STATISTIKA DAN PROBABILITAS

Raden Sri Ayu Ramadhana
Arief Aulia Rahman
Rini Yunita
Enos Lolang
Ratu Sarah Fauziah Iskandar
Retno Andriyani
Ukta Indra Nyuswantoro
Siti Aminah
Ridha Yuniara
Kusnaeni

GET PRESS INDONESIA

STATISTIKA DAN PROBABILITAS

Penulis :
Raden Sri Ayu Ramadhana
Arief Aulia Rahman
Rini Yunita
Enos Lolang
Ratu Sarah Fauziah Iskandar
Retno Andriyani
Ukta Indra Nyuswantoro
Siti Aminah
Ridha Yuniara
Kusnaeni

ISBN :

Editor : Dr. Neila Sulung, S.Pd., Ns., M.Kes.
Penyunting : Tri Putri Wahyuni,S.Pd
Desain Sampul dan Tata Letak : Atyka Trianisa, S.Pd

Penerbit : GET PRESS INDONESIA
Anggota IKAPI No. 033/SBA/2022

Redaksi :
Jln. Palarik Air Pacah No 26 Kel. Air Pacah
Kec. Koto Tangah Kota Padang Sumatera Barat
Website : www. getpress.co.id
Email : adm.getpress@gmail.com

Cetakan pertama, Desember 2023

KATA PENGANTAR

Segala Puji dan syukur atas kehadirat Allah SWT dalam segala kesempatan. Sholawat beriring salam dan doa kita sampaikan kepada Nabi Muhammad SAW. Alhamdulillah atas Rahmat dan Karunia-Nya penulis telah menyelesaikan Buku Statistika Dan Probabilitas ini.

Buku ini membahas Pengenalan Statistika dan Probabilitas, Pengumpulan dan Penyajian Data, Pengukuran Pemusatan Data, Pengukuran Penyebaran Data, Distribusi Peluang Diskrit, Variabel Acak Kontinu, Teorema Probabilitas, Rata-rata dan Varians dari Variabel Acak, Estimasi dan Interval Kepercayaan, Regresi dan Korelasi.

Proses penulisan buku ini berhasil diselesaikan atas kerjasama tim penulis. Demi kualitas yang lebih baik dan kepuasan para pembaca, saran dan masukan yang membangun dari pembaca sangat kami harapkan.

Penulis ucapkan terima kasih kepada semua pihak yang telah mendukung dalam penyelesaian buku ini. Terutama pihak yang telah membantu terbitnya buku ini dan telah mempercayakan mendorong, dan menginisiasi terbitnya buku ini. Semoga buku ini dapat bermanfaat bagi masyarakat Indonesia.

Padang, Desember 2023

Penulis

DAFTAR ISI

KATA PENGANTAR ..... **i**
DAFTAR ISI ..... **ii**
DAFTAR GAMBAR ..... **vi**
DAFTAR TABEL ..... **vii**
BAB 1 PENGENALAN STATISTIKA DAN PROBABILITAS **1**
1.1 Pendahuluan 1
1.2 Statistika 2
1.2.1 Penggolongan Statistik 3
1.2.2 Ciri Khas Statistik 4
1.2.3 Permasalahan Statistik 5
1.3 Probabilitas Statistika 7
1.3.1 Fungsi, Kegunaan dan Ruang Lingkup Statistika 7
1.3.2 Bagian Ilmu Statistika, Metodologi dan Elemen Statistika 9
1.3.3 Pengumpulan dan Tipe Data Statistika 10
DAFTAR PUSTAKA 12
BAB 2 PENGUMPULAN DAN PENYAJIAN DATA **13**
2.1 Tabel 13
2.2 Grafik 18
DAFTAR PUSTAKA 28
BAB 3 PENGUKURAN PEMUSATAN DATA **29**
3.1 Pendahuluan 29
3.2 Pengukuran Pemusatan Data 29
3.3 Rata-Rata atau Mean 29
3.3.1 Mean untuk Data Berkelompok 31
3.3.2 Rata-rata atau Mean Menggunakan Angka Terkaan 33
3.3.3 Rata-Rata Hitung Atau Mean Dengan Metode Coding 34
3.3.4 Rata-Rata Hitung Atau Mean Untuk Data Berbobot 34
3.4 Rata-rata Letak (Median) 35
3.4.1 Median untuk data tidak berkelompok 36

3.4.2 Median untuk data kelompok (dalam distribusi frekuensi) 36
3.5 Modus 38
3.5.1 Modus untuk data berkelompok (distribusi frekuensi) 38
3.6 Hubungan antara mean, median dan modus 40
3.7 Ukuran Letak 41
3.7.1 Kuartil 41
3.7.2 Desil 43
3.7.3 Persentil 44
3.8 Penutup 46
DAFTAR PUSTAKA 47
BAB 4 UKURAN PENYEBARAN DATA 49
4.1 Pendahuluan 49
4.2 Ukuran Penyebaran Data 50
4.2.1 Rentang 52
4.2.2 Varians 54
4.2.3 Simpangan Rata-rata 56
4.2.4 Standar Deviasi 57
4.2.5 Kuartil dan Rentang Antarkuartil 64
4.3 Indeks Variasi Kualitatif dan Indeks Segregasi 67
DAFTAR PUSTAKA 72
BAB 5 DISTRIBUSI PELUANG DISKRIT 73
5.1 Pendahuluan 73
5.2 Distribus Peluang Diskrit 74
5.2.1 Distribusi Bernouli 74
5.2.2 Distribusi Binomial 75
5.2.3 Distribusi Poisson 77
5.2.4 Distribusi Geometrik 79
DAFTAR PUSTAKA 82
BAB 6 VARIABEL ACAK KONTINU 83
6.1 Probabilitas Variabel Acak Interval Berbentuk Kontinu 83
6.2 Distribusi Fungsi Kumulatif Kontinyu 85
6.3 Peluang Distribusi Kontinu 87
DAFTAR PUSTAKA 92

