingkat variasi, organisasi dapat mempertimbangkan risiko, kestabilan tren, dan konsistensi performa sebelum membuat keputusan strategis. Hal ini menjadikan ukuran pemusatan dan penyebaran bukan hanya sebagai konsep matematis, tetapi juga perangkat analitis yang mendukung penilaian berbasis bukti.

3.7 Latihan Soal

1. Jelaskan perbedaan makna dan fungsi ukuran pemusatan dan ukuran penyebaran dalam statistika deskriptif.
2. Dalam kondisi seperti apa median lebih tepat digunakan daripada *mean*? Sertakan contoh situasi data.
3. Jelaskan konsep kuartil dan rentang interkuartil serta alasan mengapa IQR dianggap lebih robust terhadap *outlier*.
4. Bandingkan varians dan simpangan baku dari sisi interpretasi dan kegunaan praktisnya dalam analisis data.
5. Sebuah kelas A dan kelas B memiliki rata-rata nilai yang sama. Bagaimana Anda menggunakan ukuran penyebaran untuk menilai kelas mana yang lebih konsisten performanya?

Bab 4: Analisis Data

Kuantitatif dan Kualitatif

4.1 Konsep Analisis Data

Analisis data merupakan suatu proses sistematis yang bertujuan untuk mengolah, menafsirkan, dan menarik kesimpulan dari data yang telah dikumpulkan melalui berbagai metode. Dalam konteks ekonomi dan bisnis, analisis data tidak hanya bersifat teknis, tetapi juga strategis karena hasilnya berperan langsung dalam mendukung pengambilan keputusan yang berbasis bukti (*evidence-based decision making*). Proses analisis data melibatkan tahapan mulai dari pembersihan (*data cleaning*), transformasi, eksplorasi pola, hingga pemodelan dan validasi, dengan tujuan menghasilkan informasi yang relevan, akurat, dan dapat ditindaklanjuti. Penggunaan teknik analitik memungkinkan para pelaku bisnis dan ekonom untuk memahami dinamika pasar, memprediksi tren masa depan, serta mengidentifikasi peluang atau risiko yang tersembunyi di balik data mentah.

Dalam praktiknya, analisis data di bidang ekonomi dapat dilakukan melalui pendekatan deskriptif, diagnostik, prediktif, maupun preskriptif. Analisis deskriptif digunakan untuk memahami situasi saat ini melalui statistik dasar seperti rata-rata, median, dan distribusi. Sementara itu, analisis diagnostik mencari penyebab di

balik fenomena yang teridentifikasi. Untuk proyeksi masa depan, analisis prediktif mengandalkan model statistik atau algoritma pembelajaran mesin yang memanfaatkan data historis guna memprediksi variabel-variabel penting, seperti permintaan produk atau fluktuasi harga. Adapun analisis preskriptif berfokus pada rekomendasi tindakan optimal berdasarkan hasil prediksi dan skenario tertentu. Penerapan teknik-teknik ini semakin berkembang seiring dengan integrasi *big data* dan *machine learning* dalam analisis ekonomi modern (Ala'raj et al., 2021).

Kemampuan untuk menafsirkan data secara tepat juga memerlukan pemahaman konteks ekonomi yang melatarbelakangi fenomena yang dianalisis. Oleh karena itu, analisis data tidak dapat dipisahkan dari keterampilan berpikir kritis, pemodelan ekonomi, dan interpretasi hasil statistik. Kesalahan dalam menafsirkan hasil analisis dapat berakibat pada kebijakan atau strategi bisnis yang keliru. Di sinilah pentingnya validitas data dan pemilihan metode analisis yang tepat. Untuk memastikan keandalan hasil, analisis harus dilakukan dengan memperhatikan prinsip transparansi, replikasi, serta pengujian asumsi statistik yang relevan (Zhang & Chen, 2020). Dengan demikian, analisis data dalam ekonomi dan bisnis bukan hanya instrumen teknis, melainkan pilar utama dalam membentuk keputusan yang rasional, adaptif, dan kompetitif di tengah kompleksitas pasar global.

4.2 Analisis Data Kuantitatif

Analisis data kuantitatif merupakan pendekatan yang mengandalkan angka dan teknik statistik untuk mengidentifikasi pola, menguji hubungan antarvariabel, serta menarik kesimpulan berdasarkan bukti empiris. Pendekatan ini banyak digunakan dalam penelitian sosial, kesehatan, pendidikan, dan ekonomi karena mampu menghasilkan temuan yang objektif dan terukur. Data kuantitatif biasanya diperoleh melalui survei, eksperimen, atau data administratif yang telah terstandar sehingga dapat dianalisis secara sistematis. Analisis kuantitatif tidak hanya membantu menggambarkan fenomena, tetapi juga memprediksi kecenderungan serta menguji teori yang telah dirumuskan sebelumnya (Field, 2020).

Metode yang digunakan dalam analisis kuantitatif bervariasi sesuai tujuan penelitian, mulai dari teknik dasar hingga statistik inferensial yang kompleks.

4.2.1 Statistik deskriptif

Statistik deskriptif digunakan untuk merangkum dan menyajikan data dalam bentuk yang mudah dipahami. Teknik ini meliputi pengukuran pemusatan seperti *mean*, median, dan modus, serta ukuran penyebaran seperti rentang, varians, dan standar deviasi. Grafik dan tabel juga digunakan untuk memvisualisasikan distribusi data, membantu peneliti memahami karakteristik umum sebelum melanjutkan ke analisis lebih lanjut.

Penggunaan statistik deskriptif sangat penting sebagai langkah awal karena memberikan gambaran awal terkait pola umum, potensi

outlier, dan distribusi variabel yang dapat memengaruhi pemilihan metode analisis lanjutan (Gravetter, 2021).

4.2.2 Regresi dan korelasi

Regresi dan korelasi digunakan untuk mengidentifikasi hubungan antara dua atau lebih variabel. Korelasi mengukur kekuatan dan arah hubungan linier, sedangkan regresi memungkinkan prediksi nilai variabel dependen berdasarkan variabel independen. Model regresi linear merupakan bentuk yang paling umum, namun model lainnya seperti regresi logistik atau regresi multivariat digunakan ketika variabel dependen bersifat kategorikal atau lebih kompleks.

Analisis ini membantu menjelaskan pengaruh relatif suatu variabel dan memberikan dasar bagi pengambilan keputusan berbasis data, seperti memprediksi tingkat partisipasi, faktor penentu risiko, atau hasil kebijakan.

4.2.3 Uji hipotesis

Uji hipotesis digunakan untuk menentukan apakah perbedaan atau hubungan yang ditemukan dalam data terjadi secara kebetulan atau signifikan secara statistik. Prosedur ini melibatkan formulasi hipotesis nol (*null hypothesis*) dan hipotesis alternatif, penentuan tingkat signifikansi, serta penggunaan alat uji seperti *t-test*, *chi-square*, atau ANOVA, bergantung pada jenis dan distribusi data.

Melalui uji hipotesis, peneliti dapat menilai validitas temuan dan memastikan bahwa kesimpulan yang ditarik memiliki dasar statistik yang kuat. Dengan demikian, analisis kuantitatif mendukung penelitian yang akurat, replikatif, dan konsisten.

4.3 Analisis Data Kualitatif

4.3.1 Karakteristik dan Tujuan Analisis Kualitatif

Analisis data kualitatif merupakan pendekatan yang digunakan untuk memahami makna, pengalaman, atau fenomena yang tidak dapat diukur dengan angka. Fokus utama analisis ini adalah interpretasi mendalam terhadap data non-numerik yang diperoleh dari wawancara, observasi lapangan, catatan lapangan, maupun dokumen tertulis. Berbeda dengan pendekatan kuantitatif yang menekankan pada pengukuran dan generalisasi, pendekatan kualitatif bertujuan mengeksplorasi dinamika subjektif dan konteks sosial budaya dari suatu fenomena. Dalam penelitian sosial atau bisnis, analisis ini membantu mengungkap pola perilaku, persepsi, atau motivasi di balik data yang bersifat naratif (Creswell & Poth, 2021).

4.3.2 Teknik-Teknik Utama dalam Analisis Kualitatif

Beberapa teknik umum yang digunakan dalam analisis kualitatif meliputi *coding*, kategorisasi, analisis tema, dan interpretasi naratif. *Coding* merupakan proses identifikasi dan pelabelan unit-unit informasi penting dalam data, seperti kutipan atau kalimat bermakna. Langkah ini dilanjutkan dengan pengelompokan ke dalam kategori yang relevan, sehingga mempermudah penemuan pola atau hubungan antartema. Analisis tema kemudian digunakan untuk menggali ide sentral atau makna dominan yang muncul dari berbagai kategori tersebut. Sementara itu, interpretasi naratif melibatkan rekonstruksi pengalaman atau cerita

partisipan secara holistik, untuk memahami makna dalam konteks personal dan sosial. Proses analisis bersifat iteratif dan reflektif, menuntut peneliti untuk terus kembali ke data asli guna memastikan kedalaman dan keabsahan temuan (Nowell et al., 2017).

4.3.3 Validitas dan Kontribusi Analisis Kualitatif

Meskipun tidak menghasilkan generalisasi statistik, analisis data kualitatif memiliki kekuatan dalam mengungkap dimensi makna yang kompleks dan mendalam. Validitas dalam analisis ini dicapai melalui triangulasi data, *member checking*, serta audit trail yang transparan. Hasil analisis kualitatif sangat berharga dalam pengembangan teori, pemahaman budaya organisasi, pengembangan kebijakan berbasis konteks, maupun dalam penyusunan strategi bisnis yang responsif terhadap dinamika sosial.

4.4 Perbandingan Analisis Kuantitatif dan Kualitatif

Dalam penelitian ilmiah, analisis data merupakan komponen penting yang menentukan arah interpretasi hasil. Dua pendekatan utama yang digunakan adalah analisis kuantitatif dan kualitatif, yang memiliki perbedaan mendasar dalam paradigma, metode, serta keluaran data. Pemilihan pendekatan sangat bergantung pada rumusan masalah, tujuan penelitian, dan jenis data yang dikumpulkan. Keduanya juga dapat digabungkan dalam desain *mixed-method* untuk saling melengkapi keterbatasan masing-masing pendekatan.

4.4.1 Analisis kuantitatif: pengukuran dan generalisasi

Analisis kuantitatif berakar pada paradigma positivistik yang menekankan objektivitas, validitas pengukuran, serta kemampuan untuk digeneralisasi ke populasi yang lebih luas. Pendekatan ini menggunakan data numerik dan teknik statistik untuk menguji hipotesis, menemukan pola, serta mengukur hubungan antar variabel.

Data kuantitatif biasanya dikumpulkan melalui instrumen terstandar seperti kuesioner, survei, atau pengukuran laboratorium. Hasilnya disajikan dalam bentuk tabel, grafik, dan ukuran statistik seperti rerata, simpangan baku, regresi, atau korelasi. Keunggulan utama analisis ini adalah kemampuannya menghasilkan temuan yang replikatif dan efisien dalam menjawab pertanyaan berskala besar (Bryman, 2021).

4.4.2 Analisis kualitatif: makna dan kedalaman

Sebaliknya, analisis kualitatif berangkat dari paradigma konstruktivistik yang menekankan makna subjektif, konteks sosial, dan pemahaman mendalam terhadap pengalaman manusia. Data kualitatif berupa narasi, deskripsi, atau hasil observasi yang dianalisis secara tematik, induktif, dan interpretatif.

Metode yang digunakan mencakup wawancara mendalam, diskusi kelompok terarah (*focus group*), dan studi kasus. Analisis dilakukan dengan mengekstrak pola, makna, dan interpretasi dari konteks yang kompleks dan dinamis. Pendekatan ini sangat berguna ketika peneliti ingin memahami proses, motivasi, atau dinamika

sosial yang tidak dapat diungkap melalui angka semata (Creswell & Poth, 2020).

4.4.3 Pendekatan *mixed-method*: integrasi dan komplementaritas

Penggunaan kombinasi kuantitatif dan kualitatif dalam *mixed-method research* semakin berkembang, terutama dalam studi multidimensi yang membutuhkan data numerik sekaligus pemahaman kontekstual. Dalam pendekatan ini, data kualitatif dapat digunakan untuk menjelaskan temuan statistik, atau sebaliknya, hasil kualitatif dapat divalidasi melalui survei kuantitatif. Integrasi kedua pendekatan memungkinkan peneliti mendapatkan gambaran yang lebih utuh dan komprehensif, serta menjembatani kesenjangan antara hasil statistik dan realitas sosial.

4.5 Latihan Soal

1. Jelaskan perbedaan utama antara analisis data kuantitatif dan kualitatif.
2. Mengapa analisis kuantitatif penting dalam pengambilan keputusan bisnis?
3. Uraikan langkah-langkah umum dalam melakukan analisis kualitatif.
4. Berikan contoh situasi di mana kombinasi analisis kuantitatif dan kualitatif diperlukan.
5. Jelaskan peran kategori dan tema dalam analisis kualitatif.

Bab 5: Probabilitas dan Distribusi

5.1 Konsep Dasar Probabilitas

5.1.1 Pengertian Probabilitas

Probabilitas merupakan konsep dasar dalam matematika yang digunakan untuk mengukur tingkat kemungkinan terjadinya suatu peristiwa dalam suatu percobaan atau proses acak. Secara kuantitatif, probabilitas dinyatakan dalam nilai numerik antara 0 dan 1, di mana nilai 0 menunjukkan peristiwa yang mustahil terjadi, sedangkan nilai 1 menunjukkan peristiwa yang pasti terjadi. Konsep ini memungkinkan ketidakpastian yang melekat pada fenomena acak untuk dianalisis secara sistematis dan rasional, sehingga probabilitas menjadi fondasi penting dalam pengambilan keputusan berbasis data dan analisis risiko (Ross, 2020).

Dalam kerangka teoritis, probabilitas tidak hanya dipahami sebagai peluang intuitif, tetapi sebagai ukuran matematis yang mengikuti aturan dan aksioma tertentu. Pendekatan ini memastikan bahwa penilaian terhadap kemungkinan suatu peristiwa dilakukan secara konsisten dan dapat dipertanggungjawabkan. Probabilitas digunakan untuk memodelkan berbagai fenomena, mulai dari kejadian sederhana seperti pelemparan koin hingga sistem kompleks yang melibatkan banyak variabel acak. Oleh karena itu, pemahaman

yang baik mengenai pengertian probabilitas menjadi langkah awal yang krusial sebelum mempelajari konsep lanjutan dalam teori peluang (Grimmett & Stirzaker, 2020).

Selain berperan dalam matematika murni, probabilitas memiliki aplikasi luas dalam berbagai bidang ilmu, seperti statistika, ekonomi, teknik, dan ilmu komputer. Dalam konteks ini, probabilitas membantu menjelaskan pola ketidakpastian dan variasi yang muncul dalam data nyata. Dengan menggunakan probabilitas, suatu peristiwa tidak hanya dinilai berdasarkan kepastian atau ketidakpastian semata, tetapi berdasarkan tingkat kemungkinan yang dapat diukur dan dibandingkan. Hal ini menjadikan probabilitas sebagai alat analitis yang esensial dalam memahami dan mengelola ketidakpastian secara ilmiah (Blitzstein & Hwang, 2019).

5.1.2 Ruang Sampel dan Kejadian

Ruang sampel merupakan himpunan yang memuat seluruh kemungkinan hasil yang dapat terjadi dari suatu percobaan acak. Setiap elemen dalam ruang sampel merepresentasikan satu hasil yang mungkin muncul dan bersifat saling lepas satu sama lain. Penentuan ruang sampel secara tepat merupakan langkah awal yang sangat penting dalam analisis probabilitas, karena seluruh perhitungan peluang suatu peristiwa didasarkan pada struktur ruang sampel tersebut. Ruang sampel dapat bersifat terbatas atau tak terbatas, bergantung pada jenis percobaan yang dianalisis (Ross, 2020).

Kejadian didefinisikan sebagai himpunan bagian dari ruang sampel yang merepresentasikan hasil-hasil tertentu yang menjadi

fokus perhatian. Suatu kejadian dapat terdiri atas satu hasil tunggal atau beberapa hasil sekaligus, tergantung pada kriteria yang ditetapkan. Probabilitas suatu kejadian ditentukan berdasarkan proporsi atau ukuran kejadian tersebut relatif terhadap keseluruhan ruang sampel. Dengan demikian, hubungan antara ruang sampel dan kejadian bersifat hierarkis, di mana ruang sampel menjadi kerangka utama dan kejadian merupakan bagian spesifik di dalamnya (Grimmett & Stirzaker, 2020).

Pemahaman mengenai ruang sampel dan kejadian memungkinkan analisis probabilitas dilakukan secara terstruktur dan logis. Kesalahan dalam mendefinisikan ruang sampel atau kejadian dapat menyebabkan interpretasi peluang yang keliru. Oleh karena itu, ketelitian dalam mengidentifikasi seluruh kemungkinan hasil dan mengelompokkan kejadian secara tepat menjadi kunci dalam penerapan konsep probabilitas. Dalam praktiknya, konsep ruang sampel dan kejadian juga menjadi dasar bagi pengembangan model probabilistik yang lebih kompleks, seperti distribusi peluang dan variabel acak, yang banyak digunakan dalam analisis kuantitatif modern (Blitzstein & Hwang, 2019).

5.2 Aturan dan Hukum Probabilitas

5.2.1 Aturan Penjumlahan

Aturan penjumlahan digunakan untuk menghitung probabilitas gabungan dari dua kejadian atau lebih.

5.2.2 Aturan Perkalian

Aturan perkalian diterapkan untuk menentukan probabilitas kejadian yang terjadi secara bersamaan.

5.2.3 Probabilitas Bersyarat dan Teorema Bayes

Probabilitas bersyarat merupakan konsep probabilitas yang memperhitungkan pengaruh suatu kejadian terhadap kejadian lain. Probabilitas kejadian A dengan syarat kejadian B telah terjadi dinyatakan sebagai $P(A|B)$, yang dihitung dengan rumus:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ dengan } P(B) > 0$$

Konsep ini penting dalam analisis yang melibatkan ketergantungan antar kejadian, seperti dalam diagnosa medis, analisis risiko, dan sistem rekomendasi. Probabilitas bersyarat juga menjadi dasar dari Teorema Bayes, yang memungkinkan pembaruan probabilitas berdasarkan informasi baru. Teorema Bayes dinyatakan sebagai:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

di mana:

- a. $P(A|B)$ adalah probabilitas posterior,
- b. $P(B|A)$ adalah *likelihood*,
- c. $P(A)$ adalah probabilitas *prior*,

d. $P(B)$ adalah probabilitas marginal.

Teorema Bayes banyak diterapkan dalam statistika Bayesian, pembelajaran mesin, pengambilan keputusan, dan analisis prediktif (Gelman et al., 2020).

5.3 Variabel Acak dan Distribusi Probabilitas

5.3.1 Variabel Acak Diskrit dan Kontinu

Variabel acak merupakan konsep dasar dalam teori probabilitas yang digunakan untuk merepresentasikan hasil numerik dari suatu proses acak. Variabel ini berfungsi sebagai penghubung antara fenomena acak di dunia nyata dengan model matematis yang memungkinkan analisis kuantitatif. Berdasarkan sifat nilai yang mungkin diambalnya, variabel acak diklasifikasikan menjadi variabel acak diskrit dan variabel acak kontinu. Klasifikasi ini penting karena menentukan pendekatan matematis dan teknik analisis yang digunakan dalam perhitungan probabilitas dan inferensi statistik.

Variabel acak diskrit adalah variabel yang memiliki himpunan nilai yang dapat dihitung secara terpisah dan berhingga atau tak hingga terhitung. Nilai-nilai tersebut biasanya diperoleh dari proses pencacahan, seperti jumlah kejadian, banyaknya objek, atau frekuensi suatu peristiwa. Setiap nilai yang mungkin muncul dapat didaftarkan secara eksplisit, sehingga peluang kemunculannya dapat ditentukan secara langsung. Karakteristik ini menjadikan variabel acak diskrit relatif lebih mudah divisualisasikan dan dianalisis,

terutama pada tahap awal pembelajaran probabilitas. Dalam penerapannya, variabel acak diskrit banyak digunakan untuk memodelkan fenomena yang bersifat terhitung dan tidak berkesinambungan.

Sebaliknya, variabel acak kontinu memiliki nilai yang berada pada suatu selang bilangan real dan tidak dapat dihitung satu per satu. Variabel ini biasanya muncul dari proses pengukuran, seperti waktu, jarak, berat, atau suhu, yang secara teoretis dapat mengambil nilai apa pun dalam suatu interval tertentu. Karena sifatnya yang kontinu, probabilitas variabel acak kontinu tidak ditentukan pada satu nilai tunggal, melainkan pada rentang nilai. Hal ini menuntut pendekatan matematis yang berbeda, terutama dalam penggunaan fungsi dan integral untuk menentukan peluang. Pemahaman perbedaan antara variabel acak diskrit dan kontinu menjadi landasan penting dalam membangun model probabilitas yang sesuai dengan karakteristik data yang dianalisis (Ross, 2020).

5.3.2 Fungsi Distribusi Probabilitas

Fungsi distribusi probabilitas merupakan alat matematis yang digunakan untuk menggambarkan bagaimana peluang terdistribusi pada nilai-nilai yang mungkin dari suatu variabel acak. Distribusi ini memberikan kerangka formal untuk memahami pola ketidakpastian dan variasi yang muncul dalam fenomena acak. Melalui distribusi probabilitas, hubungan antara nilai variabel acak dan peluang kemunculannya dapat dinyatakan secara sistematis dan konsisten.

