CV. ASKARA SASTRA MEDIA

STATISTIK
DASAR

Dendi Zainuddin Hamidi, S.T., M.M
Dr. Drs. Mohzana, S.Pd., M.Pd
Ully Muzakir, M.T
Inayatul Mutmainnah, S.Sos.M.Si
Ratnasari, M.Pd
Ratih Milati Ilham, S.E.Sy., M.E., CIQnR
Dr. Agus Suprapto, S.P., M.P., IPM
Titik Suhartini S.Pd M.Si
Ir. Pengki Irawan, S.TP., M.Si
Andika Putra, S.Sos

STATISTIK DASAR

Penulis:
Dendi Zainuddin Hamidi, S.T., M.M
Dr. Drs. Mohzana, S.Pd., M.Pd
Ully Muzakir, M.T
Inayatul Mutmainnah, S.Sos., M.Si
Ratnasari, M.Pd
Ratih Milati Ilham, S.E.Sy., M.E., CIQnR
Dr. Agus Suprapto, S.P., M.P., IPM
Titik Suhartini S.Pd., M.Si
Ir. Pengki Irawan, S.TP., M.Si
Andika Putra, S.Sos

ASKARA SASTRA
MEDIA

STATISTIK
DASAR

Penulis :
Dendi Zainuddin Hamidi, S.T., M.M
Dr. Drs. Mohzana, S.Pd., M.Pd
Ully Muzakir, M.T
Inayatul Mutmainnah, S.Sos., M.Si
Ratnasari, M.Pd
Ratih Milati Ilham, S.E.Sy., M.E., CIQnR
Dr. Agus Suprapto, S.P., M.P., IPM
Titik Suhartini S.Pd., M.Si
Ir. Pengki Irawan, S.TP., M.Si
Andika Putra, S.Sos

Editor dan Desain Cover :
Lambrika Dwi

Ukuran:
vii hal + 191 hal; 14,8cm x 21cm

Diterbitkan Oleh :

ASKARA SASTRA MEDIA

Jln. Al-Hidayah, Jombang, Jawa Timur - 61481
Email : askarasastramedia@gmail.com

ISBN : 978-623-10-4975-9
Terbitan: November 2024

KATA PENGANTAR

Syukur *alhamdulillah* penulis haturkan kepada Allah Swt. yang senantiasa melimpahkan karunia dan berkah-Nya sehingga penulis mampu merampungkan karya ini tepat pada waktunya, sehingga penulis dapat menghadirkannya dihadapan para pembaca. Kemudian, tak lupa *shalawat* dan salam semoga senantiasa tercurah limpahkan kepada Nabi Muhammad SAW, para sahabat, dan ahli keluarganya yang mulia.

Buku Statistik dasar mencakup konsep seperti mean (rata-rata), median (nilai tengah), dan modus (nilai yang paling sering muncul) untuk menggambarkan data secara deskriptif. Dispersi data diukur melalui range (selisih nilai maksimum dan minimum), varians (penyebaran data dari rata-rata), dan standar deviasi (ukuran seberapa tersebar data dari rata-rata). Selain itu, distribusi data dapat dijelaskan dengan melihat skewness (kecondongan) dan kurtosis (kelancipan), yang membantu memahami bentuk distribusi data dalam suatu populasi atau sampel.

Tujuan buku statistik dasar adalah untuk memberikan pemahaman menyeluruh tentang konsep dan teknik statistik yang digunakan untuk menganalisis,

menginterpretasi, dan menyajikan data. Buku ini bertujuan membantu pembaca menguasai alat-alat statistik seperti pengukuran tendensi sentral, variabilitas, distribusi probabilitas, serta metode inferensial seperti uji hipotesis dan regresi. Dengan pemahaman tersebut, pembaca diharapkan mampu menerapkan statistik untuk pengambilan keputusan yang lebih baik dalam berbagai bidang, termasuk sains, ekonomi, bisnis, dan ilmu sosial.

Penulis menyampaikan terima kasih yang tak terhingga bagi semua pihak yang telah berpartisipasi. Terakhir seperti kata pepatah bahwa" Tiada Gading Yang Tak Retak" maka penulisan buku ini juga jauh dari kata sempurna, oleh karena itu penulis sangat berterima kasih apabila ada saran dan masukkan yang dapat diberikan guna menyempurnakan buku ini di kemudian hari.

2024

Penulis

DAFTAR ISI

BAB I
PENGERTIAN DAN RUANG LINGKUP STATISTIK DASAR

1.1. Pengertian Statistika

Kata Statistika berasal dari bahasa latin yaitu Status (dalam bahasa Inggris: *State*) yang berarti negara. Hal ini dikarenakan untuk beberapa dekade, statistics (statistika) semata-mata hanya dikaitkan dengan penyajian angka-angka tentang situasi perekonomian, kependudukan dan politik yang terjadi di suatu negara atau untuk menyatakan hal-hal yang berhubungan dengan ketatanegaraan. Selanjutnya istilah statistika menjadi lebih umum dan identik dengan kegiatan penelitian yaitu pengumpulan, pengolahan, dan analisis data. Hal ini dikarenakan proses pengumpulan, pengolahan, maupun analisis data dalam statistika memerlukan suatu metodologi yang memiliki kesamaan dengan penelitian. Oleh karena itu, saat ini statistika menjadi salah satu disiplin ilmu, yang bahkan menjadi ilmu pra syarat apabila sebuah penelitian akan menggunakan metode kuantitatif.

Beberapa ahli memberikan definisi statistika, diantaranya Freund & Perles (1974) yang mengartikan statistika sebagai ilmu yang berkaitan dengan metode pengumpulan, pengorganisasian, analisis, dan interpretasi data untuk membuat keputusan yang masuk akal berdasarkan data yang tersedia. Selanjutnya Anderson et al. (2017) mendefinisikan statistika sebagai "*a science of collecting, organizing, analyzing, interpreting, and presenting data in a meaningful way to make informed decisions*", dapat diterjemahkan ke dalam bahasa Indonesia sebagai sebuah ilmu untuk mengumpulkan, mengorganisir, menganalisis, menginterpretasikan, dan menyajikan data dengan cara yang bermakna untuk membuat keputusan yang tepat. Berdasarkan pendapat kedua ahli tersebut maka statistika memungkinkan pengambilan keputusan berdasarkan data yang dikumpulkan dan diolah.

Dalam perspektif lainnya, Walpole et al. (1993) mendefinisikan statistika sebagai "*a branch of mathematics that provides procedures for analyzing numerical data obtained from samples*", dapat diterjemahkan ke dalam bahasa Indonesia sebagai cabang matematika yang menyediakan prosedur untuk menganalisis data numerik yang diperoleh dari sampel. Dalam perspektif ini, statistika dikaitkan dengan

matematika serta teknik untuk memproses data numerik dari sejumlah sampel dan menghubungkannya dengan populasi yang lebih besar. Pendapat Walpole et al. (1993) tersebut menurut penulis dilatar belakangi oleh seringnya statistika digunakan sebagai alat analisis dalam metode penelitian kuantitatif, yaitu metode penelitian yang ditujukan untuk menguji hipotesis dan melakukan generalisasi dari kesimpulan hasil penelitian tersebut.

Bila dikaitkan antara pendapat Freund & Perles (1974); Anderson et al. (2017) dengan pendapat Walpole et al. (1993), maka dapat disimpulkan bahwa data yang diolah dan dianalisis dalam statistika adalah data yang berbentuk angka. Dengan demikian maka berdasarkan pendapat para ahli tersebut, statistika dapat diartikan sebagai disiplin ilmu ataupun kegiatan yang mengkaji tentang seluk-beluk data yang berbentuk angka mulai dari pengumpulan, pengolahan, penyajian, analisis, sampai kepada penarikan kesimpulan. Kesimpulan dari analisis data dengan menggunakan metode statistika tentu saja dapat digunakan sebagai dasar untuk pengambilan keputusan. Hal ini karena data yang diolah dengan metode statistika merupakan data empiris (fakta), baik data sekunder (data yang telah ada) maupun data primer (data yang

dikumpulkan langsung oleh peneliti dari lapangan), baik data time series (data beberapa periode dari satu entitas) maupun data *cross section* (data satu periode dari beberapa entitas), baik data yang sudah berbentuk angka (data kuantitatif) maupun data yang diangkakan (data kualitatif yang dikuantifikasi).

1.2. Ruang Lingkup Statistika

Ruang lingkup statistika mencakup berbagai aktivitas ilmiah yang berkaitan dengan pengumpulan, pengolahan, analisis, dan interpretasi data. Statistika tidak hanya terbatas pada angka, tetapi juga melibatkan teknik dan metode untuk mengekstraksi informasi dari data yang kompleks. Secara umum, statistika terbagi menjadi dua cabang utama (Hamidi, 2020; Siegel & Castellan Jr., 1988; Triola, 2019), yaitu:

1. Statistik Deskriptif

   Berfokus pada penyajian dan penggambaran data melalui tabel, grafik, dan ukuran-ukuran statistik seperti mean, median, dan modus. Tujuannya adalah menggambarkan atau merangkum karakteristik data tanpa membuat kesimpulan yang lebih luas. Jika ditinjau dari perspektif sumber pengumpulan datanya, Statistik Deskriptif terdiri dari metode-metode

yang berkaitan dengan pengumpulan dan analisis data yang bersumber dari keseluruhan elemen populasi sehingga dapat menghasilkan suatu informasi yang memiliki arti gambaran/keadaan populasi tersebut. Berdasarkan hal tersebut, maka statistik deskriptif bertujuan untuk menggambarkan atau meringkas karakteristik data tanpa melakukan generalisasi lebih lanjut. Lebih lanjut lagi, Bluman (2019) mengemukakan bahwa "statistik deskriptif terdiri dari metode-metode untuk mengatur dan meringkas data sehingga pola atau karakteristik utama dari data tersebut dapat lebih mudah dipahami."

2. Statistik Inferensial

   Statistik inferensial melibatkan metode yang memungkinkan penarikan kesimpulan atau generalisasi tentang populasi berdasarkan data yang diperoleh dari sampel. Statistik inferensial menggunakan konsep probabilitas untuk membuat prediksi dan menguji hipotesis. Statistik ini melibatkan pengujian hipotesis, estimasi parameter, dan pengujian hubungan antarvariabel dengan metode seperti uji-t, ANOVA, dan regresi. Berdasarkan hal tersebut

maka dapat disimpulkan bahwa statistik inferensial berusaha membuat inferensi atau kesimpulan tentang populasi berdasarkan sampel data. Dengan kata lain, Statistika Inferensial terdiri dari metode-metode yang berkaitan dengan pengumpulan dan analisis data yang bersumber dari sampel (sebagian populasi) sehingga informasi yang diperoleh berupa pendugaan atau generalisasi terhadap populasinya.

Dalam statistik inferensial, kita menggunakan alat seperti estimasi parameter, uji hipotesis, dan analisis regresi untuk menarik kesimpulan yang lebih luas dari data yang terbatas. McClave, Benson, & Sincich (2008) menjelaskan bahwa statistik inferensial adalah bagian dari statistika yang melibatkan pengambilan keputusan atau kesimpulan tentang suatu populasi berdasarkan informasi dari sampel.

Baik statistik deskriptif maupun statistik inferensial, keduanya dapat digunakan di berbagai bidang seperti bisnis, ekonomi, kesehatan, teknik, sosial, serta pendidikan untuk mendukung pengambilan keputusan berdasarkan bukti empiris.

Selain dua cabang utama tersebut, ruang lingkup statistika juga mencakup beberapa bidang khusus lainnya (Walpole et al., 1993), seperti:

a. Teori Probabilitas :yaitu cabang statistika yang dijadikan dasar untuk membuat prediksi dan keputusan di bawah kondisi ketidakpastian.
b. Desain Eksperimen : yaitu salah satu metode dalam statistika yang digunakan untuk merancang eksperimen yang menghasilkan data valid.
c. Analisis Data Multivariat : yaitu salah satu teknik analisis dalam statistika yang digunakan untuk menganalisis data yang melibatkan banyak variabel secara simultan.

Jika ditinjau dari sebaran data yang akan dianalisis, statistika dibagi menjadi statistik parametrik dan statistik non parameterik. Penjelasan dari kedua jenis statistik tersebut:

1. Statistik Parametrik

   Statistik parametrik adalah cabang dari statistik inferensial yang menggunakan asumsi tentang distribusi data dalam populasi yang diteliti. Salah satu asumsi utama dalam statistik parametrik adalah bahwa data mengikuti

distribusi tertentu, biasanya distribusi normal (Gaussian). Teknik-teknik statistik parametrik menggunakan parameter-parameter populasi, seperti rata-rata dan variansi, untuk membuat inferensi tentang populasi berdasarkan data sampel. Statistik parametrik biasanya lebih efisien dibandingkan metode non-parametrik ketika asumsi-asumsi dasar terpenuhi.

Ciri-ciri statistik parametrik:

a. Distribusi Data

   Dalam hal distribusi data, statistik parametrik berasumsi bahwa data yang digunakan berasal dari distribusi tertentu, biasanya distribusi normal atau distribusi dengan bentuk yang diketahui. Menurut Walpole et al. (1993), statistik parametrik bergantung pada asumsi bahwa data mengikuti distribusi probabilitas tertentu, yang memungkinkan penggunaan rumus-rumus matematika untuk pengujian hipotesis dan estimasi parameter.

b. Penggunaan Parameter Populasi

   Statistik parametrik fokus pada parameter populasi, seperti rata-rata (mean), variansi, dan standar deviasi. Data yang digunakan

diolah berdasarkan asumsi-asumsi tentang parameter-parameter ini untuk menarik kesimpulan tentang populasi.

c. Efisiensi dan Keakuratan

   Jika asumsi-asumsi yang mendasari statistik parametrik benar, metode ini lebih efisien dan akurat dibandingkan statistik non-parametrik. Sebagaimana dikemukakan oleh Montgomery & Runger (1994), statistik parametrik memberikan hasil yang lebih kuat dan lebih akurat karena mereka secara eksplisit mempertimbangkan distribusi probabilitas data.

d. Uji Parametrik

   Beberapa metode dan uji dalam statistik parametrik termasuk uji-t, ANOVA (Analisis Varians), dan regresi linier. Misalnya, uji-t digunakan untuk membandingkan rata-rata dua kelompok, sementara ANOVA digunakan untuk membandingkan rata-rata dari tiga atau lebih kelompok.

Dengan demikian, statistik parametrik memiliki kelebihan:

a. Efisiensi

   Statistik parametrik biasanya membutuhkan ukuran sampel yang lebih kecil untuk mendapatkan kesimpulan yang sama akuratnya dengan metode non-parametrik.

b. Lebih Kuat

   Bila asumsi distribusi dipenuhi, uji parametrik memiliki kekuatan statistik yang lebih tinggi.

Namun di luar kelebihannya, statistik parametrik memiliki keterbatasan, diantaranya:

a. Asumsi yang Ketat

   Salah satu kelemahan utama statistik parametrik adalah ketergantungannya pada asumsi distribusi data. Jika asumsi ini tidak terpenuhi, hasilnya bisa bias atau tidak valid.

b. Ketidakcocokan untuk Data Skewed

   Statistik parametrik kurang cocok jika data tidak berdistribusi normal atau memiliki outliers yang signifikan.

Secara keseluruhan, statistik parametrik merupakan pendekatan yang kuat dan efisien untuk analisis data ketika asumsi-asumsi yang mendasarinya terpenuhi, terutama mengenai distribusi data yang normal.

2. Statistik Non Parametrik

   Statistik non-parametrik adalah cabang statistik yang tidak bergantung pada asumsi tentang distribusi data dalam populasi. Menurut Siegel & Castellan Jr. (1988), statistik non parametrik mengacu pada seperangkat teknik statistik yang tidak bergantung pada data yang termasuk dalam distribusi tertentu. Dengan kata lain, statistik non parametrik merupakan metode analisis dalam statistika dimana parameter populasi tidak harus mengikuti distribusi normal sehingga dikatakan distribusi bebas (free distribution), serta varians tidak perlu homogen. Berbeda dengan statistik parametrik, metode non-parametrik tidak memerlukan data untuk mengikuti distribusi tertentu, seperti distribusi normal. Oleh karena itu, statistik non-parametrik sering digunakan ketika data tidak memenuhi asumsi-asumsi distribusi yang

diperlukan dalam uji parametrik atau ketika data berada dalam skala ordinal atau nominal.

Ciri-ciri Statistik Non-Parametrik:

a. Tidak Membutuhkan Asumsi Distribusi

   Statistik non-parametrik tidak mengasumsikan bahwa data berasal dari distribusi tertentu. Conover (1999) menyatakan bahwa metode non-parametrik tidak memerlukan asumsi yang ketat tentang distribusi data, yang membuatnya lebih fleksibel dalam berbagai situasi analisis. Hal ini menjadikan metode ini ideal untuk data yang tidak berdistribusi normal atau data ordinal.

b. Cocok untuk Data Skala Ordinal atau Nominal

   Statistik non-parametrik lebih sering digunakan untuk data yang tidak dalam skala interval atau rasio, seperti data ordinal atau nominal, di mana tidak ada asumsi mengenai jarak antara nilai-nilai data.

c. Fleksibilitas

   Statistik non-parametrik lebih fleksibel karena dapat digunakan dalam berbagai

kondisi data, termasuk data yang memiliki outliers atau distribusi yang tidak simetris.

d. Uji Non-Parametrik

   Beberapa contoh uji non-parametrik yang sering digunakan meliputi Uji Mann-Whitney U, Uji Kruskal-Wallis, Uji Wilcoxon, dan Uji Chi-Square. Misalnya, Uji Mann-Whitney U digunakan untuk membandingkan dua kelompok independen ketika data tidak berdistribusi normal.

Dengan demikian, statistik non parametrik memiliki kelebihan:

a. Tidak Membutuhkan Asumsi Distribusi

   Statistik non-parametrik lebih umum digunakan saat data tidak memenuhi asumsi distribusi normal. Ini adalah keuntungan besar dalam situasi di mana distribusi data sulit diidentifikasi atau data tidak berdistribusi secara normal.

b. Bisa Digunakan pada Data Kecil

   Statistik non-parametrik sering kali lebih sesuai untuk data dengan ukuran sampel yang kecil, karena tidak memerlukan estimasi parameter populasi.

Di luar kelebihannya, statistik non parametrik memiliki keterbatasan, diantaranya:

a. Kurang Efisien

   Statistik non-parametrik cenderung kurang efisien dibandingkan statistik parametrik jika asumsi parametrik terpenuhi (Siegel & Castellan Jr., 1988). Dengan kata lain, jika data berdistribusi normal, metode non parametrik tidak lebih tepat untuk digunakan karena memerlukan data sampel yang lebih banyak.

b. Kekuatan Uji yang Lebih Rendah

   Statistik non-parametrik umumnya memiliki kekuatan statistik yang lebih rendah dibandingkan metode parametrik ketika asumsi parametrik berlaku, yang berarti diperlukan sampel yang lebih besar untuk mencapai hasil yang sama.

Secara keseluruhan, statistik non-parametrik adalah alat yang sangat berguna untuk analisis data ketika asumsi distribusi normal tidak terpenuhi atau ketika data berada dalam skala ordinal atau nominal. Meskipun metode ini mungkin kurang efisien dalam kondisi di mana

asumsi parametrik terpenuhi, fleksibilitasnya membuatnya sangat diperlukan dalam banyak situasi penelitian.

Secara keseluruhan, ruang lingkup statistika mencakup berbagai metode dan teknik untuk memahami, menganalisis, dan membuat kesimpulan dari data. Statistika sangat penting dalam berbagai disiplin ilmu seperti ekonomi, kedokteran, teknik, dan ilmu sosial, karena memungkinkan peneliti untuk mengatasi ketidakpastian dan membuat keputusan yang lebih baik berdasarkan data.

1.3. Peranan Statistika

Statistika mempunyai peranan yang sangat penting dalam langkah-langkah pokok metode ilmiah pada tingkat pengumpulan informasi. Misalnya statistika memberi petunjuk kepada para peneliti bagaimana cara yang wajar dan baik untuk mengumpulan data yang informatif termasuk penentuan macam dan banyak data (jika mengambil sample), sehingga kesimpulan yang ditarik dari analisis data dapat dinyatakan dengan tingkat ketepatan (presisi) yang diinginkan.

Selain itu, statitika juga memiliki peranan penting sebagai alat bantu dalam penelitian, yaitu dalam hal

pengumpulan, pengolahan dan analisis data. Baik penelitian yang menggunakan data kuantitatif (berbentuk angka) maupun penelitian yang menggunakan data kualitatif (tidak berbentuk angka) namun hendak dikuantifikasi (diubah ke bentuk angka). Dalam studi tentang ekonomi makro maupun mikro misalnya, menyangkut tingkat pertumbuhan ekonomi, pendapatan masyarakat, jumlah penduduk, permintaan akan suatu produk, dlsb. Menurut Guilford (1950) statistika memiliki beberapa kegunaan diantaranya:

1. Statistika memungkinkan pencatatan paling eksak data penelitian
2. Memberikan cara untuk melakukan pengolahan data dalam bentuk angka
3. Memberikan arahan berpikir / tata kerja yang definitif dan eksak
4. Memberikan cara meringkas data dalam berbagai bentuk
5. Sebagai dasar menarik kesimpulan
6. Memberikan landasan untuk melakukan ramalan (prediksi)
7. Memungkinkan peneliti mampu menganalisis dan menjelaskan serta menguraikan sebab akibat yang kompleks dan rumit.

1.4. Metodologi Statistika

Yang dimaksud dengan metodologi dalam statistika adalah tahapan-tahapan yang terdiri dari:

1. Identifikasi Masalah

   Pada tahap ini, diidentifikasi masalah yang terjadi yang akan dicari solusinya dengan analisis statistika. Masalah adalah ketika kenyataan tidak sesuai harapan atau ketika realita tidak seusai dengan idealita. Setelah identifikasi masalah maka selanjutnya dapat merumuskan tujuan penelitian.

2. Pengumpulan Data

   Pada tahap ini, data yang mendukung atau bersangkutan dengan permasalahan dikumpulkan. Data yang dikumpulkan dapat berupa:

   a. Data Primer, yaitu data yang diperoleh langsung dari objek yang bermasalah.
   b. Data Sekunder, yaitu data yang diperoleh langsung dari laporan/hasil penelitian terdahulu yang memiliki permasalahan yang sama.

3. Pengolahan Data

   Bila data yang terkumpul berupa data primer, maka harus diolah terlebih dahulu. Pengolahan

dapat berupa klasifikasi data, coding, peringkasan data, distribusi frekuensi, dan sebagainya.

4. Penyajian dan Analisis Data
   a. Penyajian data umumnya ke dalam suatu tabel atau gambar (grafik).
   b. Analisis data merupakan interpretasi (penafsiran) dari data yang telah diolah secara statistika yang umumnya berbentuk ukuran-ukuran.
5. Kesimpulan dan Saran

   Kesimpulan adalah ringkasan dari poin-poin utama yang telah dibahas, sering kali menjawab pertanyaan atau masalah utama dalam suatu penelitian atau diskusi. Saran adalah rekomendasi untuk tindakan di masa depan, yang didasarkan pada temuan atau kesimpulan tersebut.

1.5. Bebarapa Konsep Dasar dalam Statistika

Terdapat beberapa konsep dasar yang harus dipahami dan dihafalkan betul dalam statistika, yaitu diantaranya:

1. Populasi : yaitu sekumpulan objek yang akan diteliti, yang memiliki satuan.

2. Sampel : yaitu bagian dari populasi yang dianggap dapat mewakili populasi tersebut.
3. Variabel: yaitu sesuatu yang kondisinya, nilainya, atau datanya bisa bervariasi (berubah-ubah). Variabel inilah yang berpotensi mengalami masalah sehingga memerlukan penelitian untuk mengetahui penyebab masalah tersebut. Suatu variabel dikatakan bermasalah apabila nilai atau kondisi realnya tidak sesuai (mengalami gap) dengan nilai atau kondisi idealnya.
4. Pembulatan Data: dalam Statistika, tidak sebagaimana dalam matematika, pembulatan data mengikuti atau disesuaikan dengan kebutuhannya, tidak mengikuti kaidah yang kaku.

   Contoh: 6,5 bisa dibulatkan menjadi 6 atau 7
5. Notasi Sigma (Σ) : adalah notasi yang melambangkan penjumlahan. Rumus-rumus pengolahan data dalam ststistika banyak yang menggunakan notasi sigma. Tips untuk memudahkan perhitungan yang menggunakan notasi sigma adalah kerjakan dalam bentuk kolom atau tabel vertikal.