BAB 7 TEOREMA PROBABILITAS.....93
7.1 Konsep dasar probabilitas.....93
7.2 Sejarah dan perkembangan teorema probabilitas.....94
7.3 Pentingnya probabilitas dalam berbagai bidang.....95
7.4 Ruang sampel dan kejadian.....96
7.5 Aksioma probabilitas.....97
7.6 Variabel Acak dan Distribusi Probabilitas.....98
7.7 Teorema Probabilitas Kondisional dan Independensi.....99
7.8 Variabel Acak Bersama dan Distribusi Gabungan.....101
7.9 Hukum Besar Bilangan dan Teorema Batas Pusat.....103
7.10 Aplikasi Teorema Probabilitas dalam Bidang Lain.....105
DAFTAR PUSTAKA.....107
BAB 8 RATA-RATA DAN VARIANS DARI VARIABEL ACAK.....109
8.1 Pendahuluan.....109
8.2 Mean Variabel Acak.....109
8.2.1 Contoh Mean Variabel Acak Diskrit.....111
8.2.2 Contoh Mean Variabel Acak Kontinu.....112
8.3 Varians Variabel Acak.....113
8.3.1 Contoh Varians Variabel Acak Diskrit.....114
8.3.2 Contoh Varians Variabel Acak Kontinu.....115
DAFTAR PUSTAKA.....116
BAB 9 ESTIMASI DAN INTERVAL KEPERCAYAAN.....117
9.1 Estimasi.....117
9.2 Interval Kepercayaan.....119
DAFTAR PUSTAKA.....124
BAB 10 REGRESI DAN KORELASI SEDERHANA.....125
10.1 Pendahuluan.....125
10.2 Pengetian Regresi.....126
10.3 Regresi Linier Sederhana.....127
10.3.1 Metode Kuadrat Terkecil.....127
10.3.2 Interpretasi terhadap nilai koefisien regresi.....129
10.3.3 Syarat dan Prosedur Regresi Linier sederhana.....129
10.4 Korelasi.....134
10.4.1 Koefisien Korelasi dan Koefisien Determinasi.....134
10.4.2 Interpretasi terhadap nilai koefisien korelasi.....135
10.4.3 Klasifikasi Koefisien Korelasi.....135

DAFTAR PUSTAKA 143
BIODATA PENULIS

DAFTAR GAMBAR

Gambar 2.1. Grafik penjualan kamus bahasa 20
Gambar 2.2. Macam-macam bentuk grafik batang (Histogram) 21
Gambar 2.3. Grafik Garis Penjualan Kamus 21
Gambar 2.4. Grafik Lingkaran Ekstrakulikuler Siswa 23
Gambar 2.5. Grafik Lingkaran derajat tentang data makanan 25
Gambar 2.6. Grafik lingkaran persen tentang data hobi olahraga siswa 27
Gambar 4.1. Perbedaan tinggi badan dua tim pemain basket 50
Gambar 4.2. Rentang tinggi badan 52
Gambar 4.3. Data pengamatan dan simpangan dari nilai rata-rata 57
Gambar 4.4. Visualisasi variasi data kelompok I dan kelompok II 63
Gambar 9.1. Estimator bias dan tak-bias 117
Gambar 10.1. Jenis hubungan antara variabel *X* dan *Y* 126
Gambar 10.2. Garis regresi hubungan antara variabel *X* dan *Y* 132

DAFTAR TABEL

Tabel 2.1. Jumlah Mahasiswa Prodi Pendidikan Matematika 14
Tabel 2.2. Tingkat Produksi Hasil Pertanian tahun 2007 di Kalimantan 14
Tabel 2.3. Tingkat Kepuasan Mahasiswa terhadap layanan akademik. 15
Tabel 2.4. Tabel Distribusi Frekuensi dengan Turus 18
Tabel 2.5. Tabel Penjualan Kamus 19
Tabel 3.1. Populasi hasil pemeriksaaan % ketidakmurnian ikan tuna 30
Tabel 3.2. Sampel hasil pemeriksaaan % ketidakmurnian ikan tuna 30
Tabel 3.3. Distribusi Frekuensi Data Tunggal 31
Tabel 3.4. Distribusi Frekuensi Data Berkelompok 32
Tabel 3.5. Cara menghitung rata-rata atau Mean Nilai Statistika Mahasiswa 32
Tabel 3.6. Cara menghitung rata-rata atau Mean Dengan Angka Terkaan 33
Tabel 3.7. Cara menghitung rata-rata atau Mean Dengan Metode Coding 34
Tabel 3.8. Cara menghitung rata-rata atau Mean untuk data berbobot 35
Tabel 3.9. Data Interval nilai untuk data berkelompok 37
Tabel 3.10. Cara menghitung Median pada data Bekelompok 37
Tabel 3.11. Data Interval nilai untuk data berkelompok 39
Tabel 3.12. Cara untuk menghitung modus pada data berkelompok 39
Tabel 3.13. Cara untuk Menghitung K1, K2 dan K3 pada Data Berkelompok 42
Tabel 3.14. Cara Untuk Menghitung D3 dan D9 pada Data Berkelompok 44
Tabel 3.15. Cara untuk Menghitung P60 dan P75 pada Data Berkelompok 45
Tabel 4.1. Simpangan dari nilai rata-rata 57