Pada variabel acak diskrit, distribusi probabilitas dinyatakan melalui fungsi massa probabilitas (*probability mass function*), yang

menetapkan peluang untuk setiap nilai diskrit yang mungkin. Jumlah seluruh peluang harus bernilai satu, mencerminkan kepastian bahwa salah satu dari nilai tersebut akan terjadi. Distribusi ini memungkinkan perhitungan berbagai ukuran penting, seperti nilai harapan dan varians, yang digunakan untuk menggambarkan kecenderungan pusat dan tingkat penyebaran variabel acak. Dengan demikian, fungsi distribusi probabilitas tidak hanya memberikan informasi tentang peluang individual, tetapi juga tentang karakteristik keseluruhan dari fenomena acak yang dimodelkan.

Untuk variabel acak kontinu, distribusi probabilitas dinyatakan melalui fungsi kerapatan probabilitas (*probability density function*). Berbeda dengan kasus diskrit, nilai fungsi ini tidak secara langsung menyatakan peluang pada satu titik, melainkan merepresentasikan kepadatan peluang pada suatu interval. Probabilitas diperoleh dengan menghitung luas daerah di bawah kurva fungsi kerapatan pada rentang nilai tertentu. Pendekatan ini mencerminkan sifat kontinu variabel yang tidak memungkinkan penetapan peluang pada satu nilai tunggal. Selain fungsi kerapatan, fungsi distribusi kumulatif (*cumulative distribution function*) digunakan untuk menyatakan peluang bahwa variabel acak mengambil nilai kurang dari atau sama dengan suatu nilai tertentu.

Distribusi probabilitas memegang peranan sentral dalam analisis statistik karena menjadi dasar bagi berbagai metode inferensi dan pengambilan keputusan di bawah ketidakpastian. Dengan memahami bentuk dan sifat distribusi probabilitas, analisis terhadap data acak dapat dilakukan secara lebih terstruktur dan

bermakna. Oleh karena itu, penguasaan konsep fungsi distribusi probabilitas merupakan prasyarat penting dalam memahami teori probabilitas dan aplikasinya dalam berbagai bidang keilmuan (Devore, 2021).

5.3.3 Nilai Harapan dan Varians

Nilai harapan (*expected value*) dari suatu variabel acak merupakan ukuran pusat yang menggambarkan rata-rata tertimbang dari semua nilai yang mungkin, dengan bobot sesuai probabilitasnya. Untuk variabel acak diskrit X dengan fungsi massa probabilitas $p(x)$, nilai harapan didefinisikan sebagai:

$$E[X] = \sum_x x \cdot p(x)$$

Sedangkan untuk variabel acak kontinu X dengan fungsi kerapatan probabilitas $f(x)$, nilai harapan didefinisikan sebagai:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Varians (σ^2) mengukur sebaran atau dispersi nilai variabel acak di sekitar nilai harapannya, dihitung sebagai:

$$\text{Var}(X) = E[(X - E[X])^2]$$

Nilai harapan dan varians merupakan parameter penting yang menggambarkan karakteristik distribusi probabilitas, seperti kecenderungan pusat dan tingkat variabilitas (Casella & Berger, 2021).

5.4 Distribusi Probabilitas Diskrit

Distribusi probabilitas diskrit merupakan konsep fundamental dalam statistika dan probabilitas yang digunakan untuk memodelkan peubah acak yang memiliki nilai terhingga atau dapat dihitung satu per satu. Distribusi ini banyak diterapkan dalam berbagai bidang, seperti sains, teknik, ekonomi, dan ilmu sosial, untuk menganalisis fenomena yang hasilnya berupa hitungan kejadian. Karakteristik utama distribusi probabilitas diskrit adalah adanya *Probability Mass Function* (PMF) yang menyatakan peluang setiap nilai peubah acak, dengan jumlah seluruh peluang sama dengan satu. Pemahaman terhadap distribusi diskrit memungkinkan analisis yang lebih sistematis terhadap ketidakpastian dan variasi hasil dalam suatu proses acak.

Dalam praktik analisis data, distribusi probabilitas diskrit sering digunakan untuk menyederhanakan realitas yang kompleks menjadi model matematis yang dapat dianalisis secara kuantitatif. Pemilihan jenis distribusi yang tepat bergantung pada sifat proses yang diamati, asumsi yang mendasari percobaan, serta tujuan analisis. Di antara berbagai distribusi diskrit yang umum digunakan, distribusi binomial dan distribusi Poisson menempati posisi penting karena kesederhanaan asumsi dan luasnya aplikasi dalam pemodelan kejadian acak.

5.4.1 Distribusi Binomial

Distribusi binomial digunakan untuk memodelkan jumlah keberhasilan yang terjadi dalam sejumlah percobaan yang bersifat independen dan identik, yang dikenal sebagai percobaan Bernoulli. Setiap percobaan hanya memiliki dua kemungkinan hasil, yaitu keberhasilan atau kegagalan, dengan peluang keberhasilan yang konstan pada setiap percobaan. Distribusi ini sangat relevan ketika fokus analisis terletak pada banyaknya keberhasilan yang dicapai dalam jumlah percobaan tertentu, bukan pada urutan terjadinya hasil tersebut.

Secara konseptual, distribusi binomial memberikan kerangka matematis untuk mengukur variasi hasil yang mungkin muncul akibat sifat acak dari percobaan. Parameter utama dalam distribusi ini adalah jumlah percobaan dan peluang keberhasilan, yang bersama-sama menentukan bentuk distribusi peluangnya. Ketika jumlah percobaan relatif kecil, distribusi binomial cenderung asimetris, sedangkan pada jumlah percobaan yang lebih besar, bentuk distribusinya dapat mendekati distribusi kontinu tertentu. Karakteristik ini menjadikan distribusi binomial fleksibel untuk berbagai skenario pemodelan kejadian diskrit.

Dalam aplikasi praktis, distribusi binomial banyak digunakan untuk analisis kualitas, evaluasi keberhasilan suatu program, serta pemodelan kejadian yang memiliki hasil biner. Kejelasan asumsi dan interpretasi parameter membuat distribusi ini mudah dipahami dan diterapkan. Dengan demikian, distribusi binomial berperan

penting sebagai alat analisis dasar dalam statistika inferensial dan pengambilan keputusan berbasis probabilitas (Ross, 2021).

5.4.2 Distribusi Poisson

Distribusi Poisson digunakan untuk memodelkan frekuensi kejadian suatu peristiwa dalam interval waktu atau ruang tertentu, dengan asumsi bahwa kejadian tersebut terjadi secara acak dan independen. Distribusi ini sangat sesuai untuk situasi di mana kejadian yang diamati relatif jarang, tetapi jumlah peluang terjadinya kejadian sangat besar. Parameter utama distribusi Poisson adalah laju kejadian rata-rata, yang merepresentasikan ekspektasi jumlah kejadian dalam satu interval pengamatan.

Keunggulan distribusi Poisson terletak pada kemampuannya menyederhanakan pemodelan kejadian acak yang kompleks tanpa memerlukan informasi rinci tentang setiap peluang kejadian individual. Distribusi ini sering digunakan untuk menganalisis kejadian seperti jumlah kedatangan, cacat produksi, atau insiden tertentu dalam rentang waktu tertentu. Bentuk distribusinya ditentukan sepenuhnya oleh satu parameter, sehingga interpretasi hasil analisis menjadi lebih langsung dan intuitif.

Selain itu, distribusi Poisson memiliki keterkaitan erat dengan distribusi binomial, terutama ketika jumlah percobaan sangat besar dan peluang keberhasilan pada setiap percobaan sangat kecil. Dalam kondisi tersebut, distribusi Poisson dapat digunakan sebagai pendekatan untuk menyederhanakan perhitungan probabilitas. Hubungan ini menunjukkan bahwa distribusi Poisson tidak berdiri sendiri, melainkan menjadi bagian dari kerangka distribusi

probabilitas diskrit yang saling terkait. Dengan penerapan yang tepat, distribusi Poisson menjadi alat yang efektif dalam analisis fenomena acak berbasis frekuensi kejadian (Blitzstein & Hwang, 2019).

5.4.3 Distribusi Geometrik dan Hipergeometrik

1. Distribusi Geometrik

Distribusi geometrik memodelkan jumlah percobaan Bernoulli yang diperlukan hingga mencapai keberhasilan pertama. Jika p adalah probabilitas keberhasilan pada setiap percobaan, maka probabilitas bahwa keberhasilan pertama terjadi pada percobaan ke- k adalah:

$$P(X=k)=(1-p)^{k-1} \cdot p$$

Distribusi ini digunakan untuk memodelkan waktu tunggu dalam proses acak, seperti analisis reliabilitas dan teori antrian (Wackerly et al., 2018).

2. Distribusi Hipergeometrik

Distribusi hipergeometrik digunakan untuk memodelkan jumlah keberhasilan dalam pengambilan sampel tanpa pengembalian dari populasi terbatas. Jika terdapat N objek dengan K berlabel "sukses", dan n objek diambil tanpa pengembalian, maka probabilitas mendapatkan k sukses adalah:

$$P(X=k)=\frac{\binom{kK}{k} \binom{n-k}{N-k}}{\binom{n}{N}}$$

Distribusi ini sering digunakan dalam pengendalian kualitas, audit sampling, dan analisis survei (Montgomery & Runger, 2018).

5.5 Distribusi Probabilitas Kontinyu

Distribusi probabilitas kontinyu memiliki peran sentral dalam analisis statistik karena digunakan untuk memodelkan variabel acak yang dapat mengambil nilai pada rentang kontinyu. Berbeda dengan distribusi diskret, distribusi kontinyu menggambarkan peluang kejadian melalui fungsi kerapatan probabilitas (*probability density function*), di mana peluang suatu nilai spesifik dinyatakan melalui luasan pada interval tertentu. Dalam berbagai bidang seperti sains, teknik, ekonomi, dan kesehatan, distribusi probabilitas kontinyu menjadi dasar dalam pengambilan keputusan berbasis data, pengujian hipotesis, serta pemodelan fenomena alam maupun sosial. Dua distribusi kontinyu yang paling sering digunakan dan memiliki karakteristik fundamental adalah distribusi normal dan distribusi eksponensial.

5.5.1 Distribusi Normal

Distribusi normal merupakan distribusi probabilitas kontinyu yang paling luas penerapannya dalam analisis statistik. Distribusi ini memiliki bentuk kurva lonceng yang simetris, dengan nilai tengah sebagai titik pusat distribusi. Karakteristik utama distribusi normal terletak pada dua parameter, yaitu rata-rata dan simpangan baku, yang secara bersama-sama menentukan lokasi dan penyebaran data. Sifat simetri distribusi normal menyebabkan nilai rata-rata, median, dan modus berada pada titik yang sama, sehingga distribusi ini sering digunakan sebagai representasi kondisi ideal atau alami dari suatu fenomena.

Distribusi normal banyak digunakan karena kemampuannya dalam merepresentasikan variasi alami yang muncul dari akumulasi berbagai faktor kecil dan independen. Dalam praktik statistik, asumsi kenormalan sering menjadi dasar dalam berbagai metode analisis inferensial, seperti estimasi parameter dan pengujian hipotesis. Selain itu, distribusi normal memiliki sifat matematis yang memudahkan analisis, termasuk keterkaitannya dengan *central limit theorem*, yang menjelaskan kecenderungan distribusi rata-rata sampel mendekati distribusi normal seiring bertambahnya ukuran sampel. Sifat ini menjadikan distribusi normal sebagai fondasi teoritis dalam banyak prosedur statistik modern (Ross, 2020).

Dalam konteks penerapan, distribusi normal digunakan untuk memodelkan berbagai fenomena, seperti pengukuran biologis, kesalahan pengukuran, dan variabel alam yang berfluktuasi secara acak di sekitar nilai rata-rata. Pemahaman terhadap distribusi normal memungkinkan analisis untuk menilai penyebaran data, mengidentifikasi nilai ekstrem, serta melakukan prediksi berbasis probabilitas. Oleh karena itu, distribusi normal tidak hanya penting secara teoritis, tetapi juga sangat relevan dalam analisis data empiris di berbagai disiplin ilmu (Devore, 2019).

5.5.2 Distribusi Eksponensial

Distribusi eksponensial merupakan distribusi probabilitas kontinu yang digunakan untuk memodelkan waktu tunggu hingga terjadinya suatu peristiwa tertentu. Distribusi ini sering diterapkan pada fenomena yang berkaitan dengan durasi atau selang waktu, seperti waktu hingga kegagalan sistem, waktu kedatangan

pelanggan, atau waktu antarperistiwa dalam suatu proses acak. Karakteristik utama distribusi eksponensial adalah sifatnya yang tidak bergantung pada waktu sebelumnya, yang dikenal sebagai sifat *memoryless*. Sifat ini menyatakan bahwa peluang terjadinya suatu peristiwa pada masa mendatang tidak dipengaruhi oleh lamanya waktu yang telah berlalu.

Distribusi eksponensial ditentukan oleh satu parameter laju, yang merepresentasikan tingkat kejadian per satuan waktu. Nilai parameter ini berpengaruh langsung terhadap bentuk distribusi, di mana laju yang lebih tinggi menunjukkan peluang kejadian yang lebih cepat. Dalam analisis statistik, distribusi eksponensial sering digunakan sebagai model sederhana namun efektif untuk sistem yang diasumsikan memiliki tingkat kejadian konstan. Kesederhanaan model ini membuat distribusi eksponensial banyak digunakan dalam analisis keandalan, teori antrian, dan pemodelan proses stokastik (Wasserman, 2021).

Penerapan distribusi eksponensial memberikan kerangka probabilistik untuk memahami dan memprediksi durasi waktu tunggu dalam berbagai sistem. Dengan menggunakan distribusi ini, analis dapat memperkirakan peluang suatu peristiwa terjadi dalam interval waktu tertentu serta mengevaluasi kinerja sistem berbasis waktu. Meskipun memiliki asumsi yang relatif sederhana, distribusi eksponensial tetap menjadi alat penting dalam analisis statistik karena kemampuannya dalam menggambarkan fenomena waktu tunggu secara praktis dan konsisten (Ross, 2020).

5.5.3 Distribusi Uniform Kontinu

Distribusi uniform kontinu menggambarkan situasi di mana semua nilai dalam interval $[a,b]$ memiliki peluang yang sama untuk terjadi. Fungsi kerapatan probabilitasnya adalah:

$$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$$

Distribusi ini digunakan untuk memodelkan ketidakpastian yang merata, seperti dalam simulasi Monte Carlo, pembangkitan bilangan acak, dan analisis risiko (Rice, 2019).

5.6 Latihan Soal

1. Jelaskan konsep dasar probabilitas dan ruang sampel.
2. Uraikan perbedaan antara variabel acak diskrit dan kontinu.
3. Bagaimana aturan penjumlahan dan perkalian digunakan dalam probabilitas?
4. Jelaskan karakteristik distribusi binomial dan Poisson.
5. Analisis peran distribusi normal dalam analisis statistik.
6. Jelaskan konsep probabilitas bersyarat dan berikan contoh penerapannya dalam kehidupan sehari-hari.
7. Hitung nilai harapan dan varians untuk variabel acak dengan distribusi binomial dengan parameter $n=10$ dan $p=0.5$.
8. Bandingkan distribusi binomial dengan distribusi hipergeometrik dalam konteks pengambilan sampel.
9. Jelaskan peran distribusi uniform kontinu dalam simulasi stokastik

10. Berikan contoh penerapan Teorema Bayes dalam pengambilan keputusan di bawah ketidakpastian.

Bab 6. Statistika Inferensial

6.1 Konsep Dasar Statistika Inferensial

6.1.1 Pengertian Statistika Inferensial

Statistika inferensial merupakan cabang statistika yang berfokus pada proses penarikan kesimpulan, estimasi, atau generalisasi mengenai karakteristik suatu populasi berdasarkan informasi yang diperoleh dari sampel. Berbeda dengan statistika deskriptif yang hanya menggambarkan data apa adanya, statistika inferensial melampaui data yang diamati dengan menggunakan prinsip probabilitas dan teori peluang untuk membuat pernyataan yang bersifat umum. Melalui pendekatan ini, ketidakpastian yang melekat pada data sampel dapat dikelola secara sistematis sehingga kesimpulan yang dihasilkan memiliki dasar matematis yang kuat dan dapat dipertanggungjawabkan (Casella & Berger, 2021).

Secara konseptual, statistika inferensial bertumpu pada asumsi bahwa sampel yang diambil mengandung informasi representatif tentang populasi asalnya. Informasi tersebut kemudian dianalisis menggunakan teknik inferensial, seperti estimasi parameter dan pengujian hipotesis, untuk memperoleh gambaran tentang kondisi populasi yang sebenarnya. Proses ini melibatkan pengukuran tingkat ketidakpastian, yang umumnya dinyatakan dalam bentuk peluang, interval kepercayaan, atau tingkat signifikansi. Dengan demikian, statistika inferensial tidak hanya

menghasilkan kesimpulan, tetapi juga menyertakan ukuran keyakinan terhadap kesimpulan tersebut (Wackerly et al., 2020).

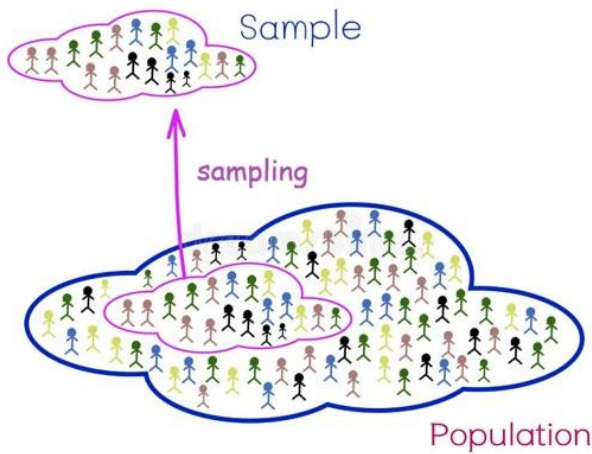
Dalam praktik keilmuan dan profesional, statistika inferensial memiliki peran penting karena pengamatan terhadap seluruh anggota populasi sering kali tidak memungkinkan akibat keterbatasan waktu, biaya, dan sumber daya. Oleh karena itu, penggunaan sampel yang dianalisis secara inferensial menjadi solusi metodologis untuk memperoleh informasi yang relevan tanpa harus melakukan pengamatan menyeluruh. Pendekatan ini banyak digunakan dalam berbagai bidang, seperti kesehatan, ekonomi, pendidikan, dan ilmu sosial, untuk mendukung pengambilan keputusan berbasis data. Keandalan statistika inferensial sangat bergantung pada ketepatan metode analisis dan kesesuaian asumsi yang digunakan terhadap karakteristik data (Devore, 2023).

6.1.2 Peran Sampel dan Populasi

Hubungan antara sampel dan populasi merupakan fondasi utama dalam statistika inferensial. Populasi didefinisikan sebagai keseluruhan objek, individu, atau kejadian yang menjadi sasaran analisis, sedangkan sampel merupakan sebagian dari populasi yang dipilih untuk diamati dan dianalisis. Proses pengambilan sampel dari populasi disebut *sampling*.

Dalam konteks inferensi statistik, sampel berfungsi sebagai wakil populasi, sehingga kualitas inferensi sangat ditentukan oleh sejauh mana sampel tersebut mampu mencerminkan karakteristik populasi secara akurat. Oleh karena itu, proses penarikan sampel

menjadi tahap krusial yang memengaruhi validitas dan reliabilitas hasil inferensi (Casella & Berger, 2021).



Gambar 1. Ilustrasi Populasi dan Sampel

Sampel yang baik umumnya diperoleh melalui prosedur pengambilan sampel yang terencana dan berbasis probabilitas, sehingga setiap anggota populasi memiliki peluang yang sama untuk terpilih. Prinsip ini bertujuan meminimalkan bias dan memastikan bahwa variasi dalam populasi tercermin secara proporsional dalam sampel. Ketika hubungan antara sampel dan populasi terjaga dengan baik, parameter populasi yang tidak diketahui, seperti rata-rata atau proporsi, dapat diperkirakan secara lebih akurat melalui statistik sampel. Estimasi ini kemudian digunakan sebagai dasar untuk menarik kesimpulan yang bersifat umum (Wackerly et al., 2020).

Peran sampel dalam statistika inferensial tidak hanya sebagai sumber data, tetapi juga sebagai dasar pengukuran ketidakpastian. Variabilitas yang muncul dalam sampel mencerminkan variabilitas populasi dan menjadi dasar dalam perhitungan kesalahan baku,

interval kepercayaan, serta pengujian hipotesis. Sementara itu, populasi berperan sebagai kerangka konseptual yang menjadi tujuan akhir inferensi. Kesimpulan yang dihasilkan dari analisis inferensial selalu diarahkan untuk menjelaskan atau memprediksi karakteristik populasi, bukan sekadar menggambarkan sampel itu sendiri (Devore, 2023).

Dengan demikian, hubungan antara sampel dan populasi harus dipahami sebagai hubungan yang saling terkait dan tidak terpisahkan. Inferensi statistik yang valid hanya dapat dicapai apabila sampel dipilih dan dianalisis secara tepat sesuai dengan tujuan dan asumsi yang mendasarinya. Pemahaman yang mendalam mengenai peran sampel dan populasi akan membantu memastikan bahwa kesimpulan yang ditarik melalui statistika inferensial bersifat sah, dapat dipercaya, dan relevan untuk mendukung pengambilan keputusan dalam berbagai konteks keilmuan dan praktis.