Contoh Soal:

$X_1 = 3$; $X_2 = -4$; $X_3 = 6$; $X_4 = -1$

$Y_1 = -2$; $Y_2 = 8$; $Y_3 = -5$; $Y_4 = 7$

Hitunglah:

a) $\sum X$
b) $\sum XY$
c) $\sum Y^2$
d) $\sum X^2Y$

Jawab:

No	**X**	**Y**	**XY**	**Y^2**	**X^2**	X2Y
1	3	-2	-6	4	9	-18
2	-4	8	-32	64	16	128
3	6	-5	-30	25	36	-180
4	-1	7	-7	49	1	7
$\sum$	**4**		**-75**	**142**		**-63**

Jadi :

a) $\sum X = 4$
b) $\sum XY = -75$
c) $\sum Y^2 = 142$
d) $\sum X^2Y = -63$

BAB II
PENYAJIAN DATA

2.1. Tabel Distribusi Frekuensi

Tabel Distribusi Frekuensi merupakan salah satu cara paling dasar untuk menyajikan data dalam bentuk yang lebih ringkas dan mudah dipahami. Tabel ini membantu mengelompokkan data berdasarkan frekuensi kemunculan setiap nilai atau kelompok nilai. Dalam statistik, distribusi frekuensi adalah alat penting yang digunakan untuk menyusun data, terutama ketika berhadapan dengan sejumlah besar data.

Tabel distribusi frekuensi adalah tabel yang menunjukkan distribusi data dari serangkaian pengamatan berdasarkan jumlah frekuensi setiap kategori atau kelompok nilai tertentu. Frekuensi adalah jumlah atau banyaknya data yang muncul dalam kategori tertentu.

1. Jenis Tabel Distribusi Frekuensi

   Terdapat beberapa jenis tabel distribusi frekuensi yang umum digunakan:

   a. Distribusi Frekuensi Sederhana

Ini adalah bentuk paling dasar dari tabel distribusi frekuensi, yang hanya mencantumkan kategori nilai dan jumlah kemunculan (frekuensi) dari setiap nilai tersebut. Bentuk ini digunakan untuk data yang tidak memiliki banyak variasi atau kisaran nilai.

Nilai	Frekuensi
10	2
20	3
30	4
40	1

b. Distribusi Frekuensi Relatif

Tabel ini menampilkan frekuensi relatif dari setiap nilai, yang biasanya dinyatakan dalam bentuk persentase atau proporsi dari total data.

Nilai	Frekuensi	Frekuensi Relatif (%)
10	2	20%
20	3	30%
30	4	40%
40	1	10%

c. Distribusi Frekuensi Kumulatif

Distribusi frekuensi kumulatif menunjukkan jumlah kumulatif dari frekuensi, yakni dengan menjumlahkan frekuensi sebelumnya hingga nilai yang bersangkutan.

Nilai	Frekuensi	Frekuensi Kumulatif
10	2	2
20	3	5
30	4	9
40	1	10

d. Distribusi Frekuensi Kelompok (*Grouped Frequency Distribution*)

Untuk data yang terdiri dari banyak nilai berbeda, nilai-nilai tersebut biasanya dikelompokkan ke dalam interval tertentu atau kelas interval. Setiap kelas memiliki rentang nilai yang diwakili oleh frekuensi masing-masing.

Interval Kelas	Frekuensi
10 - 19	2
20 - 29	3

30 - 39	4
40 - 49	1

30 Langkah-langkah Membuat Tabel Distribusi Frekuensi

Berikut adalah tahapan umum dalam menyusun tabel distribusi frekuensi, khususnya distribusi frekuensi kelompok:

a. Tentukan Jumlah Kelas

Menggunakan rumus *Sturges*:

$$K = 1 + 3.3 \log n$$

di mana K adalah jumlah kelas dan n adalah jumlah total data.

b. Tentukan Rentang Data

Rentang (*Range*) dihitung dengan cara mengurangkan nilai maksimum dengan nilai minimum:

$$Rentang = Nilai\ Maksimum - Nilai\ Minimum$$

c. Tentukan Panjang Interval Kelas

Panjang interval dihitung dengan membagi rentang data dengan jumlah kelas:

$$Panjang\ Interval = \frac{Rentang}{K}$$

d. Tentukan Batas Kelas

   Batas kelas harus ditentukan untuk setiap interval sehingga setiap data dapat ditempatkan dengan jelas dalam kelas tertentu.

e. Hitung Frekuensi Setiap Kelas

   Menghitung jumlah data yang termasuk dalam setiap kelas.

31 Keuntungan Penggunaan Tabel Distribusi Frekuensi

a. Menyederhanakan Data: Data yang besar bisa diringkas ke dalam beberapa kategori atau kelas sehingga lebih mudah dianalisis.
b. Identifikasi Pola: Tabel ini memudahkan identifikasi pola atau distribusi dari data, seperti kecenderungan data terkonsentrasi di sekitar nilai tertentu.
c. Perbandingan Data: Memudahkan dalam membandingkan berbagai kategori atau kelompok data.
d. Dasar Visualisasi: Tabel distribusi frekuensi sering digunakan sebagai dasar untuk membuat grafik seperti histogram, diagram batang, atau ogive.

32 Keterbatasan Tabel Distribusi Frekuensi

a. Hilangnya Detail Data Individu: Saat data dikelompokkan dalam interval, informasi detail dari data individu bisa hilang.
b. Tergantung pada Kelas Interval: Hasil analisis bisa berbeda tergantung pada bagaimana interval kelas ditentukan.

Tabel distribusi frekuensi adalah alat yang sangat bermanfaat dalam meringkas dan memahami distribusi suatu kumpulan data. Sebagai langkah awal dalam analisis statistik, tabel ini memberikan fondasi yang kuat untuk berbagai analisis lebih lanjut.

2.2. Diagram dan Grafik

Diagram dan Grafik merupakan alat visualisasi yang digunakan dalam statistik untuk menyajikan data secara grafis sehingga lebih mudah dipahami dan diinterpretasikan. Penggunaan diagram dan grafik memudahkan dalam melihat pola, tren, hubungan, dan distribusi data yang mungkin sulit dikenali jika hanya disajikan dalam bentuk tabel. Setiap jenis diagram atau

grafik memiliki fungsi tertentu yang disesuaikan dengan tipe data dan tujuan analisis.

1. Jenis-jenis Diagram dan Grafik
    a. Diagram Batang (*Bar Chart*)

        Diagram batang adalah representasi data dalam bentuk batang vertikal atau horizontal. Setiap batang mewakili kategori data, dan panjang batang menunjukkan frekuensi atau nilai dari data tersebut. Diagram ini cocok untuk data kategoris (kualitatif) atau diskret.

        Ciri-ciri:

        - Digunakan untuk membandingkan kategori yang berbeda.
        - Batang bisa vertikal atau horizontal.
        - Jarak antar batang biasanya sama.

        Contoh: Sebuah diagram batang yang menunjukkan frekuensi penjualan produk berbeda selama bulan tertentu.

    b. Diagram Lingkaran (*Pie Chart)*

        Diagram lingkaran digunakan untuk menyajikan data sebagai bagian dari keseluruhan dalam bentuk sektor-sektor lingkaran. Setiap sektor menunjukkan

proporsi atau persentase dari keseluruhan data.

Ciri-ciri:

- Sering digunakan untuk data kategoris.
- Menunjukkan kontribusi setiap bagian terhadap total.
- Cocok untuk membandingkan bagian-bagian dalam satu set data.

Contoh: Sebuah diagram lingkaran yang menunjukkan distribusi anggaran sebuah perusahaan.

c. Histogram

Histogram adalah diagram yang menyerupai diagram batang, tetapi digunakan untuk data kontinu. Batang dalam histogram menunjukkan frekuensi distribusi data dalam interval kelas yang ditentukan. Tidak ada jarak antara batang, karena data kontinu.

Ciri-ciri:

- Digunakan untuk data kontinu.
- Interval kelas ditampilkan pada sumbu horizontal.

- Tinggi batang menunjukkan frekuensi data dalam interval tertentu.

Contoh: Sebuah histogram yang menunjukkan distribusi berat badan sekelompok siswa dalam interval tertentu.

d. Poligon Frekuensi

Poligon frekuensi adalah grafik garis yang menggambarkan distribusi frekuensi dari data. Titik-titik pada poligon dihubungkan oleh garis lurus dan mewakili frekuensi pada titik tengah setiap kelas interval.

Ciri-ciri:

- Digunakan untuk data kontinu.
- Menyambungkan titik-titik frekuensi pada interval kelas.
- Memungkinkan perbandingan beberapa distribusi data.

Contoh: Poligon frekuensi untuk data yang menunjukkan tinggi badan dalam rentang kelas tertentu.

e. Ogive (Grafik Frekuensi Kumulatif)

Ogive adalah grafik garis yang menunjukkan frekuensi kumulatif dari data. Grafik ini membantu untuk menentukan

jumlah data yang berada di bawah atau di atas nilai tertentu.

Ciri-ciri:

- Digunakan untuk melihat frekuensi kumulatif data.
- Meningkat seiring bertambahnya frekuensi.

Contoh: Sebuah ogive yang menunjukkan berapa banyak siswa yang mendapat nilai di bawah atau di atas 70.

f. Diagram Pencar (*Scatter Plot*)

Diagram pencar digunakan untuk menyajikan hubungan antara dua variabel numerik. Setiap titik pada grafik mewakili pasangan nilai dari dua variabel tersebut. *Scatter plot* sering digunakan untuk memvisualisasikan korelasi atau hubungan antara variabel.

Ciri-ciri:

- Menunjukkan hubungan antara dua variabel.
- Cocok untuk data numerik atau kontinu.
- Dapat membantu mengidentifikasi korelasi (positif, negatif, atau tidak ada hubungan).

Contoh: *Scatter plot* yang menunjukkan hubungan antara tinggi badan dan berat badan sekelompok orang.

g. Diagram Batang Bertumpuk (*Stacked Bar Chart*)

Diagram batang bertumpuk digunakan untuk menyajikan data dalam kategori yang berbeda tetapi dengan menumpuk satu kategori di atas yang lain. Ini memudahkan untuk membandingkan kontribusi relatif setiap kategori terhadap total.

Ciri-ciri:

- Menunjukkan kontribusi berbagai komponen dalam kategori yang sama.
- Digunakan untuk data kategoris.

Contoh: Sebuah diagram batang bertumpuk yang menunjukkan penjualan berbagai jenis produk di berbagai cabang.

h. Boxplot (*Box-and-Whisker Plot*)

Boxplot adalah grafik yang menunjukkan penyebaran data berdasarkan kuartil. Ini memberikan gambaran tentang distribusi, median, dan outlier dalam data.

Ciri-ciri:

- Menunjukkan distribusi data berdasarkan lima angka penting: minimum, kuartil pertama (Q1), median (Q2), kuartil ketiga (Q3), dan maksimum.
- *Outlier* sering ditandai dengan titik di luar kotak.

Contoh: Boxplot yang menunjukkan distribusi nilai ujian dalam satu kelas.

i. Grafik Garis *(Line Chart*)

Grafik garis digunakan untuk menunjukkan perubahan nilai data dari waktu ke waktu. Setiap titik pada grafik mewakili nilai data pada saat tertentu, dan titik-titik tersebut dihubungkan oleh garis.

Ciri-ciri:

- Cocok untuk data deret waktu (*time series*).
- Sering digunakan untuk menggambarkan tren.

Contoh: Grafik garis yang menunjukkan perubahan suhu udara harian selama satu bulan.

2. Langkah-langkah Membuat Diagram dan Grafik

a. Pilih Jenis Data: Tentukan jenis data (diskrit atau kontinu, kategoris atau numerik) sebelum memilih jenis diagram atau grafik yang sesuai.
b. Tentukan Skala: Skala pada sumbu harus dipilih dengan tepat agar data dapat ditampilkan dengan jelas.
c. Label Sumbu dan Judul: Beri label yang jelas pada sumbu x dan y, dan sertakan judul yang informatif agar grafik mudah dimengerti.
d. Plot Data: Plotkan data sesuai dengan jenis grafik yang digunakan. Pastikan akurasi dalam memplot data.
e. Tinjau Grafik: Periksa kembali apakah grafik sudah menyampaikan informasi yang benar dan mudah dipahami.

3. Kelebihan dan Keterbatasan Penggunaan Diagram dan Grafik
   a. Kelebihan:
      - Memudahkan Interpretasi Data: Diagram dan grafik membantu dalam menyederhanakan data yang besar sehingga lebih mudah diinterpretasikan.

- Identifikasi Pola: Pola, tren, dan hubungan antar data lebih mudah dikenali melalui visualisasi.
- Komunikasi yang Efektif: Diagram dan grafik memudahkan dalam menyampaikan informasi kompleks secara cepat dan efektif.

b. Keterbatasan:

- Kesalahan Interpretasi: Penggunaan skala yang tidak tepat atau grafik yang tidak sesuai dapat menyebabkan kesalahan interpretasi.
- Kehilangan Detail: Informasi detail dari data individu mungkin hilang, terutama dalam grafik yang hanya menyajikan ringkasan data.

2.3. Peta Deta

Peta Data adalah representasi visual yang digunakan untuk menampilkan data geografis dalam bentuk peta, di mana data ditempatkan pada wilayah geografis yang relevan. Peta data memungkinkan kita untuk mengidentifikasi pola, distribusi, dan hubungan antar data yang terkait dengan lokasi atau wilayah. Penggunaan peta dalam analisis data sangat membantu

untuk memahami data spasial atau data yang berhubungan dengan lokasi geografis.

Peta data (*map*) adalah bentuk representasi visual yang menggabungkan peta geografis dengan informasi numerik atau kategoris. Dengan memetakan data pada lokasi tertentu, peta data memungkinkan untuk menyoroti karakteristik geografis yang relevan dengan variabel yang dianalisis, seperti kepadatan populasi, tingkat kriminalitas, jumlah kasus penyakit, distribusi ekonomi, dan lain-lain.

1. Jenis-Jenis Peta Data
    a. Peta Tematik (*Thematic Map*)

        Peta tematik adalah jenis peta yang digunakan untuk menggambarkan informasi atau tema khusus yang terikat dengan lokasi geografis tertentu. Peta ini sering digunakan dalam berbagai disiplin ilmu seperti ekonomi, ekologi, epidemiologi, dan demografi. Peta tematik memetakan data di atas peta geografis dan biasanya berfokus pada satu aspek informasi tertentu.

        Contoh: Peta tematik yang menunjukkan distribusi curah hujan di seluruh negara atau wilayah.

b. Peta Choropleth (*Choropleth Map*)

Peta *choropleth* adalah jenis peta tematik yang menunjukkan data kuantitatif atau kategoris berdasarkan wilayah geografis dengan menggunakan gradien warna. Biasanya, semakin tinggi nilai dari suatu variabel, semakin gelap atau kuat warna yang digunakan di wilayah tersebut.

- Kelebihan: Peta *choropleth* memudahkan visualisasi perbandingan antar wilayah geografis.
- Kekurangan: Data yang disajikan adalah agregat berdasarkan wilayah, yang mungkin tidak mencerminkan variasi dalam wilayah kecil.
- Contoh: Peta *choropleth* yang menampilkan kepadatan penduduk berdasarkan provinsi atau negara bagian.

c. Peta Titik (*Dot Density Map*)

Peta titik menampilkan distribusi atau kepadatan suatu fenomena atau data dengan menggunakan titik-titik di atas peta

geografis. Setiap titik biasanya mewakili sejumlah data atau nilai tertentu.

- Kelebihan: Peta ini sangat efektif untuk menunjukkan distribusi data dalam area geografis yang luas.
- Kekurangan: Untuk data dengan kepadatan tinggi, peta titik bisa menjadi sulit dibaca karena jumlah titik yang terlalu banyak.
- Contoh: Peta titik yang menunjukkan lokasi fasilitas kesehatan di suatu negara.

d. Peta Isoplet (*Isopleth Map*)

Peta isoplet adalah peta yang menggunakan garis kontur untuk menunjukkan area dengan nilai yang sama. Peta ini sering digunakan dalam geografi fisik, seperti dalam menunjukkan tekanan atmosfer, suhu, atau curah hujan.

- Kelebihan: Isoplet sangat berguna untuk menunjukkan distribusi data kontinu dalam suatu area geografis.
- Kekurangan: Membutuhkan interpolasi, sehingga mungkin kurang akurat di beberapa wilayah yang datanya terbatas.

- Contoh: Peta isoplet yang menunjukkan distribusi suhu atau tekanan udara pada wilayah tertentu.

e. Peta Simbol Proporsional (*Proportional Symbol Map*)

Peta simbol proporsional menampilkan data kuantitatif menggunakan simbol dengan ukuran yang berbeda sesuai dengan nilai data yang direpresentasikan. Biasanya, semakin besar nilai suatu variabel, semakin besar simbol yang digunakan.

- Kelebihan: Peta ini efektif untuk membandingkan jumlah relatif atau besaran suatu fenomena di berbagai wilayah.
- Kekurangan: Ukuran simbol yang terlalu besar dapat menutupi detail geografis di sekitar simbol tersebut.
- Contoh: Peta simbol proporsional yang menunjukkan volume perdagangan antar negara dengan lingkaran proporsional.

f. Peta Panas (*Heat Map*)

Peta panas adalah representasi visual data di mana intensitas warna digunakan untuk

menunjukkan konsentrasi atau kepadatan data. Biasanya, warna yang lebih gelap atau lebih cerah menunjukkan area dengan nilai yang lebih tinggi atau lebih banyak data.

- Kelebihan: Peta ini sangat intuitif untuk mengenali pola atau area konsentrasi data.
- Kekurangan: Kurang cocok untuk menunjukkan data kategoris atau nilai absolut yang tepat.
- Contoh: Peta panas yang menunjukkan kepadatan lalu lintas di kota-kota besar.

g. Peta Aliran (*Flow Map*)

Peta aliran menunjukkan gerakan atau arus dari satu tempat ke tempat lain. Peta ini sering digunakan untuk menunjukkan migrasi, arus perdagangan, aliran sungai, dan lain-lain. Garis pada peta menunjukkan arah dan volume aliran, dengan ketebalan garis menunjukkan besarnya volume.

- Kelebihan: Peta ini sangat cocok untuk menggambarkan proses dinamis, seperti pergerakan populasi atau barang.

- Kekurangan: Bisa menjadi sangat kompleks jika ada terlalu banyak jalur atau aliran yang tumpang tindih.
- Contoh: Peta aliran yang menunjukkan migrasi penduduk antar negara.

2. Langkah-langkah Membuat Peta Data
   a. Pilih Jenis Data

      Langkah pertama adalah memilih jenis data yang akan divisualisasikan. Data tersebut bisa berupa data kuantitatif (seperti jumlah penduduk, pendapatan, suhu, dll.) atau data kategoris (seperti jenis tanah, jenis iklim, dll.).

   b. Pilih Peta Dasar

      Peta dasar atau peta referensi adalah peta geografis yang akan digunakan sebagai latar belakang untuk memplot data. Peta ini bisa berupa peta dunia, negara, atau wilayah tertentu tergantung dari cakupan data yang digunakan.

   c. Plot Data pada Peta

      Setelah peta dasar dipilih, langkah selanjutnya adalah memplot data pada wilayah atau lokasi yang sesuai. Ini

dilakukan dengan cara mengasosiasikan data dengan koordinat atau area geografis pada peta.

d. Tentukan Visualisasi yang Tepat

Pilih metode visualisasi yang sesuai dengan jenis data dan tujuan analisis. Misalnya, untuk menunjukkan variasi tingkat kemiskinan antar wilayah, peta choropleth mungkin lebih tepat. Untuk menunjukkan lokasi fasilitas kesehatan, peta titik dapat digunakan.

e. Tentukan Warna atau Simbol

Warna atau simbol yang digunakan pada peta harus jelas dan informatif. Penggunaan skala warna yang konsisten, simbol yang proporsional, dan label yang jelas sangat penting untuk mempermudah interpretasi.

f. Tinjau dan Sesuaikan

Setelah peta selesai dibuat, tinjau kembali apakah peta tersebut memberikan informasi yang akurat dan mudah dipahami. Pastikan legenda, judul, dan skala peta tercantum dengan benar.

3. Keuntungan Penggunaan Peta Data
    a. Visualisasi Spasial yang Efektif: Peta data memungkinkan untuk melihat pola distribusi geografis dan tren yang mungkin sulit dikenali hanya dengan menggunakan tabel atau grafik biasa.
    b. Identifikasi Pola dan Tren: Peta data memudahkan untuk mengidentifikasi pola-pola spasial, seperti kepadatan populasi di kota-kota besar, penyebaran penyakit, atau tingkat pengangguran di berbagai wilayah.
    c. Pengambilan Keputusan: Peta data sangat membantu dalam pengambilan keputusan yang berhubungan dengan lokasi, seperti penentuan lokasi fasilitas publik atau rute distribusi logistik.
4. Keterbatasan Peta Data
    a. Kesalahan Interpretasi: Jika tidak dipilih dengan hati-hati, warna, skala, atau simbol yang digunakan pada peta dapat menyesatkan atau menyebabkan kesalahan interpretasi data.
    b. Keterbatasan Akurasi: Peta yang hanya menyajikan data agregat berdasarkan wilayah besar mungkin tidak menunjukkan

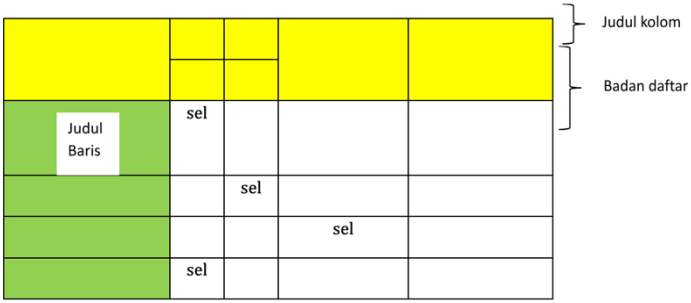
variasi detail yang ada dalam wilayah tersebut.

c. *Overplotting*: Pada beberapa jenis peta, seperti peta titik atau simbol proporsional, terlalu banyak data bisa menyebabkan peta menjadi penuh dan sulit dibaca.

BAB III
TABEL DISTRIBUSI FREKUENSI

3.1 Pendahuluan

Istilah distribusi frekuensi merupakan sebuah cara menyajikan data yang dimulai dengan membuat tabel distribusi frekuensi. Sebelum dipelajari bagaimana cara membuat tabel distribusi frekuensi kita juga harus mengetahui istilah digunakan dalam membuat tabel. Pada bab sebelumnya yang menjelaskan terkait penyajian data statistik kita sudah mengetahui terlebih dahulu terkait istilah pada tabel seperti kolom tabel, judul baris, data pada sel dan badan daftar. Berikut ini dapat kita lihat contoh tabel penyajian data yang akan kita gunakan dalam membuat tabel distribusi frekuensi.



Berdasarkan tabel di atas baru selanjutkan kita akan membuat sebuah tabel distribusi frekuensi berdasarkan data sebaran sebanyak sejumlah "n" data yang nantinya akan dijelaskan dengan contoh-contoh. Pada judul kolom pertama sekali kita akan membuat 3 judul kolom yang terdiri dari (1) Nomor; (2) Rentang Nilai, dan (3) Frekuensi. Sedangkan pada judul baris akan kita isikan banyak kelas yang terdiri dengan nomor urut untuk memudahkan kita menyajikan datanya. Selajutnya dapat kita lihat pada gambar 3.2. untuk tabel distribusi frekuensi.

No	**Rentang Nilai (Kelas Interval)**	**Frekuensi (f)**
1	$X_1 - X_2$	2
2	$X_3 - X_4$	3
3	$X_5 - X_6$	5
dst	$X_n - X_{n+1}$	1

Sebelum memulai bagaimana cara mengisi tabel di atas kita juga harus memahami di awal tentang isitilah yang sering digunakan. Dalam daftar distribusi frekuensi banyak objek yang dikumpulkan dalam kelompok-kelompok yang berbentuk nilai X_1 sampai dengan X_2 yang sering disebut dengan "*interval*". Ke

dalam kelas interval ini antara X_1 sampai dengan X_2 dimasukkan semua data yang bernilai diantara dari nilai X_1 sampai dengan X_2.

Susunan dari kelas interval ini dimulai dari urutan yang data terkecil terus selanjut ke bawah hingga ke nilai data yang terbesar. Dimulai dari atas pada nomor 1 diberikan nama kelas interval pertama, pada nomor 2 diberikan nama kelas interval kedua, dan seterusnya sampai degan yang terakhir disebutkan sebagai kelas interval terakhir. Ini semuanya berada dalam kolom pada judul kolom Rentang Nilai (Kelas Interval). Judul kolom kelas interval ini nantinya boleh digantikan dengan judul berdasarkan data yang didapatkan seperti nilai ujian, nilai berat badan, ataupun nilai-nilai lainnya berdasarkan data yang disajikan.

Sedangkan pada kolom kanan dengan judul kolom Frekuensi (f) berisikan bilangan-bilangan yang menyatakan berapa jumlah data yang terdapat dalam tiap kelas interval. Jadi pada kolom ini berisikan frekuensi (f) yang dapat saja berupa jumlah orang atau jumlah objek yang sedang diteliti. Misalnya f=2 untuk kelas interval $X_1 - X_2$, ini dapat diartikan bahwa ada 2 orang atau 2 data yang memiliki nilai antara dari nilai X_1 sampai dengan X_2, dan begitu untuk seterusnya.

Bilangan-bilangan disebelah kiri kelas interval disebut sebagai *ujung bawah* dan bilangan-bilangan disebelah kanan disebut sebagai *ujung atas*. Ujung-ujung bawah kelas interval pertama, kedua, dan seterusnya adalah seperti X_1, X_3, X_5, dan X_n. Sedangkan ujung-ujung kelas interval atasnya seperti X_2, X_4, X_6, dan X_{n+1}. Selisih positif antara kedua ujung bawah berurutan disebut *panjang kelas interval* yang disingkat dengan "p". Jadi nilai $p = X_1 - X_3 = X_3 - X_5 = X_5 - X_n$.