Tabel 4.2. Menghitung deviasi kuadrat tinggi badan Tim II 59
Tabel 4.3. Kuadrat simpangan tinggi badan Tim I 59
Tabel 4.4. Nilai $\sum x_i$ dan $(\sum x_i^2)$ data tinggi badan Tim II 62
Tabel 4.5. Perbedaan variasi data Kelompok I dan Kelompok II 62
Tabel 4.6. Tabel frekuensi kategori status perkawinan 68
Tabel 4.7. Distribusi Jenis Kelamin dan Pekerjaan 62
Tabel 4.8. Mengatur jumlah wanita untuk memperoleh proporsi yang sama 70
Tabel 4.9. Selisih proporsi sebagai ukuran penyebaran 70
Tabel 9.1. Nilai-nilai z_c yang berkorespondensi dengan berbagai tingkat kepercayaan 120
Tabel 10.1. Deskripsi besar pendapatan dan berat badan 130
Tabel 10.2. Perhitungan Unsur-unsur Persamaan Regresi 131
Tabel 10.3. Klasifikasi nilai koefisien korelasi 135
Tabel 10.4. Deskripsi Upah Pegawai dan Produksi beras perusahaan 137
Tabel 10.5. Perhitungan Unsur-unsur Persamaan Regresi 138
Tabel 10.6. Deskripsi pengalaman kerja sales dan hasil penjualan 140
Tabel 10.7. Perhitungan Unsur-unsur Koefisien Korelasi 140
Tabel 10.8. Deskripsi antara inflasi dan jumlah uang yang beredar 141
Tabel 10.9. Deskripsi peringkat jenis kopi berdasarkan rasanya 142

BAB 1
PENGENALAN STATISTIKA DAN PROBABILITAS

Oleh Raden Sri Ayu Ramadhana

1.1 Pendahuluan

Statistika memiliki peranan yang penting dalam kehidupan, tidak hanya diperlukan dalam bidang penelitian saja akan tetapi sangat diperlukan bagi bidang-bidang ilmu pengetahuan lainnya seperti: pendidikan, psikologi, antropologi, bisnis asuransi, pertanian, kedokteran, teknik, industri, astronomi, dan sebagainya.

Pada aplikasi, statistika sudah banyak digunakan dalam kehidupan sehari-hari. Sebagai contoh: Pada tahun 2022 tercatat 11.159 kasus jumlah kecelakaan lalu lintas yang terjadi di Sumatera Utara, Prakiraan cuaca di sekitar kota Medan pada pagi hari bulan Oktober mengalami hujan ringan, kemudian penghitungan pemilu dengan menggunakan data *polling* dalam memprediksi hasil pemilu, dan hasil panen sawit diperkirakan 1 ton sekali panen dalam tiap hektarnya. Hal ini sering kita dengar atau kita baca dalam surat kabar maupun berita di televisi. Pemerintah menggunakan statistika untuk menilai hasil dari pembangunan masa lalu, juga dapat merencanakan pembangunan untuk masa mendatang. Pimpinan dalam hal ini mengambil manfaat dari kegunaan statistika untuk melaksanakan tindakan yang perlu dalam menjalankan tugasnya.

Dalam dunia penelitian, statistika wajib digunakan, tujuannya ialah untuk mengetahui apakah cara yang baru ditemukan lebih baik dari cara yang lama, kemudian melalui riset yang dilakukan di laboratorium, atau penelitian yang dilakukan di lapangan, perlu diadakan penilaian dengan statistika. Begitulah pentingnya statistika dalam kehidupan.

1.2 Statistika

Kata "statistik" berasal dari bahasa latin yaitu "*status"* yang memiliki persamaan makna dalam bahasa inggris "*state*" atau "staat" dalam bahasa belanda yang dalam bahasa Indonesia diartikan negara (Sudijono, 2018). Awalnya, kata "statistik" diartikan sebagai kumpulan bahan keterangan/data, baik yang berbentuk angka maupun bukan yang berbentuk angka dan memiliki arti penting atau kegunaan besar bagi suatu negara. Dalam perkembangannya, kata "statisktik" diartikan sebagai kumpulan bahan keterangan yang berbentuk angka atau disebut dengan data kuantitatif sedangkan bahan keterangan yang bukan berbentuk angka atau disebut dengan data kualitatif tidak lagi disebut dengan statistik.

"Statistics" dan *"statistic"* memiliki arti kata yang berbeda. Kata "*statistics"* artinya ilmu statistik sedangkan kata *"statistic"* berarti ukuran yang diperoleh atau berasal dari sampel. Namun pada saat ini jika kita mendengar atau membaca istilah statistik terdapat beberapa macam pengertian. Pengertian pertama ialah statistik yang diberi pengertian sebagai *data statistik* yaitu kumpulan bahan keterangan yang berupa angka angka atau bilangan atau statistik dikenal dengan deretan atau kumpulan angka yang menunjukkan keterangan mengenai cabang kegiatan hidup tertentu. Dalam artian ini adalah seperti: statistik perdagangan, statistik pertanian, statistik penduduk, statistik pendidikan, dan lain-lain. Kita ambil satu contoh seperti dalam istilah statistik pendidikan yaitu mengandung pengertian sebagai kumpulan bahan keterangan berupa angka yang kaitannya erat dengan kegiatan di bidang pendidikan atau dalam proses belajar mengajar. Misalnya: kumpulan bahan keterangan mengenai hasil belajar yang dicapai oleh peserta didik, seperti: kumpulan nilai hasil tes formatif, kumpulan nilai hasil sumatif dan sebagainya. Dalam hal ini istilah statistik dengan pengertian sebagai data statistik atau disebut juga data kuantitatif ialah data angka yang dapat memberikan gambaran mengenai keadaan, peristiwa maupun gejala tertentu.