6.2 Estimasi Parameter

Estimasi parameter merupakan inti dari statistika inferensial karena menjadi jembatan antara data sampel dan karakteristik populasi yang ingin diketahui. Dalam banyak situasi nyata, pengamatan terhadap seluruh populasi tidak memungkinkan dilakukan, sehingga informasi yang tersedia selalu mengandung ketidakpastian. Estimasi parameter bertujuan mengelola ketidakpastian tersebut secara sistematis dengan memanfaatkan prinsip probabilitas dan teori statistika. Melalui estimasi, nilai

parameter populasi seperti rerata, proporsi, atau varians dapat ditaksir secara rasional berdasarkan data sampel, dengan tingkat ketepatan yang dapat diukur dan dijelaskan secara logis (Casella & Berger, 2021).

Dalam kerangka inferensial, estimasi parameter tidak hanya berorientasi pada hasil akhir berupa angka, tetapi juga pada proses penalaran statistik yang mendasarinya. Proses ini melibatkan pemilihan estimator yang sesuai, pemahaman sifat-sifat statistik estimator, serta interpretasi hasil estimasi dalam konteks ketidakpastian data. Oleh karena itu, estimasi parameter menjadi fondasi penting bagi analisis lanjutan, pengujian hipotesis, dan pengambilan keputusan berbasis data.

6.2.1 Estimasi Titik

Estimasi titik merupakan pendekatan yang digunakan untuk memperkirakan nilai parameter populasi melalui satu nilai numerik tunggal yang dihitung dari data sampel. Nilai ini dihasilkan oleh suatu estimator, yaitu fungsi statistik yang dirancang untuk merepresentasikan parameter populasi tertentu. Contoh umum estimasi titik adalah penggunaan rata-rata sampel (\bar{x}) sebagai penaksir rata-rata populasi (μ) atau proporsi sampel sebagai penaksir proporsi populasi. Estimator semacam ini dipilih karena memiliki hubungan matematis yang jelas dengan parameter yang diestimasi.

Dalam teori statistika, kualitas estimasi titik dinilai berdasarkan sifat-sifat tertentu, seperti ketidakbiasan, konsistensi, dan efisiensi. Estimator yang tidak bias diharapkan menghasilkan nilai rata-rata yang sama dengan parameter populasi jika

pengambilan sampel dilakukan berulang kali. Konsistensi menunjukkan bahwa estimasi akan semakin mendekati nilai parameter sebenarnya seiring bertambahnya ukuran sampel. Sementara itu, efisiensi berkaitan dengan seberapa kecil variasi estimator dibandingkan estimator lain yang sejenis.

Meskipun estimasi titik bersifat ringkas dan mudah digunakan, pendekatan ini memiliki keterbatasan karena tidak menyertakan informasi mengenai tingkat ketidakpastian estimasi. Dua sampel yang berbeda dapat menghasilkan nilai estimasi titik yang berbeda, meskipun berasal dari populasi yang sama. Oleh karena itu, estimasi titik sebaiknya dipahami sebagai perkiraan terbaik berdasarkan data yang tersedia, bukan sebagai representasi pasti dari parameter populasi. Dalam praktik, estimasi titik sering digunakan sebagai dasar awal analisis sebelum dilengkapi dengan pendekatan yang memberikan gambaran ketidakpastian secara lebih eksplisit (Wasserman, 2022).

6.2.2 Estimasi Interval

Estimasi interval merupakan pendekatan yang memberikan taksiran parameter populasi dalam bentuk rentang nilai yang disertai tingkat kepercayaan tertentu. Berbeda dengan estimasi titik yang hanya menyajikan satu angka, estimasi interval mengakui secara eksplisit adanya variasi dan ketidakpastian yang melekat pada proses pengambilan sampel. Interval yang dihasilkan dirancang sedemikian rupa sehingga, dalam jangka panjang, proporsi tertentu dari interval yang dibentuk akan mencakup parameter populasi yang sebenarnya.

Konsep tingkat kepercayaan menjadi elemen utama dalam estimasi interval. Tingkat kepercayaan mencerminkan derajat keyakinan statistik terhadap prosedur pembentukan interval, bukan kepastian bahwa parameter berada di dalam interval tertentu. Dengan demikian, estimasi interval memberikan kerangka interpretasi yang lebih realistis terhadap hasil estimasi, karena menyajikan informasi tentang presisi dan reliabilitas taksiran parameter.

Lebar interval estimasi dipengaruhi oleh beberapa faktor, antara lain ukuran sampel, variasi data, dan tingkat kepercayaan yang dipilih. Sampel yang lebih besar dan variasi data yang lebih kecil cenderung menghasilkan interval yang lebih sempit, yang menunjukkan estimasi lebih presisi. Sebaliknya, interval yang lebar mencerminkan ketidakpastian yang lebih besar. Dalam konteks pengambilan keputusan, informasi mengenai rentang nilai yang mungkin sering kali lebih bermakna dibandingkan satu nilai estimasi tunggal, karena memungkinkan penilaian risiko dan batas toleransi secara lebih komprehensif (DeGroot & Schervish, 2021).

Contoh Soal:

Seorang manajer minimarket ingin menganalisis pengeluaran pelanggan per transaksi untuk keperluan penentuan strategi promo dan penataan produk. Manajer tersebut mencatat total belanja (dalam ribu rupiah) dari 10 pelanggan yang datang secara acak pada satu hari.

Tabel 1. Data total belanja pelanggan (dalam ribu rupiah):

Pelanggan	1	2	3	4	5	6	7	8	9	10
Pengeluaran (ribu Rp)	40	45	50	48	42	46	44	47	49	43

- Tentukan estimasi titik rata-rata pengeluaran pelanggan minimarket.
- Tentukan interval kepercayaan 95% untuk rata-rata pengeluaran pelanggan minimarket.

Jawaban:

- Estimasi titik rata-rata pengeluaran pelanggan minimarket

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

$$\bar{x} = \frac{1}{10} (40 + 45 + 50 + 48 + 42 + 46 + 44 + 47 + 49 + 43)$$

$$\bar{x} = \frac{1}{10} (454) = 45,4$$

Rata-rata pengeluaran pelanggan minimarket diperkirakan sebesar Rp45.400,- per transaksi.

- Interval kepercayaan 95% untuk rata-rata pengeluaran pelanggan minimarket

- Menentukan standar deviasi atau simpangan baku sampel

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

$$s = \sqrt{\frac{(40-45,4)^2 + (45-45,4)^2 + (50-45,4)^2 + \dots + (43-45,4)^2}{10-1}}$$

$$s = \sqrt{10,27} = 3,20$$

- Menghitung interval kepercayaan

Karena ukuran sampel kecil ($n < 30$) dan simpangan baku populasi tidak diketahui, digunakan distribusi t

$$t_{\alpha/2;n-1} = t_{0,05/2;10-1} = t_{0,025;9} = 2,262$$

$$\bar{x} \pm t_{\alpha/2;n-1} \cdot \frac{s}{\sqrt{n}}$$

$$\bar{x} - t_{\alpha/2;n-1} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2;n-1} \cdot \frac{s}{\sqrt{n}}$$

$$45.4 - 2.262 \left(\frac{3.20}{\sqrt{10}} \right) \leq \mu \leq 45.4 + 2.262 \left(\frac{3.20}{\sqrt{10}} \right)$$

$$43.11 \leq \mu \leq 47.69$$

Dengan tingkat kepercayaan 95%, rata-rata pengeluaran pelanggan minimarket diperkirakan berada antara Rp43.110,- hingga Rp47.690,- per transaksi.

6.3 Pengujian Hipotesis

6.3.1 Konsep Hipotesis Statistik

Pengujian hipotesis merupakan prosedur statistika inferensial yang digunakan untuk menarik kesimpulan mengenai parameter populasi berdasarkan informasi dari data sampel. Hipotesis statistik dirumuskan sebagai pernyataan formal yang dapat diuji secara objektif melalui pendekatan probabilistik, sehingga memungkinkan pengambilan keputusan yang sistematis di bawah ketidakpastian akibat variasi sampel. Dalam konteks ini, pengujian hipotesis tidak bertujuan untuk membuktikan kebenaran mutlak suatu pernyataan, melainkan untuk menilai sejauh mana data yang diamati memberikan dukungan terhadap dugaan tertentu tentang populasi.

Secara umum, pengujian hipotesis melibatkan dua pernyataan yang saling berlawanan, yaitu hipotesis nol (H_0) yang menyatakan tidak adanya perbedaan atau pengaruh, dan hipotesis alternatif (H_1) yang merepresentasikan dugaan yang ingin didukung oleh data. Keputusan pengujian didasarkan pada statistik uji atau *p-value* yang menunjukkan kekuatan bukti data terhadap H_0 , dengan mempertimbangkan tingkat signifikansi sebagai batas risiko kesalahan. Jika *p-value* lebih kecil dari tingkat signifikansi, maka H_0 ditolak. Dengan pendekatan ini, pengujian hipotesis berperan penting sebagai alat analisis data yang mendukung pengambilan keputusan secara rasional dan berbasis probabilitas (Casella & Berger, 2021).

6.3.2 Kesalahan Tipe I dan Tipe II

Dalam proses pengujian hipotesis, selalu terdapat risiko kesalahan dalam menetapkan keputusan statistik. Kesalahan tersebut dibedakan menjadi dua kategori utama, yaitu kesalahan tipe I dan kesalahan tipe II. Kesalahan tipe I terjadi ketika hipotesis nol yang sebenarnya benar justru ditolak, sehingga peneliti menyimpulkan adanya perbedaan atau pengaruh yang sebenarnya tidak ada. Risiko terjadinya kesalahan ini dikendalikan melalui tingkat signifikansi yang ditetapkan sebelum pengujian, yang merepresentasikan batas maksimum probabilitas melakukan penolakan hipotesis nol secara keliru.

Sebaliknya, kesalahan tipe II muncul ketika hipotesis nol yang sebenarnya salah tidak berhasil ditolak, sehingga pengujian gagal mendeteksi efek atau perbedaan yang benar-benar ada dalam

populasi. Peluang untuk menghindari kesalahan ini dikenal sebagai daya uji (*power*), yang menggambarkan kemampuan suatu uji statistik dalam mengidentifikasi efek yang nyata. Daya uji dipengaruhi oleh beberapa faktor, antara lain ukuran sampel, keragaman data, tingkat signifikansi, dan besar efek yang diharapkan. Karena risiko kesalahan tipe I dan tipe II saling memengaruhi, penentuan tingkat signifikansi dan perancangan pengujian hipotesis perlu mempertimbangkan keseimbangan di antara keduanya agar keputusan berbasis data dapat diambil secara tepat, andal, dan bertanggung jawab (Wasserman, 2020).

Tabel 2. Tipe Kesalahan dalam Pengujian Hipotesis

Kondisi H_0	Tolak H_0	Gagal Tolak H_0
H_0 benar	<i>Type I Error</i> (α)	Keputusan benar
H_0 salah	Keputusan benar	<i>Type II Error</i> (β)

- a. α : peluang melakukan kesalahan tipe I
- b. β : peluang melakukan kesalahan tipe II
- c. Daya uji (Power) = $1 - \beta$

Ilustrasi:

- Kesalahan Tipe I: Perusahaan menyimpulkan strategi promosi baru meningkatkan penjualan, padahal sebenarnya tidak → berisiko pemborosan biaya.
- Kesalahan Tipe II: Perusahaan menyimpulkan promosi tidak efektif, padahal sebenarnya berdampak positif → kehilangan peluang keuntungan.

6.4 Uji Statistik Parametrik

Uji statistik parametrik merupakan metode analisis yang digunakan untuk menguji hipotesis dengan beberapa asumsi tertentu terhadap data. Umumnya, uji ini digunakan pada data berskala interval atau rasio, berdistribusi normal, dan memiliki varians yang relatif sama antar kelompok yang dibandingkan. Jika syarat-syarat tersebut terpenuhi, uji parametrik dapat memberikan hasil yang akurat dan memiliki kemampuan yang baik dalam mendeteksi perbedaan atau pengaruh yang benar-benar ada. Oleh karena itu, uji parametrik sering digunakan karena hasilnya mudah diinterpretasikan dan didukung oleh teori statistik yang kuat, namun penggunaannya tetap harus diawali dengan pengecekan karakteristik data agar kesimpulan yang diperoleh tetap valid.

6.4.1 Uji t dan Uji z

Uji z dan uji t termasuk dalam kelompok uji parametrik yang digunakan untuk mengevaluasi perbedaan rata-rata dengan berlandaskan asumsi tertentu mengenai distribusi data. Uji z umumnya diterapkan ketika ukuran sampel relatif besar dan simpangan baku populasi telah diketahui, sehingga distribusi rata-rata sampel dapat diasumsikan mengikuti distribusi normal. Dalam kondisi tersebut, uji z mampu memberikan penilaian yang akurat terhadap perbedaan rata-rata dan sering digunakan pada analisis populasi besar atau pengujian yang didukung oleh informasi parameter populasi yang lengkap.

Sementara itu, uji t lebih banyak digunakan dalam praktik karena tidak memerlukan informasi mengenai simpangan baku populasi dan cocok untuk ukuran sampel kecil hingga menengah dengan asumsi normalitas data. Uji ini memiliki beberapa bentuk, seperti uji t satu sampel, uji t dua sampel independen, dan uji t berpasangan, yang disesuaikan dengan rancangan penelitian dan tujuan analisis. Baik uji t maupun uji z berperan penting dalam menentukan apakah perbedaan rata-rata yang diamati signifikan secara statistik atau sekadar disebabkan oleh variasi acak, sehingga keduanya menjadi dasar inferensi yang andal dalam pengambilan keputusan berbasis data (Kim, 2021).

6.4.2 Analisis Varians (ANOVA)

Analisis varians (ANOVA) merupakan metode uji statistik parametrik yang digunakan untuk membandingkan rata-rata dari tiga kelompok atau lebih dalam satu prosedur pengujian. Konsep utama ANOVA adalah memisahkan keragaman data menjadi variasi yang berasal dari perbedaan antar kelompok dan variasi yang muncul di dalam masing-masing kelompok, kemudian menilai apakah perbedaan antar kelompok tersebut cukup besar secara statistik. Pendekatan ini memungkinkan pengujian perbedaan rata-rata secara efisien tanpa meningkatkan risiko kesalahan yang dapat terjadi jika perbandingan dilakukan secara berulang menggunakan uji dua kelompok.

Dalam praktik analisis data, ANOVA banyak digunakan pada penelitian yang melibatkan beberapa kategori atau perlakuan, karena mampu memberikan kesimpulan awal mengenai adanya

perbedaan rata-rata antar kelompok dengan satu uji saja. ANOVA memiliki beberapa bentuk, seperti ANOVA satu arah dan dua arah, yang dipilih berdasarkan jumlah faktor yang dianalisis. Selain itu, hasil ANOVA dapat dilanjutkan dengan uji pasca-hoc untuk mengetahui kelompok mana yang berbeda secara signifikan. Meskipun demikian, penerapan ANOVA tetap mensyaratkan terpenuhinya asumsi normalitas, kesamaan varians, dan independensi data agar kesimpulan yang dihasilkan bersifat valid dan dapat dipercaya (Field, 2020).

Contoh Soal:

Seorang manajer pemasaran ingin mengetahui apakah terdapat perbedaan yang signifikan pada rata-rata penjualan harian berdasarkan strategi promosi yang digunakan di tiga cabang toko. Adapun datanya disajikan sebagai berikut:

Tabel 3. Data Penjualan Harian (dalam juta rupiah)

Strategi Promosi	Cabang 1	Cabang 2	Cabang 3
Diskon (%)	12	15	13
<i>Cashback</i>	18	20	19
<i>Bundling</i>	14	16	15

Jawaban:

a. Merumuskan Hipotesis

H_0 : Tidak terdapat perbedaan rata-rata penjualan antar strategi promosi.

H_1 : Terdapat minimal satu strategi promosi dengan rata-rata penjualan yang berbeda

b. Menghitung Rata-rata

- Rata-rata diskon

$$\bar{x}_1 = \frac{12+15+13}{3} = 13,33$$

- Rata-rata *Cashback*

$$\bar{x}_2 = \frac{18+20+19}{3} = 19,00$$

- Rata-rata *Bundling*

$$\bar{x}_3 = \frac{14+16+15}{3} = 15,00$$

- Rata-rata Keseluruhan

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

$$\bar{x} = \frac{1}{9} \cdot (12 + 15 + 13 + 18 + \dots + 15) = 15,78 = 15,78$$

c. Menghitung Jumlah Kuadrat Antar Kelompok (SSB)

SSB (*Sum of Squares Between*): Mengukur variasi akibat perbedaan antar rata-rata kelompok.

$$SSB = \sum n_i (\bar{x}_i - \bar{x})^2$$

$$SSB = 3(13,33 - 15,78)^2 + 3(19 - 15,78)^2 + 3(15 - 15,78)^2$$

$$SSB = 18,00 + 31,07 + 1,82 = 50,89$$

d. Menghitung Jumlah Kuadrat Dalam Kelompok (SSW)

SSW (*Sum of Squares Within*): Mengukur variasi data di dalam masing-masing kelompok.

$$SSW = \sum (x_{ij} - \bar{x}_i)^2$$

Keterangan:

x_{ij} = nilai pengamatan ke- j pada kelompok ke- i

\bar{x}_i = rata-rata kelompok ke- i

$$SSW_{diskon} = (12 - 13,33)^2 + (15 - 13,33)^2 + (13 - 13,33)^2 = 4,67$$

$$SSW_{Cashback} = (18 - 19)^2 + (20 - 19)^2 + (19 - 19)^2 = 2$$

$$SSW_{Bundling} = (14 - 15)^2 + (16 - 15)^2 + (15 - 15)^2 = 2$$

$$\text{Total SSW} = 4,67 + 2 + 2 = 8,67$$

- e. Menghitung Jumlah Kuadrat Total (SST)

SST (*Sum of Squares Total*): Total variasi seluruh data

$$SST = SSB + SSW = 50,89 + 8,67 = 59,56$$

- f. Menghitung df (*degree of freedom*): Derajat bebas

$$n = 9 \text{ (total observasi)}$$

$$k = 3 \text{ (jumlah kelompok: Diskon, Cashback, dan Bundling)}$$

$$\text{df antar kelompok} = k - 1 = 3 - 1 = 2$$

$$\text{df dalam kelompok} = n - k = 9 - 3 = 6$$

- g. Menghitung Rata-rata Jumlah Kuadrat (MS)

MS (*Mean Square*): Rata-rata jumlah kuadrat

$$MS_{between} = \frac{SS}{df} = \frac{50,89}{2} = 25,45$$

$$MS_{within} = \frac{SS}{df} = \frac{8,67}{6} = 1,45$$

- h. Menghitung nilai F (Statistik Uji ANOVA)

$$F_{hitung} = \frac{MS_{between}}{MS_{within}} = \frac{25,45}{1,45} = 17,61$$

$$F_{tabel} = F_{(\alpha=0.05; df=2,6)} = 5,1433 \approx 5,14$$

Tabel 4. ANOVA

Sumber Variasi	JK (SS)	df	KT (MS)	F
Antar Kelompok (<i>Between</i>)	50,89	2	25,45	17,61
Dalam Kelompok (<i>Within</i>)	8,67	6	1,45	
Total	61,56	8		

i. Keputusan dan Kesimpulan

Nilai F_{hitung} lebih besar dari F_{tabel} yaitu $17,61 > 5,14$ maka H_0 di tolak, sehingga dapat disimpulkan bahwa terdapat perbedaan rata-rata penjualan yang signifikan di antara strategi promosi. Artinya, jenis promosi yang digunakan berpengaruh terhadap tingkat penjualan harian.

6.5 Uji Statistik Nonparametrik

Uji statistik nonparametrik adalah metode analisis yang digunakan ketika data tidak memenuhi persyaratan uji parametrik, seperti asumsi distribusi normal atau kesamaan varians antar kelompok. Uji ini sangat sesuai untuk data dengan skala pengukuran nominal dan ordinal, serta untuk data interval atau rasio yang memiliki distribusi tidak normal atau jumlah sampel yang terbatas. Karena tidak bergantung pada bentuk distribusi tertentu, metode nonparametrik bersifat lebih fleksibel dan tahan terhadap penyimpangan karakteristik data. Oleh sebab itu, dalam analisis statistik terapan, uji nonparametrik sering dipilih sebagai alternatif yang aman dan reliabel untuk memastikan kesimpulan yang

dihasilkan tetap valid ketika asumsi statistik klasik tidak dapat dipenuhi (Conover, 2020).

6.5.1 Uji *Chi-Square*

Uji *chi-square* (chi-kuadrat) merupakan metode statistik nonparametrik yang banyak digunakan untuk menganalisis data berbentuk kategori. Uji ini bertujuan untuk mengetahui apakah terdapat hubungan yang signifikan antara variabel-variabel kategorik atau untuk menilai kesesuaian antara data hasil pengamatan dengan distribusi yang diharapkan berdasarkan hipotesis nol. Dasar pengujiannya adalah membandingkan frekuensi yang diperoleh dari data dengan frekuensi teoritis yang seharusnya muncul, kemudian merangkum selisih keduanya dalam suatu statistik uji yang mengikuti distribusi *chi-square*.

Dalam praktiknya, uji *chi-square* sering diterapkan pada tabel kontingensi untuk mengevaluasi keterkaitan antarvariabel kategorik, serta pada uji kesesuaian untuk melihat apakah pola data empiris sesuai dengan distribusi tertentu. Metode ini relatif mudah digunakan dan tidak mensyaratkan asumsi kenormalan data, sehingga cocok untuk berbagai kondisi penelitian. Namun demikian, validitas hasil uji tetap bergantung pada terpenuhinya syarat frekuensi harapan yang memadai pada setiap kategori. Oleh karena itu, pemahaman terhadap struktur data dan karakteristik kategori sangat diperlukan agar kesimpulan yang dihasilkan akurat dan dapat dipertanggungjawabkan (Agresti, 2019; Conover, 2020).