Selain dari ujung kelas interval ada lagi yang biasanya disebut dengan *batas kelas interval*. Hal ini bergantung pada ketelitian data yang digunakan. Jika data di catat dengan teliti hingga satuan, maka *batas bawah kelas* sama dengan ujung bawah dikurangi dengan nilai tengah yaitu 0,5. Batas atasnya di dapat dari ujung atas ditambah dengan 0,5. Untuk data yang dicatat hingga satu desimal maka batas bawahnya sama dengan ujung bawah dikurangi dengan nilai 0,05 dan batas atas sama dengan ujung atasnya ditambahkan dengan nilai 0,05. Kalau data hingga dua desimal maka batas bawahnya sama dengan ujung bawah di kurangi dengan nilai 0,005 dan batas atas sama dengan ujung atasnya ditambahkan dengan nilai 0,005 dan begitu seterusnya.

3.2 Membuat Daftar Distribusi Frekuensi

Untuk memudahkan memahami cara membuat daftar distribusi frekuensi ada baiknya kita langsung membuatkan berdasarkan data contoh berikut ini. Pada contoh kali ini kita akan menghitung data dari hasil pengukuran berat badan walaupun banyak data yang dicontohkan pada penelitian adalah data nilai dari hasil belajar.

Berdasarkan hasil pengukuran berat badan mahasiswa dalam satuan kilo gram (kg) di sebuah universitas pada ruangan matakuliah statistika dasar dengan total mahasiswa sebanyak 40 orang (diketahui n adalah jumlah data sehingga n=40) diperoleh data sebagai berikut.

40	44	46	40	42	65
43	50	45	55	70	60
53	40	60	43	62	40
60	75	43	63	42	43
50	55	43	42	41	40
45	50	42	43	65	50
45	75	54	55		

Untuk membuat daftar distribusi frekuensi dengan panjang kelas yang sama sesuai dengan aturan yang

telah dijelaskan diatas, maka dilakukan langkah-langkah sebagai berikut :

1. Mencari nilai terbesar (NB)

   Berdasarkan data di atas kita peroleh bahwa nilai terbesar adalah 75.

2. Mencari nilai terkecil (NK)

   Berdasarkan data di atas juga kita peroleh bahwa nilai terkecil adalah 40.

3. Menghitung nilai rentang (R)

   R = Nilai Terbesar – Nilai Terkecil

   R = NB – NK = 75 – 40 = 35.

   Maka diperoleh nilai rentang sebesar 35.

4. Menghitung banyak kelas interval (BK), dimana "n" adalah jumlah data sehingga n=40.

   $BK = 1 + \{(3,3) log\ (n)\}$

   $BK = 1 + \{(3,3\} log\ (40)\}$

   $BK = 1 + \{(3,3)\ (1,6)\}$

   $BK = 1 + \{5,28\}$

   $BK = 6,26 \approx 6$

   Hasil perhitungan diperoleh nilai banyak kelas sebesar 6 sehingga nilai ini menjadi acuan untuk membuat baris pada kolom nomor sebenyak 6

baris.

5. Menghitung panjang kelas interval (P)

$$P = \frac{Rentang}{Banyak\ Kelas\ interval} = \frac{R}{BK}$$

Dari hasil perhitungan poin 3 diperoleh nilai rentang sebesar 35 dan nilai banyak kelas pada poin 4 sebesar BK = 6, maka di peroleh P adalah:

$$P = \frac{36}{6} = 6$$

Maka dapat di simpulkan bahwa panjang interval kelas sepanjang 6 angka, sehingga untuk nilai selisih positif antara kedua ujung bawah berurutan adalah 6 atau dapat dituliskan dengan persamaan yaitu :

$$p = 6 = X_1 - X_3 = X_3 - X_5 = X_5 - X_n$$

Berikut tabel distribusi yang dibuatkan berdasarkan hasil data yang telah diperoleh menurut tahapannya, yaitu :

No.	Rentang Nilai (Kelas Interval)	Frekuensi (f)
1	40 – 45	20
2	46 – 51	5
3	52 – 57	5

4	58 – 63	5
5	64 – 69	2
6	70 – 75	3
Total		**40**

Adapun penjelasan dalam penyusunan tabelnya adalah sebagai berikut :

1. Nomor urut sebanyak 6 baris diperoleh berdasarkan nilai banyak kelas interval yaitu BK=6.
2. Nilai ujung bawah kelas interval yang dimulai dengan nilai 40 dipilih dikarenakan nilai terkecil dari data adalah sebesar NK = 40.
3. Nilai ujung atas kelas interval yang diakhiri dengan nilai 45 dikarenakan hasil perhitungan dari panjang kelas interval di peroleh sebesar P=6, sehingga nilai 40 sampai dengan 46 adalah angka yang memiliki panjang interval 6 buah data yaitu 40, 41, 42, 43, 44 dan 45. Begitu juga seterusnya untuk baris ke 2 sampai dengan baris ke 6 pada judul kolom Rentang Nilai (Kelas Interval).
4. Untuk nilai yang diisikan pada judul kolom frekuensi adalah pada baris pertama nilai 20 diperoleh dari jumlah data yang ada sebanyak

20 orang yang memiliki berat badannya pada nilai 40 sampai dengan 45 kg. Sedangkan pada baris kedua nilai 5 diperoleh dari jumlah data yaitu sebanyak 5 orang memiliki berat badan antara 46 sampai dengan 51 kg. Begitu juga serterusnya hingga nomor baris ke-6 dimana nilai 3 pada kolom frekuensi tersebut adalah 3 diperoleh dari data dimana ada 3 orang yang memiliki berat badan dari nilai 70 sampai dengan 75 kg. Maka setelah di total diperoleh frekuensi sebanyak 40 data.

3.3 Distribusi Frekuensi Relatif

Pada pembahasan di atas sebelumnya telah dijelaskan cara membuat tabel distribusi frekuensi yang dinyatakan dengan banyak data yang terdapat di dalam kelas interval, sehingga dinyatakan dalam bentuk frekuensi absolut. Jika frekuensinya dinyakan dalam bentuk persentase maka dapat dikatakan sebagai Distribusi Frekuensi Relatif.

Tentu saja untuk frekuensi absolut dan frekuensi relatif dapat disajikan dalam sebuah daftar dan bentuk tabel. Adapun cara membuat tabel daftar distribusi frekuensi relatif kita menggunakan rumus menghitung

presentase dari jumlah frekuensi absolut terhadap total data yang dilambangkan dengan f_{rel} atau f (%), yaitu :

$$f_{rel} = f(\%) = \frac{f_i}{n} x100\%$$

Berikut ini tabel antara distribusi frekuensi absolut dan juga distribusi frekuensi relatif.

No.	Rentang Nilai (Kelas Interval)	Frekuensi Absolut (f)	Frekuensi Relatif f(%)
1	40 – 45	20	
2	46 – 51	5	
3	52 – 57	5	
4	58 – 63	5	
5	64 – 69	2	
6	70 – 75	3	
Total		**40**	**100**

Dari tabel diatas dapat kita selesaikan untuk kolom pada frekuensi relatif dengan tahapan pada nomor baris sebagai berikut :

1. Pada baris pertama dimana nilai f (absolut) adalah 20, sehingga nilai frekuensi relatifnya adalah

$$f\ (1) = \frac{20}{40} x100\% = 0{,}5\ x\ 100\% = 50\%$$

2. Pada baris ke-2, 3 dan ke-4 dimana nilai f (absolut) adalah 5, sehingga nilai frekuensi relatifnya adalah

$$f\ (1) = \frac{5}{40} x100\% = 0{,}125\ x\ 100\% = 12{,}5\%$$

3. Pada baris ke-5 dimana nilai f (absolut) adalah 2, sehingga nilai frekuensi relatifnya adalah

$$f\ (1) = \frac{2}{40} x100\% = 0{,}05\ x\ 100\% = 5\%$$

4. Pada baris ke-6 dimana nilai f (absolut) adalah 3, sehingga nilai frekuensi relatifnya adalah

$$f\ (1) = \frac{3}{40} x100\% = 0{,}075\ x\ 100\% = 7{,}5\%$$

Berdasarkan data perhitungan persentase yang diperoleh, maka tabel distribusi frekuensi relatif dapat dituliskan sebagai berikut :

No.	**Rentang Nilai (Kelas Interval)**	**Frekuensi Absolut (f)**	**Frekuensi Relatif f(%)**
1	40 – 45	20	50

2	46 – 51	5	12,5
3	52 – 57	5	12,5
4	58 – 63	5	12,5
5	64 – 69	2	5
6	70 – 75	3	7,5
Total		**40**	**100**

3.4 Distribusi Frekuensi Kumulatif

Setelah kita mengetahui distribusi frekuensi relatif maka ada satu lagi daftar yang biasa dinamakan dengan daftar distribusi frekuensi kumulatif atau penjumlahan secara keseluruhan. Daftar distribusi frekuensi kumulatif dapat dibentuk dari daftar distribusi frekuensi biasa yaitu dengan menjumlahkan frekuensi demi frekuensi. Saat ini dikenal dengan dua macam distribusi frekuensi kumulatif yaitu (1) kurang dari; dan (2) lebih dari sama dengan. Tentu saja untuk kedua hal ini terdapat pula frekuensi-frekuensi absolut dan frekuensi relatifnya. Berikut kita perhatikan daftar distribusi frekuensi kumulatif kurang dari.

No.	Nilai Berat Badan Kurang Dari	Frekuensi Kumulatif (f)
1	< 40	0
2	< 46	20
3	< 52	25

4	< 58	30
5	< 64	35
6	< 70	37
7	< 76	40

Pada gambar 3.7 dapat terlihat bahwa jumlah data orang yang berat badannya kurang dari 40 kg sebanyak 0 orang, begitu juga untuk data selanjutnya yang nilai kurang dari 46 sebanyak 20 orang. Begitu seterusnya sampai ada nilai kurang dari 76 terdapat sebanyak 40 data.

Selanjutnya kita akan memperhatikan daftar distribusi frekuensi lebih dari dan sama dengan, yaitu :

No.	**Nilai Berat Badan Kurang Dari**	**Frekuensi Kumulatif (f)**
1	≥ 40	40
2	≥ 46	20
3	≥ 52	15
4	≥ 58	10
5	≥ 64	5
6	≥ 70	3
7	≥ 76	0

Pada gambar 3.8 di atas dapat terlihat bahwa jumlah data orang yang berat badannya lebih dari dan

sama dengan 40 kg sebanyak 40 orang, begitu juga untuk data selanjutnya yang lebih dari dan sama dengan 46 sebanyak 20 orang. Begitu seterusnya sampai pada nilai lebih dari dan sama dengan 76 terdapat sebanyak 0 data.

BAB IV
UKURAN GEJALA PUSAT

4.1 Pengertian Ukuran Gejala Pusat

Ukuran gejala pusat, atau sering disebut sebagai measures of central tendency, merupakan nilai yang mewakili pusat atau nilai tengah dari sekumpulan data. Dalam analisis statistik, ukuran gejala pusat sangat penting untuk merangkum data dalam satu nilai yang menunjukkan titik keseimbangan atau karakteristik umum dari data tersebut. Nilai ini membantu kita untuk memahami data dalam bentuk yang lebih sederhana dan memungkinkan kita untuk melakukan perbandingan antara kelompok data yang berbeda.

1. Tujuan dan Fungsi

   Ukuran gejala pusat memiliki beberapa tujuan utama:

   a. Menyederhanakan Data: Dengan mengurangi sekumpulan data menjadi satu nilai representatif, ukuran gejala pusat membuat interpretasi data lebih mudah.

b. Perbandingan Antardata: Memungkinkan perbandingan antar kelompok data atau set data yang berbeda.
c. Prediksi dan Pengambilan Keputusan: Memberikan dasar bagi estimasi dan keputusan dalam berbagai bidang, termasuk ekonomi, pendidikan, kesehatan, dan ilmu sosial.

2. Jenis-Jenis Ukuran Gejala Pusat

Ukuran gejala pusat terdiri dari tiga jenis utama, yaitu:

a. *Mean* (Rata-rata): Mean atau rata-rata aritmatika adalah jumlah seluruh data dibagi dengan banyaknya data. Mean sering digunakan untuk data kuantitatif dan cocok untuk distribusi data yang simetris. Namun, mean kurang tahan terhadap data ekstrem (*outliers*) yang bisa menyebabkan distorsi.
b. Median: Median adalah nilai tengah dalam kumpulan data yang diurutkan. Untuk data yang berjumlah ganjil, median adalah nilai yang tepat di tengah; untuk data yang berjumlah genap, median adalah rata-rata dari dua nilai tengah. Median lebih tahan terhadap outliers dan sering digunakan

dalam distribusi data yang tidak simetris atau data ordinal.

c. Modus: Modus adalah nilai yang paling sering muncul dalam suatu kumpulan data. Modus sangat berguna untuk data kategorik atau data diskrit, terutama ketika kita ingin mengetahui kategori atau nilai yang paling umum.

3. Manfaat dan Keterbatasan

   Masing-masing ukuran gejala pusat memiliki kelebihan dan keterbatasan dalam analisis data:

   a. Mean: Representatif untuk distribusi simetris tetapi rentan terhadap outliers.
   b. Median: Cocok untuk data yang tidak simetris atau mengandung outliers, tetapi tidak mempertimbangkan seluruh nilai dalam data.
   c. Modus: Ideal untuk data kategorik, tetapi bisa tidak unik jika data memiliki lebih dari satu modus.

Ukuran gejala pusat sering dikombinasikan dengan ukuran dispersi, seperti simpangan baku (standar deviasi) atau rentang, untuk memberikan gambaran yang lebih komprehensif tentang karakteristik data.

4.2 Kegunaan Ukuran Gejala Pusat dalam Statistik

Ukuran gejala pusat sangat penting dalam statistik karena memberikan gambaran tentang karakteristik umum dari data yang dikumpulkan, seperti nilai rata-rata atau titik keseimbangan dari data tersebut. Kegunaan utama ukuran gejala pusat adalah untuk meringkas informasi kompleks ke dalam bentuk yang lebih sederhana, membantu dalam pengambilan keputusan, serta memungkinkan perbandingan dan analisis lebih lanjut.

1. Menyederhanakan Data

   Ukuran gejala pusat memungkinkan data yang besar dan kompleks diringkas ke dalam satu nilai representatif, seperti rata-rata, median, atau modus. Ini membantu kita memahami data secara keseluruhan tanpa harus menganalisis setiap nilai individu. Dengan demikian, ukuran gejala pusat sering digunakan dalam laporan atau presentasi untuk menyajikan data dalam bentuk yang lebih mudah dipahami.

2. Perbandingan Antara Kelompok Data

   Dalam statistik deskriptif, ukuran gejala pusat memungkinkan perbandingan yang mudah antara kelompok data yang berbeda. Misalnya, rata-rata pendapatan per kapita antara dua

negara atau rerata nilai siswa di dua sekolah dapat dibandingkan untuk memahami perbedaan atau persamaan di antara mereka. Median dan modus juga berguna untuk membandingkan data yang memiliki outliers atau distribusi yang tidak simetris.

3. Dasar Pengambilan Keputusan

   Ukuran gejala pusat memberikan dasar yang kuat untuk pengambilan keputusan berbasis data di berbagai bidang, seperti ekonomi, pendidikan, kesehatan, dan riset pasar. Misalnya:

   a. Dalam bisnis, rata-rata pengeluaran pelanggan dapat digunakan untuk merencanakan strategi penjualan.
   b. Dalam pendidikan, median nilai siswa dapat membantu guru menilai kinerja keseluruhan kelas.
   c. Dalam epidemiologi, rata-rata insiden penyakit per tahun dapat digunakan untuk membuat kebijakan kesehatan.

4. Menyediakan Informasi untuk Inferensi Statistik

   Ukuran gejala pusat tidak hanya digunakan dalam statistik deskriptif, tetapi juga berperan dalam inferensi statistik. Rata-rata atau median

dari sampel dapat digunakan untuk memperkirakan atau menginferensikan rata-rata populasi. Contoh kasusnya adalah analisis tren di mana rata-rata digunakan untuk memprediksi atau memperkirakan kejadian di masa depan.

5. Mengidentifikasi Pola dan Tren

   Ukuran gejala pusat membantu mengidentifikasi pola atau tren dalam data. Dengan mengamati perubahan rata-rata dari waktu ke waktu, kita dapat memahami tren pertumbuhan atau penurunan dalam bidang tertentu, seperti tren penjualan, tingkat kelahiran, atau hasil ujian. Misalnya, peningkatan rata-rata nilai ujian dari tahun ke tahun dapat mengindikasikan peningkatan kualitas pendidikan.

6. Mengatasi Ketidaksimetrisan Data dan Outliers

   Ukuran gejala pusat, seperti median dan modus, memungkinkan analisis yang lebih tepat pada data yang tidak simetris atau mengandung outliers. Median sangat berguna dalam mengatasi masalah yang disebabkan oleh data ekstrem yang dapat mempengaruhi rata-rata. Misalnya, untuk data pendapatan di mana

terdapat outliers berupa penghasilan sangat tinggi, median lebih representatif dibandingkan rata-rata.

Contoh kasus aplikasi ukuran gejala pusat misalnya, dalam penelitian tentang kesehatan masyarakat, rata-rata dan median digunakan untuk mengukur indeks massa tubuh (BMI) di berbagai wilayah. Informasi ini dapat membantu dalam merancang program diet atau kesehatan di wilayah tertentu berdasarkan rata-rata BMI penduduknya.

4.3 Penerapan dalam Data Kontinu dan Data Diskrit

Ukuran gejala pusat, seperti mean, median, dan modus, digunakan baik dalam data kontinu maupun data diskrit. Namun, cara penerapan dan interpretasi masing-masing ukuran gejala pusat berbeda antara dua jenis data ini, bergantung pada sifat data dan tujuan analisis.

1. Data Kontinu

   Data kontinu adalah data yang dapat mengambil setiap nilai dalam rentang tertentu, termasuk nilai desimal atau pecahan. Contoh umum dari

data kontinu meliputi tinggi badan, berat badan, waktu, suhu, dan pengukuran lain yang dapat memiliki nilai dengan pecahan atau nilai antara. Penerapan ukuran gejala pusat pada data kontinu biasanya bertujuan untuk menyajikan representasi umum dari rentang nilai yang dihasilkan oleh proses pengukuran tersebut.

Penerapan Ukuran Gejala Pusat dalam Data Kontinu:

a. Rata-rata (*Mean*): Mean sering kali digunakan untuk data kontinu karena memberikan nilai rata-rata yang memperhitungkan setiap observasi. Misalnya, rata-rata tinggi badan siswa di kelas memberikan informasi tentang tinggi tubuh rata-rata dari semua siswa.

b. Median: Median juga sering diterapkan untuk data kontinu, terutama ketika data memiliki outliers atau distribusi yang tidak simetris. Sebagai contoh, untuk data pendapatan tahunan, yang biasanya memiliki outliers (orang dengan pendapatan sangat tinggi), median akan lebih representatif karena menunjukkan

nilai tengah tanpa dipengaruhi oleh nilai ekstrem.

c. Modus: Modus pada data kontinu digunakan untuk mengidentifikasi nilai yang paling sering muncul, namun tidak selalu seumum mean dan median. Modus lebih bermanfaat ketika data kontinu dikategorikan atau diubah menjadi interval, seperti rentang suhu tertentu yang paling sering terjadi.

Contoh Aplikasi: Dalam statistik kesehatan, misalnya, mean dan median berat badan bayi baru lahir di rumah sakit dapat digunakan untuk memahami kondisi kesehatan bayi. Mean memberikan rata-rata keseluruhan, sedangkan median mungkin lebih mencerminkan kondisi umum jika terdapat beberapa bayi dengan berat badan ekstrem (sangat kecil atau sangat besar).

2. Data Diskrit

Data diskrit adalah data yang hanya dapat mengambil nilai tertentu, biasanya dalam bentuk bilangan bulat. Contoh data diskrit meliputi jumlah anak dalam keluarga, jumlah mobil yang terjual per hari, dan skor ujian yang diukur dalam angka bulat.

Data diskrit seringkali terbatas pada kategori atau angka yang spesifik dan tidak memiliki nilai antara.

Penerapan Ukuran Gejala Pusat dalam Data Diskrit:

a. Rata-rata (Mean): Mean pada data diskrit memberikan nilai rata-rata yang bisa tidak menjadi bilangan bulat. Misalnya, rata-rata jumlah kendaraan yang terjual per hari di sebuah dealer mungkin adalah 7,5, meskipun pada kenyataannya hanya 7 atau 8 kendaraan yang bisa terjual dalam sehari.

b. Median: Median dalam data diskrit dihitung dengan cara yang sama seperti pada data kontinu, yaitu dengan mencari nilai tengah. Jika data memiliki jumlah bilangan yang ganjil, nilai median adalah nilai tengah. Median sangat berguna dalam data diskrit dengan distribusi yang tidak simetris.

c. Modus: Modus sangat sering digunakan dalam data diskrit, terutama ketika data bersifat kategorikal atau frekuensi tertentu sangat tinggi. Modus dalam data diskrit memberikan nilai atau kategori yang paling sering muncul. Sebagai contoh, dalam survei mengenai

preferensi merek ponsel, modus menunjukkan merek yang paling banyak dipilih oleh responden.

Contoh Aplikasi: Dalam pendidikan, jumlah siswa yang memperoleh nilai tertentu pada ujian dapat dianalisis menggunakan ukuran gejala pusat. Misalnya, mean skor ujian menunjukkan performa rata-rata seluruh siswa, sementara modus menunjukkan nilai yang paling sering diperoleh siswa, yang berguna untuk melihat apakah ada kecenderungan pada nilai tertentu.

Perbedaan Pendekatan dalam Data Kontinu dan Data Diskrit

a. Pengaruh *Outliers*: Pada data kontinu, median sering dipilih untuk mengurangi pengaruh outliers. Pada data diskrit, outliers tidak selalu memiliki pengaruh signifikan, terutama pada data diskrit dengan jumlah kategori yang terbatas.

b. Interprestasi Mean: Mean pada data kontinu lebih tepat mewakili nilai tengah karena mengikutsertakan setiap data dalam perhitungan. Dalam data diskrit, mean bisa jadi tidak realistis jika hasilnya bukan bilangan

bulat, meskipun tetap bisa memberikan gambaran rata-rata.

4.4 Hubungan dengan Ukuran Dispersi

Ukuran gejala pusat dan ukuran dispersi adalah dua konsep dasar dalam statistik deskriptif yang saling melengkapi untuk menggambarkan karakteristik data. Sementara ukuran gejala pusat, seperti mean, median, dan modus, menunjukkan lokasi pusat data atau nilai representatif dari dataset, ukuran dispersi memberikan informasi tentang seberapa tersebar data tersebut di sekitar nilai pusatnya. Keduanya penting untuk memahami distribusi data secara menyeluruh dan membuat interpretasi yang lebih akurat.

1. Ukuran Gejala Pusat

   Ukuran gejala pusat berfungsi untuk menunjukkan nilai representatif atau nilai yang "mewakili" sekumpulan data. Mean, median, dan modus adalah ukuran gejala pusat yang paling umum digunakan. Namun, ukuran gejala pusat saja tidak cukup untuk memahami keseluruhan sifat data karena nilai-nilai data

mungkin sangat bervariasi di sekitar nilai pusat tersebut.

2. Ukuran Dispersi

   Ukuran dispersi, seperti rentang (*range*), simpangan baku (*standard deviation*), variansi (*variance*), dan jangkauan interkuartil (*interquartile range*), menunjukkan tingkat penyebaran atau keragaman data di sekitar nilai gejala pusat. Ukuran ini menjelaskan seberapa dekat atau jauh nilai-nilai data dari nilai pusat, memberikan konteks tambahan tentang data.

   a. Rentang (*Range*): Mengukur perbedaan antara nilai tertinggi dan terendah, menunjukkan variasi total dalam dataset.

   b. Variansi (*Variance*) dan Simpangan Baku (Standard Deviation): Mengukur penyimpangan rata-rata dari nilai gejala pusat, memberikan gambaran tentang penyebaran data di sekitar mean.

   c. Jangkauan Interkuartil (*Interquartile Range*/IQR): Mengukur rentang antara kuartil pertama (Q1) dan kuartil ketiga (Q3), memberikan informasi tentang

penyebaran data di sekitar median, terutama untuk data dengan distribusi tidak simetris.

Hubungan antara Ukuran Gejala Pusat dan Dispersi

Ukuran gejala pusat dan ukuran dispersi bekerja bersama untuk memberikan gambaran lengkap tentang data:

a. Menginterpretasi Rata-Rata dengan Simpangan Baku

- Rata-rata (*mean*) sering dipadukan dengan simpangan baku (*standard deviation*) dalam analisis data. Rata-rata memberikan nilai pusat, sementara simpangan baku menunjukkan seberapa jauh data menyebar di sekitar rata-rata.

- Jika simpangan baku kecil, data cenderung terkonsentrasi di sekitar rata-rata, menunjukkan bahwa nilai-nilai data relatif konsisten.