Untuk pengertian kedua dalam istilah statistik diartikan sebagai *kegiatan statistik* ialah seperti yang termaktub dalam Undang-Undang No. 16 Tahun 1997 tentang Statistik yang memiliki

tiga dimensi yaitu (1) data atau informasi yang berupa angka; (2) sistem yang memadukan penyelenggaraan statistik; (3) ilmu yang mempelajari cara pengumpulan, pengolahan, penyajian dan analisis data (Depdiknas, 1997). Istilah statistik adalah pengertian kedua dapat di istilahkan sebagai "Departemen Pusat Statistik" yang mengandung pengertian memiliki tugas pokok atau spesifik dalam bidang tertentu dalam hal ini mencakup pengumpulan data, penyajian data, dan penganalisissan data.

Dalam pengertian ketiga, istilah statistik diartikan sebagai *metode statistik* yaitu cara yang ditempuh dalam menyusun, mengumpulkan, menyajikan dan menganalisis serta memberikan interpretasi terhadap sekumpulan bahan keterangan berupa angka bisa memberikan makna tertentu.

Selanjutnya untuk pengertian keempat istilah statistik diartikan sebagai *ilmu statistik*, yaitu ilmu pengetahuan yang mempelajari dan mengembangkan secara ilmiah tahapan-tahapan yang ada dalam kegiatan statistik. Dalam hal ini ilmu statistik ialah ilmu pengetahuan yng berhubungan dengan bagaimana cara mengumpulkan data, mengolah dan menganalisisnya dan menarik kesimpulan yang didasarkan atas kumpulan data dan analisis yang dilakukan (Sudjana, 2009). Adapun hal-hal yang perlu diperhatikan dalam mempelajari ilmu statistik ialah mengembangkan prinsip-prinsip, metode dan prosedur yang perlu dilalui atau digunakan dalam rangka mengumpulkan data angka, menyusun atau mengatur data angka, menyajikan atau menggambarkan data angka, menganalisis data angka, dan kemudian menarik kesimpulan, membuat perkiraan, serta menyusun ramalan secara ilmiah atas dasar dari pengumpulan data tersebut.

1.2.1 Penggolongan Statistik

Berdasarkan tahapan-tahapan dalam kegiatan statistik sebagai ilmu pengetahuan dibedakan menjadi dua, yaitu Statistik Inferensial dan Statistiktik Deskriptif (Sudijono, 2018). Statistik Inferensial ialah statistik yang menggunakan cara yang dapat digunakan sebagai alat dalam mengambil keputusan yang sifatnya umum dari kumpulan data yang sudah disusun atau di olah. Statistik Inferensial juga memiliki aturan tertentu dalam membuat

kesimpulan, menyusun dan membuat ramalan, taksiran dan semacamnya.

Selanjutnya untuk Stautistik Deskriptif ialah statistik yang cara kerjanya meliputi cara-cara dalam mengumpulkan, menyusun, mengolah dan menyajikan serta menganalisis data atau angka dengan tujuan memberikan gambaran yang rapi dan jelas tentang suatu keadaan atau peristiwa sehingga dapat ditarik maknanya atau pengertiannya dari suatu keadaan tersebut.

Dalam hal ini statistik deskriptif merupakan dasar dari ilmu statistik secara universal juga merupakan hal yang paling fundamental dalam struktur ilmu statistik, sedangkan statistik inferensial sifatnya lebih mendalam. Artinya untuk dapat memahami statistik inferensial, kita harus lebih dahulu memahami statistik deskriptif.

1.2.2 Ciri Khas Statistik

Umumnya dalam statistik yang fungsinya sebagai ilmu pengetahuan memiliki ciri yaitu, (1) statistik bekerja dengan angka; (2) Statistik bersifat objektif; dan (3) Statistik bersifat universal.

Statistik bekerja dengan angka atau bilangan, berarti dalam melaksanakan tugasnya statistik memerlukan bahan berupa keterangan yang bersifat kuantitatif. Apabila statistika digunakan untuk menganalisis data yang bersifat kualitatif yaitu bahan keterangan yang tidak berbentul angka atau bilangan) maka sebagai langkah awalnya data kualitatif tersebut harus dikonversi terlebih dahulu menjadi data kuantitatif atau disebut juga dengan istilah proses kuantifikasi. Sebagai contoh: diberikan sebuah angket dengan beberapa pertanyaan, dimana setiap pertanyaan di labeli dengan “Sangat Pintar”, “Pintar”, “Cukup” dan “Kurang” dalam hal ini ialah bahan keterangan yang sifatnya kualitatif tentang prestasi belajar peserta didik. Agar bisa dianalisis secara statistik, data kualitatif tersebut harus dikonversi menjadi data kuantitatif. Kita ambil contoh, peserta didik dikatakan “ Sangat Pintar” jika peserta didik mendapat nilai “91-100”, “Pintar” jika peserta didik mendapat nilai “81-90”, “Cukup Pintar” jika peserta didik mendapat nilai “61-80” dan “Kurang” jika peserta didik mendapat nilai “0-60” atau peserta didik dengan kriteria “sangat pintar” berjumlah 4 orang,

peserta didik dengan kriteria "pintar" berjumlah 18 orang, peserta didik dengan kriteria "cukup pintar" berjumlah 4 orang, peserta didik dengan kriteria "kurang pintar" berjumlah 2 orang.

Statistik bersifat Objektif atau sering disebut dengan alat penilai secara nyata artinya statistik bekerja sesuai dengan fakta yang terjadi dilapangan. Dalam hal ini, kesimpulan yang dihasilkan dan ramalan yang dinyatakan oleh statistik sebagai bagian dari ilmu pengetahuan didasari oleh data angka yang ditemukan dan diolah berdasarkan kondisi faktual dilapangan bukan berdasarkan pada pengaruh luar atau subjektivitas.