Contoh Soal:

Sebuah perusahaan ritel ingin mengetahui apakah jenis metode pembayaran berhubungan dengan kategori usia pelanggan. Data hasil survei 120 pelanggan disajikan dalam tabel berikut:

Tabel 5. Data jenis metode pembayaran

Kategori Usia	Tunai	Non-Tunai	Total
< 30 tahun	20	30	50
\geq 30 tahun	40	30	70
Total	60	60	120

Lakukan analisis *chi-square* untuk melihat apakah terdapat hubungan antara kategori usia dan metode pembayaran.

Jawaban:

a. Rumusan Hipotesis Uji

H_0 : Tidak ada perbedaan distribusi tingkat kesembuhan berdasarkan jenis terapi.

H_1 : Ada perbedaan distribusi tingkat kesembuhan berdasarkan jenis terapi.

b. Tingkat Signifikansi

$$\alpha = 5\% = 0,05$$

$$db = (r-1)(c-1) = (2-1)(2-1) = 1$$

$$\chi^2(0,05; db=1) = 3,84 \text{ (tabel } \chi\text{-square)}$$

c. Statistik Uji

- Hitung Frekuensi Harapan (*Expected Frequency*)

$$E_{ij} = \frac{(\text{Total Baris}) \cdot (\text{Total Kolom})}{\text{Total Keseluruhan}}$$

Frekuensi harapan (E_{ij}):

$$E_{11} = (50 \times 60) / 120 = 25$$

$$E_{12} = (50 \times 60) / 120 = 25$$

$$E_{21} = (70 \times 60) / 120 = 35$$

$$E_{22} = (70 \times 60) / 120 = 35$$

- Hitung Statistik Uji *Chi-Square*

$$\chi^2_{hitung} = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$\chi^2_{hitung} = \frac{(20-25)^2}{25} + \frac{(30-25)^2}{25} + \frac{(30-35)^2}{35} + \frac{(30-35)^2}{35} = 3,42$$

- d. Keputusan dan Kesimpulan

Nilai chi-square hitung (χ^2_{hitung}) kurang dari chi-square tabel yaitu $3,42 < 3,84$ artinya H_0 gagal di tolak atau H_1 diterima. Sehingga dapat disimpulkan bahwa tidak terdapat bukti statistik yang cukup untuk menyatakan adanya hubungan antara kategori usia pelanggan dan metode pembayaran. Dengan kata lain, perbedaan pola pembayaran yang terlihat masih dapat dianggap terjadi secara kebetulan.

6.5.2 Uji Mann-Whitney

Uji Mann-Whitney merupakan metode statistik nonparametrik yang digunakan untuk membandingkan dua kelompok data yang bersifat independen ketika asumsi uji parametrik, khususnya uji t, tidak terpenuhi. Uji ini tidak menggunakan nilai data secara langsung, melainkan memanfaatkan peringkat gabungan dari kedua kelompok, sehingga lebih tahan terhadap keberadaan pencilan dan

distribusi data yang tidak normal. Melalui pendekatan berbasis peringkat tersebut, uji Mann–Whitney menilai apakah dua kelompok berasal dari populasi dengan kecenderungan pusat yang sama atau berbeda. Oleh karena itu, uji ini banyak digunakan pada data sosial, ekonomi, dan perilaku yang umumnya tidak memenuhi asumsi kenormalan (Field, 2020).

Contoh Soal:

Sebuah perusahaan ingin mengetahui apakah tingkat kepuasan pelanggan berbeda antara pelanggan toko fisik dan pelanggan toko online. Data kepuasan diukur menggunakan skala ordinal 1–10. Adapun datanya sebagai berikut:

Tabel 6. Data tingkat kepuasan pelanggan

Toko Fisik	6	7	5	6	7
Toko Online	8	9	7	8	9

Lakukan uji Mann-Whitney untuk melihat apakah terdapat perbedaan tingkat kepuasan pelanggan toko fisik dengan pelanggan toko online.

Jawaban:

a. Rumusan Hipotesis Uji

H_0 : Tidak terdapat perbedaan tingkat kepuasan pelanggan antara toko fisik dan toko online.

H_1 : Terdapat perbedaan tingkat kepuasan pelanggan antara toko fisik dan toko online.

b. Tingkat Signifikansi

$\alpha = 0,05$ (dua sisi) $n_1 = 5$ $n_2 = 5$

Nilai kritis dari tabel Mann-Whitney adalah:

$$U_{tabel} = U_{(n1=5, n2 = 5; \alpha/2=0.025)} = 2$$

c. Statistik Uji

Tabel 7. Perangkingan Data gabungan

Nilai	Kelompok	Ranking
5	Fisik	1
6	Fisik	2,5
6	Fisik	2,5
7	Fisik	5
7	Fisik	5
7	Online	5
8	Online	7,5
8	Online	7,5
9	Online	9,5
9	Online	9,5
Total		55

Jumlah Peringkat Tiap Kelompok

$$R_1 (\text{Fisik}) = 1 + 2,5 + 2,5 + 5 + 5 = 16$$

$$R_2 (\text{Online}) = 5 + 7,5 + 7,5 + 9,5 + 9,5 = 39$$

Menghitung uji Mann-Whitney (U)

$$U_1 = n_1 \cdot n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_1 = (5)(5) + \frac{5(5+1)}{2} - 16 = 24$$

$$U_2 = n_1 \cdot n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$U_1 = (5)(5) + \frac{5(5+1)}{2} - 39 = 1$$

$$U = \min(U_1; U_2) = \min(24; 1) = 1$$

Sehingga $U_{hitung} = 1$

d. Kriteria Uji

Tolak H_0 jika U_{hitung} kurang dari U_{tabel} atau gagal tolak H_0 jika U_{hitung} lebih besar dari U_{tabel}

e. Keputusan dan Kesimpulan

Tolak H_0 karena nilai U_{hitung} lebih kecil dari U_{tabel} yaitu $1 < 2$. Sehingga dapat disimpulkan bahwa terdapat perbedaan yang signifikan secara statistik antara tingkat kepuasan pelanggan toko fisik dengan toko *online*, yang artinya terdapat perbedaan kinerja penjualan metode *offline* dan metode *online*.

6.5.3 Uji Wilcoxon

Uji Wilcoxon merupakan uji nonparametrik yang digunakan untuk membandingkan dua kondisi data yang saling berpasangan, seperti pengukuran sebelum dan sesudah suatu perlakuan pada subjek yang sama. Uji ini mengevaluasi perbedaan peringkat antar pasangan data untuk menentukan apakah perubahan yang terjadi bersifat signifikan secara statistik. Keunggulan uji Wilcoxon terletak pada kemampuannya mendeteksi perubahan sistematis tanpa mensyaratkan distribusi normal pada selisih data. Dengan demikian, uji ini sangat sesuai digunakan ketika data berpasangan tidak memenuhi asumsi uji parametrik, namun analisis inferensial tetap diperlukan secara valid dan andal (Field, 2020).

Contoh Soal:

Sebuah perusahaan ritel ingin mengevaluasi dampak kebijakan diskon baru terhadap rata-rata nilai transaksi harian

(dalam juta rupiah). Nilai transaksi dicatat sebelum dan sesudah kebijakan diskon diberlakukan pada 8 gerai yang sama. Adapun datanya sebagai berikut:

Tabel 8. Data transaksi (juta rupiah)

Gerai	Sebelum	Sesudah
1	12	14
2	15	16
3	10	12
4	18	19
5	20	22
6	14	15
7	16	18
8	13	14

Lakukan analisis untuk melihat apakah terdapat perbedaan rata-rata nilai transaksi harian sebelum dan sesudah kebijakan diskon.

Jawaban:

a. Rumusan Hipotesis Uji

H_0 : Tidak terdapat perbedaan median nilai transaksi sebelum dan sesudah kebijakan diskon.

H_1 : Terdapat perbedaan median nilai transaksi sebelum dan sesudah kebijakan diskon.

b. Tingkat Signifikansi

$\alpha = 5\% = 0,05$ (dua sisi) dan $n = 8$

$T_{\text{tabel}} = T_{(n=8; \alpha/2=0.025)} = 4$

c. Statistik Uji

Tabel 9. Menghitung selisih dan tanda

Gerai	Sebelum	Sesudah	selisih (d)	$ d $	rank	Tanda
1	12	14	+ 2	2	6,5	+
2	15	16	+ 1	1	2,5	+
3	10	12	+ 2	2	6,5	+
4	18	19	+ 1	1	2,5	+
5	20	22	+ 2	2	6,5	+
6	14	15	+ 1	1	2,5	+
7	16	18	+ 2	2	6,5	+
8	13	14	+ 1	1	2,5	+

- Ranking berdasarkan $|d|$ dari yang terkecil ke terbesar. Empat nilai $|d| = 1 \rightarrow$ seharusnya ranking 1, 2, 3, 4 namun karena nilainya sama, maka:

$$\text{Ranking rata - rata} = \frac{1+2+3+4}{4} = 2,5$$

- Empat nilai $|d| = 2 \rightarrow$ seharusnya ranking 5, 6, 7, 8 namun karena nilainya sama, maka:

$$\text{Ranking rata - rata} = \frac{5+6+7+8}{4} = 6,5$$

- Jumlah Ranking Positif dan Negatif

$$T^+ = (4 \times 2,5) + (4 \times 6,5) = 10 + 26 = 36$$

$$T^- = 0$$

- Statistik uji Wilcoxon:

$$T = \min (T^+ ; T^-) = \min (36 ; 0) = 0$$

$$\text{Sehingga } T_{\text{hitung}} = 0$$

d. Kriteria Uji

Tolak H_0 jika T_{hitung} kurang dari T_{tabel} atau gagal tolak H_0 jika T_{hitung} lebih besar T_{tabel} .

e. Keputusan dan Kesimpulan

Tolak H_0 karena nilai T_{hitung} lebih kecil dari T_{tabel} yaitu $0 < 4$. Sehingga dapat disimpulkan bahwa terdapat perbedaan yang signifikan secara statistik antara nilai transaksi sebelum dan sesudah penerapan kebijakan diskon. Hasil ini menunjukkan bahwa kebijakan diskon memberikan dampak nyata terhadap peningkatan nilai transaksi harian pada gerai ritel.

6.6 Latihan Soal

1. Jelaskan perbedaan statistika deskriptif dan inferensial.
2. Uraikan perbedaan antara estimasi titik dan estimasi interval.
3. Jelaskan konsep hipotesis nol dan hipotesis alternatif.
4. Apa perbedaan utama antara uji parametrik dan nonparametrik?
5. Analisis peran ANOVA dalam pengujian perbedaan rata-rata.

Bab 7: Analisis Korelasi dan Regresi

7.1 Pengantar Korelasi dan Regresi

Analisis korelasi dan regresi merupakan dua pendekatan statistik yang fundamental dalam memahami dan menginterpretasi hubungan antara dua atau lebih variabel. Meskipun sering digunakan secara bersamaan, keduanya memiliki tujuan dan fungsi yang berbeda namun saling melengkapi dalam praktik analisis data. Korelasi digunakan untuk mengukur sejauh mana hubungan linier antara dua variabel, dengan hasil berupa koefisien korelasi yang bernilai antara -1 dan $+1$. Nilai mendekati $+1$ menunjukkan hubungan positif yang kuat, di mana peningkatan salah satu variabel cenderung diikuti peningkatan variabel lainnya. Sebaliknya, nilai mendekati -1 menunjukkan hubungan negatif, di mana peningkatan satu variabel berkorelasi dengan penurunan variabel lain. Jika nilai mendekati nol, hal ini menandakan bahwa tidak terdapat hubungan linier yang signifikan antara kedua variabel tersebut.

Di sisi lain, regresi memiliki tujuan yang lebih kompleks, yaitu untuk memodelkan hubungan antara variabel dependen dan satu atau lebih variabel independen, serta untuk memprediksi nilai dari variabel dependen tersebut. Bentuk paling sederhana adalah regresi linier sederhana, di mana hanya terdapat satu variabel

independen yang mempengaruhi variabel dependen. Model ini menghasilkan sebuah persamaan linier yang menggambarkan bagaimana perubahan pada variabel independen memengaruhi nilai variabel dependen. Selain regresi linier sederhana, terdapat pula regresi linier berganda yang melibatkan lebih dari satu variabel independen, serta regresi nonlinier yang digunakan ketika hubungan antar variabel tidak bersifat linier.

Korelasi dan regresi sangat penting dalam berbagai disiplin ilmu, mulai dari ekonomi, psikologi, hingga ilmu kesehatan. Dalam bidang ekonomi, misalnya, korelasi dapat digunakan untuk mengetahui sejauh mana harga saham berkorelasi dengan suku bunga, sementara regresi dapat digunakan untuk memprediksi pendapatan berdasarkan tingkat pendidikan atau pengalaman kerja. Dalam ilmu kesehatan, regresi sering digunakan untuk mengidentifikasi faktor risiko dari suatu penyakit dengan menganalisis hubungan antara variabel seperti usia, indeks massa tubuh, dan gaya hidup dengan kejadian penyakit tertentu. Dalam konteks ini, analisis regresi tidak hanya memungkinkan prediksi, tetapi juga menginformasikan kebijakan berbasis data dengan mengungkap variabel-variabel yang paling berpengaruh (Schober et al., 2018).

Penting untuk dipahami bahwa korelasi tidak selalu menyiratkan kausalitas. Dua variabel dapat menunjukkan korelasi yang tinggi tanpa adanya hubungan sebab-akibat. Oleh karena itu, analisis regresi sering kali lebih disukai dalam penelitian kausal karena memungkinkan kontrol terhadap variabel lain dan

memberikan informasi mengenai kekuatan serta arah pengaruh variabel independen terhadap variabel dependen. Meski demikian, interpretasi hasil regresi juga harus dilakukan dengan hati-hati, memperhatikan asumsi-asumsi dasar seperti linearitas, homoskedastisitas, independensi residual, dan normalitas error. Pelanggaran terhadap asumsi-asumsi ini dapat menyebabkan hasil analisis yang bias dan tidak dapat diandalkan (Yoo et al., 2020).

Dalam praktiknya, proses analisis korelasi dan regresi juga bergantung pada visualisasi data seperti diagram pencar (*scatter plot*) yang membantu melihat pola hubungan antar variabel secara intuitif sebelum dilakukan perhitungan statistik. Perangkat lunak statistik seperti SPSS, R, atau Python kini banyak digunakan untuk memudahkan perhitungan koefisien korelasi maupun estimasi model regresi, termasuk pengujian signifikansi melalui *t-test* atau *F-test*. Penggunaan alat-alat ini telah memperluas penerapan metode korelasi dan regresi ke dalam praktik bisnis, riset ilmiah, serta pengambilan keputusan berbasis data.

Dengan demikian, pemahaman mendalam mengenai analisis korelasi dan regresi menjadi sangat krusial bagi siapa pun yang bekerja dengan data kuantitatif. Metode ini tidak hanya memungkinkan eksplorasi hubungan antar variabel secara deskriptif dan prediktif, tetapi juga menjadi dasar bagi inferensi statistik yang valid dan dapat ditindaklanjuti dalam konteks nyata.

7.2 Analisis Korelasi

7.2.1 Koefisien Korelasi Pearson

Koefisien korelasi Pearson (*Pearson's correlation coefficient*), yang sering dilambangkan dengan r , digunakan untuk mengukur hubungan linier antara dua variabel kontinu. Nilainya berkisar antara -1 hingga +1. Nilai +1 menunjukkan hubungan linier positif sempurna, sementara -1 menunjukkan hubungan linier negatif sempurna. Nilai 0 menandakan tidak adanya korelasi linier. Koefisien ini sangat bergantung pada asumsi normalitas distribusi data serta hubungan linier antarvariabel. Oleh karena itu, validitas penggunaannya memerlukan pengujian awal terhadap distribusi data (Schober et al., 2021).

7.2.2 Korelasi Spearman dan Kendall

Ketika data tidak memenuhi asumsi normalitas atau bersifat ordinal, koefisien korelasi nonparametrik seperti Spearman (*Spearman's rho*) dan Kendall (*Kendall's tau*) digunakan sebagai alternatif. Korelasi Spearman mengukur asosiasi monotonik berdasarkan peringkat data, sedangkan Kendall memberikan ukuran kekonsistenan peringkat antara dua variabel. Keduanya lebih tahan terhadap pencilan dan distribusi data yang tidak simetris, sehingga cocok untuk digunakan dalam penelitian sosial atau data survei dengan skala Likert. Selain itu, korelasi Kendall cenderung lebih konservatif dibanding Spearman dalam menghasilkan nilai koefisien (Arifin & Yusoff, 2020).

7.2.3 Interpretasi Korelasi

Interpretasi nilai korelasi harus dilakukan secara kontekstual dan berhati-hati. Korelasi positif menunjukkan bahwa peningkatan satu variabel diikuti oleh peningkatan variabel lainnya (hubungan searah), sedangkan korelasi negatif mengindikasikan hubungan berlawanan arah. Namun, kekuatan hubungan juga perlu dipertimbangkan. Misalnya, nilai korelasi antara 0,1 hingga 0,3 dianggap lemah, 0,3 hingga 0,5 sedang, dan di atas 0,5 kuat. Penting pula untuk menyadari bahwa korelasi tinggi tidak selalu berarti hubungan sebab-akibat, sehingga analisis lanjutan seperti regresi sering kali diperlukan untuk inferensi yang lebih mendalam.

7.3 Analisis Regresi Linier

7.3.1 Model Regresi Linier Sederhana

Model regresi linier sederhana digunakan untuk menganalisis hubungan antara satu variabel independen (X) dan satu variabel dependen (Y). Persamaan umum model ini adalah $Y = a + bX$, di mana a merupakan intersep yang menunjukkan nilai Y saat $X = 0$, dan b adalah koefisien kemiringan yang menggambarkan perubahan rata-rata Y untuk setiap satu unit perubahan X . Model ini sering diterapkan dalam berbagai bidang, seperti ekonomi, ilmu sosial, dan ilmu eksakta, untuk memprediksi nilai dependen berdasarkan variabel prediktor tunggal (Montgomery et al., 2021).

7.3.2 Estimasi Parameter dan Interpretasi

Estimasi parameter a (intersep) dan b (kemiringan) dilakukan menggunakan metode kuadrat terkecil (*Ordinary Least Squares/OLS*). Metode ini meminimalkan jumlah kuadrat selisih antara nilai observasi dan nilai yang diprediksi oleh model. Interpretasi parameter a adalah nilai awal atau dasar dari variabel dependen ketika variabel independen bernilai nol, sedangkan b menunjukkan arah dan besarnya pengaruh variabel independen terhadap variabel dependen. Nilai b positif menandakan hubungan langsung, sedangkan nilai negatif menunjukkan hubungan invers. Keandalan interpretasi parameter juga tergantung pada validitas asumsi model (Wooldridge, 2020).

7.3.3 Asumsi Model Regresi

Model regresi linier sederhana didasarkan pada beberapa asumsi penting untuk memastikan estimasi parameter tidak bias dan inferensi statistik valid. Asumsi tersebut meliputi: (1) linearitas, yaitu hubungan antara X dan Y bersifat linier; (2) normalitas residual, artinya sisa model terdistribusi normal; (3) homoskedastisitas, yaitu varians residual konstan untuk semua nilai X ; dan (4) independensi error, yang menyatakan bahwa residual tidak saling bergantung. Pelanggaran terhadap asumsi-asumsi ini dapat menyebabkan estimasi parameter yang tidak akurat dan kesimpulan yang menyesatkan (Montgomery et al., 2021).

7.3.4 Model Regresi Linier Berganda

Regresi linier berganda (*multiple linear regression*) merupakan pengembangan dari regresi linier sederhana dengan melibatkan lebih dari satu variabel independen. Model ini digunakan ketika variabel dependen dipengaruhi oleh beberapa faktor secara simultan. Persamaan umum model regresi linier berganda adalah:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

di mana:

- a. Y adalah variabel dependen,
- b. β_0 adalah intersep,
- c. $\beta_1, \beta_2, \dots, \beta_k$ adalah koefisien regresi untuk masing-masing variabel independen,
- d. X_1, X_2, \dots, X_k adalah variabel independen,
- e. ε adalah error term.

Model ini memungkinkan analisis untuk mengontrol pengaruh beberapa variabel sekaligus dan mengidentifikasi kontribusi relatif masing-masing prediktor terhadap variabel dependen. Namun, penerapannya memerlukan perhatian terhadap masalah multikolinieritas, yaitu korelasi tinggi antar variabel independen yang dapat mengganggu estimasi parameter (James et al., 2021).

7.3.5 Regresi Nonlinier

Tidak semua hubungan antar variabel bersifat linier. Regresi nonlinier digunakan ketika hubungan antara variabel dependen dan independen lebih sesuai dijelaskan dengan bentuk fungsi lain, seperti polinomial, eksponensial, atau logaritmik. Contoh model regresi polinomial orde dua:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

Model nonlinier dapat memberikan *goodness-of-fit* yang lebih baik jika pola data memang tidak linier, namun interpretasi koefisien menjadi lebih kompleks dan memerlukan metode estimasi seperti *nonlinear least squares* (NLS). Penggunaan model nonlinier juga memerlukan pemeriksaan diagnostik model yang lebih ketat (Fox & Weisberg, 2019).

7.4 Penilaian dan Aplikasi Model

7.4.1 Koefisien Determinasi (R^2)

Koefisien determinasi, disimbolkan sebagai R^2 , merupakan salah satu indikator utama dalam evaluasi model regresi. Nilai ini menunjukkan seberapa besar proporsi variasi dalam variabel dependen yang dapat dijelaskan oleh variabel independen dalam model. Sebagai contoh, nilai R^2 sebesar 0,80 mengindikasikan bahwa 80% variasi variabel dependen dapat dijelaskan oleh model regresi yang digunakan, sedangkan sisanya dipengaruhi oleh faktor lain di luar model. Semakin tinggi nilai R^2 , maka semakin baik model dalam menjelaskan hubungan antara variabel. Namun demikian, nilai R^2 yang tinggi tidak selalu menjamin bahwa model bebas dari *overfitting*, sehingga perlu didampingi dengan evaluasi tambahan seperti *adjusted R^2* atau uji statistik lainnya (Sarstedt et al., 2020).