- Sebaliknya, jika simpangan baku besar, data lebih tersebar, menunjukkan

variabilitas yang tinggi di sekitar rata-rata.

b. Median dan Jangkauan Interkuartil

- Median sangat cocok dipadukan dengan jangkauan interkuartil (IQR) untuk data yang tidak simetris atau memiliki outliers. Median memberikan nilai tengah, sedangkan IQR menunjukkan rentang dari data yang sebagian besar terletak di tengah (antara kuartil pertama dan kuartil ketiga).

- Jika IQR kecil, data sekitar median lebih homogen, tetapi jika IQR besar, ada penyebaran yang lebih luas di sekitar median.

c. Pengaruh *Outliers* dan Distribusi Data

- Ukuran gejala pusat dan ukuran dispersi juga saling mempengaruhi dalam kaitannya dengan outliers. Mean sangat dipengaruhi oleh *outliers*, dan dalam kasus ini, simpangan baku juga akan tinggi karena adanya nilai

ekstrem. Sebaliknya, median dan IQR lebih stabil terhadap outliers dan cenderung memberikan gambaran yang lebih akurat dalam situasi data yang memiliki *outliers*.

- Dalam distribusi normal (simetris), mean dan median hampir sama, dan simpangan baku memberikan informasi yang akurat tentang penyebaran data di sekitar mean.

Penggunaan dalam Analisis Risiko dan Keseimbangan

Ukuran dispersi digunakan dalam berbagai bidang untuk mengukur risiko atau ketidakpastian yang terkait dengan nilai gejala pusat. Misalnya, dalam keuangan, rata-rata return suatu investasi menunjukkan potensi profitabilitas, sementara simpangan baku dari return tersebut menunjukkan risiko atau volatilitas. Kombinasi ini memberikan gambaran lebih lengkap tentang potensi dan risiko dari sebuah investasi.

Dalam analisis nilai ujian siswa, rata-rata menunjukkan nilai tengah dari seluruh skor, tetapi simpangan baku menunjukkan variasi antara skor siswa. Jika simpangan baku rendah, sebagian besar siswa memiliki nilai yang dekat dengan rata-rata, menunjukkan bahwa tingkat pemahaman materi cenderung seragam di antara siswa. Namun, jika simpangan baku tinggi, berarti ada variabilitas yang besar dalam pemahaman siswa terhadap materi, sehingga langkah remedial mungkin diperlukan untuk siswa yang berada jauh dari rata-rata.

BAB V
UKURAN LETAK

Ukuran letak dalam statistik adalah nilai-nilai tertentu yang digunakan untuk menunjukkan posisi relatif data dalam distribusi atau penyebaran data. Ukuran letak memberikan gambaran di mana suatu data berada dalam suatu himpunan data, baik secara keseluruhan maupun dalam kelompok tertentu. Ukuran letak yang paling umum digunakan meliputi kuartil, desil, persentil, median, dan modus. Ukuran letak membantu memahami distribusi data dengan lebih baik, terutama dalam menentukan apakah suatu data berada di atas, di bawah, atau di tengah dari data lainnya.

5.1 Ukuran Letak Tertentu

Ukuran Letak Tertentu dalam statistik adalah konsep yang berkaitan dengan pengukuran posisi data dalam suatu distribusi. Ukuran ini memberikan informasi mengenai letak suatu nilai relatif terhadap nilai-nilai lain dalam data tersebut. Ada beberapa ukuran letak yang umum digunakan, antara lain kuartil,

desil, dan persentil. Setiap ukuran ini membagi data menjadi beberapa bagian yang sama besar dan membantu memahami distribusi data secara lebih rinci. Berikut penjelasan setiap ukuran:

1. Kuartil

   Kuartil membagi data menjadi empat bagian yang sama besar. Ada tiga kuartil utama yang biasa digunakan, yaitu:

   a. Kuartil pertama (Q1): Nilai yang membagi 25% data di bawahnya dan 75% data di atasnya.
   b. Kuartil kedua (Q2): Sama dengan median, membagi 50% data di bawahnya dan 50% di atasnya.
   c. Kuartil ketiga (Q3): Nilai yang membagi 75% data di bawahnya dan 25% di atasnya.

   Rumus kuartil untuk data berurutan:

   - $Q1 = \frac{n+1}{4}$
   - $Q2 = \frac{n+1}{2}$
   - $Q3 = \frac{3(n+1)}{4}$

   Dimana n adalah jumlah data.

2. Desil

   Desil membagi data menjadi sepuluh bagian yang sama besar. Ada sembilan desil yang digunakan, yaitu D1 hingga D9.

   a. Desil pertama (D1): 10% data di bawahnya dan 90% di atasnya.
   b. Desil kedua (D2): 20% data di bawahnya dan 80% di atasnya, dan seterusnya hingga desil ke-9 (D9).

   Rumus desil untuk data yang terurut:

$$Dk = \frac{k(n+1)}{10}$$

   Dimana k adalah nomor desil yang diinginkan dan n adalah jumlah data.

3. Persentil

   Persentil membagi data menjadi seratus bagian yang sama besar. Terdapat 99 persentil, yang biasa dilambangkan dengan P1 hingga P99.

   a. Persentil pertama (P1): 1% data di bawahnya dan 99% di atasnya.
   b. Persentil ke-50 (P50): Sama dengan median.

Rumus persentil:

$$Pk = \frac{k(n+1)}{100}$$

Dimana k adalah nomor persentil yang diinginkan dan n adalah jumlah data.

4. Interprestasi Ukuran Letak

   Ukuran letak digunakan untuk memahami bagaimana suatu nilai atau sekelompok nilai dalam kumpulan data terdistribusi. Ukuran letak dapat menunjukkan:

   a. Seberapa jauh nilai-nilai data tersebar.
   b. Posisi relatif suatu data dibandingkan data lainnya.
   c. Apakah distribusi data simetris atau miring.

Misalnya, jika kita ingin mengetahui bagaimana nilai ujian siswa berada dalam distribusi, kita bisa menggunakan kuartil, desil, atau persentil untuk melihat posisi nilai mereka dibandingkan dengan seluruh siswa.

5.2 Ukuran Letak Berdasarkan Posisi

Ukuran Letak Berdasarkan Posisi dalam statistik adalah metode untuk menggambarkan posisi suatu data

dalam distribusi. Ukuran ini membantu dalam memahami seberapa baik nilai tertentu mewakili keseluruhan data dan bagaimana data tersebut terdistribusi. Ukuran letak ini mencakup beberapa indikator, yaitu *mean* (rata-rata), median, modus, dan ukuran letak lainnya seperti kuartil, desil, dan persentil. Berikut adalah penjelasan masing-masing ukuran:

1. *Mean* (Rata-rata)

   Mean adalah ukuran pusat data yang dihitung dengan menjumlahkan semua nilai data dan membaginya dengan jumlah nilai tersebut.

   Rumus:

$$\text{Mean} = \frac{\sum X}{n}$$

   $\sum$x = Jumlah semua nilai

   n = Jumlah nilai

2. Median

   Median adalah nilai yang berada di tengah data ketika data tersebut diurutkan. Jika jumlah data ganjil, median adalah nilai tengah. Jika jumlah data genap, median adalah rata-rata dari dua nilai tengah.

   Contoh:

a. Untuk data 1, 3, 3, 6, 7, 8, 9 (jumlah ganjil), median adalah 6.
b. Untuk data 1, 2, 3, 4, 5, 6 (jumlah genap), median adalah (3+4)/2=3.5(3 + 4) / 2 = 3.5(3+4)/2=3.5.

3. Modus

   Modus adalah nilai yang paling sering muncul dalam sekumpulan data. Data bisa memiliki satu modus (unimodal), lebih dari satu modus (bimodal atau multimodal), atau tidak memiliki modus sama sekali.

   Contoh:

   a. Dalam data 1, 2, 2, 3, 4, 4, 4, 5, modusnya adalah 4 (muncul paling sering).
   b. Dalam data 1, 2, 3, 4, tidak ada modus karena semua nilai muncul sekali.

4. Kuartil

   Kuartil membagi data menjadi empat bagian yang sama. Ada tiga kuartil:

   a. Kuartil pertama (Q1): 25% data berada di bawahnya.
   b. Kuartil kedua (Q2): Median atau 50% data berada di bawahnya.
   c. Kuartil ketiga (Q3): 75% data berada di bawahnya.

5. Desil

   Desil membagi data menjadi sepuluh bagian yang sama. Terdapat sembilan desil yang digunakan untuk menandai posisi data.

6. Persentil

   Persentil membagi data menjadi seratus bagian yang sama. Ini sering digunakan untuk menentukan posisi relatif suatu nilai dalam populasi.

7. *Range* (Jarak)

   Range adalah selisih antara nilai maksimum dan nilai minimum dalam suatu data. Ini memberikan gambaran tentang rentang data.

   Rumus:

   $$Range = Nilai\ Maksimum - Nilai\ Minimum$$

8. *Interquartile Range* (IQR)

   IQR adalah selisih antara kuartil ketiga (Q3) dan kuartil pertama (Q1). Ini menunjukkan rentang di mana 50% nilai tengah dari data terletak.

   Rumus:

   $$\text{IQR} = Q3 - Q1$$

Ukuran letak berdasarkan posisi memberikan informasi yang penting dalam analisis data. Masing-masing ukuran memiliki kegunaan dan konteksnya

sendiri. Penggunaan ukuran ini sangat bergantung pada sifat data yang sedang dianalisis dan tujuan dari analisis tersebut.

5.3 Penerapan Ukuran Letak dalam Berbagai Bidang

Ukuran letak, yang meliputi kuartil, desil, dan persentil, adalah alat penting dalam statistik untuk mengetahui posisi relatif data dalam kumpulan data yang lebih besar. Penerapan ukuran letak digunakan dalam berbagai bidang untuk memberikan informasi yang lebih mendalam mengenai distribusi data. Berikut adalah beberapa penerapan ukuran letak dalam beberapa bidang penting.

1. Bidang Pendidikan
    a. Evaluasi Prestasi Siswa: Dalam sistem pendidikan, persentil sering digunakan untuk menilai posisi relatif prestasi seorang siswa dibandingkan dengan siswa lainnya. Misalnya, jika seorang siswa berada pada persentil ke-90 dalam ujian, ini berarti dia lebih baik daripada 90% siswa lainnya.
    b. Penempatan Siswa: Data hasil ujian yang diukur berdasarkan kuartil atau desil dapat membantu dalam pengelompokan siswa

berdasarkan kemampuan mereka. Misalnya, siswa di kuartil pertama dapat menerima bimbingan tambahan, sementara siswa di kuartil keempat mungkin ditempatkan di program yang lebih maju.

c. Pengukuran Standar Tes: Tes standar nasional atau internasional seperti SAT atau GRE menggunakan ukuran letak seperti persentil untuk menunjukkan bagaimana hasil ujian seorang siswa dibandingkan dengan peserta lainnya.

2. Bidang Ekonomi

a. Analisis Distribusi Pendapatan: Ukuran letak seperti desil dan kuartil sering digunakan untuk menganalisis distribusi pendapatan dalam suatu populasi. Misalnya, ekonom dapat membagi distribusi pendapatan menjadi desil untuk melihat proporsi populasi yang berada dalam kelompok penghasilan tertinggi atau terendah. Hal ini berguna untuk mengevaluasi tingkat ketimpangan pendapatan.

b. Penentuan Tingkat Kemiskinan: Pemerintah sering menggunakan ukuran

persentil untuk menentukan batasan-batasan yang digunakan untuk mendefinisikan kemiskinan. Sebagai contoh, rumah tangga yang memiliki pendapatan di bawah persentil ke-10 mungkin dikategorikan sebagai rumah tangga miskin.

c. Studi Pasar: Dalam riset pasar, ukuran letak digunakan untuk menganalisis distribusi penjualan atau pengeluaran konsumen. Persentil dapat menunjukkan bagaimana produk tertentu diterima oleh konsumen dan berapa banyak konsumen yang membeli pada tingkat harga tertentu.

3. Bidang Kesehatan

a. Penilaian Kesehatan Anak: Persentil sering digunakan dalam kesehatan anak untuk menilai pertumbuhan dan perkembangan fisik. Misalnya, grafik pertumbuhan bayi menggunakan persentil untuk menunjukkan bagaimana berat atau tinggi seorang bayi dibandingkan dengan bayi lain pada usia yang sama. Jika bayi berada di persentil ke-25, ini berarti berat badan atau

tinggi badannya lebih rendah dari 75% bayi lainnya.

b. Distribusi Penyakit: Ukuran letak juga digunakan untuk menganalisis distribusi penyakit dalam populasi. Misalnya, persentil dapat digunakan untuk memantau distribusi tekanan darah dalam populasi untuk menentukan proporsi individu yang berisiko terkena hipertensi.

c. Penggunaan Data dalam Epidemiologi: Data epidemiologi mengenai penyebaran penyakit atau faktor risiko kesehatan sering dianalisis dengan menggunakan kuartil dan desil. Dengan membagi populasi berdasarkan persentil, ahli epidemiologi dapat mengidentifikasi kelompok yang paling rentan terhadap penyakit tertentu.

4. Bidang Sosial

a. Studi Sosial Ekonomi: Dalam studi sosial-ekonomi, ukuran letak seperti desil dan kuartil membantu memahami distribusi aset, pengeluaran, atau faktor-faktor sosial lainnya dalam populasi. Misalnya, peneliti dapat membagi populasi berdasarkan kuartil pendapatan untuk menganalisis

bagaimana akses terhadap pendidikan atau layanan kesehatan bervariasi di antara kelompok pendapatan yang berbeda.

b. Analisis Mobilitas Sosial: Persentil digunakan untuk mempelajari mobilitas sosial, misalnya untuk menentukan berapa proporsi individu yang berhasil naik dari persentil bawah dalam hal penghasilan atau pendidikan ke persentil yang lebih tinggi selama hidup mereka.

5. Bidang Manajemen dan Bisnis

a. Evaluasi Kinerja Karyawan: Dalam manajemen sumber daya manusia, perusahaan sering menggunakan ukuran letak seperti kuartil untuk mengevaluasi kinerja karyawan. Misalnya, karyawan yang berada di kuartil keempat mungkin dianggap sebagai kinerja terbaik dan berhak atas promosi atau bonus.

b. Segmentasi Pasar: Persentil dapat digunakan dalam analisis data pelanggan untuk segmentasi pasar. Misalnya, pelanggan di persentil tertinggi berdasarkan pengeluaran bulanan dapat dianggap sebagai "pelanggan premium,"

yang bisa mendapatkan perlakuan khusus atau layanan VIP.

c. Analisis Risiko: Dalam keuangan perusahaan, ukuran letak digunakan untuk menganalisis risiko. Misalnya, dalam pengelolaan portofolio, nilai-nilai kuartil dari distribusi pengembalian investasi dapat digunakan untuk mengukur volatilitas atau risiko dari suatu investasi.

6. Bidang Olahraga

a. Penilaian Performa Atlet: Dalam olahraga, persentil digunakan untuk membandingkan performa atlet. Misalnya, data dari perlombaan lari atau tes kebugaran atlet dapat dikategorikan ke dalam persentil untuk menentukan seberapa baik kinerja atlet dibandingkan dengan atlet lainnya di kelompok umur yang sama.

b. Seleksi Atlet Berprestasi: Kuartil dapat digunakan dalam seleksi atlet untuk kompetisi, dengan atlet di kuartil keempat dianggap sebagai yang terbaik dan dipilih untuk mengikuti turnamen atau mewakili tim nasional.

Ukuran letak adalah alat penting dalam berbagai bidang untuk memahami distribusi data dan membantu dalam pengambilan keputusan. Baik dalam pendidikan, ekonomi, kesehatan, maupun bisnis, ukuran letak memberikan wawasan tentang bagaimana individu atau data berhubungan dengan keseluruhan distribusi. Dengan mengetahui posisi relatif seseorang atau entitas dalam konteks yang lebih luas, keputusan yang lebih akurat dan berbasis data dapat dibuat.

BAB VI
UKURAN SIMPANGAN BAKU DAN VARIANS

Analisis data dalam statistik sangat bergantung pada kemampuan untuk memahami variasi yang ada di dalam data tersebut. Setiap himpunan data memiliki karakteristik yang unik, dan salah satu cara untuk menggambarkan karakteristik tersebut adalah dengan mengukur seberapa tersebar atau bervariasi data itu. Dua ukuran utama yang digunakan dalam statistik untuk menggambarkan penyebaran atau variasi data adalah simpangan baku (standar deviasi) dan varians. Kedua konsep ini saling berkaitan dan memberikan informasi tentang sejauh mana data tersebar dari nilai rata-rata. Dalam bab ini, kita akan menjelaskan definisi, rumus, serta peran varians dan simpangan baku dalam analisis data.

6.1 Ukuran Simpangan Baku

Menurut Gunawan (2015) simpangan baku adalah akar dari tengah kuadrat simpangan dari nilai tengah atau akar simpangan rata-rata kuadrat. Untuk sampel simpangan baku disimbolkan dengan s. Sedangkan

simpangan baku populasi disimbolkan dengan σ (baca sigma) (Andi Asari, et. al. 2023).

1. Simpangan Baku Data Tunggal

   Simpangan Baku data tunggal dapat dicari dengan :

   a. Metode biasa

      Untuk sampel besar (n > 30)

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

      Untuk sampel kecil (n < 30)

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

      Contoh Soal:

      Tentukan ukuran simpangan baku dari nilai ujian dari kelima anak: 6,7, 8,9,10!

      Jawab:

      Gunakan table bantuan

x	**x^2**
6	36
7	49
8	64
9	81
10	100

40	**30**

$\bar{x}$ = 8 didapat dari 33040= 8,25

x	$x - x$	$(x - \bar{x})^2$
6	-2	4
7	-1	1
8	0	0
9	1	1
10	2	4
		10

Metode biasa

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

$$s = \sqrt{\frac{10}{5-1}} = \sqrt{2,5} = 1,58$$

2. Simpangan Baku Data Berkelompok
   a. Simpangan Baku data berkelompok dapat dicari dengan rumus:

      Untuk sampel besar (n > 30)

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

b. Untuk sampel kecil (n < 30)

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

Contoh Soal:

Tentukan Simpangan Baku dari nilai matematika dari 80 siswa SMA Nusa Bangsa!

Nilai	fi	xi	$fi.xi$	x-x	$(x-\bar{x})^2$	$fi(x-\bar{x})^2$
53 – 58	2	55,5	111	-19,5	380,25	760,5
59 – 64	12	61,5	738	-13,5	182,25	2187
65 – 70	10	67,5	675	-7,5	56,25	562,5
71 – 76	23	73,5	1.690,5	-1,5	2,25	51,75
77 – 82	14	79,5	1113	4,5	20,25	283,5
83 – 88	10	85,5	855	10,5	110,25	1102,5
89 – 94	5	91,5	457,5	16,5	272,25	1361,25
95 - 100	4	97,5	390	22,5	506,25	2025
	80		6.030			8.334

$\bar{x} = 75$ didapat dari $\frac{6.030}{80} = 75,375$

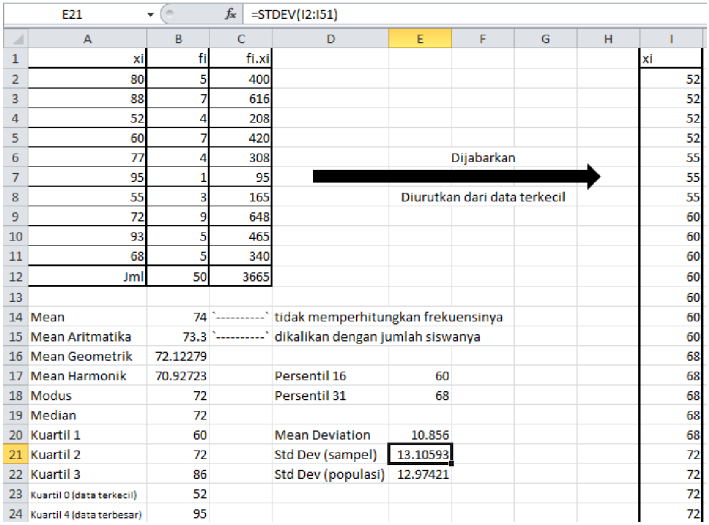
Metode Biasa

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

$$s = \sqrt{\frac{8.334}{80-1}} = \sqrt{\frac{8.334}{79}} = \sqrt{105,5} = 10,27$$

Perhitungan Simpangan Baku atau Standar Deviasi (Sampel) menggunakan excel

Terdapat dua cara untuk mencari simpangan baku. Pertama mencari simpangan baku yang berasal dari data sampel dengan bantuan program MS. Excel, yaitu rumus berikut : =STEDEV(x_1:x_n)



6.2 Ukuran Varians

Menurut Iesyah Rodliyah (2021) Varians (ragam) merupakan rata-rata dari jumlah kuadrat simpangan tiap data. Varians adalah alat ukut variabilitas serangkaian data yang dihitung dengan mencari rata-rata selisih/beda kuadrat antara data observasi dengan pusat datanya (Kustituanto & Badrudin, 1994). Varians untuk data populasi disimbolkan σ^2 dan untuk data sampel disimbolkan dengan s^2. $(X - \mu)^2$

1. Varians data tunggal

   Rumus untuk varians data tunggal adalah:

   a. Untuk populasi yang kurang dari 30 (n>30)

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

   b. Untuk sampel yang lebih dari 30 (n<30)

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

   Contoh Soal:

   Tentukan ukuran simpangan baku dari nilai ujian dari kelima anak: 6,7, 8,9,10

   Jawab:

x	x^2
6	36
7	49
8	64

9	81
10	100
40	330

$\bar{x}$ = 8 didapat dari $\frac{330}{40} = 8{,}25$

x	x- $\bar{x}$	$(x-\bar{x})^2$
6	-2	4
7	-1	1
8	0	0
9	1	1
10	2	4
		10

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

$$s^2 = \frac{10}{5-1} = 2{,}5$$

2. Varians data berkelompok

   Varians data berkelompok dapat dicari dengan rumus:

   a. Untuk sampel besar (n > 30)

$$\sigma^2 = \frac{\sum f(x-\mu)^2}{n}$$

b. Untuk sampel kecil (n < 30)

$$S^2 = \frac{\sum f(x-\bar{x})^2}{n-1}$$

Contoh Soal:

Tentukan Varians dari nilai matematika dari 80 siswa SMA Nusa Bangsa.

Nilai	fi	xi	$fi.xi$	x-x	$(x-\bar{x})^2$	$fi(x-\bar{x})^2$
53 – 58	2	55,5	111	-19,5	380,25	760,5
59 – 64	12	61,5	738	-13,5	182,25	2187
65 – 70	10	67,5	675	-7,5	56,25	562,5
71 – 76	23	73,5	1.690,5	-1,5	2,25	51,75
77 – 82	14	79,5	1113	4,5	20,25	283,5
83 – 88	10	85,5	855	10,5	110,25	1102,5
89 – 94	5	91,5	457,5	16,5	272,25	1361,25
95 - 100	4	97,5	390	22,5	506,25	2025
	80		6.030			8.334

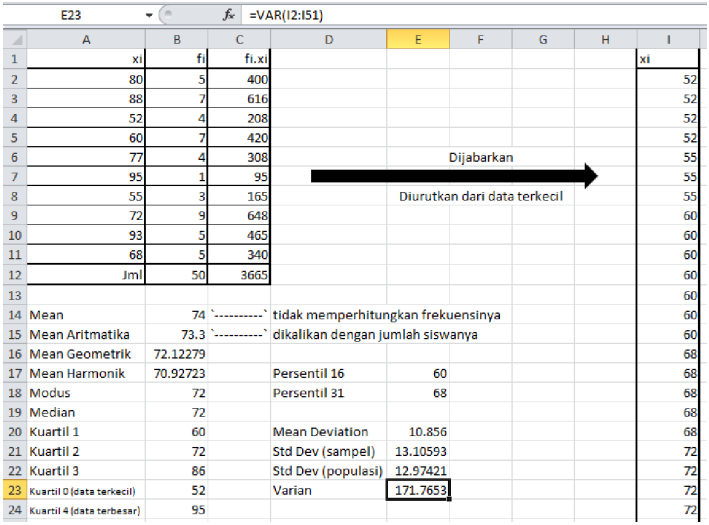
$\bar{x}$ = 75 didapat dari $\frac{6.30}{80} = 75{,}375$

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

$$s^2 = \frac{8.334}{80-1} = \frac{8.334}{79} = 105{,}5$$

Perhitungan Varian menggunakan excel

Cara mencari varian dengan bantuan MS. Excel menggunakan rumus : =VAR(x1,xn).