Statistik bersifat universal. Dalam hal ini garapan dari statistik ini luas dan menyeluruh, artinya statistik bisa digunakan hampir dalam semua cabang ilmu dan kegiatan hidup manusia. Misalnya dalam bidang kependudukan kita kenal adanya statistik kelahiran, statistik kematian, statistik nikah, statistik talak, statistik cerai dan sebagainya. Dalam bidang ekonomi misalnya dikenal dengan statistik perdagangan, statistik pertanian dan sebagainya. Kemudian contoh lainnya adanya statistik kecelakaan lalu lintas, statistik pendidikan, statistik kriminalitas dan sebagainya.

1.2.3 Permasalahan Statistik

Anas Sudijono dalam bukunya *Pengantar Statistik Pendidikan* (Sudijono, 2018) mengemukakan bahwasannya ada tiga permasalahan dasar dalam statistik, yakni (1) permasalahan tentang rata-rata, (2) permasalahan tentang penyebaran, dan (3) permasalahan tentang saling-hubungan/korelasi.

Permasalahan tentang rata-rata *(average)* sering kita temui dalam kehidupan sehari-hari. Kita ambil contoh yaitu seorang guru mengambil nilai rata-rata yang diperoleh siswa dengan tujuan untuk mengetahui kualitas dari siswanya, seorang sarjana ekonomi mencari rata-rata pendapatan per KK dalam suatu wilayah untuk menentukan tingkat perekonomian dalam wilayah tersebut, suatu lembaga pendidikan mencari nilai rata-rata assesmen kompetensi minimum tingkat SD/MI disuatu provinsi. Pada dasarnya setiap orang, baik itu lembaga maupun perseorangan secara sadar maupun tidak sadar telah berpikir dan menggunakan statistik baik

digunakan untuk tujuan yang sederhana maupun untuk tujuan yang kompleks.

Permasalahan lainnya yakni mengenai penyebaran atau lebih dikenal dengan variabilitas. Dalam kata bahasa Indonesia kita lebih mengenal dengan kata "variasi" yang berarti "keanekaragaman". Dalam kehidupan sehari-hari akan menyenangkan jika banyak sesuatu yang banyak variasinya sehingga tidak membosankan, namun dalam statistik kita mengusahakan agar sesuatu itu tidak banyak variasinya, agar variabilitasnya kecil. Karena jika ukuran variabilitasnya kecil akan menunjukkan kualitas yang tinggi. Sebagai contoh, seorang guru menyatakan tingkat kepintaran siswa kelas A lebih merata atau homogen dari pada siswa kelas B. Artinya murid pada kelas B memiliki perbedaan kepintaran antara satu dengan yang lainnya lebih tajam daripada antar siswa dalam kelas A. Contoh lain ialah, seorang produsen bola lampu listrik berharap kualitas bola lampu listrik yang diproduksinya seragam, yaitu memiliki ketahanan/umur yang sama antara bola lampu yang satu dengan yang lainnya atau tidak ada perbedaan yang signifikan antara umur bola lampu listrik yang satu dengan yang lainnya.

Permasalahan yang selanjutnya ialah persoalan tentang korelasi atau hubungan. Sebagai contoh, seorang guru menyatakan bahawa jika siswa pintar dalam matematika maka akan pintar dalam ilmu fisika dikarenakan dalam ilmu fisika juga ada hal yang sifatnya menghitung, contoh lainnya yaitu seorang produsen menyatakan bahwa jika inflasi makin meningkat, maka akan banyak perusahaan yang gulung tikar, kementerian kesehata menyatakan jika protokol kesehatan diabaikan, maka kasus Covid-19 akan meningkat, dan lainnya.

Dalam tiga persoalan statistik diatas mengenai, rata-rata, variabilitas dan korelasi merupakan soalan yang mendasar dalam statistik dan sudah sering ditemui dalam kehidupan. Dalam persoalan tersebut, dapat dinyatakan dalam besaran bilangan, dan dapat juga dianalisis lebih lanjut dan akan dibahas dalam pembahasan selanjutnya.

1.3 Probabilitas Statistika

Supranto dalam bukunya menyatakan bahwasannya probabilitas ialah peluang bahwa sesuatu akan terjadi. Dengan adanya suatu nilai yang digunakan untuk mengukur tingkat terjadinya suatu kejadian secara acak (Supranto, 2000). Dalam cabang matematika dikenal dengan istilah probabilitas, probabilitas jika didekati secara rumus matematika dan secara data statistika maka muncullah istilah probabilitas matematik dan probabilitas statistik. Probabilitas matematik ialah angka yang menyatakan kemungkinan terjadinya suatu kejadian sedangkan probabilitas statistik ialah data yang sudah dikumpulkan dari lapangan dan dianalisis menggunakan rumus matematika. Saat ini statistika yang kita kenal sekarang adalah perkembangan dari probabilitas statistika.

1.3.1 Fungsi, Kegunaan dan Ruang Lingkup Statistika

Adapun fungsi dari statistika ialah (a) mendeskripsikan data dalam bentuk tertentu; (b) penyederhanaan data yang rumit menjadi lebih simpel; (c) menentukan hubungan sebab akibat; (d) untuk menggambarkan suatu perbandingan; (e) mengukur besaran dari suatu gejala; (f) dapat memperluas pengalaman individu dengan mempelajari kesimpulan dari data penilaiannya.