7.4.2 Uji Signifikansi dan *P-Value*

Uji signifikansi digunakan untuk menguji pengaruh variabel independen terhadap variabel dependen secara statistik. Dalam konteks regresi, *p-value* berfungsi sebagai dasar untuk menolak atau menerima hipotesis nol (H_0) yang menyatakan bahwa tidak terdapat pengaruh signifikan. Umumnya, ambang batas signifikansi yang digunakan adalah 0,05. Jika *p-value* lebih kecil dari 0,05, maka variabel tersebut dianggap memiliki pengaruh yang signifikan terhadap variabel dependen. Uji *t* digunakan untuk menilai signifikansi masing-masing koefisien regresi, sedangkan uji *F* menguji signifikansi model secara keseluruhan. Hasil uji ini penting dalam menyaring variabel yang benar-benar relevan dan menghindari kesalahan pengambilan keputusan (Ghasemy et al., 2021).

7.4.3 Diagnostik Residual dan Validasi Model

Setelah model regresi diestimasi, penting untuk melakukan diagnostik residual guna memverifikasi asumsi model. Beberapa alat diagnostik yang umum digunakan antara lain:

1. Plot Residual vs *Fitted*: untuk mendeteksi pola nonlinier dan heteroskedastisitas.
2. Q-Q Plot: untuk memeriksa normalitas residual.
3. Plot Residual vs Variabel Independen: untuk mendeteksi ketidakkonstanan varians.
4. Uji Durbin-Watson: untuk menguji autokorelasi residual pada data *time series*.

Selain itu, validasi model dapat dilakukan dengan metode seperti *cross-validation* atau pembagian data menjadi *training set* dan *test set* untuk menghindari *overfitting*. Model yang baik harus mampu melakukan prediksi yang akurat pada data baru, bukan hanya pada data yang digunakan untuk estimasi (Hastie et al., 2009).

7.4.4 Aplikasi Praktis

Model regresi merupakan alat analisis yang sangat luas penerapannya dalam berbagai bidang. Di bidang bisnis, regresi digunakan untuk memprediksi permintaan, penjualan, dan tren pasar. Dalam ekonomi, regresi membantu mengestimasi pengaruh kebijakan terhadap variabel makroekonomi seperti inflasi atau pengangguran. Sektor kesehatan menggunakan regresi untuk mempelajari hubungan antara gaya hidup dan risiko penyakit, sementara ilmu sosial memanfaatkannya untuk mengevaluasi pengaruh pendidikan, pendapatan, atau lingkungan terhadap perilaku masyarakat. Keandalan regresi dalam melakukan peramalan dan mengidentifikasi hubungan antarvariabel menjadikannya metode yang penting dalam pengambilan keputusan berbasis data.

7.5 Latihan Soal

1. Buatlah sebuah studi kasus sederhana yang melibatkan prediksi menggunakan regresi linier satu variabel.
2. Bagaimana cara menginterpretasikan nilai koefisien determinasi (R^2) dalam regresi?
3. Sebutkan dan jelaskan asumsi-asumsi dasar dalam model regresi linier sederhana.
4. Apa arti nilai koefisien korelasi +1, -1, dan 0? Berikan contohnya masing-masing.
5. Jelaskan perbedaan antara analisis korelasi dan regresi dalam konteks data kuantitatif.

Bab 8: Analisis Deret Waktu

(Time Series)

8.1 Konsep Dasar Deret Waktu

Analisis deret waktu (*time series analysis*) merupakan bagian penting dalam statistika terapan yang fokus pada pengolahan dan interpretasi data yang dikumpulkan secara berurutan berdasarkan waktu. Data deret waktu tersusun secara kronologis, baik dalam satuan harian, bulanan, triwulanan, hingga tahunan, dan mengandung karakteristik yang khas karena adanya keterkaitan setiap pengamatan dengan waktu sebelumnya. Berbeda dengan data *cross-sectional* yang hanya menggambarkan kondisi pada satu titik waktu, data deret waktu memungkinkan peneliti menelusuri dinamika perubahan fenomena sepanjang waktu. Oleh karena itu, analisis deret waktu memiliki peran yang sangat penting dalam mengidentifikasi tren historis, pola musiman atau siklus, serta melakukan peramalan masa depan. Analisis ini tidak hanya mempertimbangkan hubungan antar variabel, tetapi juga pola/struktur yang membentuk fenomena tersebut (Wei, 2019; Cryer & Chan, 2008).

Metode analisis deret waktu banyak dimanfaatkan di beragam sektor. Di bidang ekonomi dan keuangan, metode ini digunakan untuk memprediksi inflasi, produk domestik bruto

(PDB), nilai tukar, tingkat pengangguran, dan harga saham. Dalam konteks bisnis, perusahaan memanfaatkan data penjualan bulanan untuk menyusun strategi pemasaran, menentukan stok, serta merencanakan anggaran. Bahkan dalam sektor kesehatan dan meteorologi, deret waktu digunakan untuk memantau penyebaran penyakit atau meramalkan curah hujan. Dengan demikian, karakteristik utama dari analisis ini adalah memberikan wawasan berbasis historis guna mendukung pengambilan keputusan yang lebih tepat dan terukur.

Secara umum, analisis deret waktu melibatkan beberapa komponen utama, yaitu tren (*trend*), pola musiman (*seasonality*), siklus (*cycle*), dan variasi acak (*random variation*). Tren menggambarkan arah umum data dalam jangka panjang seperti meningkat, menurun, atau stabil. Pola musiman menunjukkan fluktuasi periodik yang terjadi secara konsisten dalam interval waktu tertentu, seperti kenaikan penjualan menjelang hari raya. Sementara itu, komponen siklus mencerminkan fluktuasi jangka panjang yang terkait dengan kondisi ekonomi atau bisnis yang lebih luas. Adapun variasi acak menunjukkan fluktuasi yang bersifat tidak teratur dan sulit diprediksi karena dipengaruhi oleh faktor-faktor luar yang tidak dapat dikendalikan.

Salah satu pendekatan klasik dari metode deret waktu adalah model ARIMA (*Autoregressive Integrated Moving Average*). Model ini efektif untuk memodelkan data stasioner maupun non stasioner dengan mempertimbangkan korelasi waktu dan perbedaan antar observasi. Selain itu, pendekatan seperti *exponential smoothing*,

model musiman, serta metode berbasis *machine learning* seperti *Recurrent Neural Network* (RNN) atau *Long Short-Term Memory* (LSTM) juga mulai banyak digunakan, terutama ketika data memiliki struktur yang kompleks dan jumlah observasi besar (Hyndman & Athanasopoulos, 2018).

Keunggulan analisis deret waktu terletak pada kemampuannya dalam mengidentifikasi keteraturan atau pola tersembunyi dalam data masa lalu untuk memprediksi masa depan. Namun demikian, akurasi peramalan sangat bergantung pada kualitas data dan pemilihan model yang sesuai. Oleh karena itu, sebelum melakukan peramalan, perlu dilakukan eksplorasi data awal, uji stasioneritas, serta evaluasi performa model melalui metode seperti *time series cross-validation* atau analisis kesalahan peramalan (*forecast error*). Dengan penerapan yang cermat, analisis deret waktu dapat berfungsi sebagai instrumen yang efektif untuk mendukung perencanaan strategis baik dalam jangka pendek maupun jangka panjang.

8.2 Komponen Deret Waktu (*Trend*, Musiman, Siklus, dan *Irregular*)

Analisis deret waktu (*time series analysis*) merupakan metode statistik yang digunakan untuk mempelajari pola perubahan data berdasarkan urutan waktu. Tujuannya adalah memahami dinamika historis suatu fenomena dan melakukan peramalan (*forecasting*) terhadap kejadian di masa depan. Setiap deret waktu

pada dasarnya terdiri dari beberapa komponen utama, yaitu *trend* (T), musiman (S), siklus (C), dan irregular (I). Pemisahan atau dekomposisi komponen-komponen ini memungkinkan analisis untuk mengidentifikasi sumber variasi data dan memilih model peramalan yang paling sesuai.

8.2.1 Komponen *Trend* (T)

Komponen *trend* menggambarkan pola data dalam jangka panjang yang meningkat, menurun, atau stabil. *Trend* biasanya disebabkan oleh faktor struktural seperti pertumbuhan ekonomi, perkembangan teknologi, atau perubahan demografi.

Sebagai contoh, peningkatan penjualan kendaraan listrik dari tahun ke tahun mencerminkan adanya *upward trend* akibat transisi energi global. Dalam analisis statistik, *trend* dapat diestimasi menggunakan metode seperti regresi linier sederhana, *moving average*, atau *smoothing eksponensial*.

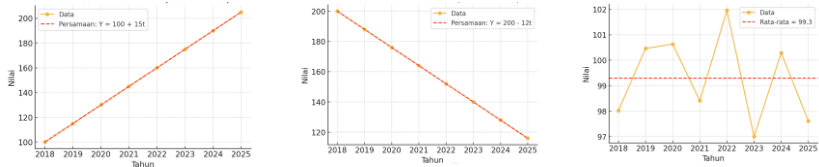
Model matematis umum untuk *trend linier* yang melibatkan data deret waktu/ nilai variabel pada waktu t (Y_t) adalah:

$$Y_t = a + bt$$

Model tersebut melibatkan nilai konstanta a (intersep), dan b menunjukkan arah serta laju perubahan *trend*. Nilai $b > 0$ menunjukkan *trend* meningkat, sedangkan $b < 0$ menunjukkan penurunan jangka panjang.

Menurut Hyndman dan Athanasopoulos (2021), pengenalan pola *trend* merupakan langkah pertama dalam analisis deret waktu karena kesalahan identifikasi dapat menyebabkan bias dalam

peramalan jangka panjang. Ilustrasi dari *trend* naik, *trend* turun, ataupun stabil dapat dilihat pada Gambar berikut



Gambar 1. Ilustrasi tren naik, turun dan stabil

8.2.2 Komponen Musiman (S)

Komponen musiman menunjukkan fluktuasi periodik yang berulang secara teratur dalam interval waktu tertentu, seperti harian, bulanan, atau triwulanan. Pola ini sering kali dipengaruhi oleh faktor kalender atau kondisi sosial-budaya. Contohnya, peningkatan konsumsi listrik pada musim panas, lonjakan penjualan ritel menjelang hari raya, atau penurunan jumlah wisatawan pada musim hujan.

Komponen musiman bersifat tetap dalam jangka waktu panjang dan dapat diestimasi dengan metode dekomposisi musiman atau model aditif dan multiplikatif. Model tersebut melibatkan nilai *trend* (T_t), musiman/periode yang berulang (S_t), siklus/fluktuasi jangka panjang yang tidak tetap periodenya (C_t) dan komponen acak (I_t).

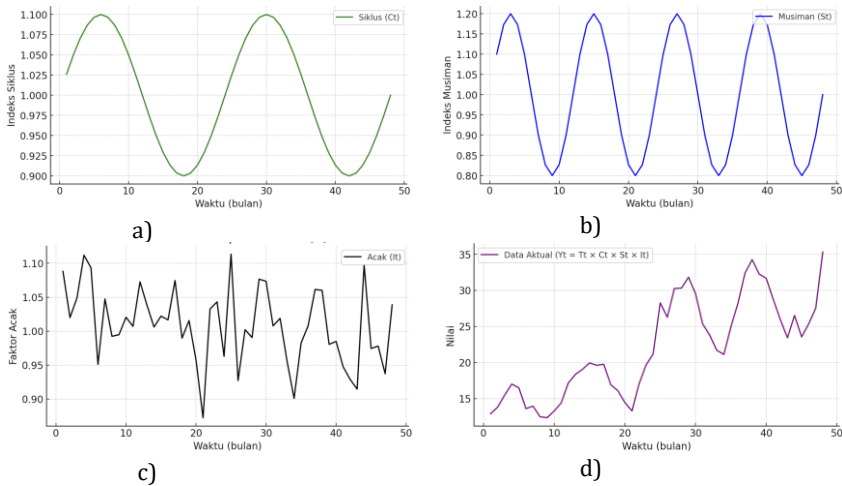
Secara umum, model deret waktu aditif dinyatakan sebagai:

$$Y_t = T_t + S_t + C_t + I_t$$

Sedangkan untuk model multiplikatif:

$$Y_t = T_t \times S_t \times C_t \times I_t$$

Pemilihan bentuk model bergantung pada konsistensi fluktuasi musiman; jika variasi musiman konstan, digunakan model aditif, sementara bila bervariasi proporsional terhadap nilai rata-rata data maka model multiplikatif lebih sesuai. Ilustrasi setiap komponen dapat dilihat pada Gambar berikut.



Gambar 1. Ilustrasi data penjualan deret waktu d) dengan setiap komponen a)-c)

8.2.3 Komponen Siklus (C)

Komponen siklus mencerminkan fluktuasi jangka panjang yang tidak beraturan tetapi berulang secara berkala, biasanya berlangsung selama lebih dari satu tahun. Faktor penyebab utama adalah perubahan ekonomi makro seperti inflasi, resesi, atau ekspansi bisnis.

Siklus ekonomi sering digambarkan dalam empat fase: ekspansi, puncak, kontraksi, dan resesi. Analisis siklus membantu

perencana kebijakan ekonomi dan pelaku bisnis untuk menentukan waktu yang tepat dalam mengambil keputusan investasi.

Deteksi komponen siklus sering kali dilakukan dengan filter statistik, seperti *Hodrick-Prescott filter* atau *Baxter-King filter*, yang memisahkan komponen jangka panjang dan jangka pendek dari data ekonomi.

8.2.4 Komponen *Irregular (I)*

Komponen *irregular* menggambarkan perubahan acak atau fluktuasi tidak teratur yang tidak dapat dijelaskan oleh faktor *trend*, *musiman*, atau *siklus*. Komponen ini biasanya muncul akibat peristiwa tak terduga seperti bencana alam, krisis politik, pandemi, atau fluktuasi pasar mendadak.

Dalam analisis statistik, *irregular component* dianggap sebagai *random noise* yang diharapkan berdistribusi acak dan tidak memiliki pola sistematis. Meskipun demikian, pada masa modern, analisis *irregular shock* menjadi penting karena guncangan eksternal dapat memiliki dampak signifikan terhadap pola jangka panjang, terutama dalam konteks ekonomi dan kesehatan masyarakat (Wei, 2019).

8.2.5 Integrasi dan Aplikasi Analisis Komponen Deret Waktu

Dengan mengenali keempat komponen utama tersebut, peneliti dapat membangun model peramalan yang lebih akurat dan interpretatif. Misalnya, model ARIMA (*AutoRegressive Integrated Moving Average*) digunakan ketika data menunjukkan *trend* dan *irregular noise*, sementara model SARIMA (*Seasonal ARIMA*) digunakan ketika terdapat pola musiman.

Pemahaman yang mendalam terhadap struktur deret waktu membantu pengambil keputusan dalam merancang strategi berbasis data yang adaptif terhadap perubahan jangka pendek maupun jangka panjang.

8.3 Model Peramalan (*Autoregressive, Moving Average, Exponential Smoothing, ARIMA*)

Peramalan deret waktu (*time series forecasting*) merupakan metode analisis statistik yang digunakan untuk memprediksi nilai di masa depan berdasarkan pola data historis. Metode ini memiliki peran penting dalam pengambilan keputusan bisnis, ekonomi, dan manajemen operasional, karena memungkinkan perencanaan yang lebih akurat dan responsif terhadap perubahan.

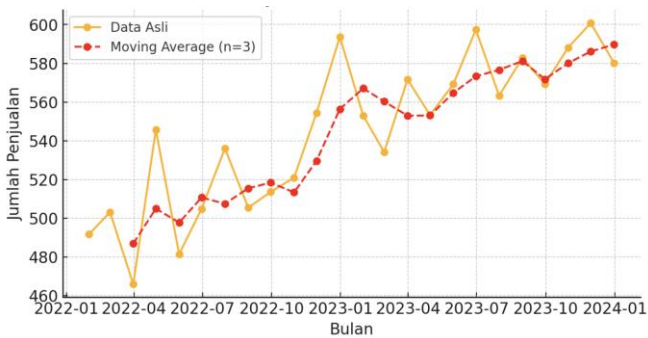
8.3.1 *Moving Average (MA)*

Model *Moving Average* bekerja dengan menghitung rata-rata dari sejumlah observasi dalam rentang waktu tertentu, sehingga fluktuasi jangka pendek dapat dihaluskan dan tren jangka panjang menjadi lebih jelas. Pendekatan ini sederhana namun efektif untuk data yang relatif stabil tanpa tren musiman yang kompleks. MA digunakan untuk identifikasi pola dasar dan sebagai dasar perbandingan dengan model yang lebih kompleks. Meskipun bersifat sederhana, metode ini cukup efektif untuk menganalisis data yang relatif stabil dan tidak memiliki pola musiman yang rumit. Pendekatan MA sering digunakan untuk mengidentifikasi pola umum dalam data serta menjadi acuan awal sebelum penerapan

model deret waktu yang lebih kompleks. Persamaan *Moving Average* dengan menerapkan penghalusan ke- n dapat dituliskan berdasarkan persamaan berikut.

$$\bar{Y}_t = \frac{Y_{t-1} + Y_{t-2} + \dots + Y_{t-n}}{n}$$

Sebagai contoh diberikan data penjualan mobil bulanan selama dua tahun (2022–2023) yang menunjukkan fluktuasi yang cukup tajam dari bulan ke bulan. Dengan menerapkan rata-rata bergerak tiga bulan, variasi jangka pendek dapat dihaluskan, sehingga arah kenaikan (tren naik) menjadi lebih jelas terlihat seperti yang diilustrasikan pada Gambar berikut.



Gambar 2. Ilustrasi data penjualan mobil bulanan dan model MA

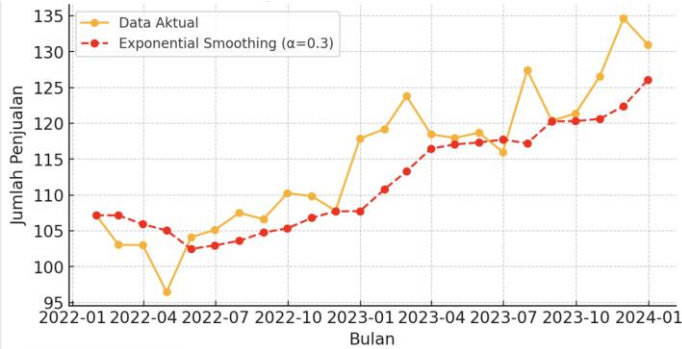
8.3.2 Exponential Smoothing (ES)

Metode *Exponential Smoothing* memberikan bobot yang lebih besar pada data terbaru dibandingkan dengan data lama, sehingga model ini lebih sensitif dan cepat menyesuaikan diri terhadap perubahan pola dalam data. Pendekatan ini sangat sesuai untuk deret waktu yang menunjukkan tren yang relatif konsisten, karena kemampuannya menghasilkan prediksi yang adaptif terhadap

dinamika baru. Jenis lain dari metode *Exponential Smoothing* adalah *Holt's Linear Trend Method* dan *Holt–Winters Seasonal Method*. Model tersebut memungkinkan pemodelan tren linear maupun pola musiman yang menjadikannya metode fleksibel dan efisien dalam berbagai konteks peramalan deret waktu. Model *Exponential Smoothing* yang paling sederhana, tanpa melibatkan tren dan musiman, dapat dituliskan berdasarkan persamaan berikut

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)Y_t$$

Nilai α menyatakan *smoothing constant*) yang terbatas antara $0 < \alpha < 1$. Semakin besar α semakin besar bobot pada data terbaru sehingga model lebih responsif terhadap perubahan. Kebalikannya, semakin kecil α model lebih halus sehingga perubahan jangka pendek kurang direspons.



Gambar 2. Ilustrasi data penjualan mobil bulanan dan model *Exponential Smoothing*

8.3.3 ARIMA (*AutoRegressive Integrated Moving Average*)

Model ARIMA (*Autoregressive Integrated Moving Average*) merupakan pendekatan statistik yang menggabungkan tiga komponen utama, yaitu autokorelasi (*autoregressive*), pembedaan

(integrated), dan rata-rata bergerak (moving average). Model ini dirancang untuk menangkap pola ketergantungan jangka pendek maupun jangka panjang dalam data, termasuk tren dan fluktuasi acak yang kompleks. ARIMA dapat diterapkan pada data stasioner maupun nonstasioner melalui proses diferensiasi untuk menghilangkan tren.

Dalam penerapannya, pembangunan model ARIMA melibatkan beberapa tahap penting: pengujian stasioneritas, identifikasi dan estimasi parameter (p, d, q), serta validasi model melalui analisis residual dan pengukuran kesalahan prediksi. Karena fleksibilitas dan keakuratannya, ARIMA menjadi salah satu model paling populer dalam analisis deret waktu ekonomi dan bisnis, terutama untuk peramalan penjualan, permintaan pasar, harga saham, dan berbagai indikator makroekonomi (Box, Jenkins, Reinsel, & Ljung, 2015).