E23 =VAR(I2:I51)

	A	B	C	D	E	F	G	H	I
1	xi	fi	fi.xi						xi
2	80	5	400						52
3	88	7	616						52
4	52	4	208						52
5	60	7	420						52
6	77	4	308		Dijabarkan				55
7	95	1	95						55
8	55	3	165		Diurutkan dari data terkecil				55
9	72	9	648						60
10	93	5	465						60
11	68	5	340						60
12	Jml	50	3665						60
13									60
14	Mean	74	`----------`	tidak memperhitungkan frekuensinya					60
15	Mean Aritmatika	73.3	`----------`	dikalikan dengan jumlah siswanya					60
16	Mean Geometrik	72.12279							68
17	Mean Harmonik	70.92723		Persentil 16	60				68
18	Modus	72		Persentil 31	68				68
19	Median	72							68
20	Kuartil 1	60		Mean Deviation	10.856				68
21	Kuartil 2	72		Std Dev (sampel)	13.10593				72
22	Kuartil 3	86		Std Dev (populasi)	12.97421				72
23	Kuartil 0 (data terkecil)	52		Varian	171.7653				72
24	Kuartil 4 (data terbesar)	95							72

6.3 Peran Varians dan Simpangan Baku dalam Analisis Data

Varians dan simpangan baku memainkan peran penting dalam berbagai aspek analisis data. Kedua ukuran ini digunakan untuk memahami distribusi, pola, dan kecenderungan data yang lebih dalam, serta membantu dalam pengambilan keputusan berbasis data. Berikut adalah beberapa peran utama dari varians dan simpangan baku dalam analisis data:

1. Mengukur Penyebaran Data

Varians dan simpangan baku adalah ukuran penyebaran yang paling umum digunakan. Keduanya membantu untuk memahami seberapa jauh data menyebar dari rata-ratanya. Semakin besar nilai varians atau simpangan baku, semakin tersebar data, yang berarti bahwa data memiliki nilai yang lebih beragam. Sebaliknya, jika nilai varians atau simpangan baku kecil, ini menunjukkan bahwa data lebih terpusat di sekitar rata-rata.

2. Mendeteksi Anomali dan Outlier

   Dengan menggunakan simpangan baku, dapat mengidentifikasi apakah ada nilai data yang sangat jauh dari rata-rata, yang dikenal sebagai outlier atau nilai pencilan. Jika suatu nilai data terletak jauh lebih besar atau lebih kecil daripada simpangan baku dari rata-rata, nilai tersebut mungkin dianggap sebagai outlier. Hal ini penting dalam banyak aplikasi, seperti dalam deteksi kecurangan, analisis risiko, atau ketika mencoba memahami perilaku ekstrem dalam data.

3. Memahami Dispersi dalam Distribusi Data

   Dalam banyak kasus, data tidak selalu terdistribusi secara merata. Varians dan

simpangan baku memberikan informasi tentang distribusi tersebut. Sebagai contoh, dalam distribusi normal, sekitar 68% data berada dalam satu simpangan baku dari rata-rata, 95% dalam dua simpangan baku, dan 99,7% dalam tiga simpangan baku. Ini membantu dalam membuat prediksi dan inferensi statistik.

4. Pengambilan Keputusan dalam Risiko dan Ketidakpastian

   Dalam konteks bisnis dan keuangan, simpangan baku dan varians digunakan untuk mengukur risiko. Dalam analisis portofolio investasi, misalnya, simpangan baku dari return investasi digunakan untuk mengukur volatilitas aset. Semakin besar simpangan baku return, semakin besar risiko terkait dengan investasi tersebut. Investor seringkali membuat keputusan berdasarkan pada ukuran variabilitas return ini untuk mengelola risiko portofolio mereka.

5. Dasar untuk Uji Statistik

   Varians dan simpangan baku berfungsi sebagai dasar dalam banyak metode uji statistik, seperti ANOVA (Analisis Varians) dan t-test. Dalam uji ini, varians digunakan untuk memahami apakah perbedaan antara kelompok data adalah

signifikan secara statistik atau tidak. Dengan mengukur seberapa besar penyebaran data, kita dapat menentukan apakah hasil yang diperoleh adalah karena perbedaan yang nyata atau hanya akibat kebetulan semata.

6. Penyederhanaan Model Statistik

   Dalam model statistik, varians digunakan untuk memahami kesalahan model atau residu. Sebagai contoh, dalam regresi linear, varians residu (penyimpangan antara nilai aktual dan nilai prediksi) membantu dalam mengevaluasi seberapa baik model tersebut sesuai dengan data. Simpangan baku juga digunakan dalam analisis regresi untuk mengevaluasi *goodness-of-*fit dari suatu model.

7. Penentuan Keakuratan Estimasi

   Dalam banyak kasus, kita tidak selalu memiliki akses ke seluruh populasi data dan harus bekerja dengan sampel. Varians sampel membantu mengestimasi varians populasi. Semakin kecil varians sampel, semakin akurat estimasi kita tentang penyebaran data dalam populasi. Simpangan baku digunakan dalam interval kepercayaan untuk menentukan

seberapa jauh nilai rata-rata populasi mungkin berbeda dari rata-rata sampel yang diambil.

8. Evaluasi Kinerja Algoritma

   Dalam pembelajaran mesin dan kecerdasan buatan, varians dan simpangan baku dapat digunakan untuk mengevaluasi kinerja algoritma. Mereka membantu dalam menilai apakah model yang dibangun terlalu banyak mengakomodasi noise (*overfitting*) atau apakah model tersebut dapat digeneralisasi dengan baik ke data yang belum pernah dilihat.

BAB VII
POPULASI SAMPEL DAN TEKNIK SAMPEL

7.1 Populasi dan Sampel

1. Definisi Populasi dan Sampel dalam Penelitian

   Dalam penelitian ilmiah, istilah "populasi" dan "sampel" adalah konsep dasar yang sangat penting untuk dipahami. Populasi mengacu pada kumpulan individu, objek, atau elemen yang memiliki karakteristik tertentu yang menjadi objek penelitian. Sedangkan, sampel adalah bagian dari populasi yang diambil untuk tujuan penelitian. Penggunaan sampel ini memungkinkan peneliti untuk menarik kesimpulan tentang populasi tanpa harus meneliti keseluruhan populasi, yang sering kali tidak mungkin atau terlalu mahal untuk dilakukan. Populasi bisa berupa manusia, hewan, tanaman, perusahaan, atau objek lain tergantung dari fokus penelitian. Misalnya, jika penelitian dilakukan tentang preferensi konsumen terhadap produk elektronik, populasi bisa mencakup semua konsumen elektronik di

suatu wilayah tertentu. Namun, untuk membuat penelitian lebih praktis, peneliti memilih sampel yang representatif dari populasi tersebut untuk dianalisis.

2. Pentingnya Memilih Sampel yang Representatif
   Pemilihan sampel yang representatif sangat penting untuk menghasilkan kesimpulan yang valid dan akurat. Sampel yang representatif adalah sampel yang mencerminkan karakteristik utama dari populasi, sehingga hasil penelitian dari sampel dapat digeneralisasikan ke populasi secara keseluruhan. Ketidakrepresentatifan dalam pengambilan sampel dapat menyebabkan bias yang signifikan dalam hasil penelitian, yang pada akhirnya mengurangi validitas kesimpulan. Oleh karena itu, teknik sampling yang digunakan harus dipilih secara hati-hati untuk memastikan bahwa setiap anggota populasi memiliki peluang yang adil untuk dipilih sebagai bagian dari sampel.

3. Konsep dan Ruang Lingkup Populasi dan Sampel

   Dalam penelitian, ada beberapa konsep dasar yang harus dipahami terkait dengan populasi dan sampel:

   a. Populasi Target: Populasi yang menjadi fokus utama penelitian dan kepada siapa hasil penelitian akan digeneralisasikan.
   b. Populasi Akses: Subset dari populasi target yang secara praktis dapat diakses oleh peneliti. Ini sering kali ditentukan oleh keterbatasan geografis, waktu, atau sumber daya.
   c. Unit Sampel: Elemen individu dari populasi yang dipilih untuk diikutsertakan dalam sampel.
   d. Sampling Frame: Daftar atau kerangka yang digunakan untuk memilih sampel dari populasi.

   Populasi dan sampel dapat bervariasi tergantung pada jenis penelitian. Penelitian survei, eksperimen, maupun kuasi-eksperimen semuanya memerlukan pemahaman yang kuat tentang bagaimana populasi dan sampel berinteraksi dalam konteks penelitian tersebut.

4. Jenis-jenis Populasi

   Dalam penelitian, terdapat beberapa jenis populasi yang dapat diidentifikasi berdasarkan ruang lingkup penelitian:

   a. Populasi Terbatas: Populasi yang memiliki batasan geografis atau demografis yang jelas, misalnya siswa di satu sekolah atau penduduk suatu kota.
   b. Populasi Tak Terbatas: Populasi yang tidak memiliki batasan yang jelas atau yang terus berkembang seiring waktu, misalnya semua konsumen online di dunia.
   c. Populasi Finit: Populasi yang memiliki jumlah anggota yang terbatas dan dapat dihitung, seperti jumlah pegawai di suatu perusahaan.
   d. Populasi Infinit: Populasi yang anggotanya sangat banyak sehingga tidak bisa dihitung secara praktis, misalnya jumlah bakteri dalam suatu sampel tanah.

5. Kapan Sampel Dibutuhkan?

   Dalam banyak penelitian, penggunaan sampel menjadi sangat diperlukan karena faktor-faktor berikut:

a. Keterbatasan Waktu: Meneliti seluruh populasi akan membutuhkan waktu yang sangat lama.
b. Keterbatasan Biaya: Sumber daya finansial sering kali tidak mencukupi untuk melibatkan seluruh populasi.
c. Kemudahan dalam Pengumpulan Data: Mengumpulkan data dari sampel jauh lebih mudah dan praktis dibandingkan mengumpulkan data dari populasi penuh.
d. Ketersediaan Populasi: Dalam beberapa kasus, seluruh populasi mungkin tidak dapat diakses atau tidak dapat diidentifikasi secara lengkap.

Dengan alasan-alasan tersebut, penelitian ilmiah sering kali bergantung pada penggunaan sampel yang mewakili populasi.

7.2 Jenis-Jenis Populasi dan Sampel

1. Jenis-Jenis Populasi dan Sampel

   Dalam penelitian, pemahaman tentang berbagai jenis populasi dan sampel sangat penting agar hasil yang diperoleh dapat akurat dan sesuai dengan tujuan penelitian. Populasi mencakup seluruh elemen yang menjadi objek penelitian,

sedangkan sampel adalah bagian dari populasi yang dipilih untuk dianalisis lebih lanjut. Pemilihan sampel yang tepat dari populasi merupakan kunci untuk mendapatkan hasil yang dapat digeneralisasikan. Jenis-jenis populasi dan sampel beragam, tergantung dari karakteristik, cakupan, serta tujuan penelitian.

2. Jenis-Jenis Populasi

   Populasi dapat diklasifikasikan berdasarkan berbagai kriteria, seperti cakupan geografis, batasan waktu, dan sifat elemen-elemen dalam populasi. Berikut adalah beberapa jenis populasi yang sering digunakan dalam penelitian:

   a. Populasi Terbatas (*Finite Population*)

      Populasi terbatas adalah populasi yang memiliki jumlah elemen yang pasti dan dapat dihitung. Elemen-elemen dalam populasi ini biasanya memiliki batasan geografis atau demografis yang jelas. Misalnya, populasi mahasiswa di sebuah universitas atau populasi pegawai di sebuah perusahaan.

      Contoh:

- Populasi siswa di Sekolah Menengah Atas XYZ pada tahun ajaran 2024.
- Populasi petani di provinsi Jawa Barat.

b. Populasi Tak Terbatas (*Infinite Population*) Populasi tak terbatas adalah populasi yang memiliki jumlah elemen yang sangat besar sehingga tidak dapat dihitung atau terlalu besar untuk dihitung. Biasanya, populasi ini digunakan dalam studi teoretis atau simulasi matematis.

   Contoh:

   - Jumlah mikroorganisme dalam sampel tanah.
   - Jumlah pengguna internet di seluruh dunia.

c. Populasi Target Populasi target adalah populasi yang menjadi fokus utama penelitian, yaitu kelompok yang peneliti ingin tarik kesimpulannya. Populasi target bisa berupa individu, kelompok, organisasi, atau elemen lainnya yang relevan dengan tujuan penelitian.

   Contoh:

   - Populasi target dari penelitian tentang perilaku konsumen adalah semua pengguna

produk tertentu di wilayah yang ditentukan.

- Populasi target untuk studi kesehatan adalah seluruh penduduk di suatu negara.

d. Populasi Akses (*Accessible Population*) Populasi akses adalah bagian dari populasi target yang dapat diakses oleh peneliti untuk dijadikan sampel. Populasi akses biasanya lebih kecil dari populasi target karena keterbatasan geografis, waktu, atau sumber daya.

   Contoh:

   Jika populasi target adalah seluruh pelajar SMA di Indonesia, maka populasi akses bisa jadi hanya mencakup pelajar di beberapa provinsi yang dapat diakses oleh peneliti.

3. Jenis-Jenis Sampel

   Sampel adalah bagian dari populasi yang dipilih untuk dianalisis. Teknik pengambilan sampel yang tepat sangat penting untuk memastikan bahwa sampel yang diambil representatif dari populasi yang lebih luas. Berikut adalah beberapa jenis sampel yang umum digunakan dalam penelitian:

a. Sampel Acak Sederhana (*Simple Random Sampling*) Sampel acak sederhana adalah metode pengambilan sampel di mana setiap elemen dalam populasi memiliki peluang yang sama untuk dipilih. Teknik ini sering dianggap sebagai metode yang paling sederhana dan efektif untuk memastikan representativitas sampel.

   Contoh: Memilih 100 siswa secara acak dari seluruh siswa di sebuah sekolah untuk diikutsertakan dalam survei.

b. Sampel Sistematis (*Systematic Sampling)* Sampel sistematis adalah metode pengambilan sampel dengan memilih elemen-elemen pada interval tetap dari populasi. Setelah elemen pertama dipilih secara acak, elemen-elemen berikutnya diambil pada jarak tertentu (misalnya setiap elemen ke-5 atau ke-10).

   Contoh: Dari daftar 1000 karyawan, peneliti memilih setiap karyawan ke-10 untuk diikutsertakan dalam survei.

c. Sampel Bertingkat (*Stratified Sampling*) Sampel bertingkat digunakan ketika populasi terbagi dalam kelompok-kelompok (strata) yang berbeda. Peneliti memilih sampel dari setiap

strata untuk memastikan representasi yang proporsional dari kelompok-kelompok tersebut. Teknik ini efektif ketika populasi heterogen dan peneliti ingin memastikan bahwa setiap kelompok diwakili dalam sampel.

Contoh: Dalam penelitian tentang pendapatan, peneliti mungkin membagi populasi menjadi strata berdasarkan kelompok pendapatan rendah, menengah, dan tinggi, lalu mengambil sampel dari masing-masing kelompok.

d. Sampel Kluster (*Cluster Sampling)* Sampel kluster digunakan ketika populasi terbagi ke dalam kelompok-kelompok yang secara alami terbentuk, seperti kota, desa, atau kelas. Peneliti memilih sejumlah kluster secara acak dan kemudian mengumpulkan data dari semua elemen dalam kluster tersebut.

   Contoh: Dalam studi pendidikan, peneliti dapat memilih beberapa sekolah secara acak dan mengumpulkan data dari semua siswa di sekolah-sekolah tersebut.

e. Sampel Bertujuan (*Purposive Sampling*) Sampel bertujuan adalah teknik pengambilan sampel yang dilakukan berdasarkan pertimbangan atau kriteria tertentu yang ditentukan oleh peneliti.

Sampel ini dipilih karena elemen-elemen tersebut dianggap memiliki informasi yang relevan dengan penelitian.

Contoh: Memilih dokter spesialis tertentu untuk diwawancarai tentang pengobatan penyakit langka.

f. Sampel Bola Salju (Snowball Sampling) Sampel bola salju digunakan ketika populasi sulit dijangkau atau tidak dapat diidentifikasi dengan jelas. Peneliti memulai dengan sejumlah kecil responden dan meminta mereka untuk merekomendasikan individu lain yang relevan untuk diikutsertakan dalam penelitian.

Contoh: Penelitian tentang pengguna narkoba ilegal yang dilakukan dengan meminta peserta yang sudah diwawancarai untuk memperkenalkan peneliti kepada pengguna lainnya.

4. Perbandingan Antara Jenis-Jenis Sampel

Pemilihan jenis sampel sangat bergantung pada tujuan penelitian, sifat populasi, serta sumber daya yang tersedia. Sampel acak sederhana ideal untuk populasi yang homogen, sementara sampel bertingkat lebih efektif untuk populasi

yang heterogen. Sampel sistematis mudah diterapkan, tetapi mungkin tidak selalu representatif jika pola tertentu ada dalam populasi. Teknik-teknik non-probabilitas seperti purposive sampling dan snowball sampling berguna untuk penelitian eksploratif atau ketika akses ke populasi sulit diperoleh, tetapi hasil penelitian dari sampel ini lebih sulit untuk digeneralisasikan.

7.3 Teknik Sampling Probabilitas Sampel

1. Pengantar Teknik Sampling Probabilitas

   Teknik sampling probabilitas adalah metode pengambilan sampel di mana setiap elemen dalam populasi memiliki peluang yang sama atau yang diketahui untuk dipilih menjadi bagian dari sampel. Teknik ini dianggap lebih akurat dan representatif karena secara statistik memungkinkan hasil penelitian untuk digeneralisasikan ke seluruh populasi. Sampling probabilitas sering digunakan dalam penelitian ilmiah dan survei karena lebih objektif dibandingkan teknik non-probabilitas. Berikut ini adalah beberapa teknik sampling probabilitas yang umum digunakan: sampling

acak sederhana, sampling sistematis, sampling bertingkat, dan sampling kluster.

2. Sampling Acak Sederhana (*Simple Random Sampling*)

   Sampling acak sederhana adalah metode di mana setiap elemen dalam populasi memiliki peluang yang sama untuk dipilih sebagai sampel. Teknik ini sering digunakan dalam penelitian kuantitatif karena kesederhanaannya. Pengambilan sampel dapat dilakukan dengan dua cara: dengan pengembalian (*sampling with replacement*) atau tanpa pengembalian (*sampling without replacement*).

   a. Kelebihan:
      - Teknik ini dianggap paling objektif karena menghilangkan bias dalam pemilihan sampel.
      - Dapat digunakan dengan berbagai jenis populasi dan memiliki generalisasi yang baik.

   b. Kekurangan:
      - Sulit dilakukan ketika populasi sangat besar atau tersebar secara geografis.

- Memerlukan daftar populasi lengkap, yang sering kali sulit diakses.

c. Contoh: Misalkan seorang peneliti ingin meneliti kebiasaan membaca buku mahasiswa di sebuah universitas yang memiliki 10.000 mahasiswa. Peneliti bisa memilih sampel acak dari daftar mahasiswa menggunakan tabel angka acak atau perangkat lunak pengacakan angka.

3. Sampling Sistematis (*Systematic Sampling*)

   Sampling sistematis adalah teknik pengambilan sampel dengan memilih elemen-elemen pada interval tetap dari populasi. Setelah elemen pertama dipilih secara acak, elemen-elemen berikutnya diambil dengan jarak tertentu, yang disebut interval pengambilan sampel (sampling interval). Interval ini dihitung dengan membagi jumlah total populasi dengan ukuran sampel yang diinginkan.

   a. Kelebihan:
      - Lebih mudah dan cepat daripada sampling acak sederhana, terutama ketika bekerja dengan populasi besar.

- Dapat menghasilkan sampel yang seimbang jika tidak ada pola dalam populasi.

b. Kekurangan:
   - Jika terdapat pola dalam populasi, hasilnya mungkin tidak representatif. Misalnya, jika pola pengambilan sampel bertepatan dengan pola dalam populasi, bisa menyebabkan bias.

c. Contoh: Seorang peneliti ingin memilih sampel dari populasi 1000 orang dengan ukuran sampel 100. Peneliti memilih elemen pertama secara acak, kemudian mengambil setiap orang ke-10 dari daftar hingga mencapai 100 orang.

4. Sampling Bertingkat (*Stratified Sampling*)
   Sampling bertingkat digunakan ketika populasi dibagi ke dalam kelompok-kelompok yang homogen, yang disebut strata, berdasarkan karakteristik tertentu (misalnya usia, jenis kelamin, pendapatan). Dari setiap strata, diambil sampel secara acak, dan ukuran sampel dari setiap strata bisa proporsional atau tidak proporsional terhadap ukuran populasi strata

tersebut. Teknik ini bertujuan untuk memastikan bahwa setiap strata dalam populasi diwakili dalam sampel.

a. Kelebihan:
    - Teknik ini menghasilkan sampel yang lebih representatif, terutama ketika populasi memiliki kelompok yang berbeda secara signifikan.
    - Meningkatkan presisi hasil penelitian karena memperhitungkan perbedaan antar strata.
b. Kekurangan:
    - Memerlukan pengetahuan yang mendalam tentang karakteristik populasi sebelum membagi strata.
    - Memakan waktu dan lebih kompleks dalam pelaksanaannya.
c. Contoh: Misalkan seorang peneliti ingin meneliti tingkat kepuasan pelanggan di sebuah kota besar. Pelanggan dibagi ke dalam strata berdasarkan zona geografis (pusat kota, pinggiran, desa), dan dari setiap zona diambil sampel acak yang proporsional terhadap jumlah pelanggan di setiap zona.

5. Sampling Kluster (*Cluster Sampling*)

   Sampling kluster digunakan ketika populasi secara alami terbagi dalam kelompok-kelompok atau kluster, seperti kelas di sekolah atau rumah tangga di sebuah wilayah. Peneliti memilih beberapa kluster secara acak, kemudian mengambil sampel dari seluruh anggota dalam kluster yang terpilih, atau memilih sampel acak dari setiap kluster.

   a. Kelebihan:
      - Lebih efisien dan murah dibandingkan sampling acak sederhana atau bertingkat ketika populasi tersebar secara geografis.
      - Memungkinkan peneliti untuk bekerja dengan unit yang lebih mudah dikelola (misalnya, rumah tangga atau kelompok kerja).

   b. Kekurangan:
      - Risiko bias yang lebih tinggi jika kluster yang dipilih tidak cukup representatif dari seluruh populasi.
      - Teknik ini mungkin memerlukan sampel yang lebih besar untuk

mencapai tingkat presisi yang sama dengan teknik lain.

c. Contoh: Peneliti ingin meneliti perilaku konsumsi listrik di sebuah provinsi. Mereka membagi populasi menjadi kluster-kluster berdasarkan wilayah atau desa, lalu memilih beberapa desa secara acak untuk dijadikan sampel, kemudian meneliti semua rumah tangga di desa tersebut.

6. Kombinasi Teknik Sampling Probabilitas

   Dalam beberapa kasus, peneliti dapat menggabungkan beberapa teknik sampling probabilitas untuk memperoleh hasil yang lebih baik. Misalnya, peneliti dapat menggunakan stratifikasi terlebih dahulu, kemudian melakukan sampling acak sederhana dalam setiap strata. Pendekatan ini sering digunakan dalam survei skala besar atau penelitian nasional yang kompleks.

7. Pertimbangan dalam Memilih Teknik Sampling Probabilitas

   Pemilihan teknik sampling probabilitas harus mempertimbangkan beberapa faktor, termasuk:

a. Tujuan penelitian: Jika tujuan penelitian adalah untuk menghasilkan hasil yang dapat digeneralisasikan, sampling probabilitas adalah pilihan yang tepat.
b. Karakteristik populasi: Populasi yang heterogen lebih baik dianalisis menggunakan sampling bertingkat, sementara populasi yang homogen dapat diwakili dengan sampling acak sederhana atau sistematis.
c. Sumber daya dan waktu: Beberapa teknik sampling seperti sampling acak sederhana memerlukan lebih banyak waktu dan sumber daya dibandingkan teknik seperti sampling sistematis atau kluster.
d. Aksesibilitas data: Jika populasi sulit dijangkau atau tersebar, sampling kluster mungkin lebih efisien.

7.4 Teknik Sampling Non-Probabilitas

1. Pengantar Teknik Sampling Non-Probabilitas

   Sampling non-probabilitas adalah metode pengambilan sampel di mana setiap elemen

dalam populasi tidak memiliki peluang yang sama atau yang diketahui untuk terpilih menjadi bagian dari sampel. Pada metode ini, pemilihan sampel tidak dilakukan secara acak, melainkan berdasarkan penilaian atau kriteria tertentu dari peneliti. Oleh karena itu, hasil yang diperoleh tidak bisa digeneralisasikan secara langsung ke populasi yang lebih luas. Meski begitu, teknik ini tetap berguna, terutama dalam penelitian eksploratif, kualitatif, atau ketika kondisi populasi atau sumber daya penelitian tidak memungkinkan penggunaan teknik sampling probabilitas. Teknik sampling non-probabilitas meliputi beberapa metode, seperti sampling kuota, purposive sampling, snowball sampling, dan convenience sampling.

2. *Convenience Sampling* (Sampling Mudah)

   Convenience sampling atau sampling mudah adalah teknik di mana peneliti memilih sampel yang paling mudah diakses atau tersedia pada saat penelitian. Dalam metode ini, tidak ada kriteria khusus dalam pemilihan sampel, dan keputusan peneliti sering didasarkan pada

ketersediaan subjek atau elemen yang paling mudah dijangkau.

a. Kelebihan:
    - Mudah dan cepat dilakukan, terutama ketika sumber daya dan waktu terbatas.
    - Dapat digunakan dalam tahap eksplorasi awal penelitian untuk memahami fenomena dasar sebelum dilanjutkan dengan teknik sampling yang lebih ketat.
b. Kekurangan:
    - Rentan terhadap bias karena sampel mungkin tidak representatif terhadap populasi.
    - Hasil yang diperoleh tidak bisa digeneralisasikan ke populasi yang lebih luas dengan tingkat akurasi yang tinggi.
c. Contoh: Peneliti melakukan survei terhadap orang-orang yang lewat di pusat perbelanjaan untuk menanyakan kebiasaan belanja mereka. Pemilihan responden didasarkan pada siapa pun yang tersedia dan bersedia untuk diwawancarai.