Kegunaan dari statistika ialah (a) menganalisis data; dalam hal ini bertugas membantu penelitian dalam hal membaca data yang sudah dikumpulkan sehinnga bisa diambil suatu keputusan, kemudian membantu peneliti dalam menginterpretasi data yang sudah dikumpulkan; (b) membuat ramalan; dalam hal ini bertugas untuk membantu peneliti dalam memprediksi waktu yang akan datang. (c) menguji hipotesa; dalam hal ini membantu penelitian dalam menggunakan sampel sehingga bisa bekerja secara efisien hasilnya sesuai dengan objek yang ingin diteliti, kemudian dapat membantu peneliti dalam melihat terjadi/tidak terjadinya perbedaan antara kelompok yang satu dengan kelompok lainnya berdasarkan objek yang diteliti, juga membantu peneliti dalam melihat ada/tidaknya hubungan antara variabel yang satu dengan variabel lainnya. (d) rangkaian ilmu statistik; dalam hal ini digunakan sebagai bagian dari ilmu pengetahuan seperti pemerintah menggunakan statistika dalam melakukan

perbandingan terhadap hasil pembangunan, pendidik menggunakan statistika untuk melihat prestasi belajar siswa maupun metode pembelajarannya.

Sugiyono dalam bukunya menyatakan bahwa peranan statistika (Sugiyono, 2007) ialah: (a) sebagai alat penghitung besarnya anggota sampel yang diambil dari suatu populasi, sehingga jumlah sampel yang dibutuhkan akan lebih bisa dipertanggungjawabkan; (b) sebagai alat dalam menguji validitas dan realibilitas instrumen sebelum instrumen tersebut digunakan dalam penelitian; (c) sebagai teknik dalam menyajikan data, sehingga data lebih komunikatif. Seperti membuat tabel, grafik maupun diagram; (d) sebagai alat dalam menganalisis data yakni menguji hipotesis yang diajukan dalam penelitian. Jadi dapat disimpulkan bahwasannya statistika memiliki peranan dalam kehidupan yaitu:

1. Menawarkan pemberian teknik-teknik yang sederhana dalam pengkalsifikasian data dan penyajian agar dipahami dengan mudah
2. Membantu dalam menyimpulkan dari perbedaan yang didapat benar-benar signifkan
3. Digunakan dalam berbagai bidang ilmu, seperti ilmu ekonomi, astronomi, biologi, pendidikan, kedokteran, asuransi, farmasi, geologi dan sebagainya
4. Dalam teknik statistiknya dapat digunakan dalam menguji hipotesa
5. Memberikan informasi terkait karakteristik distribusi dalam populasi tertentu baik yang sifatnya diskrit maupun kontiniu. Hal ini berguna dalam melihat perilaku populasi yang sedang diamati.
6. Terdapat prosedural yang praktis untuk melaksanakan survei dalam pengumpulan data melalui teknik sampling, gunanya untuk mendapatkan hasil pengukuran yang terpercaya
7. Terdapat prosedural yang praktis dalam menduga karakteristik suatu populasi dengan menggunakan pendekatan karakteristik sampel dengan menggunakan metode penaksiran, pengujian hipotesis, analisis varians

dan sebagainya yang berguna untuk mengetahui ukuran pemusatan dan penyebaran data juga mengenai perbedaan dan kesamaan dalam populasi.

8. Terdapat prosedural praktis dalam meramalkan kejadian pada suatu objek tertentu pada masa yang akan datang dengan didasari dengan keadaan masa lampau dan masa sekarang. Dengan menggunakan metode regresi dan metode deret waktu pengetahuan ini berguna untuk memperkecil risiko dari akibat ketidakpastian yang nantinya akan dihadapu dimasa yang akan datang.
9. Terdapat prosedural praktis dalam melakukan pengujian yang bersifat kualitatif melalui statistik non parametrik

Ruang lingkup statistika ialah, (a) Ekonomi dan Bisnis; (b) Teknik dan Mekanika; (c) Sipil dan Perencanaan; (d) Sosial dan Budaya; (e) Pemerintahan; (d) Komputer dan Informasi; (e) Psikologi dan Komunikasi; dan (f) Matematika dan Pengetahuan Alam.

1.3.2 Bagian Ilmu Statistika, Metodologi dan Elemen Statistika

Dalam pembagian ilmu statistika bisa dibedakan menjadi, (1) Statistik deskriptif; membahas tentang penjelasan dari berbagai karakteristik dalam suatu data (2) Statistik Induktif-Inferensi; membahas tentang pernyataan dalam suatu populasi yang berdasarkan informasi dari sampel yang random yang diambil dari populasi itu. (3) Teori Probabilitas; membahas tentang suatu angka yangmenunjukkan tingkat keyakinan terjadinya suatu peristiwa dan (4) Analisis Keputusan; membahas tentang pengambilan keputusan jika alternatif tindakan diketahui, namun hasil dari tiap tindakan berbeda.

Metodologi dari statistika ialah, (1) mengidentifikasi persoalan-persoalan; (2) mengumpulkan fakta; (3) mengumpulkan data asli yang bersifat baharu; (4) mengkalsifikasi data; (4) menyajikan data; (5) menyajikan data; dan (6) menganalisis data