Secara umum model ARIMA(p, d, q) yang melibatkan nilai deret waktu pada waktu ke-t (Y_t) ditulis sebagai berikut:

$$\Phi_p(B)(1 - B)^d Y_t = \Theta_q(B)e_t$$

Keterangan:

B: operator *backshift* ($BY_t = Y_{t-1}$)

$\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$: formula *autoregressive*

$\Theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$: formula *moving average*

e_t : error acak (*white noise*)

d: orde diferensiasi (jumlah kali data di-*difference* agar stasioner)

Langkah-langkah dalam menerapkan model ARIMA dapat ditulis sebagai berikut:

a. Uji stasioneritas

Tahapan ini memastikan bahwa data deret waktu memiliki sifat stasioner, yaitu nilai rata-rata dan variansnya tidak berubah sepanjang waktu. Adapun ciri-ciri data stasioner adalah tidak terdapat tren, varians relatif konstan, dan autokorelasi menurun seiring meningkatnya lag. Peneliti dapat melakukan uji stasioner dengan cara melakukan plot apakah data berfluktuasi di sekitar nilai rata-rata tetap atau melakukan uji statistik seperti *Augmented Dickey–Fuller* (ADF), *Phillips–Perron* (PP), dan *Kwiatkowski–Phillips–Schmidt–Shin* (KPSS). Jika tidak stasioner dalam rata-rata dapat dilakukan *differencing* atau transformasi box-cox untuk stasioner dalam varians.

b. Identifikasi orde model p dan q

Orde model p dan q dari model ARIMA dapat ditentukan menggunakan plot ACF (*Autocorrelation Function*) dan PACF (*Partial Autocorrelation Function*) dari data yang sudah stasioner.

Model	ACF	PACF
AR(p)	Menurun bertahap pada lag ke-p	Terpotong pada lag ke-q
MA(q)	Terpotong pada lag ke-q	Menurun bertahap pada lag ke-p
ARMA(p,q)	Menurun keduanya	Menurun keduanya

Sebagai contoh jika ACF terpotong pada lag 1 dan PACF menurun bertahap pada lag 1 maka dugaan awal ARIMA(0,0,1).

c. Estimasi parameter

Menaksir parameter pada model ARIMA(p,d,q) bertujuan untuk meminimalkan kesalahan model terhadap data aktual. Adapun metode yang biasanya digunakan adalah *Maximum Likelihood Estimation* (MLE) dan *Least Squares Estimation* (LSE)

d. Diagnostik model

Lanngkah ini bertujuan untuk memastikan bahwa model yang dipilih layak dan residual-nya bersifat *white noise* (acak, tidak berkorelasi). Pengujiannya dapat dilakukan dengan plot ACF dan PACF residul (pola acak), menggunakan *uji Ljung-Box Q* dan uji normalitas residual.

e. Peramalan

Menggunakan model ARIMA terbaik untuk dilakukan tahapan di masa mendatang.

8.3.4 Pemilihan Model

Pemilihan model peramalan yang tepat bergantung pada karakteristik data, kestasioneran, dan tujuan peramalan. Data dengan pola sederhana dapat menggunakan MA atau ES, sedangkan data dengan tren kompleks, musiman, atau autokorelasi tinggi lebih cocok menggunakan ARIMA. Evaluasi performa model melalui *mean absolute error* (MAE), *root mean square error* (RMSE), dan

Akaike Information Criterion (AIC) menjadi langkah penting untuk memastikan akurasi prediksi.

8.4 Penilaian Model Peramalan

Setelah model peramalan dikembangkan, evaluasi menjadi tahap krusial untuk menilai akurasi dan keandalan prediksi. Evaluasi ini memastikan bahwa model dapat digunakan secara efektif untuk pengambilan keputusan dan perencanaan berdasarkan data historis. Penggunaan metrik evaluasi yang tepat dan metode validasi yang sistematis memungkinkan pemilihan model terbaik dan mencegah *overfitting*.

8.4.1 Metrik Evaluasi

Beberapa metrik evaluasi umum digunakan untuk menilai kualitas prediksi model peramalan:

1. *Mean Absolute Error* (MAE): Mengukur rata-rata absolut selisih antara nilai aktual (Y_t) dan nilai prediksi (\hat{Y}_t). MAE memberikan indikasi seberapa besar kesalahan prediksi secara rata-rata tanpa memperhitungkan arah kesalahan.

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$

2. *Mean Squared Error* (MSE): Menghitung rata-rata kuadrat selisih antara nilai aktual (Y_t) dan prediksi (\hat{Y}_t). MSE menekankan kesalahan besar lebih daripada kesalahan kecil karena adanya kuadrat, sehingga berguna untuk menilai stabilitas model.

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$$

3. *Root Mean Squared Error* (RMSE): merupakan akar kuadrat dari MSE, sehingga mengembalikan satuan hasil ke dalam satuan asli data. Semakin kecil nilai RMSE, semakin baik kinerja model namun tetap memperhatikan satuan dari data yang digunakan.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$$

4. *Mean Absolute Percentage Error* (MAPE): Menyatakan kesalahan prediksi sebagai persentase dari nilai aktual, sehingga mempermudah interpretasi dan perbandingan antar-model atau dataset yang berbeda.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right|$$

Penggunaan MAPE lebih baik untuk membandingkan akurasi antar-model di dataset berbeda skala, karena bersifat persentase (%).

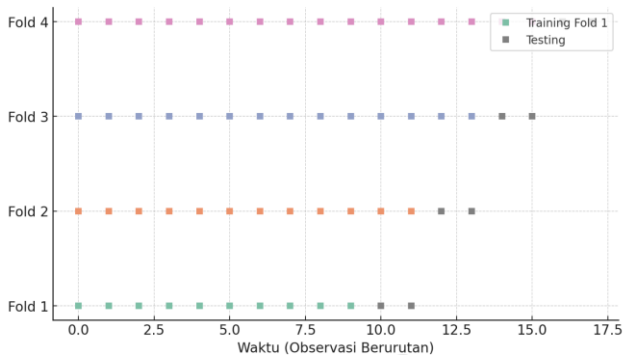
Penggunaan kombinasi metrik ini membantu analisis lebih komprehensif terhadap performa model, karena masing-masing metrik memiliki sensitivitas berbeda terhadap variasi data dan *outlier*.

8.4.2 Validasi Silang (*Cross-Validation*)

Selain metrik evaluasi, validasi silang (*cross-validation*) menjadi metode penting untuk menguji kestabilan model terhadap data baru. Proses validasi silang (*cross-validation*) juga merupakan langkah krusial dalam menilai keandalan dan kestabilan model peramalan terhadap data baru. Teknik ini bekerja dengan cara membagi dataset menjadi beberapa subset di mana model secara bergantian dilatih pada sebagian data (*training set*) dan diuji pada subset lainnya (*validation set*). Melalui pendekatan ini, setiap bagian data berkesempatan menjadi data uji, sehingga evaluasi yang dihasilkan menjadi lebih objektif dan representatif.

Validasi silang berperan penting dalam menghindari *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri dengan data pelatihan sehingga kehilangan kemampuan generalisasi pada data baru. Dengan menerapkan *k-fold cross-validation* atau *rolling-origin evaluation* khusus untuk deret waktu, analisis dapat mengukur seberapa stabil performa model dalam berbagai skenario temporal. Pendekatan ini memastikan bahwa model tidak hanya bekerja baik pada data historis, tetapi juga mampu menghasilkan prediksi yang konsisten dan andal ketika diterapkan pada data aktual di masa depan.

Dengan demikian, validasi silang tidak hanya meningkatkan kepercayaan terhadap hasil peramalan, tetapi juga memperkuat dasar pengambilan keputusan berbasis data di berbagai bidang seperti ekonomi, keuangan, dan kebijakan publik.



Gambar 3. Ilustrasi *Cross Validation*

8.4.3 Implikasi Evaluasi Model

Proses evaluasi yang dilakukan secara tepat dan terukur memungkinkan analisis untuk menentukan model yang paling akurat dan efisien dalam merepresentasikan pola data yang sebenarnya. Model yang telah melalui tahap evaluasi sistematis dapat dimanfaatkan baik untuk peramalan jangka pendek maupun jangka panjang dengan tingkat kesalahan yang relatif rendah. Lebih jauh lagi, hasil evaluasi menjadi landasan penting dalam proses optimasi model, seperti pemilihan parameter yang optimal, penerapan transformasi data yang sesuai, hingga pengembangan model hibrida guna meningkatkan kinerja prediksi.

Dengan penerapan kombinasi metrik evaluasi yang relevan serta validasi silang yang memadai, model peramalan dapat diandalkan sebagai alat bantu pengambilan keputusan yang efektif, presisi, dan adaptif terhadap dinamika perubahan data.

8.5 Latihan Soal

1. Jelaskan pengertian analisis deret waktu dan manfaatnya dalam dunia bisnis.
2. Sebutkan dan jelaskan empat komponen utama dalam deret waktu.
3. Apa perbedaan antara metode *Moving Average* dan *Exponential Smoothing*?
4. Bagaimana prinsip kerja model ARIMA dalam peramalan ekonomi?
5. Tuliskan model ARIMA(0,0,1), ARIMA(1,1,0), dan ARIMA(1,0,0)
6. Sebutkan tiga ukuran yang digunakan dalam mengevaluasi ketepatan model peramalan.

Bab 9: Analisis Data dengan *Software* Statistik

9.1 Pendahuluan

Dalam era ekonomi digital yang ditandai oleh ledakan informasi dan pertumbuhan data yang eksponensial, analisis data dengan bantuan *software* statistik menjadi komponen esensial dalam pengambilan keputusan yang berbasis bukti. Di sektor ekonomi dan bisnis, pengolahan data bukan lagi sekadar aktivitas pendukung, melainkan telah menjadi inti dari proses perencanaan strategis, evaluasi kinerja, hingga perumusan kebijakan. Penggunaan perangkat lunak statistik seperti SPSS, Stata, R, atau Python memungkinkan para analis dan peneliti untuk mengelola data dalam skala besar dengan efisiensi tinggi, mengidentifikasi pola dan tren tersembunyi, serta menyusun model prediktif yang kompleks. Keunggulan teknologi ini terletak pada kemampuannya dalam mempercepat proses analisis secara sistematis dan mengurangi risiko kesalahan manual yang umum terjadi dalam pengolahan data konvensional.

Di dunia akademik, penguasaan *software* statistik menjadi keterampilan dasar yang harus dimiliki oleh mahasiswa dan peneliti di berbagai disiplin ilmu, terutama di bidang ekonomi, manajemen, dan ilmu sosial. Analisis data kuantitatif yang valid dan reliabel

mebutuhkan alat bantu yang mampu menangani berbagai teknik statistik, mulai dari uji deskriptif, regresi, analisis multivariat, hingga *time series forecasting*. Selain itu, perangkat lunak statistik juga memungkinkan visualisasi data yang lebih informatif sehingga mempermudah interpretasi hasil bagi pengambil keputusan. Seiring dengan meningkatnya tuntutan transparansi dan akuntabilitas dalam pengelolaan bisnis, keahlian dalam mengolah dan menyajikan data statistik menjadi nilai tambah yang sangat signifikan dalam dunia profesional. Penelitian menunjukkan bahwa penggunaan *software* statistik secara efektif dapat meningkatkan kualitas riset, memperkuat validitas temuan, dan mempercepat siklus inovasi dalam organisasi (Grolemond & Wickham, 2017).

Dalam praktik bisnis, data tidak hanya digunakan untuk melihat kondisi saat ini, tetapi juga sebagai dasar perumusan strategi yang bersifat adaptif terhadap perubahan pasar. Dengan dukungan perangkat statistik, perusahaan dapat melakukan segmentasi pelanggan, analisis perilaku konsumen, peramalan penjualan, dan evaluasi efisiensi operasional secara lebih akurat. Oleh karena itu, integrasi antara pemahaman statistik dan kemampuan teknis dalam penggunaan *software* menjadi kompetensi utama yang dibutuhkan di berbagai sektor, baik swasta maupun publik (Kelleher & Tierney, 2018). Secara keseluruhan, kemajuan dalam teknologi pengolahan data telah mengubah paradigma analisis dari yang bersifat deskriptif menuju ke arah prediktif dan preskriptif, menjadikan *software* statistik sebagai alat yang tak tergantikan dalam mendukung keunggulan kompetitif dan pengambilan keputusan berbasis data.

9.2 Jenis-Jenis Software Statistik

9.2.1 SPSS dan STATA: Kemudahan dan Efisiensi Analisis Data Sosial

Statistical Package for the Social Sciences (SPSS) dan STATA merupakan dua perangkat lunak statistik yang banyak digunakan dalam penelitian sosial, pendidikan, dan kesehatan. SPSS dikenal karena antarmuka grafisnya yang ramah pengguna dan kemampuan analisis deskriptif serta inferensial yang cepat, seperti uji *t*, ANOVA, korelasi, dan regresi linier. Fitur *point-and-click* pada SPSS sangat membantu pengguna non-programmer dalam mengolah data tanpa harus menulis sintaks rumit. Sementara itu, STATA menawarkan efisiensi dalam mengelola data berskala besar dan mendalam, terutama dalam bidang ekonomi dan epidemiologi. Keunggulan STATA terletak pada dokumentasi yang sistematis serta struktur perintah yang konsisten, sehingga cocok untuk analisis panel, *survival analysis*, dan pemodelan ekonomi (Acock, 2018).

9.2.2 R dan Python: Fleksibilitas untuk Analisis Lanjutan

R dan Python merupakan *open-source software* yang sangat fleksibel untuk analisis statistik dan pemodelan data tingkat lanjut. R unggul dalam visualisasi data dan pengembangan metode statistik baru. Dengan ribuan paket tambahan yang tersedia di CRAN (*Comprehensive R Archive Network*), R dapat digunakan untuk analisis multivariat, *time series*, hingga *machine learning*. Python, di sisi lain, memiliki keunggulan sebagai bahasa pemrograman general-purpose yang juga sangat kuat dalam analisis data berkat

pustaka seperti *pandas*, *NumPy*, *scikit-learn*, dan *matplotlib*. Python juga unggul dalam integrasi dengan teknologi web dan big data, menjadikannya pilihan populer di bidang data science dan kecerdasan buatan (Ozdemir & Susarla, 2022). Kedua perangkat lunak ini membutuhkan keterampilan pemrograman yang lebih tinggi, tetapi memberikan kebebasan penuh dalam eksplorasi dan replikasi analisis.

9.2.3 SAS: Standar Industri dalam Analisis Statistik Terstruktur

SAS (*Statistical Analysis System*) banyak digunakan di industri dan institusi pemerintah karena stabilitas, keamanan data, dan keandalannya dalam pengolahan data skala besar. SAS unggul dalam pemrosesan data terstruktur dan sering digunakan dalam penelitian klinis, aktuarial, dan lembaga finansial. Fitur SAS mencakup manajemen data yang kuat, analisis prediktif, dan pelaporan otomatis. Meskipun tidak sepopuler R atau Python dalam komunitas *open-source*, SAS tetap menjadi standar industri karena reputasinya yang andal dan dukungan teknis yang komprehensif. Biaya lisensinya yang relatif tinggi menjadi pertimbangan utama dalam adopsi, terutama di sektor pendidikan dan organisasi non-profit.

9.3 Prosedur Umum Analisis Data

Analisis data merupakan proses sistematis untuk mengubah data mentah menjadi informasi yang bermakna dalam rangka mendukung pengambilan keputusan atau pengujian hipotesis. Dalam praktik penelitian kuantitatif maupun kegiatan operasional berbasis data, prosedur analisis harus dilakukan secara terstruktur dan akurat. Prosedur umum ini biasanya mengikuti tahapan teknis tertentu, dimulai dari impor data ke dalam perangkat lunak analisis hingga tahap interpretasi hasil. Pemilihan metode analisis bergantung pada jenis data, tujuan analisis, dan kerangka teoritis yang mendasari (Hair et al., 2019).

9.3.1 Impor dan Pembersihan Data

Langkah pertama dalam analisis data adalah mengimpor data dari berbagai sumber ke dalam *software* statistik seperti SPSS, R, Python, atau Excel. Sumber data dapat berupa *spreadsheet*, basis data relasional, atau hasil survei daring. Setelah data berhasil diimpor, dilakukan pembersihan data (*data cleaning*) untuk mengatasi nilai kosong, kesalahan input, atau duplikasi. Pembersihan data bertujuan memastikan keakuratan dan konsistensi, sehingga tidak terjadi bias dalam proses analisis.

9.3.2 Eksplorasi dan Analisis Statistik Deskriptif

Setelah data dibersihkan, dilakukan eksplorasi data untuk memahami pola umum, distribusi variabel, serta hubungan awal antar variabel. Tahap ini mencakup visualisasi menggunakan grafik batang, histogram, diagram kotak, dan tabel frekuensi. Selanjutnya,

analisis statistik deskriptif dilakukan untuk menggambarkan karakteristik utama data melalui ukuran seperti rata-rata, median, modus, standar deviasi, dan rentang. Analisis ini memberikan pemahaman dasar terhadap dataset sebelum dilakukan analisis lanjutan (Field, 2018).

9.3.3 Analisis Inferensial dan Interpretasi

Analisis inferensial dilakukan untuk menguji hipotesis dan menarik kesimpulan yang dapat digeneralisasi ke populasi. Metode yang umum digunakan meliputi uji t, ANOVA, regresi linear, korelasi, dan uji *chi-square*, tergantung pada jenis data dan desain penelitian. Hasil analisis kemudian diinterpretasikan dalam konteks tujuan studi. Visualisasi data melalui grafik dan diagram sangat membantu dalam menyampaikan hasil kepada pemangku kepentingan secara jelas dan informatif. Tahap akhir ini penting untuk mengaitkan temuan statistik dengan implikasi praktis atau teoritis.

9.4 Contoh Kasus Penggunaan *Software* Statistik

Penggunaan *software* statistik menjadi bagian penting dalam analisis data ekonomi, sosial, maupun penelitian terapan lainnya. *Software* seperti SPSS, Stata, atau R memungkinkan peneliti melakukan pengolahan data secara sistematis dan akurat, mulai dari deskripsi data hingga analisis inferensial. Studi kasus berikut menggambarkan penerapan *software* statistik dalam menilai dampak

kebijakan fiskal terhadap perilaku konsumsi rumah tangga. Analisis ini penting untuk memahami efektivitas kebijakan pemerintah dalam meningkatkan kesejahteraan masyarakat serta mengoptimalkan alokasi sumber daya (Field, 2018).

9.4.1 Persiapan Data dan Variabel Penelitian

Analisis dimulai dengan pengumpulan data dari Badan Pusat Statistik (BPS), termasuk informasi mengenai konsumsi rumah tangga, pengeluaran pemerintah, dan tingkat pajak. Variabel dependen dalam penelitian ini adalah konsumsi rumah tangga, sedangkan variabel independen mencakup pengeluaran pemerintah dan tingkat pajak. Data yang dikumpulkan harus terlebih dahulu dibersihkan dan diperiksa kualitasnya, termasuk menangani *missing values*, *outlier*, dan kesalahan *input*. Tahap ini krusial untuk memastikan hasil analisis dapat dipercaya dan akurat.

9.4.2 Analisis Regresi Linear menggunakan SPSS

Setelah data siap, analisis dilakukan menggunakan SPSS melalui prosedur regresi linear. Model regresi digunakan untuk mengestimasi hubungan antara konsumsi rumah tangga dengan pengeluaran pemerintah dan tingkat pajak. Hasil regresi meliputi koefisien regresi, nilai *p-value*, dan *R-squared*, yang membantu peneliti menilai signifikansi statistik dan kekuatan hubungan antarvariabel. Interpretasi hasil ini memungkinkan evaluasi dampak kebijakan fiskal terhadap perilaku konsumsi secara empiris.

9.4.3 Implikasi Hasil dan Pengambilan Keputusan

Hasil analisis dapat digunakan sebagai dasar dalam pengambilan keputusan kebijakan ekonomi. Misalnya, jika regresi

menunjukkan bahwa peningkatan pengeluaran pemerintah secara signifikan meningkatkan konsumsi rumah tangga, pemerintah dapat mempertimbangkan penyesuaian anggaran untuk memaksimalkan efek stimulasi ekonomi. Selain itu, penggunaan *software* statistik seperti SPSS memungkinkan evaluasi skenario alternatif melalui *sensitivity analysis*, sehingga kebijakan yang dirancang lebih responsif terhadap kondisi ekonomi dan sosial (Pallant, 2020).

9.5 Tantangan dan Solusi dalam Analisis Data Statistik

9.5.1 Kesulitan Teknis dalam Penggunaan Software Statistik

Penggunaan perangkat lunak statistik seperti SPSS, R, atau Stata sering kali menimbulkan tantangan teknis, khususnya bagi pengguna yang tidak memiliki latar belakang pemrograman. Kesulitan umum meliputi pemahaman terhadap sintaks, antarmuka pengguna (*interface*), serta struktur perintah yang kompleks. Di sisi lain, pembaruan versi perangkat lunak juga dapat memengaruhi kompatibilitas data dan fungsi analisis tertentu. Tantangan ini menjadi hambatan signifikan dalam proses eksplorasi dan analisis data, terutama ketika analisis yang dilakukan bersifat multivariat atau berskala besar (Sarkar & Rashid, 2016).

9.5.2 Manajemen Data Besar dan Kualitas Data

Tantangan lain terletak pada pengelolaan *big data*, yang mencakup volume besar, variasi jenis data, serta kecepatan pemrosesan. Data yang berasal dari berbagai sumber seperti sistem

informasi rumah sakit, survei daring, atau sensor medis sering kali tidak terstruktur dan memerlukan proses *cleaning* dan transformasi sebelum dapat dianalisis. Kualitas data yang rendah, seperti duplikasi, *missing values*, dan inkonsistensi format, dapat menghasilkan interpretasi statistik yang menyesatkan. Oleh karena itu, penting bagi analis data untuk menerapkan prinsip manajemen data yang ketat serta menggunakan perangkat lunak yang mendukung pemrosesan data dalam jumlah besar secara efisien (Kitchin, 2015).