3. *Purposive* Sampling (Sampling Bertujuan)

   Purposive sampling adalah teknik di mana peneliti memilih sampel berdasarkan penilaian tertentu yang dianggap relevan untuk tujuan penelitian. Peneliti secara sengaja memilih elemen atau individu yang memiliki karakteristik atau pengalaman spesifik yang berkaitan dengan topik penelitian.

   a. Kelebihan:
      - Memungkinkan peneliti untuk fokus pada subjek yang paling relevan dengan penelitian.
      - Berguna dalam penelitian kualitatif atau eksploratif yang memerlukan pemahaman mendalam dari sudut pandang atau pengalaman spesifik.

   b. Kekurangan:
      - Tidak mewakili populasi secara umum, sehingga hasil tidak dapat digeneralisasikan.
      - Ada risiko bias subjektif dalam pemilihan sampel karena tergantung pada penilaian peneliti.

c. Contoh: Dalam penelitian tentang pengalaman kerja wanita di industri teknologi, peneliti secara sengaja memilih wanita yang bekerja sebagai insinyur perangkat lunak untuk diwawancarai, dengan asumsi bahwa mereka memiliki pengalaman yang paling relevan.

4. *Quot*a Sampling (Sampling Kuota)

   Quota sampling adalah teknik di mana peneliti membagi populasi ke dalam subkelompok atau kategori (misalnya, berdasarkan jenis kelamin, usia, atau pendapatan), kemudian mengambil sampel dari setiap subkelompok sampai mencapai kuota yang telah ditentukan sebelumnya. Kuota yang ditetapkan bertujuan untuk memastikan bahwa sampel memiliki proporsi elemen dari setiap subkelompok yang sesuai dengan proporsi populasi.

   a. Kelebihan:
      - Memungkinkan peneliti untuk memastikan bahwa subkelompok penting dari populasi terwakili dalam sampel.

- Lebih efisien dan fleksibel dibandingkan dengan sampling bertingkat dalam teknik probabilitas.

b. Kekurangan:
   - Proses pemilihan sampel dalam setiap kuota sering kali subjektif dan rentan terhadap bias.
   - Hasil tidak dapat digeneralisasikan dengan tingkat akurasi yang tinggi karena pemilihan sampel tidak dilakukan secara acak.

c. Contoh: Dalam survei mengenai kebiasaan belanja konsumen, peneliti menetapkan kuota berdasarkan kategori usia (misalnya, 20% responden harus berusia 18–25 tahun, 30% berusia 26–35 tahun, dan seterusnya), kemudian memilih individu dalam setiap kelompok umur hingga kuota tercapai.

5. *Snowball* Sampling

   Snowball sampling adalah teknik di mana peneliti memulai dengan sejumlah kecil responden yang dipilih secara purposif, kemudian meminta mereka untuk merujuk orang lain yang mungkin memiliki karakteristik

serupa atau relevan untuk penelitian. Teknik ini sering digunakan dalam populasi yang sulit dijangkau atau yang tidak memiliki daftar anggota yang jelas.

a. Kelebihan:
    - Berguna untuk menemukan responden dari populasi yang sulit diakses, seperti kelompok minoritas atau populasi yang tersembunyi.
    - Membantu membangun jaringan responden secara cepat melalui rujukan pribadi.

b. Kekurangan:
    - Rentan terhadap bias karena sampel bergantung pada jaringan sosial responden awal.
    - Tidak dapat digeneralisasikan secara luas karena proses pemilihan sampel tidak acak dan sering kali terbatas pada lingkaran tertentu.

c. Contoh: Peneliti yang ingin meneliti kelompok imigran ilegal dapat memulai dengan beberapa responden awal yang bersedia diwawancarai, lalu meminta mereka untuk merujuk teman atau kenalan

mereka yang juga termasuk dalam kelompok yang sama.

6. Kombinasi Teknik Sampling Non-Probabilitas

   Terkadang peneliti dapat menggunakan kombinasi berbagai teknik sampling non-probabilitas untuk memenuhi kebutuhan penelitian. Misalnya, peneliti dapat memulai dengan purposive sampling untuk memilih responden awal yang relevan, kemudian melanjutkan dengan snowball sampling untuk mengembangkan jaringan responden lebih lanjut.

7. Pertimbangan dalam Memilih Teknik Sampling Non-Probabilitas

   Pemilihan teknik sampling non-probabilitas harus mempertimbangkan beberapa faktor, seperti:

   a. Tujuan penelitian: Teknik non-probabilitas lebih sesuai untuk penelitian eksploratif atau kualitatif di mana generalisasi ke populasi tidak diperlukan.
   b. Keterbatasan sumber daya: Teknik ini lebih hemat waktu dan biaya, terutama ketika

data populasi lengkap sulit diperoleh atau ketika penelitian dilakukan dengan skala kecil.

c. Aksesibilitas populasi: Jika populasi sulit dijangkau, seperti kelompok marjinal atau populasi tersembunyi, teknik seperti snowball sampling dapat menjadi solusi yang efektif.

d. Representasi dan bias: Meskipun teknik non-probabilitas dapat memberikan wawasan yang berharga, peneliti harus menyadari keterbatasan dalam hal representasi populasi dan potensi bias yang dapat mempengaruhi hasil penelitian.

7.5 Menentukan Ukuran Sampel

Menentukan ukuran sampel yang tepat adalah langkah penting dalam proses penelitian karena akan memengaruhi validitas hasil penelitian. Ukuran sampel yang terlalu kecil mungkin tidak mampu mewakili populasi secara akurat, sementara ukuran yang terlalu besar dapat menghabiskan sumber daya tanpa memberikan manfaat tambahan yang signifikan. Oleh karena itu, penting untuk memahami cara menentukan

ukuran sampel secara tepat berdasarkan tujuan penelitian, metode analisis, dan karakteristik populasi.

Faktor yang Mempengaruhi Ukuran Sampel Ada beberapa faktor utama yang harus dipertimbangkan ketika menentukan ukuran sampel:

a. Tingkat Keyakinan (*Confidence Level*): Biasanya dinyatakan dalam persentase (misalnya, 95% atau 99%), tingkat keyakinan menunjukkan seberapa yakin kita terhadap hasil penelitian kita. Tingkat keyakinan yang lebih tinggi membutuhkan ukuran sampel yang lebih besar.
b. Margin Kesalahan (*Margin of Error):* Margin kesalahan merupakan rentang kesalahan yang dapat diterima dalam estimasi parameter populasi. Semakin kecil margin kesalahan yang diinginkan, semakin besar ukuran sampel yang diperlukan.
c. Variabilitas dalam Populasi: Semakin besar variabilitas atau heterogenitas dalam populasi, semakin besar sampel yang dibutuhkan untuk mendapatkan hasil yang representatif.
d. Ukuran Populasi: Jika populasi cukup kecil, rumus penyesuaian ukuran sampel seperti "finite population correction" mungkin diperlukan. Namun, untuk populasi besar,

ukuran populasi tidak mempengaruhi secara signifikan.

e. Tingkat Respon: Dalam penelitian survei, perkiraan tingkat respon (*response rate*) penting untuk diperhitungkan. Jika tingkat respon diperkirakan rendah, ukuran sampel awal harus lebih besar untuk mengatasi non-respon.

Rumus Penentuan Ukuran Sampel Terdapat beberapa rumus yang digunakan untuk menentukan ukuran sampel, tergantung pada jenis penelitian dan sifat populasi yang diteliti. Berikut rumus yang umum digunakan:

a. Rumus Slovin sering digunakan untuk menentukan ukuran sampel ketika populasi diketahui. Rumus ini sederhana dan sering digunakan dalam penelitian sosial:

$$n = \frac{N}{1 + N(e^2)}$$

di mana:

n = ukuran sampel,

N = ukuran populasi,

e = tingkat kesalahan yang diinginkan.

b. Rumus Cochran digunakan untuk menghitung ukuran sampel dalam survei dengan populasi besar (biasanya >10.000):

$$n_0 = \frac{Z^2 p(1-p)}{e^2}$$

di mana:

n_0 = ukuran sampel awal,

Z = nilai Z berdasarkan tingkat keyakinan (misalnya, 1,96 untuk tingkat keyakinan 95%),

p = proporsi populasi yang memiliki karakteristik yang diteliti (jika tidak diketahui, sering diasumsikan 0,5),

e = margin kesalahan yang diinginkan.

Contoh Aplikasi Penentuan Ukuran Sampel. Misalkan seorang peneliti ingin mengetahui pendapat mahasiswa tentang kebijakan baru di universitas dengan populasi 10.000 mahasiswa. Peneliti ingin margin kesalahan 5% dan tingkat keyakinan 95%. Dengan rumus Slovin, ukuran sampel dihitung sebagai berikut:

$$n = \frac{10.000}{1 + 10.000(0,05^2)} = 385$$

Peneliti membutuhkan sampel sekitar 385 mahasiswa untuk hasil yang representatif.

BAB VIII
HIPOTESIS PENELITIAN

8.1 Pengertian Hipotesis Penelitian

Hipotesis penelitian adalah pernyataan atau asumsi sementara yang dibuat oleh peneliti mengenai hubungan antara dua atau lebih variabel dalam penelitian. Hipotesis ini diusulkan untuk diuji melalui data empiris dan berfungsi sebagai panduan atau dasar bagi proses penelitian ilmiah. Hipotesis merupakan komponen penting dalam metode ilmiah karena menjadi dasar untuk pengujian teori, eksperimen, dan validasi data.

Hipotesis biasanya berbentuk dugaan yang dapat diuji secara objektif dan dinyatakan dalam bentuk proposisi yang jelas dan dapat diukur. Dalam penelitian kuantitatif, hipotesis membantu peneliti untuk menentukan apakah ada hubungan yang signifikan antara variabel yang diteliti atau apakah ada perbedaan antara kelompok yang diuji. Pada penelitian kualitatif, hipotesis bisa berbentuk lebih fleksibel, karena seringkali peneliti kualitatif mencari pola atau tema daripada hubungan yang eksak antara variabel.

1. Fungsi Hipotesis Penelitian

   Beberapa fungsi utama dari hipotesis penelitian adalah sebagai berikut:

   a. Mengarahkan Penelitian: Hipotesis memberikan arahan bagi peneliti untuk fokus pada tujuan penelitian yang spesifik dan menentukan data apa yang perlu dikumpulkan.
   b. Membuat Prediksi: Hipotesis adalah prediksi atau dugaan peneliti mengenai hubungan antara variabel, yang kemudian diuji untuk membuktikan kebenarannya.
   c. Menguji Teori: Dalam banyak penelitian, hipotesis digunakan untuk menguji teori yang sudah ada. Jika hipotesis diterima, maka teori tersebut diperkuat; jika ditolak, maka diperlukan revisi atau pengembangan lebih lanjut terhadap teori tersebut.
   d. Mempermudah Analisis Data: Hipotesis yang baik dapat membantu peneliti dalam memilih metode analisis data yang tepat dan dalam menyusun hasil penelitian dengan lebih sistematis.

2. Karakteristik Hipotesis yang Baik

   Hipotesis yang baik harus memenuhi beberapa kriteria berikut:

   a. Dapat Diuji (*Testable*): Hipotesis harus dapat diuji melalui pengumpulan dan analisis data empiris.
   b. Spesifik: Hipotesis harus jelas dan spesifik tentang hubungan antara variabel. Pernyataan yang terlalu umum atau kabur tidak akan memberikan hasil yang bermanfaat.
   c. Didasarkan pada Teori atau Bukti Sebelumnya: Hipotesis yang kuat biasanya didasarkan pada teori ilmiah atau hasil penelitian sebelumnya yang telah terbukti.
   d. Dapat Diukur: Variabel dalam hipotesis harus dapat diukur dengan metode penelitian yang valid dan reliabel.
   e. Berorientasi pada Hubungan Sebab-Akibat atau Asosiasi: Dalam banyak penelitian, hipotesis menyatakan hubungan sebab-akibat atau asosiasi antara variabel.

8.2 Langkah-Langkah Pengujian Hipotesis

Pengujian hipotesis merupakan proses untuk memeriksa apakah dugaan yang dibuat tentang suatu populasi dapat didukung atau ditolak berdasarkan data sampel. Berikut adalah langkah-langkah umum dalam pengujian hipotesis:

1. Merumuskan Hipotesis Nol (H_0) dan Hipotesis Alternatif (H_1)

   Langkah pertama dalam pengujian hipotesis adalah merumuskan dua hipotesis:

   a. Hipotesis Nol (H_0): Hipotesis ini adalah pernyataan yang menyatakan bahwa tidak ada efek atau perbedaan yang signifikan antara kelompok atau variabel yang diuji. Ini adalah asumsi dasar yang akan diuji.

   b. Hipotesis Alternatif (H_1 atau Ha): Hipotesis ini adalah kebalikan dari hipotesis nol dan menyatakan adanya efek atau perbedaan signifikan. Hipotesis alternatif adalah hipotesis yang ingin diuji oleh peneliti.

   Contoh:

   - H_0: "Tidak ada perbedaan rata-rata nilai ujian antara siswa yang belajar dengan metode A dan metode B."

- H_1: "Ada perbedaan rata-rata nilai ujian antara siswa yang belajar dengan metode A dan metode B."

2. Menentukan Tingkat Signifikansi (α)

   Tingkat signifikansi atau α adalah probabilitas membuat kesalahan tipe I, yaitu menolak hipotesis nol padahal sebenarnya benar. Nilai α yang sering digunakan dalam penelitian adalah 0.05 (5%) atau 0.01 (1%).

   Artinya, jika α = 0.05, ada kemungkinan 5% bahwa peneliti akan membuat kesalahan dalam menolak hipotesis nol.

3. Menentukan Statistik Uji

   Statistik uji adalah metode matematika yang digunakan untuk menguji hipotesis berdasarkan data sampel. Pemilihan statistik uji tergantung pada jenis data dan distribusi variabel yang diuji.

   Beberapa uji statistik yang umum digunakan:

   a. Uji Z: Untuk sampel besar (n > 30) dan distribusi data normal.
   b. Uji t: Untuk sampel kecil (n < 30) atau data yang tidak diketahui distribusinya.
   c. Uji Chi-Square: Untuk menguji hubungan antar variabel kategori.

d. Uji ANOVA: Untuk membandingkan rata-rata lebih dari dua kelompok.

4. Menentukan Daerah Penolakan (Critical Region)

   Daerah penolakan adalah nilai-nilai dari statistik uji yang mengarahkan pada penolakan hipotesis nol. Daerah ini ditentukan oleh nilai kritis yang bergantung pada tingkat signifikansi (α) yang dipilih.

   a. Jika hasil statistik uji jatuh dalam daerah penolakan, maka hipotesis nol akan ditolak.
   b. Jika hasil statistik uji tidak jatuh dalam daerah penolakan, maka hipotesis nol tidak ditolak.

5. Mengumpulkan dan Menganalisis Data

   Pada tahap ini, data dari sampel dikumpulkan melalui survei, eksperimen, atau metode pengumpulan data lainnya. Data tersebut kemudian dianalisis menggunakan teknik statistik yang sesuai (misalnya uji t, uji Z, atau uji Chi-Square) untuk menghitung statistik uji yang akan dibandingkan dengan nilai kritis.

6. Menghitung Nilai Statistik Uji

   Setelah data dianalisis, peneliti akan menghitung nilai statistik uji yang digunakan

untuk memutuskan apakah hipotesis nol dapat diterima atau ditolak.

Misalnya, dalam uji t, nilai t-statistik dihitung dengan rumus:

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

Di mana:

X = rata-rata sampel,

μ = rata-rata populasi (yang diasumsikan oleh H_0),

S = standar deviasi sampel,

n = ukuran sampel.

7. Menentukan Nilai P (P-Value)

   Nilai p adalah probabilitas mendapatkan hasil statistik uji yang sama atau lebih ekstrem dari hasil yang diperoleh, jika hipotesis nol benar. Jika nilai p lebih kecil dari tingkat signifikansi (α), maka hipotesis nol ditolak.

   a. Jika $p \leq \alpha$, tolak H_0 (ada cukup bukti untuk mendukung H_1).
   b. Jika $p > \alpha$, gagal menolak H_0 (tidak ada cukup bukti untuk mendukung H_1).

8. Keputusan: Menerima atau Menolak Hipotesis Nol

   Keputusan dibuat berdasarkan perbandingan antara statistik uji dan nilai kritis atau nilai p. Hasilnya adalah salah satu dari dua kemungkinan:

   a. Menolak H_0: Ini berarti ada cukup bukti statistik untuk mendukung hipotesis alternatif (H_1), bahwa ada perbedaan atau hubungan yang signifikan.
   b. Gagal Menolak H_0: Ini berarti tidak ada cukup bukti statistik untuk menolak hipotesis nol, sehingga H_0 diterima atau tetap tidak ditolak.

9. Menyimpulkan dan Menyajikan Hasil

   Langkah terakhir adalah menyusun kesimpulan berdasarkan hasil pengujian hipotesis. Kesimpulan ini harus dijelaskan dalam konteks penelitian dan dikaitkan dengan hipotesis yang diuji.

   Jika H_0 ditolak, peneliti menyatakan bahwa ada bukti yang mendukung adanya hubungan atau perbedaan. Sebaliknya, jika H_0 tidak ditolak, peneliti menyatakan bahwa tidak ada cukup bukti untuk mendukung klaim tersebut.

8.3 Jenis-Jenis Pengujian Hipotesis

Pengujian hipotesis adalah metode statistik yang digunakan untuk menguji pernyataan atau dugaan mengenai parameter populasi berdasarkan data sampel. Berdasarkan jenis hipotesis dan metode uji yang digunakan, terdapat beberapa jenis pengujian hipotesis yang penting untuk dipahami. Berikut adalah penjelasan lengkap tentang jenis-jenis pengujian hipotesis:

1. Pengujian Hipotesis Berdasarkan Arah
    a. Uji Satu Arah (*One-Tailed Test*)

        Pengujian hipotesis satu arah dilakukan ketika hipotesis alternatif (H_1) menyatakan bahwa efek atau hubungan antara variabel hanya terjadi pada satu arah (positif atau negatif). Dengan kata lain, uji satu arah digunakan ketika peneliti hanya tertarik untuk mengetahui apakah suatu parameter populasi lebih besar atau lebih kecil daripada nilai tertentu, tetapi tidak keduanya.

        - H_0: Tidak ada perbedaan atau pengaruh.
        - H_1: Ada perbedaan atau pengaruh pada satu arah (lebih besar atau lebih kecil).

Contoh: Seorang peneliti ingin menguji apakah rata-rata nilai siswa yang menggunakan metode A lebih tinggi dari 70.

- H_0: Rata-rata nilai = 70.
- H_1: Rata-rata nilai > 70 (uji satu arah ke kanan).

b. Uji Dua Arah (*Two-Tailed Test*)

Pengujian hipotesis dua arah dilakukan ketika hipotesis alternatif (H_1) menyatakan bahwa efek atau hubungan dapat terjadi di kedua arah (baik lebih besar maupun lebih kecil). Uji ini digunakan ketika peneliti tertarik untuk mengetahui apakah suatu parameter populasi berbeda dari nilai tertentu tanpa peduli arahnya.

- H_0: Tidak ada perbedaan atau pengaruh.
- H_1: Ada perbedaan atau pengaruh, tanpa menyebutkan arahnya (lebih besar atau lebih kecil).

Contoh: Seorang peneliti ingin menguji apakah rata-rata nilai siswa yang menggunakan metode A berbeda dari 70.

- H_0: Rata-rata nilai = 70.
- H_1: Rata-rata nilai ≠ 70 (uji dua arah).

2. Pengujian Hipotesis Berdasarkan Jenis Data dan Tujuan

   a. Uji Z

      Uji Z digunakan ketika data memiliki distribusi normal, ukuran sampel besar ($n > 30$), dan varians populasi diketahui. Uji ini umumnya digunakan untuk menguji hipotesis mengenai rata-rata populasi atau proporsi.

      Contoh Penggunaan: Menguji apakah rata-rata tinggi badan pria di suatu kota lebih besar dari 170 cm dengan data sampel besar (misalnya, 100 pria).

   b. Uji t

      - Uji t digunakan ketika ukuran sampel kecil ($n < 30$) atau varians populasi tidak diketahui. Ada beberapa jenis uji t, tergantung pada situasi penelitian:
      - Uji t untuk Satu Sampel: Digunakan untuk menguji apakah rata-rata satu sampel berbeda dari rata-rata populasi tertentu.

      Contoh: Apakah rata-rata nilai ujian siswa berbeda dari 75?

- Uji t Dua Sampel Berpasangan (*Paired Sample t-Test*): Digunakan ketika dua pengamatan dibuat dari sampel yang sama (misalnya, sebelum dan sesudah perlakuan pada satu kelompok).

  Contoh: Apakah ada perbedaan berat badan sebelum dan sesudah program diet?

- Uji t Dua Sampel Tidak Berpasangan (*Independent t-Test*): Digunakan ketika dua kelompok independen dibandingkan, misalnya untuk mengetahui apakah ada perbedaan rata-rata antara dua kelompok yang berbeda.

  Contoh: Apakah ada perbedaan nilai rata-rata antara siswa yang belajar menggunakan metode A dan metode B?

c. Uji ANOVA (*Analysis of Variance*)

   Uji ANOVA digunakan untuk membandingkan rata-rata lebih dari dua kelompok. Jika uji t hanya bisa membandingkan dua kelompok, ANOVA memungkinkan peneliti untuk menguji

apakah ada perbedaan signifikan antara beberapa kelompok sekaligus.

Contoh: Apakah ada perbedaan rata-rata nilai siswa di antara tiga metode pengajaran yang berbeda?

d. Uji Chi-*Square* (χ^2)

Uji Chi-*Square* digunakan untuk menguji hubungan antara dua variabel kategori. Uji ini sering digunakan dalam analisis data kategorikal untuk melihat apakah ada hubungan signifikan antara variabel-variabel tersebut.

Contoh: Apakah ada hubungan antara jenis kelamin (laki-laki dan perempuan) dan pilihan jurusan (IPA atau IPS) di sekolah?

e. Uji F

Uji F digunakan dalam analisis varians untuk menguji apakah dua populasi memiliki varians yang sama. Uji F sering digunakan dalam ANOVA dan dalam perbandingan model regresi.

Contoh: Apakah varians nilai matematika sama di dua sekolah yang berbeda?

f. Uji Proporsi

Uji proporsi digunakan untuk menguji hipotesis mengenai proporsi populasi, misalnya untuk menentukan apakah proporsi orang yang memilih kandidat tertentu berbeda dari nilai yang diharapkan.

Contoh: Apakah proporsi pemilih yang mendukung kandidat A lebih besar dari 50%?

3. Pengujian Hipotesis Nonparametrik

Pengujian hipotesis nonparametrik digunakan ketika data tidak memenuhi asumsi distribusi normal atau ketika data berskala ordinal. Beberapa uji nonparametrik yang umum digunakan antara lain:

a. Uji Mann-Whitney U

Uji ini adalah alternatif dari uji t tidak berpasangan dan digunakan untuk membandingkan dua kelompok yang independen ketika data tidak terdistribusi normal.

Contoh: Apakah ada perbedaan skor kepuasan pelanggan antara dua toko?

b. Uji Wilcoxon

Uji Wilcoxon adalah alternatif dari uji t berpasangan dan digunakan untuk membandingkan dua sampel yang berpasangan ketika data tidak terdistribusi normal.

Contoh: Apakah ada perbedaan berat badan sebelum dan sesudah program diet pada sekelompok peserta?

c. Uji Kruskal-Wallis

Uji ini adalah alternatif dari ANOVA dan digunakan untuk membandingkan lebih dari dua kelompok ketika data tidak terdistribusi normal.

Contoh: Apakah ada perbedaan tingkat kebahagiaan di antara tiga kelompok usia yang berbeda?

d. Uji Chi-*Square* untuk Kebaikan Kecocokan (*Goodness of Fit Test*)

Uji ini digunakan untuk menguji apakah distribusi sampel sesuai dengan distribusi teoretis tertentu.

Contoh: Apakah distribusi warna bola di dalam kantong sesuai dengan distribusi

yang diharapkan (misalnya, proporsi warna merah, hijau, dan biru)?

4. Pengujian Hipotesis Berdasarkan Tipe Kesalahan
    a. Kesalahan Tipe I (α)

       Kesalahan Tipe I terjadi ketika peneliti menolak hipotesis nol padahal hipotesis tersebut benar. Tingkat signifikansi (α) digunakan untuk mengukur probabilitas kesalahan ini.
    b. Kesalahan Tipe II (β)

       Kesalahan Tipe II terjadi ketika peneliti gagal menolak hipotesis nol padahal hipotesis tersebut salah. Probabilitas kesalahan ini disebut sebagai β, dan kekuatan uji (power) adalah 1 – β.