Elemen- elemen dari statistika terdiri dari: (1) populasi; (2) sampel; dan (3) variabel; (4) statistik inferensi; (5) pengukuran reabilitas dan statistik inferensi; populasi merupakan kumpulan

data yang mengidentifikasi suatu fenomena atau gejala. Populasi mempengaruhi pada kegunaan dan hubungan maupun relevansi dari data yang dikumpulkan, contohnya: "semua mahasiswa di Medan", "semua pekerja di seluruh Indonesia". Sedangkan sampel merupakan kumpulan data yang diambil dari suatu populasi. Sampel merupakan bagian dari populasi. Contohnya: populasi: "semua mahasiswa di Medan" Sampel: "mahasiswa semester VIII Program Studi Pendidikan Matematika". Selanjutnya, variabel merupakan sebuah simbol yang menggandeng setiap nilai dari suatu himpunan nilai yang disebut dengan domain dari variabel. Dalam melakukan kesimpulan terhadap populasi, tidak semua ciri populasi harus diketahui, ambil satu atau beberapa karakteristik populasi saja yang perlu diketahui itulah yang disebut sebagai variabel. Dalam variabel dikenal dengan variabel kontinu dan diskrit. Variabel kontinu diperoleh dari hasil pengukuran, contohnya tinggi badan, suhu badan, dan berat badan. Sedangkan variabel diskrit diperoleh dari hasil perhitungan, contohnya jumlah mahasiswa yang mengambil mata kuliah statistik dasar. Statistik inferensi ialah suatu keputusan, perkiraan ataupun menggeneralisasikan mengenai suatu populasi yang didasarkan dengan informasi yang terkandung dari sampel. Selanjutnya untuk menganalisa statistik yang diambil dari data sampel dalam suatu populasi, maka digunakanlah pengukuran reabilitas dari setiap kesimpulan yang akan dibuat.

1.3.3 Pengumpulan dan Tipe Data Statistika

Menurut Kusmayadi (Kusmayadi, 2004) dalam bukunya menyatakan sebagai suatu bidang ilmu statistika dibagi atas empat bagian, yaitu statistika deskriptif, probabilitas, analisis pengambilan keputusan dan statistika inferensia. Kemudian berdasarkan tahapannya statistika dibagi menjadi dua tahap yaitu statistika deskriptif dan statistika induktif.

Statistika juga dapat diklasifikasikan menjadi beberapa kelompok yaitu (1) pengelompokan statistika berdasarkan isi yang dipelajari terbagi menjadi dua yaitu (a) statistik teoritis dan (b) statistik terapan. Untuk statistik teoritis pembahasannya lebih mendalam danmatematis, dalam mempelajari statistik ini

diperlukan dasar matematika yang kuat dan mendalam. Materi yang dibahas pada bagaian ini ialah perumusan sifat-sifat, dalil-dalil, rumus-rumus dan menciptakan model-model segi-segi lainnya yang teoritis dan matematis. Sedangkan untuk statistika terapan atau yang sering kita kenal dengan metode statistika. Aturan-aturan, sifat-sifat, dalil-dalil, rumus-rumusyang telah diciptakan oleh statistik teoritis diambil dan digunakan pada bagian yang diperlukan dalam bidang pengetahuan yang sedang digeluti. Pada statistika terapan tidak dipermasalahkan dari mana didapatkannya sifat-sifat, dalil-dalil, rumus-rumus tersebut, yang penting ialah penggunaan dari metode atau cara-cara yang digunakan dalam metode statistika.

Pengelompokan statistika berdasarkan aktivitas yang dilakukan terbagi atas (a) statistika deskriptif; (b) statistika inferensial. Menurut Burhan (Nurgiyantoro, et al., 2000) bahwasannya statistika deskriptif digunakan untuk menyajikan dan menganalisis data agar lebih bermakna dan komunikatif yang disertai dengan perhitungan yang sederhana yang sifatnya memperjelas suatu keadaan. Sedangkan statistika inferensial menurut Subana (Subana, et al., 2000) merupakan statistik yang berhubungan dengan penarikan kesimpulan yang bersifat umum dari data yang telah disusun dan diolah. Untuk bagian lebih mendalamnya akan dibahas pada bab-bab selanjutnya.

DAFTAR PUSTAKA

Depdiknas, 1997. Undang-Undang Republik Indonesia No. 16 Tahun 1997 tentang Statistik. s.l.:s.n.

Kusmayadi, 2004. *Statistika Pariwisata Deskriptif.* Jakarta: Gramedia Pustaka Utama.

Nurgiyantoro, B., Gunawan & Marzuki, 2000. *Statistik terapan untuk penelitian ilmu-ilmu sosial.* Yogyakarta: Gadjah Mada University Press.

Subana, Rahadi, M. & Sudrajat, 2000. *Statistik Pendidikan.* Bandung: Pustaka Setia.

Sudijono, A., 2018. *Pengantar Statistik Pendidikan.* Depok: Rajawali Press.

Sudjana, 2009. *Metode Statistika.* Bandung: Tarsito.

Sugiyono, 2007. *Statistika Untuk Penelitian.* Bandung: Alfabeta.

Supranto, 2000. *Statistik: Teori dan Aplikasi.* 6 ed. Jakarta: Erlangga.

BAB 2
PENGUMPULAN DAN PENYAJIAN DATA

Oleh Arief Aulia Rahman

Pengumpulan dan Penyajian data merupakan penyusunan dari data kasar menjadi data kelompok yang lebih informatif dan lebih mudah untuk dilihat. Pada dasarnya, penyajian data haruslah bersifat komunikatif, rinci dan sederhana, dalam arti data yang disajikan dapat dipahami pembaca. Penyajian data dapat dilakukan dengan menggunakan tabel, grafik dan diagram lingkaran.

2.1 Tabel

Penggunaan tabel sebagai penyaji data sangatlah efisien dan cukup sering digunakan, tabel haruslah diberi judul, keterangan pada tiap kolom dan baris serta sumber data yang diambil, tabel juga digunakan sebagai dasar dalam pembuatan grafik. Ada beberapa macam jenis tabel, yaitu (1) tabel biasa yang terdiri dari tabel data nominal, tabel data ordinal dan tabel data interval dan (2) tabel distribusi frekuensi.