9.5.3 Strategi Penguatan Kompetensi dan Dukungan Analitis

Untuk mengatasi berbagai tantangan tersebut, pendekatan solutif mencakup pelatihan intensif yang berfokus pada peningkatan literasi statistik dan keterampilan teknis dalam penggunaan perangkat lunak. Pelatihan berbasis kasus (*case-based training*) dinilai efektif karena memungkinkan peserta memahami konteks analisis nyata dan meningkatkan kemampuan interpretatif. Selain itu, penggunaan dokumentasi resmi dan tutorial daring dari pengembang perangkat lunak dapat membantu pengguna memecahkan masalah secara mandiri. Komunitas daring seperti forum diskusi dan grup pengguna juga menyediakan ruang kolaboratif untuk berbagi solusi dan praktik terbaik. Pendekatan ini tidak hanya meningkatkan efisiensi analisis, tetapi juga memperluas kemampuan adaptasi terhadap dinamika kebutuhan data dan teknologi statistik.

9.6 Latihan Soal

1. Sebutkan lima *software* statistik yang umum digunakan dalam analisis data ekonomi dan bisnis.
2. Apa saja tahapan dasar dalam melakukan analisis data menggunakan *software* statistik?
3. Jelaskan secara singkat studi kasus penggunaan SPSS untuk analisis regresi.
4. Apa saja tantangan umum yang dihadapi dalam menggunakan *software* statistik, dan bagaimana mengatasinya?
5. Bandingkan kelebihan dan kekurangan penggunaan R dan SPSS.

Bab 10: Statistika dalam Pengambilan Keputusan Bisnis

10.1 Pengertian dan Cakupan Statistika

Statistika merupakan cabang ilmu yang berfokus pada pengumpulan, pengolahan, analisis, interpretasi, dan penyajian data untuk membuat keputusan yang lebih tepat dan informatif. Dalam konteks bisnis, statistika berfungsi sebagai alat penting dalam pengambilan keputusan yang didasarkan pada data empiris. Oleh karena itu, pemahaman yang mendalam mengenai statistika sangat diperlukan oleh para profesional bisnis untuk memahami pola dan tren yang ada dalam data serta membuat keputusan yang lebih rasional dan berbasis bukti.

10.1.1 Pengertian Statistika dalam Bisnis

Secara umum, statistika dapat didefinisikan sebagai ilmu yang mempelajari cara-cara untuk mengumpulkan, menganalisis, dan menginterpretasi data untuk membuat keputusan yang lebih tepat. Dalam konteks bisnis, statistika digunakan untuk memahami berbagai fenomena yang terjadi dalam operasional perusahaan, seperti tren pasar, perilaku konsumen, atau efisiensi operasional. Data yang diolah melalui metode statistik memungkinkan perusahaan untuk mengambil keputusan berdasarkan informasi yang

lebih objektif, mengurangi ketidakpastian, dan memprediksi kondisi yang akan datang.

Pentingnya statistika dalam bisnis adalah kemampuannya untuk mengubah data mentah menjadi informasi yang dapat digunakan untuk merencanakan strategi, mengidentifikasi peluang, dan mengelola risiko. Dengan pendekatan yang berbasis data, pengambilan keputusan tidak hanya mengandalkan intuisi atau pengalaman semata, melainkan juga informasi yang terukur dan dapat dipertanggungjawabkan (Ghozali, 2020).

10.1.2 Ruang Lingkup Statistika dalam Pengambilan Keputusan Bisnis

Ruang lingkup statistika dalam pengambilan keputusan bisnis sangat luas, mencakup berbagai metode dan teknik yang digunakan untuk mengumpulkan dan menganalisis data. Secara umum, statistika dalam bisnis dapat dibagi menjadi dua kategori besar: statistika deskriptif dan inferensial. Statistika deskriptif berfungsi untuk menyajikan data dalam bentuk yang lebih mudah dipahami, seperti menggunakan grafik, tabel, atau ukuran pusat tendensi seperti rata-rata, median, dan modus. Teknik ini berguna untuk memberikan gambaran umum tentang keadaan yang ada dalam perusahaan atau pasar.

Di sisi lain, statistika inferensial digunakan untuk membuat kesimpulan atau prediksi berdasarkan data sampel yang diambil dari populasi yang lebih besar. Dengan menggunakan metode ini, perusahaan dapat membuat generalisasi yang lebih luas atau meramalkan kondisi di masa depan, seperti estimasi permintaan

produk atau proyeksi keuntungan. Metode inferensial seperti uji hipotesis, analisis regresi, dan analisis varians memungkinkan manajer untuk menguji hubungan antar variabel, mengidentifikasi faktor-faktor yang mempengaruhi kinerja bisnis, dan menentukan langkah-langkah yang perlu diambil untuk meningkatkan hasil (Tufféry, 2021).

Selain itu, ruang lingkup statistika juga mencakup analisis multivariat, yang digunakan untuk mengidentifikasi hubungan kompleks antara berbagai variabel dalam satu waktu. Teknik ini sangat berguna dalam bisnis ketika banyak faktor yang saling terkait mempengaruhi hasil yang diinginkan, seperti dalam analisis segmentasi pasar atau evaluasi kinerja produk.

10.1.3 Penggunaan Statistika untuk Pengambilan Keputusan yang Lebih Baik

Penerapan statistika dalam pengambilan keputusan bisnis memungkinkan perusahaan untuk membuat keputusan yang lebih rasional dan berbasis data. Dengan menggunakan data yang dianalisis secara statistik, perusahaan dapat mengevaluasi berbagai alternatif yang ada dan memilih opsi terbaik yang memberikan hasil optimal. Misalnya, dalam keputusan pemasaran, perusahaan dapat menggunakan analisis regresi untuk memahami hubungan antara harga dan permintaan produk, atau menggunakan analisis cluster untuk mengidentifikasi segmen pasar yang paling potensial.

Selain itu, dengan statistika, manajer dapat mengukur dan memantau kinerja perusahaan secara berkelanjutan. Indikator kinerja utama (KPI) yang diukur menggunakan data statistik

membantu manajer untuk mengevaluasi apakah strategi yang diterapkan efektif atau perlu dilakukan perubahan. Dengan demikian, statistika berfungsi tidak hanya sebagai alat untuk pengambilan keputusan jangka pendek, tetapi juga sebagai dasar untuk perencanaan dan evaluasi jangka panjang dalam bisnis (Ghozali, 2020).

10.2 Pengumpulan dan Penyajian Data

Pengumpulan dan penyajian data merupakan tahap krusial dalam proses penelitian dan pengambilan keputusan berbasis bukti. Data yang valid, akurat, dan relevan menjadi fondasi bagi analisis, interpretasi, serta rekomendasi yang dapat diandalkan. Proses ini mencakup identifikasi sumber data, metode pengumpulan, serta teknik penyajian yang mempermudah pemahaman dan pengambilan keputusan oleh pemangku kepentingan.

Data dapat dikategorikan menjadi data primer, yang diperoleh langsung dari sumber asli melalui observasi, wawancara, survei, atau eksperimen, dan data sekunder, yang diperoleh dari dokumen, laporan, basis data, atau publikasi ilmiah yang sudah ada. Pemilihan jenis data dan metode pengumpulan harus mempertimbangkan tujuan penelitian, ketersediaan sumber daya, dan karakteristik populasi yang menjadi fokus studi.

10.2.1 Metode Pengumpulan Data Primer

Pengumpulan data primer memungkinkan peneliti mendapatkan informasi langsung dari subjek atau fenomena yang diteliti. Metode ini meliputi:

- a. Observasi: Pencatatan perilaku, kejadian, atau kondisi lingkungan secara sistematis. Observasi dapat bersifat partisipatif atau non-partisipatif, tergantung pada tujuan penelitian. Observasi memungkinkan peneliti memperoleh data nyata dan konteks sosial atau fisik yang relevan.
- b. Wawancara dan Kuesioner: Teknik ini digunakan untuk mengumpulkan informasi persepsi, opini, atau pengalaman individu. Wawancara dapat bersifat terstruktur, semi-terstruktur, atau tidak terstruktur, sedangkan kuesioner memungkinkan pengumpulan data dari sampel yang lebih besar. Validitas dan reliabilitas instrumen sangat penting untuk memastikan kualitas data.
- c. Eksperimen: Digunakan untuk menguji hipotesis dengan memanipulasi variabel dan mengamati efeknya. Eksperimen memberikan data yang bersifat kausal dan dapat digunakan untuk inferensi tentang hubungan antarvariabel.

Pengumpulan data primer memberikan kontrol lebih besar terhadap kualitas dan relevansi data, meskipun memerlukan waktu, biaya, dan sumber daya yang lebih besar dibandingkan data sekunder.

10.2.2 Metode Pengumpulan Data Sekunder

Data sekunder diperoleh dari sumber yang sudah tersedia, seperti jurnal ilmiah, laporan penelitian, dokumen organisasi, basis data statistik, dan arsip pemerintah. Penggunaan data sekunder memiliki keuntungan berupa efisiensi waktu dan biaya, serta memungkinkan analisis data historis atau komparatif. Namun, peneliti harus menilai keandalan dan relevansi data, termasuk validitas metode pengumpulan asli, kualitas dokumentasi, dan kesesuaian dengan tujuan penelitian.

Integrasi data sekunder dengan data primer sering digunakan untuk memperkuat analisis, memvalidasi temuan, dan memberikan konteks yang lebih luas bagi interpretasi hasil.

10.2.3 Teknik Penyajian Data

Penyajian data yang efektif memudahkan pembaca atau pemangku kepentingan dalam memahami informasi dan menarik kesimpulan yang tepat. Teknik penyajian data meliputi:

- a. **Tabel:** Menyajikan data secara terstruktur dengan baris dan kolom untuk membandingkan kategori atau variabel tertentu. Tabel cocok untuk data numerik atau kualitatif yang membutuhkan perbandingan jelas.
- b. **Grafik dan Diagram:** Meliputi diagram batang, diagram garis, diagram lingkaran, atau histogram. Representasi visual membantu mengkomunikasikan tren, distribusi, dan hubungan antarvariabel secara cepat dan intuitif.
- c. **Ringkasan Statistik:** Penyajian data dalam bentuk mean, median, modus, standar deviasi, atau persentase membantu

menyoroti pola utama dalam dataset. Teknik ini memudahkan interpretasi data yang besar atau kompleks.

Penyajian data harus mempertimbangkan audiens, tujuan analisis, dan konteks informasi agar hasil dapat diinterpretasikan dengan benar dan digunakan untuk pengambilan keputusan yang efektif (Creswell & Creswell, 2021).

Secara keseluruhan, pengumpulan dan penyajian data yang sistematis, valid, dan terstruktur menjadi pondasi bagi analisis yang akurat. Kombinasi metode data primer dan sekunder, disertai teknik penyajian yang tepat, memungkinkan peneliti dan organisasi untuk memahami fenomena, mengambil keputusan berbasis bukti, dan menghasilkan rekomendasi yang efektif.

10.3 Analisis Deskriptif

Analisis deskriptif merupakan salah satu jenis statistik yang digunakan untuk meringkas dan menggambarkan karakteristik dasar dari data yang dikumpulkan. Dalam penelitian dan analisis data, statistik deskriptif berfungsi untuk memberikan gambaran umum mengenai distribusi dan kecenderungan data tanpa melakukan pengujian hipotesis atau inferensi lebih lanjut. Dengan menggunakan statistik deskriptif, peneliti dapat dengan mudah memahami pola atau kecenderungan dalam data yang ada. Statistik deskriptif meliputi beberapa ukuran yang umum digunakan, seperti *mean*, *median*, *modus*, dan distribusi frekuensi, yang masing-masing memberikan informasi penting mengenai data yang dianalisis.

Penggunaan statistik deskriptif sangat berguna dalam berbagai bidang, baik itu dalam penelitian sosial, ekonomi, kesehatan, maupun pendidikan. Alat ini memungkinkan peneliti untuk membuat keputusan awal tentang data yang dikumpulkan sebelum melanjutkan dengan analisis lebih lanjut atau uji statistik yang lebih kompleks.

10.3.1 Ukuran Pemusatan Data: Mean, Median, dan Modus

Ukuran pemusatan data menggambarkan nilai tengah atau nilai yang mewakili suatu distribusi data. Tiga ukuran pemusatan data yang paling umum digunakan adalah *mean*, median, dan modus. Masing-masing ukuran ini memberikan informasi yang berbeda mengenai bagaimana data tersebar dan di mana pusat distribusi data tersebut.

- a. **Mean (Rata-rata):** *Mean* adalah ukuran pemusatan data yang paling sering digunakan. Ini dihitung dengan menjumlahkan semua nilai dalam data set dan membaginya dengan jumlah observasi. Rata-rata memberikan gambaran tentang nilai tengah dari data secara keseluruhan, namun sangat sensitif terhadap nilai ekstrim (*outlier*). Sebagai contoh, jika ada satu nilai yang sangat tinggi atau rendah dalam data set, *mean* dapat terdistorsi.

Formula untuk menghitung *mean* adalah:

$$\text{Mean} = \frac{\sum_{i=1}^n x_i}{n}$$

di mana x_i adalah setiap nilai dalam data set dan n adalah jumlah observasi.

- b. Median (Nilai Tengah):** Median adalah nilai yang membagi dataset menjadi dua bagian yang sama, dengan setengah data berada di atas nilai median dan setengah lainnya di bawahnya. Jika data set memiliki jumlah data yang ganjil, *median* adalah nilai yang berada di posisi tengah. Namun, jika jumlah data genap, *median* dihitung sebagai rata-rata dari dua nilai tengah.

Median tidak dipengaruhi oleh nilai ekstrem, menjadikannya lebih representatif dari distribusi data, terutama ketika data memiliki *outlier*.

- c. Modus (Nilai yang Paling Sering Muncul):** *Modus* adalah nilai yang paling sering muncul dalam data set. Dalam beberapa kasus, data dapat memiliki lebih dari satu modus jika ada beberapa nilai yang muncul dengan frekuensi yang sama. Data yang memiliki dua modus disebut sebagai bimodal, sedangkan data dengan lebih dari dua modus disebut multimodal.

Modus sangat berguna ketika ingin mengetahui nilai yang paling sering terjadi dalam data, tetapi tidak memberikan informasi tentang sebaran atau rentang data lainnya.

10.3.2 Ukuran Penyebaran Data: Rentang, Varians, dan Standar Deviasi

Selain ukuran pemusatan data, ukuran penyebaran data juga sangat penting untuk memahami variabilitas atau keragaman dalam dataset. Ukuran penyebaran yang umum digunakan meliputi rentang, varians, dan standar deviasi.

- a. Rentang (*Range*): Rentang adalah selisih antara nilai tertinggi dan nilai terendah dalam dataset. Meskipun mudah dihitung, rentang hanya memberikan gambaran kasar tentang sebaran data dan dapat sangat dipengaruhi oleh nilai ekstrem (outlier). Rentang dihitung dengan cara:

$$\text{Rentang} = X_{\max} - X_{\min}$$

di mana X_{\max} adalah nilai maksimum dan X_{\min} adalah nilai minimum dalam dataset.

- b. Varians (*Variance*): Varians mengukur seberapa jauh data tersebar dari rata-rata. Ini dihitung dengan cara menghitung kuadrat selisih antara setiap nilai dengan rata-rata, kemudian membaginya dengan jumlah observasi. Varians memberikan gambaran yang lebih detail tentang penyebaran data, tetapi satuannya tidak sama dengan satuan data asli, karena nilai varians dihitung dalam kuadrat unit.

Formula untuk menghitung varians adalah:

$$\text{Mean} = \frac{\sum_{i=1}^n (xi - \text{Mean})^2}{n}$$

- c. Standar Deviasi (*Standard Deviation*): Standar deviasi adalah akar kuadrat dari varians dan mengukur penyebaran data dalam satuan yang sama dengan data aslinya. Semakin besar standar deviasi, semakin besar variasi dalam data, sedangkan semakin kecil standar deviasi, semakin homogen data tersebut. Standar deviasi sangat berguna untuk memberikan gambaran yang lebih mudah dipahami tentang penyebaran data dibandingkan dengan varians.

Formula untuk menghitung standar deviasi adalah:

$$\text{Standar Deviasi} = \sqrt{\frac{\sum_{i=1}^n (x_i - \text{Mean})^2}{n}}$$

10.3.3 Distribusi Frekuensi

Distribusi frekuensi adalah cara untuk menggambarkan bagaimana data tersebar dalam berbagai interval atau kategori. Distribusi frekuensi menunjukkan berapa kali setiap nilai atau kelompok nilai muncul dalam dataset. Ini sangat berguna untuk memahami pola distribusi data, terutama ketika bekerja dengan data besar atau kategori.

Distribusi frekuensi sering digambarkan dengan tabel atau histogram. Tabel distribusi frekuensi mengelompokkan data ke dalam interval tertentu (misalnya, 0-10, 11-20, dll.) dan menghitung jumlah data yang berada dalam setiap interval. Histogram adalah representasi grafis dari distribusi frekuensi, di mana tinggi batang mewakili jumlah frekuensi dalam setiap interval.

Distribusi frekuensi membantu dalam identifikasi pola data, seperti bentuk distribusi (misalnya, normal, miring ke kiri, atau

miring ke kanan), serta adanya nilai ekstrim atau ketidakseimbangan dalam data.

10.4 Analisis Inferensial

Analisis inferensial merupakan cabang statistik yang digunakan untuk membuat kesimpulan atau generalisasi tentang populasi berdasarkan data sampel. Berbeda dengan statistik deskriptif yang hanya menggambarkan karakteristik sampel, analisis inferensial memungkinkan peneliti melakukan prediksi, pengambilan keputusan, dan pengujian hipotesis dengan tingkat keyakinan tertentu. Metode ini sangat penting dalam penelitian ilmiah, bisnis, dan berbagai disiplin ilmu lainnya karena membantu menilai hubungan, perbedaan, dan tren yang bersifat populasi secara keseluruhan.

10.4.1 Uji Hipotesis

Uji hipotesis adalah metode inferensial yang digunakan untuk menilai klaim atau asumsi tentang parameter populasi berdasarkan data sampel. Proses dimulai dengan perumusan hipotesis nol (*null hypothesis*, H_0) yang mewakili tidak adanya efek atau perbedaan, dan hipotesis alternatif (*alternative hypothesis*, H_1) yang menunjukkan adanya efek atau perbedaan.

Setelah itu, data sampel dianalisis menggunakan uji statistik yang sesuai, seperti uji *t*, uji *chi-square*, atau uji *ANOVA*, tergantung pada jenis data dan desain penelitian. Hasil uji kemudian dibandingkan dengan nilai kritis atau *p-value* untuk menentukan

apakah hipotesis nol dapat ditolak. Uji hipotesis memberikan kerangka objektif untuk membuat keputusan ilmiah dan meminimalkan subjektivitas dalam interpretasi data (Field, 2020).

10.4.2 Interval Kepercayaan

Interval kepercayaan (*confidence interval*) merupakan metode lain dalam analisis inferensial yang memberikan rentang nilai di mana parameter populasi diperkirakan berada dengan tingkat keyakinan tertentu, misalnya 95% atau 99%. Interval ini tidak hanya menunjukkan estimasi titik, tetapi juga ketidakpastian terkait data sampel.

Contoh penerapannya adalah estimasi rata-rata populasi dari sampel. Jika rata-rata sampel adalah 50 dengan interval kepercayaan 95% sebesar 48–52, maka peneliti dapat menyimpulkan dengan 95% keyakinan bahwa rata-rata populasi berada dalam rentang tersebut. Interval kepercayaan membantu peneliti menilai presisi estimasi dan mengkomunikasikan ketidakpastian secara kuantitatif.

10.4.3 Implikasi Analisis Inferensial

Analisis inferensial memiliki implikasi penting dalam penelitian dan pengambilan keputusan. Dengan metode inferensial, peneliti dapat:

- a. Menggeneralisasi temuan dari sampel ke populasi yang lebih luas.
- b. Menilai hubungan sebab-akibat atau asosiasi antar variabel.
- c. Mengukur ketidakpastian dan risiko kesalahan melalui *p-value* dan interval kepercayaan.

Penerapan analisis inferensial memungkinkan pembuatan kesimpulan berbasis data yang lebih akurat dan dapat dipertanggungjawabkan. Integrasi uji hipotesis dan interval kepercayaan memberikan dasar ilmiah untuk keputusan manajerial, kebijakan publik, dan penelitian akademik, sehingga meminimalkan bias dan meningkatkan validitas hasil.

Secara keseluruhan, analisis inferensial adalah alat penting dalam statistik yang memungkinkan peneliti membuat kesimpulan tentang populasi berdasarkan data sampel. Uji hipotesis dan interval kepercayaan merupakan metode utama yang memberikan kerangka kuantitatif untuk pengambilan keputusan yang rasional, objektif, dan dapat dipertanggungjawabkan.

10.5 Aplikasi Statistika dalam Pengambilan Keputusan

Statistika memiliki peran penting dalam pengambilan keputusan bisnis karena menyediakan alat kuantitatif untuk menganalisis data, mengidentifikasi pola, dan memprediksi tren. Dengan menggunakan teknik statistika, manajer dapat membuat keputusan yang lebih objektif, meminimalkan risiko, dan meningkatkan efektivitas strategi organisasi. Aplikasi statistika mencakup perencanaan operasional, evaluasi kinerja, pengembangan produk, hingga perencanaan strategis jangka panjang.

10.5.1 Perencanaan Bisnis

Dalam perencanaan bisnis, statistika digunakan untuk mengestimasi permintaan pasar, menentukan kapasitas produksi, dan mengalokasikan sumber daya secara optimal. Misalnya, analisis regresi dapat memprediksi tren penjualan berdasarkan data historis dan faktor eksternal seperti musiman atau kondisi ekonomi. Teknik proyeksi ini membantu manajer menyusun rencana produksi, stok, dan distribusi yang efisien, sehingga organisasi dapat memenuhi kebutuhan pasar tanpa pemborosan.