Berbagai jenis pengujian hipotesis membantu peneliti untuk memilih metode yang sesuai dengan data dan tujuan penelitian. Memilih uji statistik yang tepat sangat penting untuk memastikan hasil penelitian yang akurat dan valid.

BAB IX
REGRESI LINEAR SEDERHANA

9.1 Teori dan Konsep Dasar Regresi Linear Sederhana

Regresi linear sederhana adalah metode analisis statistik yang digunakan untuk memahami dan memodelkan hubungan antara dua variabel, yaitu satu variabel independen (x) dan satu variabel dependen (y) (Marill, 2004) (Prion et al., 2020). Metode ini sangat berguna untuk memperkirakan atau memprediksi nilai variabel dependen berdasarkan informasi dari variabel independen. Salah satu alasan utama penggunaan regresi linear sederhana adalah kemampuannya untuk memberikan model yang mudah dipahami dan diinterpretasikan.

Garis regresi, yang dihasilkan dari model ini, memberikan representasi terbaik dari hubungan linear antara dua variabel tersebut. Metode Least Squares atau metode kuadrat terkecil adalah teknik paling umum yang digunakan untuk menentukan garis regresi ini, dengan tujuan untuk meminimalkan jumlah kuadrat selisih antara nilai yang diobservasi dan nilai yang diprediksi (Islas Toski et al., 2020; Marill, 2004).

Regresi linear sederhana juga bergantung pada beberapa asumsi penting, yaitu linearitas, independensi, homoskedastisitas, dan normalitas. Linearity mengasumsikan bahwa hubungan antara variabel independen dan dependen bersifat linear, sementara independensi mengharuskan setiap observasi tidak saling bergantung satu sama lain. Homoskedastisitas mengacu pada konsistensi varian error dalam model, dan normalitas berarti error dari model didistribusikan secara normal (Marill, 2004). Jika salah satu asumsi ini dilanggar, maka validitas hasil regresi bisa dipertanyakan. Untuk mengatasi hubungan yang tidak linear, sering kali dilakukan transformasi variabel atau penggunaan variabel dummy untuk menangani data kategorikal (Marill, 2004).

Di luar teori, regresi linear sederhana memiliki banyak aplikasi dalam berbagai bidang. Dalam konteks pendidikan, alat simulasi dan kegiatan interaktif membantu siswa memahami hubungan antara parameter dan statistik sampel, serta bagaimana model dibentuk dan diuji (Jones et al., 2004; Richardson et al., 2004). Secara praktis, regresi ini sering digunakan untuk memecahkan masalah-masalah nyata, seperti memprediksi jumlah kecelakaan lalu lintas berdasarkan berbagai faktor lingkungan, atau menganalisis

hubungan antara polusi udara dan variabel-variabel penyebab lainnya (Clausel et al., 2014; Islas Toski et al., 2020).

Selain manfaatnya, regresi linear sederhana juga memiliki tantangan. Meskipun metodologinya terbilang sederhana, beberapa konsep, seperti pengujian inferensi dan identifikasi leverage (data yang berpengaruh tinggi), sering kali memerlukan penjelasan lebih lanjut dan contoh nyata agar lebih mudah dipahami oleh pemula(Marill, 2004). Oleh karena itu, pengajaran regresi linear sederhana biasanya melibatkan metode-metode pembelajaran yang interaktif dan eksperiensial untuk memastikan pemahaman yang lebih mendalam(Prion et al., 2020; Richardson et al., 2004).

Dengan memahami regresi linear sederhana dan berbagai asumsi serta aplikasinya, peneliti dapat membangun dasar yang kuat sebelum melanjutkan ke analisis yang lebih kompleks, seperti regresi linear berganda yang melibatkan lebih dari satu variabel independen(Marill, 2004).

9.2 Variabel Dependen dan Independen

Dalam konteks statistika, variabel dependen dan variabel independen memainkan peran kunci dalam

proses analisis data. Variabel dependen, yang sering juga disebut sebagai variabel respon, adalah variabel yang menjadi fokus utama penelitian. Variabel ini adalah nilai yang ingin diprediksi atau dijelaskan berdasarkan pengaruh dari variabel lain. Sementara itu, variabel independen, atau disebut juga variabel prediktor atau eksplanatori, adalah variabel yang dianggap memengaruhi atau memiliki hubungan dengan variabel dependen. Dengan kata lain, variabel independen merupakan faktor-faktor yang diduga mempengaruhi hasil atau respons yang diukur melalui variabel dependen.

Sebagai contoh, dalam penelitian yang bertujuan untuk menganalisis hubungan antara jumlah jam belajar siswa dan hasil ujian, hasil ujian merupakan variabel dependen, karena hasil ujian tersebut merupakan outcome yang ingin diprediksi atau dijelaskan. Sementara itu, jumlah jam belajar bertindak sebagai variabel independen, karena dihipotesiskan mempengaruhi atau menjelaskan variasi pada hasil ujian tersebut. Dalam model regresi, variabel independen digunakan untuk memprediksi nilai variabel dependen, di mana hubungan keduanya dapat direpresentasikan dalam persamaan regresi.

Dalam penelitian eksperimen, variabel independen biasanya merupakan variabel yang dimanipulasi oleh peneliti untuk melihat bagaimana perubahan tersebut mempengaruhi variabel dependen. Misalnya, dalam penelitian laboratorium yang mengevaluasi efek dari berbagai dosis obat terhadap tekanan darah, dosis obat adalah variabel independen yang dimanipulasi, sedangkan tekanan darah adalah variabel dependen yang diukur untuk melihat dampak dari manipulasi dosis obat.

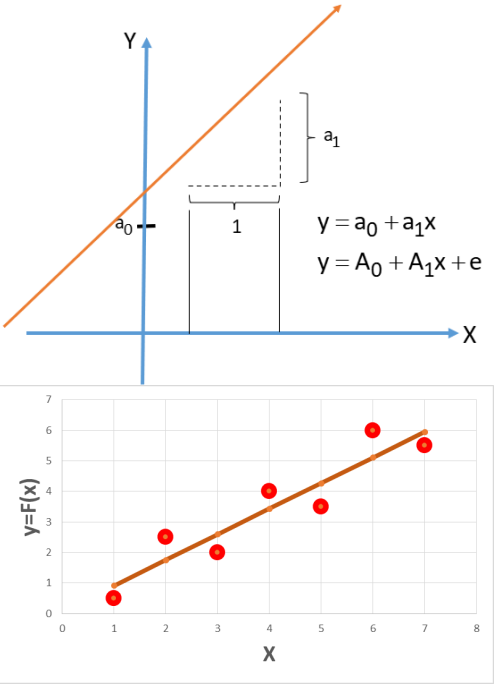
Namun, penting untuk diingat bahwa hubungan antara variabel independen dan dependen tidak selalu bersifat kausal. Dalam beberapa analisis observasional, hubungan yang ditemukan mungkin hanya bersifat asosiatif. Misalnya, jika dalam sebuah studi ditemukan korelasi antara pendapatan seseorang dengan tingkat pendidikan, itu tidak selalu berarti bahwa pendidikan secara langsung menyebabkan peningkatan pendapatan, karena banyak faktor lain yang bisa mempengaruhi hubungan tersebut, seperti lingkungan sosial, akses ke peluang ekonomi, atau jaringan profesional.

9.3 Hubungan antara Linear dan Garis Lurus

Linear sering dihubungkan dengan persamaan garis lurus. Persamaan linear adalah jenis hubungan antara dua variabel di mana perubahan pada satu variabel menghasilkan perubahan yang proporsional pada variabel lainnya. Dalam konteks matematis, hubungan ini dapat direpresentasikan dengan persamaan garis lurus, yang dinyatakan dalam bentuk ($y = a_1x + a_0$), di mana (y) adalah variabel dependen, (x) adalah variabel independen, (a_1) adalah kemiringan (*slope*) garis, dan (a0) adalah titik potong dengan sumbu y (intercept). Kemiringan garis (m) menunjukkan tingkat perubahan variabel dependen (y) untuk setiap unit perubahan dalam variabel independen (x). Hubungan linear dicirikan oleh pola yang konsisten di mana kedua variabel naik atau turun secara bersamaan dalam proporsi yang tetap, menciptakan grafik garis lurus.

Setelah mendapatkan persamaan garis linear, garis tersebut mewakili nilai yang ada pada data yang akan dibuatkan regresi linear, bahkan garis linear yang didapatkan tidak pernah melalui nilai data (Y). kondisi ini akan menghasilkan residual (e) mengindikasikan adanya masalah seperti heteroskedastisitas, di mana variabilitas residual berubah-ubah untuk nilai prediksi

yang berbeda, atau adanya autokorelasi dalam data, yang berarti residual tidak independen satu sama lain. Hal ini melanggar asumsi dasar regresi linear yang menyatakan bahwa residual harus memiliki distribusi yang identik dan independent.



Fungsi garis lurus merupakan representasi grafis dari hubungan linear antara dua variabel. Fungsi ini memiliki bentuk sederhana yang menunjukkan bahwa perubahan dalam variabel independen akan selalu menghasilkan perubahan yang tetap dan konsisten

dalam variabel dependen. Misalnya, dalam kasus analisis keteknikan seperti hubungan antara kecepatan kendaraan dan waktu yang diperlukan untuk menempuh jarak tertentu (dengan asumsi kecepatan konstan), hubungan ini bisa digambarkan dengan garis lurus. Fungsi garis lurus mempermudah prediksi dan analisis karena pola perubahan yang jelas dan mudah dihitung, menjadikannya alat yang sangat berguna dalam berbagai bidang teknik dan ilmu pengetahuan.

9.4 Asumsi dalam Regresi Linear

Pada regresi linear, asumsi-asumsi dasar memiliki peran penting dalam memastikan bahwa model yang dibangun memberikan estimasi yang valid dan dapat diandalkan. Terdapat empat asumsi utama yang harus dipenuhi, yaitu linearitas, independensi, homoskedastisitas, dan normalitas residual. Setiap asumsi ini mendukung integritas model regresi, dan pelanggaran terhadap salah satunya dapat menghasilkan model yang bias atau tidak akurat.

1. Asumsi Linearitas: Asumsi ini menyatakan bahwa terdapat hubungan linear antara variabel independen (x) dan variabel dependen (y). Ini berarti bahwa perubahan pada variabel independen menghasilkan perubahan yang

proporsional dalam variabel dependen. Grafik dari hubungan ini akan menunjukkan pola garis lurus pada scatter plot antara x dan y. Jika hubungan tidak linear, misalnya berbentuk kurva, maka model regresi linear tidak akan menangkap pola tersebut dengan baik. Untuk memeriksa linearitas, biasanya dilakukan analisis plot residual terhadap variabel prediktor. Jika terdapat pola non-linear dalam residual (misalnya pola melengkung), ini menunjukkan pelanggaran asumsi linearitas. Jika pelanggaran ini terjadi, metode alternatif seperti transformasi variabel (log atau kuadrat) atau penggunaan model regresi non-linear mungkin diperlukan.

2. Asumsi Independensi: Asumsi ini mengharuskan bahwa observasi satu dengan lainnya harus independen, atau tidak saling berkaitan. Ini berarti bahwa error yang dihasilkan oleh satu observasi tidak boleh bergantung pada error dari observasi lain. Misalnya, dalam data runtun waktu (time series), data biasanya saling berkaitan antara satu waktu dengan waktu berikutnya, sehingga asumsi independensi sering kali dilanggar.

Untuk mengatasi masalah ini, metode seperti regresi dengan variabel dummy waktu atau model autoregresi bisa digunakan.

3. Asumsi Homoskedastisitas: Homoskedastisitas berarti bahwa varian dari error (residual) dalam model regresi harus konstan di seluruh nilai variabel independen. Dengan kata lain, sebaran error harus sama pada semua tingkat prediktor. Jika varian error berubah-ubah atau tidak konstan (heteroskedastisitas), maka model regresi linear sederhana tidak lagi memberikan estimasi yang efisien, dan variabel independen mungkin memiliki pengaruh yang berbeda tergantung pada nilai tertentu dari x. Ini dapat menyebabkan estimasi koefisien yang bias serta hasil uji statistik yang salah.
4. Asumsi Normalitas Residual: Asumsi ini mengharuskan bahwa error (residual) yang dihasilkan oleh model regresi mengikuti distribusi normal. Hal ini penting untuk memastikan bahwa estimasi parameter model regresi (seperti koefisien regresi) dan hasil uji hipotesis yang dilakukan valid dan dapat diandalkan. Normalitas residual mempengaruhi

ketepatan uji statistik, terutama dalam pengujian signifikansi parameter model.

9.5 Persamaan Regresi Linear

Bentuk Umum Persamaan Regresi dalam regresi linear sederhana dituliskan sebagai $y = a + bX$ di mana Y adalah variabel dependen (variabel yang diprediksi), X adalah variabel independen (variabel yang digunakan untuk memprediksi), a adalah intersep atau titik potong garis dengan sumbu y, dan b adalah koefisien regresi atau kemiringan garis yang menunjukkan seberapa besar perubahan Y ketika X berubah satu unit. Persamaan ini digunakan untuk memahami dan memprediksi hubungan antara dua variabel, yaitu bagaimana perubahan pada X (variabel independen) mempengaruhi nilai Y (variabel dependen).

Dari Persamaan tersebut, a adalah intersep atau titik potong antara gari linear dengan sumbu Y, hal ini adalah nilai Y ketika variabel X bernilai nol, nilai intersep juga yang menunjukkan titik awal dari hubungan antara variabel-variabel tersebut. Misalnya, jika X adalah jumlah unit produksi dan Y adalah total biaya produksi, maka a mewakili biaya tetap (fixed cost) yang dikeluarkan meskipun tidak ada unit yang diproduksi. Di sisi lain, koefisien b, atau kemiringan

garis, menggambarkan seberapa besar perubahan Y untuk setiap unit perubahan pada X. Misalnya, jika b bernilai positif, maka ada hubungan positif antara X dan Y, artinya setiap kali X meningkat, Y juga akan meningkat secara proporsional. Sebaliknya, jika b negatif, peningkatan pada X akan mengurangi nilai Y.

Model ini sering digunakan dalam berbagai bidang, seperti ekonomi, teknik, atau manajemen, untuk memprediksi hasil atau mengevaluasi dampak dari perubahan satu variabel terhadap variabel lain. Misalnya, dalam analisis bisnis, model ini bisa digunakan untuk memprediksi pendapatan (Y) berdasarkan anggaran iklan (X), atau dalam teknik, model ini digunakan untuk memprediksi performa sistem berdasarkan parameter teknis tertentu.

9.6 Pengertian Koefisien Intersep (a) dan Kemiringan (b)

Dalam kehidupan sehari-hari, koefisien intersep (a) dan kemiringan (b) dalam persamaan regresi linear sering digunakan untuk memahami hubungan antara dua variabel dan bagaimana perubahan satu variabel mempengaruhi yang lain. Misalnya, dalam konteks biaya perjalanan menggunakan taksi, kita dapat memodelkan hubungan antara jarak tempuh (X) dan

total biaya perjalanan (Y). Dalam persamaan $y = a + bX$, Y adalah total biaya perjalanan, X adalah jarak yang ditempuh, a adalah biaya tetap (intersep), dan b adalah tarif per kilometer (kemiringan).

Intersep (a), dalam hal ini, mewakili biaya awal yang harus dibayar penumpang sebelum perjalanan dimulai. Biaya ini tidak tergantung pada seberapa jauh jarak yang ditempuh, dan biasanya disebut sebagai biaya buka pintu atau biaya minimum. Misalnya, jika biaya awal naik taksi adalah Rp10.000, maka itulah nilai intersep (a). Bahkan jika penumpang tidak menempuh jarak yang jauh, biaya ini tetap harus dibayar. Jadi, (a) adalah biaya tetap yang selalu ada terlepas dari variabel lain, dalam hal ini jarak.

Kemiringan (b), di sisi lain, menunjukkan bagaimana total biaya berubah sesuai dengan perubahan jarak yang ditempuh. Misalnya, jika tarif taksi adalah Rp5.000 per kilometer, maka (b) adalah Rp5.000. Ini berarti setiap kilometer tambahan yang ditempuh akan menambah Rp5.000 ke total biaya perjalanan. Jadi, nilai (b) menunjukkan tingkat perubahan dari variabel dependen (total biaya) setiap kali variabel independen (jarak) berubah satu satuan. Dalam contoh ini, semakin jauh penumpang bepergian,

semakin besar biaya perjalanan, dengan peningkatan yang proporsional dengan tarif per kilometer.

9.7 Regresi dengan Metode Kuadrat Terkecil

Regresi dengan Metode Kuadrat Terkecil (*Least Squares Regression*) adalah teknik yang digunakan untuk menentukan hubungan linear antara satu atau lebih variabel independen (prediktor) dan variabel dependen (respon). Tujuan utama metode ini adalah menemukan garis regresi yang meminimalkan jumlah kuadrat dari selisih antara nilai observasi aktual dan nilai yang diprediksi oleh model. Dalam kasus regresi linear sederhana, yang melibatkan satu variabel independen, model dapat dinyatakan sebagai $Y = a_0 + a_1 X + e$

di mana (Y) adalah variabel dependen, (X) adalah variabel independen, (a_0) adalah intersep (nilai (Y) saat (X=0)), (a^1) adalah kemiringan atau koefisien regresi yang menunjukkan perubahan (Y) untuk setiap satu unit perubahan di (X), dan (e) adalah error atau residu. Proses menghitung parameter (a_0) dan (a_1) dimulai dengan mendefinisikan fungsi kesalahan (S), yang merupakan jumlah kuadrat dari selisih antara nilai (Y_i) aktual dan nilai prediksi ($\overline{Y_i}$), atau secara matematis dinyatakan sebagai

$$S = \sum_{i=1}^{n} \left(Y_i - \overline{Y}_i \right)^2$$

dengan (n) sebagai jumlah data. Untuk meminimalkan fungsi kesalahan ini, kita menggunakan turunan parsial dari (S) terhadap (a_0) dan (a_1), kemudian menyetarakannya dengan nol untuk memperoleh solusi optimal. Hasil dari proses ini memberikan rumus parameter estimasi:

$$a_1 = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X} \right) \left(Y_i - \overline{Y}_i \right)}{\sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2}$$

dan

$$a_0 = \overline{Y} - a_1 \overline{X}$$

di mana ($\overline{X}$ dan $\overline{Y}$) adalah rata-rata dari data (X) dan (Y). Persamaan ini menunjukkan bahwa (a_1) dihitung berdasarkan kovariansi antara (X) dan (Y) dibagi dengan variansi dari (X), yang mengukur kekuatan dan arah hubungan linear antara variabel. Setelah parameter diperoleh, model regresi dapat digunakan untuk memprediksi nilai baru dari variabel dependen dengan memasukkan nilai variabel independen ke dalam persamaan regresi.

9.8 Interpretasi Koefisien Regresi

Interpretasi koefisien regresi adalah langkah penting dalam analisis regresi linear karena membantu memahami bagaimana variabel independen mempengaruhi variabel dependen dalam sebuah model. Dalam persamaan regresi linear sederhana, terdapat dua koefisien yang harus diinterpretasikan, yaitu koefisien intersep (a_0) dan koefisien kemiringan (a_1). Kedua koefisien ini memberikan informasi yang berbeda tentang hubungan antara variabel-variabel yang dianalisis.

Koefisien intersep (a_0) adalah nilai (Y) ketika variabel independen (X) bernilai nol. Dalam banyak kasus, ini dapat diartikan sebagai nilai dasar atau titik awal dari variabel dependen sebelum terjadi perubahan pada variabel independen. Misalnya, dalam konteks biaya tetap dan variabel pada produksi, jika (X) mewakili jumlah unit produksi dan (Y) mewakili total biaya produksi, maka koefisien intersep (a_1) akan menggambarkan biaya tetap yang harus dikeluarkan meskipun tidak ada unit yang diproduksi. Ini mungkin mencakup biaya seperti sewa gedung atau gaji tetap karyawan. Dalam kasus lain, nilai intersep bisa saja tidak memiliki interpretasi yang praktis, terutama jika nilai nol pada (X) tidak masuk akal dalam konteks data.

Koefisien kemiringan (a_1), atau disebut juga koefisien regresi, mengukur seberapa besar perubahan pada variabel dependen (Y) akibat perubahan satu unit pada variabel independen (X). Jika (a_1) bernilai positif, ini menunjukkan bahwa peningkatan nilai (X) akan menyebabkan peningkatan nilai (Y), yang menggambarkan hubungan positif. Sebaliknya, jika (a_1) bernilai negatif, peningkatan (X) akan menyebabkan penurunan (Y), menunjukkan hubungan negatif. Contohnya, dalam analisis hubungan antara jam belajar siswa (X) dan nilai ujian (Y), koefisien (b) positif menunjukkan bahwa setiap tambahan jam belajar akan meningkatkan nilai ujian, dengan tingkat peningkatan sesuai nilai (a_1). Ini adalah informasi kunci dalam analisis regresi karena menjelaskan kekuatan dan arah pengaruh dari variabel independen terhadap variabel dependen.

Penting untuk memahami bahwa interpretasi koefisien regresi (a_1) juga harus mempertimbangkan konteks data dan satuan pengukuran dari variabel. Misalnya, jika (X) diukur dalam satuan ribuan dan (Y) dalam satuan rupiah, koefisien (b) menunjukkan berapa rupiah perubahan pada (Y) untuk setiap ribu unit perubahan pada (X). Selain itu, signifikansi statistik dari koefisien (a_1) perlu diuji untuk memastikan bahwa

hubungan yang ditemukan tidak terjadi secara kebetulan. Ini biasanya dilakukan dengan uji t atau pengujian hipotesis untuk menentukan apakah koefisien tersebut secara signifikan berbeda dari nol, yang berarti bahwa ada hubungan yang nyata antara variabel (X) dan (Y).

Contoh perhitungan

Terdapat data disajikan pada tabel nilai antara X_i dan Y_i, dimana terdapat 7 set data, i adalah urutan data, kemudian diminta untuk menncari regresi antara X dan Y.

$\boldsymbol{i}$	$\boldsymbol{Xi}$	$\boldsymbol{Yi = f(xi)}$
1	1	0.5
2	2	2.5
3	3	2.0
4	4	4.0
5	5	3.5
6	6	6.0
7	7	5.5
$\sum$	28	24.0

Untuk menjawab soal di atas, perlu mencari nilai (a_0) dan (a_1),

$$a_1 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y_i}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

dan $a_0 = \overline{Y} - a_1 \overline{X}$ untuk itu kita perlu menambah kolom di memakai perhitungan sederhana dan lebih mudah. Tabel diatas dapat kita ubah menjadi Tabel berikut

i	x_i	Y_i=	$X_i - \overline{X}$	$Y_i - \overline{Y_i}$	$\left(X_i - \overline{X}\right)^2$	$\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y_i}\right)$
0	1	0.5	-3.0	-2.9	9.0	8.79
1	2	2.5	-2.0	-0.9	4.0	1.86
2	3	2.0	2.0	-1.4	1.0	1.43
3	4	4.0	4.0	0.6	0.0	0.00
4	5	3.5	3.5	0.1	1.0	0.07
5	6	6.0	6.0	2.6	4.0	5.14
6	7	5.5	5.5	2.1	9.0	6.21
∑	28.0	24.0	24.0	0.0	28.0	23.5
Avarage	4.0	3.4	3.4			

Dengan membuatkan table sederhana seperti ti atas, kita dapat dengan mudah menghitung parameter yang dibutuhkan dan disjaikan pada persamaan berikut

$$a_1 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y_i}\right)}{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} = \frac{23.5}{28.0} = 0.839$$

$$Y_i = a_0 + 0.839X$$

Dengan mengambil nilai rata-rata variable X dan Y, kita dapat menghitung nilai a_0

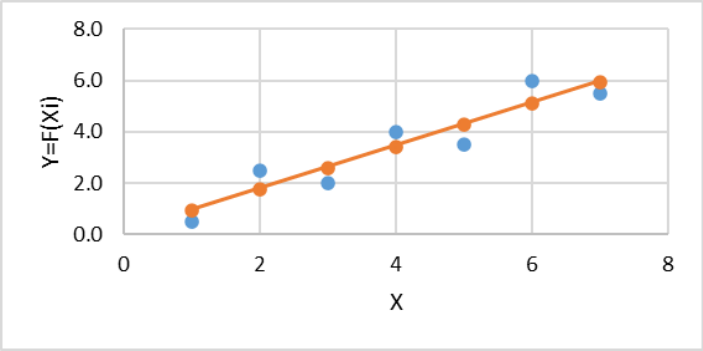
$$Y_i = a_0 + 0.839X$$

$$3.4 = a_0 + 0.839(4)$$

$$a_0 = 0.071$$

Dengan demikian persamaan regresi linear

$$Y_i = 0.071 + 0.839X$$



BAB X
ANALISIS REGRESI LINIER BERGANDA

10.1 Pengertian dan Dasar-Dasar Regresi Linier Berganda

Regresi Linier Berganda adalah metode statistik yang digunakan untuk menguji hubungan antara satu variabel dependen (Y) dengan lebih dari satu variabel independen (X_1, X_2, X_3, ..., X_n). Regresi ini merupakan perpanjangan dari regresi linier sederhana, di mana hanya satu variabel independen digunakan. Dalam regresi linier berganda, peneliti mencoba memahami seberapa besar perubahan dalam variabel dependen dipengaruhi oleh perubahan dalam beberapa variabel independen secara simultan.

Regresi linier berganda mencoba memodelkan hubungan linier antara variabel dependen (Y) dan variabel-variabel independen (X_1, X_2, ..., X_n). Persamaan umum untuk regresi linier berganda ditulis sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon$$

Di mana:

Y = variabel dependen,

$X_1, X_2, ..., X_n$ = variabel independen,

β_0 = konstanta (intersep),

$\beta_1, \beta_2, ..., \beta_n$ = koefisien regresi yang menunjukkan besarnya pengaruh masing-masing variabel independen terhadap Y,

ε = error atau residu, yang mencerminkan perbedaan antara nilai yang diamati dengan nilai yang diprediksi.

1. Tujuan Regresi Linier Berganda

   Tujuan utama dari regresi linier berganda adalah untuk memprediksi nilai variabel dependen berdasarkan nilai variabel-variabel independen. Selain itu, metode ini digunakan untuk:

   a. Mengidentifikasi variabel mana yang secara signifikan mempengaruhi variabel dependen.
   b. Menilai seberapa besar pengaruh masing-masing variabel independen terhadap variabel dependen.
   c. Menentukan model terbaik yang dapat menjelaskan hubungan antara variabel-variabel dalam penelitian.

2. Asumsi Dasar dalam Regresi Linier Berganda

   Agar analisis regresi linier berganda menghasilkan model yang valid dan dapat diandalkan, beberapa asumsi dasar harus dipenuhi, yaitu:

a. Hubungan Linier

   Terdapat hubungan linier antara variabel dependen dan variabel-variabel independen. Hal ini dapat diuji melalui plot scatter dan uji statistik.

b. Tidak Ada Multikolinearitas

   Multikolinearitas terjadi ketika ada hubungan yang sangat kuat antara dua atau lebih variabel independen. Jika ini terjadi, maka koefisien regresi menjadi tidak stabil dan sulit diinterpretasikan. Multikolinearitas dapat diuji dengan melihat nilai Variance Inflation Factor (VIF).

c. Distribusi Normal dari Error (Residual)

   Sisa-sisa atau residu (ε) harus terdistribusi normal. Asumsi ini dapat diuji dengan melihat plot histogram dari residu atau melalui uji statistik seperti uji Kolmogorov-Smirnov.

d. Homoskedastisitas

   Homoskedastisitas berarti bahwa varian error adalah konstan di seluruh nilai variabel independen. Jika varian error tidak konstan (heteroskedastisitas), model regresi bisa menjadi bias. Asumsi ini dapat diuji melalui plot residual versus nilai prediksi.

e. Tidak Ada Autokorelasi

   Autokorelasi berarti adanya hubungan antara residual dari pengamatan yang berurutan. Dalam regresi linier, residual harus independen. Autokorelasi dapat diuji dengan menggunakan uji Durbin-Watson.

10.2 Prosedur Analisis Regresi Linier Berganda

Regresi linier berganda merupakan teknik analisis statistik yang digunakan untuk menjelaskan hubungan antara satu variabel terikat (dependent) dengan dua atau lebih variabel bebas (independent). Prosedur ini diawali dengan mengidentifikasi variabel-variabel yang akan dianalisis. Variabel terikat (Y) adalah variabel yang ingin diprediksi atau dijelaskan, sementara variabel bebas (X_1, X_2, dst.) merupakan faktor-faktor yang dihipotesiskan mempengaruhi variabel terikat. Contohnya, jika variabel terikat adalah tingkat

pendapatan seseorang, variabel bebasnya bisa berupa tingkat pendidikan, pengalaman kerja, dan usia.

Tahap berikutnya adalah pengumpulan data yang relevan untuk semua variabel tersebut. Data yang digunakan harus bersifat numerik agar dapat dianalisis menggunakan metode statistik. Selain itu, perlu dipastikan bahwa jumlah sampel yang digunakan cukup untuk menghasilkan kesimpulan yang valid dan menghindari kesalahan generalisasi. Setelah data dikumpulkan, dilakukan uji asumsi awal untuk memvalidasi kelayakan model regresi. Asumsi tersebut antara lain linearitas (hubungan linear antara variabel bebas dan terikat), independensi data, homoskedastisitas (variansi residual yang konstan), dan normalitas residual. Juga penting untuk menguji multikolinearitas, yaitu hubungan linier antarvariabel bebas yang harus dihindari.

Setelah asumsi-asumsi tersebut terpenuhi, model regresi linier berganda dapat dibangun menggunakan persamaan dasar: $Y=\beta_0 + \beta_1X_1 + \beta_2X_2 + \cdots \beta_2X_2 + \beta_nX_n + \varepsilon$, Dimana Y adalah variable terikat, X_1, X_2, …. X_n adalah variable bebas, β_0 adalah konstanta (intercept), β_1, β_2, ….. βn adalah koefisien regresi, dan ε adalah residual (error). Koefisien regresi dihitung menggunakan metode *Ordinary Least Squares* (OLS), yang

meminimalkan jumlah kuadrat dari selisih antara nilai aktual dan nilai prediksi.

Setelah model terbentuk, dilakukan uji signifikansi untuk menilai seberapa baik model tersebut menjelaskan hubungan antarvariabel. Uji F digunakan untuk menguji signifikansi keseluruhan model, yaitu apakah variabel bebas secara simultan mempengaruhi variabel terikat. Uji t digunakan untuk menguji signifikansi koefisien regresi masing-masing variabel bebas, guna mengetahui variabel mana yang berkontribusi signifikan. Koefisien determinasi R^2 $2R^2$ digunakan untuk melihat seberapa besar proporsi variansi variabel terikat yang dijelaskan oleh variabel bebas.

Tahap berikutnya adalah uji asumsi model untuk memastikan model yang dibangun sudah sesuai. Uji multikolinearitas dilakukan dengan menghitung *Variance Inflation Factor* (VIF), di mana nilai VIF lebih dari 10 mengindikasikan adanya masalah multikolinearitas. Selain itu, uji normalitas residual dan homoskedastisitas dilakukan untuk memeriksa apakah model regresi mematuhi asumsi-asumsi yang mendasar. Jika semua asumsi terpenuhi, model tersebut dapat diinterpretasikan dan digunakan untuk memprediksi nilai variabel terikat.

Setelah model dianggap valid, langkah terakhir adalah melakukan prediksi menggunakan persamaan regresi yang telah diperoleh. Nilai variabel bebas baru dapat dimasukkan ke dalam model untuk memperkirakan nilai variabel terikat. Pada tahap ini, penting untuk memahami bahwa hasil prediksi selalu memiliki tingkat ketidakpastian yang tergantung pada kesalahan prediksi atau residual. Oleh karena itu, interpretasi hasil harus dilakukan dengan hati-hati dan mempertimbangkan asumsi-asumsi yang telah diuji sebelumnya.

10.3 Metode Estimasi Parameter Regresi

Metode estimasi parameter regresi bertujuan untuk menentukan nilai koefisien regresi dalam sebuah model regresi, yang merepresentasikan hubungan antara variabel bebas dan variabel terikat. Koefisien regresi ini akan memberikan informasi tentang seberapa besar pengaruh variabel bebas terhadap variabel terikat. Dalam analisis regresi, salah satu metode yang paling umum digunakan untuk estimasi parameter adalah Metode Kuadrat Terkecil OLS (Ordinary Least Squares). Selain itu, terdapat juga metode lain seperti Metode Likelihood Maksimum (Maximum Likelihood Estimation/ MLE) yang banyak

digunakan dalam konteks regresi non-linier atau untuk model yang lebih kompleks.

1. Metode Kuadrat Terkecil (*Ordinary Least Squares*/ OLS)

   Metode OLS adalah metode yang paling sering digunakan dalam regresi linier. Tujuan utama OLS adalah untuk meminimalkan jumlah kuadrat dari selisih antara nilai yang diobservasi dari variabel terikat dan nilai yang diprediksi oleh model regresi. Persamaan umum model regresi linier sederhana adalah:

   $$Y = \beta_0 + \beta_1 X + \varepsilon$$

   Pada regresi linier berganda:

   $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$$

   Dalam metode OLS, nilai koefisien (β_0,β_1,...,β_n) diestimasi dengan meminimalkan fungsi kesalahan yang dinyatakan sebagai jumlah kuadrat dari residual atau error, yaitu perbedaan antara nilai aktual dan nilai prediksi (Y):

   $$\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - (\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n))^2$$

   Langkah-langkah metode OLS meliputi:

a. Menghitung turunan parsial dari fungsi kuadrat residual terhadap masing-masing koefisien regresi.
b. Menyelesaikan sistem persamaan linear yang dihasilkan untuk mendapatkan nilai koefisien yang meminimalkan jumlah kuadrat residual.

Keuntungan utama dari OLS adalah bahwa metode ini memberikan estimator yang unbiased dan efisien jika asumsi model regresi terpenuhi, seperti tidak adanya autokorelasi, multikolinearitas, dan heteroskedastisitas.

2. Metode Likelihood Maksimum (*Maximum Likelihood Estimation*/ MLE)

   Metode Likelihood Maksimum digunakan untuk mengestimasi parameter dalam berbagai model regresi, terutama dalam model non-linier atau model dengan distribusi khusus (misalnya regresi logistik atau probit). Prinsip MLE adalah menemukan nilai koefisien yang memaksimalkan likelihood, atau probabilitas bahwa data yang diamati dapat dihasilkan oleh model dengan parameter yang diestimasi.

   Jika kita memiliki sekumpulan data observasi, Y_1,Y_2 .... Y_n dengan distribusi tertentu, misalnya

normal, likelihood dari parameter yang diestimasi adalah produk dari probabilitas terjadinya observasi tersebut, yaitu:

$$L(\theta) = P(Y_1, Y_2, \ldots, Y_n|\theta)$$

MLE mencoba mencari parameter θ yang memaksimalkan likelihood tersebut. Untuk regresi linier dengan distribusi normal, MLE menghasilkan estimator yang sama dengan OLS. Namun, dalam konteks regresi yang lebih kompleks, seperti model dengan distribusi non-normal, MLE sering kali lebih efektif.

3. *Generalized Least Squares* (GLS)

   Metode *Generalized Least Squares* (GLS) merupakan pengembangan dari OLS yang digunakan ketika asumsi homoskedastisitas (variansi residual yang konstan) dan independensi tidak terpenuhi. Pada GLS, metode estimasi parameter mempertimbangkan adanya autokorelasi dan heteroskedastisitas dalam residual.

   Jika OLS diimplementasikan pada data yang mengalami heteroskedastisitas atau autokorelasi, maka hasil estimasi tidak lagi

efisien meskipun tetap unbiased. GLS memperbaiki masalah ini dengan memberikan bobot yang berbeda pada observasi yang berbeda, tergantung pada variansinya, sehingga memperhitungkan struktur korelasi dan heteroskedastisitas yang ada dalam data.

4. *Weighted Least Squares* (WLS)

   Metode WLS mirip dengan GLS, namun secara khusus digunakan untuk mengatasi heteroskedastisitas. Jika variabilitas error bervariasi di seluruh observasi, maka metode WLS memberikan bobot pada masing-masing observasi, di mana observasi dengan varians yang lebih rendah diberi bobot lebih besar dalam estimasi parameter. Ini membuat estimasi parameter menjadi lebih efisien.

   Dalam WLS, setiap observasi memiliki bobot yang merupakan kebalikan dari variansi residualnya. Hal ini memungkinkan model untuk memperhitungkan heteroskedastisitas dan memberikan hasil estimasi yang lebih baik dibandingkan OLS dalam situasi ini.

5. *Ridge Regression* dan Lasso (*Penalized Regression*)

Dalam situasi di mana terdapat multikolinearitas tinggi antara variabel bebas, metode *Ridge Regression* atau Lasso sering digunakan. *Ridge Regression* menambahkan penalti pada besarnya koefisien regresi, yang membantu mengurangi masalah multikolinearitas. *Ridge Regression* meminimalkan fungsi kuadrat residual seperti OLS, namun dengan tambahan penalti (regularisasi) berupa jumlah kuadrat dari koefisien regresi:

$$\sum(Y_i - \hat{Y}_i)^2 + \lambda \sum \beta_j^2$$

Sedangkan Lasso menambahkan penalti berupa jumlah nilai absolut koefisien regresi:

$$\sum(Y_i - \hat{Y}_i)^2 + \lambda \sum |\beta_j|$$

Ridge Regression cenderung mengecilkan koefisien yang besar tetapi tidak membuatnya nol, sedangkan Lasso dapat mendorong koefisien yang tidak signifikan menjadi nol, sehingga juga berfungsi sebagai metode seleksi variabel.

6. Metode Non-Linear *Least Squares (*NLS)

   Metode NLS digunakan untuk mengestimasi parameter dalam model non-linier. Jika

hubungan antara variabel terikat dan variabel bebas tidak bisa diwakili oleh fungsi linear, metode NLS diperlukan untuk meminimalkan kuadrat residual pada model non-linier. Prosedur NLS lebih kompleks dibanding OLS, karena melibatkan iterasi dan metode numerik untuk mencari solusi optimal dari parameter model.

10.4 Masalah dan Penyelesaian dalam Regresi Linier Berganda

Regresi linier berganda adalah alat yang sangat berguna untuk menganalisis hubungan antara beberapa variabel. Namun, dalam praktiknya, terdapat berbagai masalah yang dapat mempengaruhi hasil analisis. Salah satu masalah yang paling umum adalah multikolinearitas, yaitu situasi di mana terdapat hubungan linear yang tinggi antara dua atau lebih variabel independen. Multikolinearitas dapat menyebabkan estimasi koefisien regresi menjadi tidak stabil, sulit untuk diinterpretasikan, dan meningkatkan variansi dari koefisien. Untuk mengatasi masalah ini, peneliti dapat menggunakan metode *Variance Inflation Factor* (VIF) untuk mendeteksi tingkat multikolinearitas. Jika VIF untuk variabel tertentu lebih

dari 10, variabel tersebut dapat dipertimbangkan untuk dihapus atau dikombinasikan dengan variabel lain.

Masalah lain yang sering muncul dalam regresi linier berganda adalah heteroskedastisitas, di mana variansi dari residual tidak konstan di seluruh rentang nilai variabel independen. Heteroskedastisitas dapat menyebabkan estimasi koefisien yang tidak efisien dan dapat mempengaruhi hasil uji signifikansi. Untuk mendeteksi heteroskedastisitas, peneliti dapat menggunakan uji Breusch-Pagan atau plot residual. Jika heteroskedastisitas terdeteksi, solusi yang dapat diterapkan termasuk penggunaan *Weighted Least Squares* (WLS) untuk memberikan bobot pada variabel, atau transformasi variabel yang relevan untuk menstabilkan variansi.

Autokorelasi adalah masalah lain yang mungkin muncul, terutama dalam data waktu. Ini terjadi ketika residual dari satu pengamatan terkait dengan residual dari pengamatan lainnya. Autokorelasi dapat menyebabkan kesalahan dalam estimasi parameter dan uji signifikansi. Untuk mendeteksi autokorelasi, peneliti dapat menggunakan uji Durbin-Watson. Jika autokorelasi terdeteksi, metode yang dapat digunakan untuk memperbaiki model adalah dengan menerapkan Generalized Least Squares (GLS), yang

mempertimbangkan adanya autokorelasi dalam model dan menghasilkan estimasi yang lebih efisien.

Masalah normalitas residual juga penting dalam analisis regresi linier berganda. Residual yang tidak terdistribusi normal dapat memengaruhi validitas uji hipotesis. Untuk menguji normalitas residual, peneliti dapat menggunakan uji Kolmogorov-Smirnov atau visualisasi menggunakan plot normal probability (P-P plot). Jika residual tidak terdistribusi normal, transformasi pada variabel terikat atau penerapan metode non-parametrik dapat menjadi solusi untuk memastikan bahwa asumsi normalitas terpenuhi.

Terakhir, masalah spesifikasi model juga harus diperhatikan. Ini terjadi ketika model yang digunakan tidak mencakup semua variabel penting atau sebaliknya, memasukkan variabel yang tidak relevan. Spesifikasi model yang buruk dapat menghasilkan estimasi yang bias dan tidak konsisten. Untuk mengatasi masalah ini, penting untuk melakukan eksplorasi awal data secara menyeluruh dan menggunakan metode seperti model seleksi (misalnya, stepwise regression) untuk menentukan variabel mana yang harus disertakan dalam model. Dengan demikian, peneliti dapat memastikan bahwa model yang

dihasilkan akurat dan dapat diandalkan dalam memprediksi variabel terikat.

DAFTAR PUSTAKA

Aczel, A. D., & Sounderpandian, J. (2012). Complete Business Statistics (7th ed.). McGraw-Hill Education.

Agresti, A., & Finlay, B. (2018). Statistical Methods for the Social Sciences (5th ed.). Pearson.

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., & Cochran, J. J. (2017). Statistics for Business & Economics 13e. In Cengage Learning (Vol. 16, Issue 4).

Asari, Andi. et, al. 2023. Pengantara Statistika. Sumatera Barat: PT MAFY MEDIA LITERASI INDONESIA

Babbie, E. (2013). The Practice of Social Research (13th ed.). Wadsworth Cengage Learning.

Black, K. (2019). Business Statistics: For Contemporary Decision Making. John Wiley & Sons.

Bluman, A. G. (2019). Elementary Statistics: A Step by Step Approach: A Brief Version. In McGraw Hill Education. McGraw-Hill Education.

Cherry, Kendra. (2014). Why are statistics necessary in psychology? About.com psychology.

Clausel, M., dan Grégoire, G. (2014). Practical Session: Multiple Linear Regression. European

Astronomical Society Publications Series,Vol. 66, 73-75. doi:10.1051/eas/1466006

Conover, W. J. (1999). Practical nonparametric statistics. John Wiley & Sons, Inc.

Creswell, J. W. (2014). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. 4th Edition. SAGE Publications.

de Smith, M. J., Goodchild, M. F., & Longley, P. (2020). Geospatial Analysis: A Comprehensive Guide to Principles, Techniques, and Software Tools (6th ed.). The Winchelsea Press.

Devore, J. L. (2011). Probability and Statistics. 8th Edition. Cengage Learning.

Draper, N. R., & Smith, H. (1998). Applied Regression Analysis. 3rd Edition. John Wiley & Sons.

Freedman, D., Pisani, R., & Purves, R. (2014). Statistics (4th ed.). W.W. Norton & Company.

Freund, R. J., & Wilson, W. J. (2006). Regression Analysis: Statistical Modeling of a Response Variable. 2nd Edition. Elsevier.

Ghozali, I. (2018). Aplikasi Analisis Multivariate dengan Program IBM SPSS 25. Badan Penerbit Universitas Diponegoro.

Gravetter, F. J., & Wallnau, L. B. (2013). Statistics for the Behavioral Sciences. Belmont: Wadsworth Cengage Learning.

Greene, W. H. (2012). Econometric Analysis (7th ed.). Pearson Education.

Guilford, J. P. (1950). Fundamental statistics in psychology and education. McGraw-Hill.

Gujarati, D. N., & Porter, D. C. (2009). Basic Econometrics (5th ed.). McGraw-Hill/Irwin.

Gunawan, Muhammad Ali. 2015. Statistik Penelitian Bidang Pendidikan, Psikologi dan Sosial.

Gupta, S. C. (2010). Fundamentals of Statistics. 6th Edition. Himalaya Publishing House.

Hadjar, Ibnu. 2019, Statistik Untuk Ilmu Pendidikan, Sosial dan Humaniora. Bandung, PT Remaja Rosdakarya.

Hamidi, D. Z. (2020). Konsep dan Ruang Lingkup Statistika Part #1 - YouTube. https://www.youtube.com/watch?v=QNNkWIzwYIc

Hidayati, Tri. (2019). Statistika Dasar. Banten: UNPAM Press

Islas Toski, M., Avila-Cardenas, K., dan Gálvez, J. (2020). Linear Regression Techniques for Car Accident Prediction. In D. Oliva dan S. Hinojosa (Eds.),

Applications of Hybrid Metaheuristic Algorithms for Image Processing (pp. 261-284). Cham: Springer International Publishing.

Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. 6th Edition. Pearson.

Jones, G. T., Hagtvedt, R., dan Jones, K. (2004). A VBA-based Simulation for Teaching Simple Linear Regression. Teaching Statistics,Vol. 26(2), 36-41. doi:10.1111/j.1467-9639.2004.00162.x

Judge, G. G., Griffiths, W. E., Hill, R. C., & Lutkepohl, H. (1985). The Theory and Practice of Econometrics (2nd ed.). Wiley.

Kerlinger, F. N. (2000). Foundations of Behavioral Research. 4th Edition. Wadsworth Publishing.

Kirkwood, B. R., & Sterne, J. A. C. (2003). Essential Medical Statistics (2nd ed.). Wiley-Blackwell.

Kraak, M.-J., & Ormeling, F. (2011). Cartography: Visualization of Spatial Data (3rd ed.). Routledge.

Kustituanto, B., & Badrudin, R. (1994). Statistika 1 (Deskriptif). Penerbit Gunadarma

Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). Applied Linear Regression Models. 4th Edition. McGraw-Hill.

Levine, D. M., Stephan, D. F., Szabat, K. A., & Berenson, M. L. (2020). Statistics for Managers Using Microsoft Excel (9th ed.). Pearson.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). Geographic Information Systems and Science (4th ed.). Wiley.

Marill, K. A. (2004). Advanced statistics: linear regression, part I: simple linear regression. Academic emergency medicine,Vol. 11(1), 87-93. doi:10.1197/j.aem.2003.09.005

McClave, J. T., Benson, P. G., & Sincich, T. (2018). Statistics for Business and Economics (13th ed.). Pearson.

Monmonier, M. (1996). How to Lie with Maps (2nd ed.). University of Chicago Press.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis. 5th Edition. John Wiley & Sons.

Moore, D. S., McCabe, G. P., Alwan, L. C., Craig, B. A., & Duckworth, W. M. (2017). The Practice of Statistics for Business and Economics (4th ed.). W. H. Freeman.

Newbold, P., Carlson, W. L., & Thorne, B. M. (2013). Statistics for Business and Economics. Boston: Pearson.

Patton, M. Q. (2014). Qualitative Research & Evaluation Methods (4th ed.). Sage Publications.

Prion, S. K., dan Haerling, K. A. (2020). Making sense of methods and measurements: simple linear regression. Clinical Simulation in Nursing,Vol. 48, 94-95. doi:10.1016/j.ecns.2020.07.004

Richardson, M., Gabrosek, J., Reischman, D., dan Curtiss, P. (2004). Morse code, scrabble, and the alphabet. Journal of Statistics Education,Vol. 12(3). doi:10.1080/10691898.2004.11910628

Robinson, A. H., Morrison, J. L., Muehrcke, P. C., Kimerling, A. J., & Guptill, S. C. (1995). Elements of Cartography (6th ed.). Wiley.

Rodliyah, Iesyah. 2021. PENGANTAR DASAR STATISTIKA Dilengkapi Analisis dengan Bantuan Software SPSS. Jombang: LPPM UNHASY TEBUIRENG

Saunders, M., Lewis, P., & Thornhill, A. (2016). Research Methods for Business Students (7th ed.). Pearson Education.

Sekaran, U. & Bougie, R. (2016). Research Methods for Business: A Skill Building Approach. 7th Edition. John Wiley & Sons.

Siegel, A. F. (2016). Practical Business Statistics (7th ed.). Academic Press.

Siegel, S., & Castellan, N. J. (1988). Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2009). Thematic Cartography and Geovisualization (3rd ed.). Pearson.

Stock, J. H., & Watson, M. W. (2020). Introduction to Econometrics (4th ed.). Pearson.

Sudjana. 2005, Metode Statistika. Bandung, PT Tarsito Bandung.

Sugiyono. (2012). Metode Penelitian Kuantitatif, Kualitatif dan R&D. Alfabeta.

Supranto, J. (2010). Statistika: Teori dan Aplikasi. Erlangga.

Teddlie, C., & Yu, F. (2007). Mixed Methods Sampling: A Typology with Examples. Journal of Mixed Methods Research, 1(1), 77-100.

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2017). Probability & Statistics for Engineers & Scientists (9th ed.). Pearson.

Weiss, N. A. (2015). Introductory Statistics. Pearson Education.

Wooldridge, J. M. (2016). Introductory Econometrics: A Modern Approach. 6th Edition. Cengage Learning.

STATISTIK DASAR

Buku "Statistik Dasar" memperkenalkan konsep-konsep fundamental statistik yang digunakan untuk menganalisis dan menginterpretasikan data. Buku ini mencakup topik-topik seperti pengumpulan data, penyajian data dalam bentuk tabel dan grafik, ukuran pemusatan (mean, median, modus), ukuran penyebaran (variansi, standar deviasi), serta konsep probabilitas dasar. Selain itu, buku ini juga membahas distribusi probabilitas, uji hipotesis, regresi, dan korelasi, yang semuanya penting untuk pengambilan keputusan berdasarkan data. Dengan pendekatan yang sederhana dan sistematis, buku ini membantu pembaca memahami konsep statistik yang sering digunakan dalam berbagai bidang penelitian.

CV. ASKARA SASTRA MEDIA

ISBN: 978-6-23104-975-9

9 786231 049759