Selain itu, statistika memungkinkan perencanaan keuangan dengan menganalisis variabilitas biaya dan pendapatan. Dengan pendekatan kuantitatif, keputusan terkait investasi, penganggaran, dan penentuan harga produk dapat dilakukan dengan dasar yang lebih kuat dan risiko lebih terkendali (Hair et al., 2020).

10.5.2 Evaluasi Kinerja

Statistika juga digunakan untuk evaluasi kinerja organisasi, baik pada level operasional maupun strategis. Analisis data kinerja, seperti tingkat produksi, kualitas produk, penjualan, dan kepuasan pelanggan, memungkinkan identifikasi area yang memerlukan perbaikan.

Metode *control chart*, analisis variansi (ANOVA), atau *Key Performance Indicators* (KPI) berbasis data statistik membantu manajer memantau konsistensi proses dan efektivitas strategi. Dengan evaluasi berbasis statistik, organisasi dapat melakukan tindakan korektif secara tepat waktu, meningkatkan kualitas, dan mempertahankan daya saing di pasar.

10.5.3 Strategi Bisnis

Dalam pengembangan strategi, statistika membantu mengidentifikasi peluang pasar, preferensi konsumen, dan potensi risiko. Analisis cluster, segmentasi pasar, atau model prediktif memungkinkan organisasi memahami perilaku pelanggan dan tren industri.

Selain itu, simulasi statistik dapat digunakan untuk mengevaluasi berbagai skenario bisnis, seperti dampak perubahan harga, kampanye pemasaran, atau investasi baru. Dengan demikian, keputusan strategis dapat didasarkan pada analisis kuantitatif yang mendalam, bukan sekadar intuisi atau asumsi, sehingga meningkatkan peluang keberhasilan jangka panjang.

10.5.4 Implikasi untuk Pengambilan Keputusan

Penerapan statistika dalam bisnis tidak hanya meningkatkan akurasi pengambilan keputusan, tetapi juga mengurangi subjektivitas dan bias. Keputusan yang didukung data memungkinkan alokasi sumber daya yang lebih efisien, identifikasi risiko lebih cepat, dan perencanaan yang lebih matang.

Secara keseluruhan, statistika berfungsi sebagai alat penting dalam setiap tahap manajemen bisnis: perencanaan, evaluasi, dan strategi. Dengan analisis yang tepat, manajer dapat memahami data secara komprehensif, membuat keputusan yang rasional, dan meningkatkan efektivitas organisasi dalam mencapai tujuan jangka pendek maupun jangka panjang.

10.6 Latihan Soal

1. Jelaskan perbedaan antara statistik deskriptif dan statistik inferensial dalam pengambilan keputusan bisnis.
2. Bagaimana cara mengumpulkan data yang valid untuk analisis bisnis? Berikan contohnya.
3. Hitung mean, median, dan modus dari data penjualan berikut: 50, 60, 55, 70, 60, 65, 60.
4. Jelaskan bagaimana interval kepercayaan dapat membantu pengambilan keputusan manajerial.
5. Berikan contoh penerapan uji hipotesis untuk menentukan strategi pemasaran yang efektif.

Profil Penulis



Isti Kamila, S.Pd., M.Si. lahir di Medan pada 3 Mei 1989 dan saat ini berdomisili di Balikpapan. Ia dikenal memiliki hobi membaca dan menulis, dua kegiatan yang memperkaya wawasan dan mendukung aktivitas akademik maupun kreatifnya. Selamat membaca.



Nurul Maqfirah Rauf, S.Pd., M.Mat. lahir di Limbung pada 8 Juni 1995. Saat ini, ia berdomisili di Jl. Giri Rejo, Karang Joang, Balikpapan. Ia memiliki minat besar terhadap kegiatan traveling dan kulineran, yang mencerminkan semangat eksplorasi dan kecintaannya

terhadap keberagaman budaya dan rasa. Baginya, “Dengan ilmu, kita tidak sekadar menebak, kita jadi lebih tahu arah.” Sebuah pesan yang menekankan pentingnya pengetahuan dalam menentukan langkah hidup dengan lebih tepat dan terarah.



Mushliha, S.Pd., M.Si. lahir di Medan pada 20 April 1989 dan saat ini berdomisili di Jakarta Barat. Ia memiliki hobi membaca, sebuah aktivitas yang terus mendukungnya dalam menulis, meneliti, dan mengembangkan pengetahuan di bidang yang ia geluti. Dalam karyanya berjudul *Statistika Terapan untuk Ekonomi dan*

Bisnis, Mushliha menyampaikan rasa terima kasih kepada para pembaca yang telah memilih buku tersebut sebagai bahan pembelajaran. Ia menulis buku ini dengan harapan agar pembaca dapat memahami statistika secara lebih mudah dan relevan dengan konteks ekonomi dan bisnis yang tengah dipelajari atau dijalani. Ia berharap setiap konsep, contoh, dan latihan dalam buku ini dapat memperkuat kemampuan analitis pembaca serta menjadi bekal dalam pengambilan keputusan berbasis data. Mushliha juga terbuka terhadap segala bentuk masukan demi penyempurnaan edisi mendatang. Ia menutup pesannya dengan ucapan semangat: selamat belajar, semoga buku ini dapat menjadi bagian kecil yang bermanfaat dalam perjalanan akademik dan profesional pembaca.



Indrawan, S.Pd., M.Si. lahir di Kota Solok pada 12 September 1991. Saat ini, ia berdomisili di Perumahan Mentari Village Blok F3 No.33, Kelurahan Karang Joang, Kecamatan Balikpapan Utara, Kota Balikpapan, Kalimantan Timur, 76127. Membaca buku merupakan hobinya, yang

mencerminkan minat kuat terhadap pengembangan wawasan dan ilmu pengetahuan. Melalui pesannya, ia mengajak para pembaca untuk terus belajar, karena menurutnya, "di balik setiap angka ada cerita dan di balik setiap risiko ada peluang." Sebuah ajakan untuk menggali lebih dalam makna di balik data dan tantangan, serta melihat potensi yang tersembunyi di dalamnya.



Dulhadi, ST. lahir di Kampung Baru pada 18 Oktober 1976. Saat ini, ia berdomisili di Sambas, Kalimantan Barat. Ia dikenal sebagai pribadi yang aktif, dengan hobi berolahraga sebagai salah satu cara menjaga kesehatan dan semangat dalam menjalani aktivitas sehari-hari. Melalui pesannya yang singkat namun bermakna, ia mengingatkan pembaca untuk tidak pernah berhenti belajar. Sebuah ajakan yang merefleksikan pentingnya pendidikan dan pengembangan diri sebagai proses seumur hidup.



Wiwit Pura Nurmayanti, S.Si., M.Si. lahir di Rempung pada 29 Mei 1990. Ia merupakan dosen Statistika di Universitas Mulawarman, Samarinda, Kalimantan Timur, dengan bidang keahlian dalam statistika terapan. Pendidikan sarjana (S1) di bidang Statistika ia tempuh di Universitas Islam Indonesia, Yogyakarta. Kemudian, ia melanjutkan studi magister (S2) pada program Statistika Terapan di Institut Pertanian Bogor (IPB). Latar belakang akademik yang kuat ini membentuk pendekatan pengajaran dan penulisannya yang seimbang antara pemahaman konsep teoritis dan penerapan dalam permasalahan empiris. Melalui karyanya berjudul *Statistika Terapan untuk Ekonomi dan Bisnis*, Wiwit menyusun materi yang bertujuan untuk membantu pembaca memahami statistika sebagai alat analisis yang sistematis dan relevan dalam pengambilan keputusan berbasis data. Buku ini diharapkan menjadi pendamping yang berguna bagi mahasiswa, dosen, dan praktisi dalam mengembangkan kemampuan analisis data yang kritis, sehingga mampu mengambil keputusan yang tepat dalam berbagai konteks ekonomi dan bisnis.



Minharwan, S.E., S.H. lahir di Tebas pada 19 Desember 1977 dan saat ini berdomisili di kota yang sama. Ia dikenal sebagai sosok yang aktif dan energik, dengan hobi bermain sepak bola yang mencerminkan semangat kompetitif dan kerja sama tim dalam kehidupan sehari-hari. Melalui pesan

mendalam yang ia sampaikan, Minharwan mengungkapkan harapannya bahwa buku *Statistika Terapan untuk Ekonomi dan Bisnis* dapat menjadi referensi akademik yang komprehensif dan bermakna, baik bagi mahasiswa, peneliti, maupun praktisi yang ingin memperdalam pemahaman mereka tentang penerapan metode statistik dalam konteks ekonomi dan bisnis. Di tengah era ketika data menjadi aset strategis dan analisis kuantitatif berperan penting dalam pengambilan keputusan, penguasaan statistik dianggapnya bukan lagi sekadar pelengkap, melainkan fondasi utama dalam memahami dinamika ekonomi modern. Buku ini disusun dengan pendekatan sistematis, jelas, dan aplikatif agar mampu menjembatani teori dan praktik. Setiap bab dirancang untuk membantu pembaca memahami konsep, menilai metode, dan menerapkannya dalam analisis fenomena ekonomi, mengidentifikasi pola, serta merumuskan strategi berbasis data yang akurat dan dapat dipertanggungjawabkan. Ia mendorong para pembaca untuk mempelajari isi buku ini dengan ketekunan dan kecermatan, layaknya seorang peneliti dan analis profesional. Menurutnya, pembelajaran statistika adalah juga

pembelajaran tentang cara berpikir logis, sistematis, dan terukur—bukan hanya sekadar soal angka dan rumus. Lebih dari itu, Minharwan berharap buku ini dapat memperluas wawasan, memotivasi pembaca untuk terus berkembang, mengejar keunggulan akademik, serta memberikan kontribusi nyata dalam dunia ekonomi dan bisnis. Ia menutup pesannya dengan semangat bahwa ilmu bukan hanya untuk diketahui, tetapi untuk digunakan demi menghasilkan keputusan yang lebih baik, wawasan yang lebih dalam, dan nilai yang lebih tinggi.



Indah Gumala Andirasdini, M.Si. lahir di Pariaman pada 16 Desember 1994 dan saat ini berdomisili di Bandar Lampung. Ia dikenal sebagai pribadi yang ceria dan dinamis, dengan hobi bermain yang mencerminkan semangat eksploratif dan kreativitas dalam menjalani aktivitas sehari-hari. Dalam pesannya kepada pembaca, ia

menyampaikan harapan yang inspiratif: *"May each page offer new insights, inspire critical thinking, and foster a lifelong passion for learning."* Sebuah ajakan untuk menjadikan setiap halaman bacaan sebagai sumber pemahaman baru, pemicu daya pikir kritis, serta penguat semangat belajar yang terus tumbuh sepanjang hayat.



Rani, S.Pd., M.M. lahir di Jakarta pada 25 Agustus 1990 dan saat ini berdomisili di kota kelahirannya. Ia memiliki beragam hobi yang mencerminkan kepribadian aktif dan penuh semangat, di antaranya adalah jalan-jalan, menonton film, dan membaca buku. Melalui pesannya kepada para pembaca, ia mengucapkan terima kasih atas pilihan menjadikan buku ini sebagai bahan bacaan. Ia berharap buku tersebut dapat memberikan informasi dan pengetahuan yang bermanfaat bagi setiap pembaca, serta menjadi bagian dari perjalanan belajar yang menyenangkan dan bermakna.

Daftar Pustaka

- Acock, A. C. (2018). *A gentle introduction to STATA* (6th ed.). Stata Press.
- Agresti, A. (2019). *An introduction to categorical data analysis*. Wiley.
- Ala'raj, M., Abbasi, G. A., & Alsmadi, M. (2021). Predictive analytics in economics and business using machine learning. *Journal of Big Data*, 8(1), 54–67.
- Analysis and Its Applications: With R Examples* (4th ed.). Springer.
- Arifin, W. N., & Yusoff, M. S. B. (2020). The fundamentals and principles of correlation analysis in medical research. *Education in Medicine Journal*, 12(1), 1–7.
- Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability* (2nd ed.). Chapman and Hall/CRC.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons.
- Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time*
- Bryman, A. (2021). *Social research methods* (6th ed.). Oxford University Press.
- Casella, G., & Berger, R. L. (2021). *Statistical inference* (3rd ed.). Cengage Learning.

- Casella, G., & Berger, R. L. (2021). *Statistical inference*. Cengage Learning.
- Conover, W. J. (2020). *Practical nonparametric statistics*. Wiley.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications.
- Creswell, J. W., & Creswell, J. D. (2021). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.
- Creswell, J. W., & Creswell, J. D. (2021). *Research design: Qualitative, quantitative, and mixed methods approaches* (6th ed.). Sage Publications.
- Creswell, J. W., & Poth, C. N. (2020). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis*.
- DeGroot, M. H., & Schervish, M. J. (2021). *Probability and statistics* (5th ed.). Pearson.
- Devore, J. L. (2019). *Probability and statistics for engineering and the sciences*. Cengage Learning.
- Devore, J. L. (2020). *Probability and statistics for engineering and the sciences* (9th ed.). Cengage Learning.
- Devore, J. L. (2021). *Probability and statistics for engineering and the sciences*. Cengage Learning.
- Durbin, J., & Koopman, S. J. (2012). *Time Series Analysis by Economic Time Series* (2nd ed.). Academic Press.

ed.). Wiley.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1–4.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Field, A. (2020). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.

Field, A. (2022). *Discovering statistics using R*. Sage Publications.
Triola, M. F. (2020). *Elementary statistics* (13th ed.). Pearson.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). Sage Publications.

Frankfort-Nachmias, C., & Leon-Guerrero, A. (2020). *Social statistics for a diverse society* (9th ed.). Sage Publications.

Franses, P. H., & van Dijk, D. (2000). *Non-Linear Time*

Freund, J. E., & Perles, B. M. (2021). *Modern Elementary Statistics* (12th ed.). Pearson Education.

Freund, J. E., & Perles, B. M. (2021). *Modern Elementary Statistics* (12th ed.). Pearson Education.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2020). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC

Ghasemy, M., Hussin, S., Daud, M. A. K. M., & Ahmad, N. A. (2021). Regression analysis in educational research: A practical review. *Heliyon*, 7(5), e07019.

- Ghozali, I. (2020). *Statistika bisnis: Teori dan aplikasi dalam analisis data bisnis*. Yogyakarta: Penerbit Andi.
- Ghozali, I. (2020). *Statistika bisnis: Teori dan aplikasi dalam analisis data bisnis*. Yogyakarta: Penerbit Andi.
- Gould, R., & Sun, D. (2021). Data visualization literacy: Understanding how students interpret graphical displays. *Statistics Education Research Journal*, 20(1), 6–20.
- Granger, C. W. J., & Newbold, P. (2014). *Forecasting*
- Gravetter, F. J., & Wallnau, L. B. (2021). *Statistics for the behavioral sciences* (11th ed.). Cengage Learning.
- Grimmett, G., & Stirzaker, D. (2020). *Probability and random processes*. Oxford University Press.
- Grolemund, G., & Wickham, H. (2017). *R for data science: Import, tidy, transform, visualize, and model data*. O'Reilly Media.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2020). *Multivariate data analysis* (8th ed.). Cengage Learning.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton
- Harvey, A. C. (1990). *Forecasting, Structural Time Series*
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- Hossain, M. M., & Williams, G. M. (2021). Statistical Distribution Characteristics and Their Applications in Research. *Journal of Applied Quantitative Methods*, 16(4), 52–65.

- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). Melbourne: OTexts.
- Introduction* (6th ed.). Chapman & Hall/CRC.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning* (2nd ed.). Springer.
- Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2020). *Nonlinear analyses in organizational research: A primer using hypothetical data*. Organizational Research Methods.
- Kelleher, J. D., & Tierney, B. (2018). *Data science: An introduction*. CRC Press.
- Kim, T. K. (2021). *Understanding one-way ANOVA using conceptual figures*. *Korean Journal of Anesthesiology*, 74(1), 1–6.
- Kitchin, R. (2015). The opportunities, challenges and risks of big data for official statistics. *Statistical Journal of the IAOS*, 31(3), 471–481.
- Marshall, C., & Rossman, G. B. (2020). *Designing qualitative research* (6th ed.). Sage Publications.
- Mayer-Schönberger, V., & Cukier, K. (2021). *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.
- Models and the Kalman Filter*. Cambridge University Press.
- Montgomery, D. C., & Runger, G. C. (2018). *Applied statistics and probability for engineers* (7th ed.). Wiley.

- Montgomery, D. C., & Runger, G. C. (2018). *Applied statistics and probability for engineers* (7th ed.). Wiley.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). Hoboken, NJ: Wiley.
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2015). *Introduction to the practice of statistics* (8th ed.). W. H. Freeman.
- Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1), 1–13.
- Ozdemir, S., & Susarla, D. (2022). *Mastering data analysis with Python*. Packt Publishing.
- Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. McGraw-Hill Education.
- Pratama, F., & Lukman, A. (2021). Central tendency analysis in descriptive statistics. *Journal of Quantitative Methods*, 3(2), 87–96.
- Rahardjo, S. (2022). Evaluating the sensitivity of mean and median in skewed data. *Indonesian Journal of Data Analytics*, 5(1), 25–34.
- Rahman, M., & Abdullah, N. (2022). Descriptive Statistics and the Importance of Variability in Quantitative Research. *International Journal of Research Methodology*, 7(2), 101–117.

- Rice, J. A. (2019). *Mathematical statistics and data analysis* (3rd ed.). Cengage Learning
- Ross, S. M. (2020). *Introduction to probability models*. Academic Press.
- Ross, S. M. (2021). *A first course in probability* (10th ed.). Pearson.
- Sarkar, I. N., & Rashid, M. A. (2016). Navigating unstructured big data in health care. *Perspectives in Health Information Management*, 13(Fall), 1–10.
- Sarstedt, M., Hair, J. F., Cheah, J. H., Becker, J. M., & Ringle, C. M. (2020). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal*, 28(3), 197–211.
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5), 1763–1768.
- Schober, P., Boer, C., & Schwarte, L. A. (2021). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, 132(1), 39–45.
- Series and Forecasting* (3rd ed.). Springer.
- Series Models in Empirical Finance*. Cambridge University Press.
- Shumway, R. H., & Stoffer, D. S. (2017). *Time Series State Space Methods* (2nd ed.). Oxford University Press.
- stationary Economic Time Series*. MIT Press.
- Sulaiman, M. (2023). Statistical considerations in selecting appropriate measures of central tendency. *International Review of Applied Statistics*, 8(1), 41–52.

- Taherdoost, H. (2016). Sampling methods in research methodology: How to choose a sampling technique for research. *International Journal of Academic Research in Management*, 5(2), 18–27.
- Triola, M. F. (2018). *Elementary statistics* (13th ed.). Pearson.
- Triola, M. F. (2020). *Elementary Statistics* (13th ed.). Pearson Education.
- Triola, M. F. (2020). *Elementary Statistics* (13th ed.). Pearson Education.
- Tsay, R. S. (2010). *Analysis of Financial Time Series* (3rd ed.). Wiley.
- Tufféry, S. (2021). *Data mining and statistics for decision making*. Wiley.
- Tufféry, S. (2021). *Data mining and statistics for decision making*. Wiley.
- University Press.
- Varian, H. R. (2021). *Artificial intelligence, economics, and industrial organization*. National Bureau of Economic Research.
- Wackerly, D. D., Mendenhall, W., & Scheaffer, R. L. (2018). *Mathematical statistics with applications* (9th ed.). Cengage Learning.
- Wasserman, L. (2020). *All of statistics: A concise course in statistical inference*. Springer.
- Wasserman, L. (2021). *All of statistics: A concise course in statistical inference*. Springer.

- Wasserman, L. (2022). *All of statistics: A concise course in statistical inference* (2nd ed.). Springer.
- Wei, W. W. S. (2019). *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Wilcox, R. R. (2020). *Introduction to robust estimation and hypothesis testing* (4th ed.). Academic Press.
- With Applications in R*. Springer.
- Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Boston, MA: Cengage Learning.
- World Health Organization. (2020). WHO guidelines on sterile pharmaceutical preparations. WHO Press.
- Yoo, W., Mayberry, R., Bae, S., Singh, K., Peter He, Q., & Lillard, J. W. (2020). A study of effects of multicollinearity in the multivariable analysis. *International Journal of Applied Science and Technology*, 10(2), 1–9.
- Zhang, X., & Chen, Y. (2020). Data-driven decision making in business: A framework for effective analysis. *International Journal of Information Management*, 52, 102075.
- Zhang, Z., & Ma, Z. (2020). *Quantile and percentile methods for data exploration: Applications in statistical analysis*. Statistics Education Research Journal.
- Zhou, Y., Xu, Z., & Gong, X. (2020). Best practices in data visualization for effective communication. *Journal of Data Science and Analytics*, 8(3), 157–168.

Buku ajar Statistika Terapan untuk Ekonomi dan Bisnis membahas penggunaan statistik sebagai alat bantu penting dalam memahami fenomena ekonomi dan aktivitas bisnis. Materi disajikan secara ringkas dan mudah dipahami, sehingga pembaca dapat mengenali konsep statistik dasar hingga penerapannya dalam analisis data keuangan, pasar, dan kinerja usaha.

Dengan pendekatan yang praktis dan relevan dengan kondisi nyata, buku ini membantu pembaca mengolah data menjadi informasi yang bernilai untuk mendukung pengambilan keputusan. Buku ini dijual untuk masyarakat umum yang ingin memperluas wawasan tentang statistik tanpa memerlukan latar belakang khusus.