1. Tabel Data Nominal

   Data jumlah mahasiswa pada program studi pendidikan matematika pada tiap tahun telah didata oleh biro akademik. Hasilnya didapat bahwa :

   a. Pada Tahun 2013 jumlah mahasiswa prodi pendidikan Matematika 65 orang dengan rincian laki-laki 15 orang dan perempuan 40 orang
   b. Pada Tahun 2014 jumlah mahasiswa prodi pendidikan Matematika 31 orang dengan rincian laki-laki 7 orang dan perempuan 12 orang
   c. Pada Tahun 2015 jumlah mahasiswa prodi pendidikan Matematika 17 orang dengan rincian laki-laki 5 orang

dan perempuan 12 orang

d. Pada Tahun 2016 jumlah mahasiswa prodi pendidikan Matematika 9 orang dengan rincian laki-laki 1 orang dan perempuan 8 orang

Buatlah tabel data nominal untuk data diatas !
Berdasarkan data diatas, dapat disusun tabel seperti berikut !

Tabel 2.1. Jumlah Mahasiswa Prodi Pendidikan Matematika

Tahun	Jenis Kelamin		Total
	Laki-laki	Perempuan	
2013	15	40	65
2014	7	24	31
2015	5	12	17
2016	1	8	9

Sumber : Biro Administrasi X

2. Tabel Data Ordinal

Data ordinal merupakan data berbentuk peringkat yang dibuat dalam persentase, seperti data produksi hasil pertanian pada tahun 2007 dipulau kalimantan, didapat data bahwa :

Tabel 2.2. Tingkat Produksi Hasil Pertanian tahun 2007 di Kalimantan

Hasil Pertanian	Banyaknya Produksi	
	Percentase	Peringkat
Kubis	14,78 %	1
Bawang Merah	14,68 %	2
Sawi	5,74 %	3
Bawang Putih	4,97 %	4
Bawang Daun	0,65 %	5
Kacang Merah	0,15 %	6

Sumber : newspaper

Berdasarkan tabel diatas, hasil produksi kubis pada tahun 2007 mendapat peringkat 1 dengan total percentase 14,78 %.

3. Tabel Data Interval

Data interval merupakan kumpulan data yang didapat dari hasil angket dengan skala likert, yaitu skala interval 1 s/d 4 dimana skor 1 berarti sangat tidak puas, 2 tidak puas, 3 puas, 4 sangat puas. Seperti contoh data tingkat kepuasan mahasiswa terhadap layanan akademik, disebarkan angket skala likert kepada 100 mahasiswa tentang layanan :
(1) proses pembelajaran, (2) kemahasiswaan, (3) perpustakaan, (4) administrasi akademik, dan (5) keuangan, berikut data disajikan dalam bentuk tabel data interval.

Tabel 2.3. Tingkat Kepuasan Mahasiswa terhadap layanan akademik.

Aspek yang diamati	**Tingkat Kepuasan**
Proses Pembelajaran	75,43
Kemahasiswaan	46,75
Perpustakaan	82,41
Administrasi Akademik	65,43
Keuangan	75,76

Berdasarkan tabel diatas tingkat kepuasan mahasiswa terhadap layanan akademik paling tinggi yaitu terletak pada perpustakaan dengan skor 82,41.

4. Tabel Distribusi Frekuensi

Tabel distribusi frekuensi merupakan salah satu bentuk penyajian data statistik yang dibuat agar memudahkan pembaca ketika data statistik yang disajikan dalam jumlah yang sangat banyak. Tabel distribusi frekuensi dapat menyederhanakan bentuk dan jumlah data sehingga lebih komunikatif.

Dalam tabel distribusi frekuensi, terdapat beberapa istilah yang perlu kita ketahui. Yaitu :

a. Tabel distribusi frekuensi memiliki **Rentang (R)**. Yaitu selisih antara data terbesar dan data terkecil. Rentang dapat dihitung dengan cara ;

$$R = X_{maks} - X_{min}$$

Dimana :

R	=	Rentang
X_{maks}	=	Data dengan nilai tertinggi
X_{min}	=	Data dengan nilai terendah

b. Tabel distribusi memiliki **Kelas (K),** yaitu banyaknya baris yang dapat dibuat untuk mempresentasikan data**.** Rumus yang digunakan untuk menghitung banyak kelas yaitu dengan menggunakan aturan sturges :

$$K = 1 + 3{,}3 \log n$$

Dimana :

K	=	Interval Kelas
n	=	Banyak data

c. Tabel distribusi memiliki **Panjang Kelas/Interval kelas,** yaitu jumlah interval pada tiap kelompok kelas. Rumus untuk menghitung panjang kelas/interval kelas yaitu :

$$P = R/K$$

Dimana :

P = Panjang Kelas R = Rentang Kelas K = Interval Kelas

Contoh 2.1 Terdapat 60 data nilai prestasi siswa dalam mata pelajaran matematika di sekolah A. Tentukan kategori prestasi siswa tersebut (tinggi, rendah, sedang) terhadap mata pelajaran matematika.

52	56	62	48	93	88
42	53	61	61	71	64
53	51	58	63	71	57
58	63	88	62	67	56
56	47	63	78	67	53
33	80	45	55	37	42
50	42	56	58	67	22
28	56	31	71	50	25
50	41	35	79	69	46
47	26	47	51	67	42

Langkah awal yang diperlukan untuk menyusun data diatas menjadi tabel distribusi frekuensi adalah sebagai berikut :

1. Menghitung Rentang data (R)
   Menghitung selisih antara data tertinggi dan data terendah yaitu : 93 – 22 = 71
2. Menghitung Interval Kelas (K) K = 1 + 3,3 Log n
   K = 1 + 3,3 Log 60
   K = 1 + 3,3 (1,8)
   K = 6,94 ≈ 7, jadi interval kelas pada data diatas adalah 7
3. Menghitung Panjang Kelas (P)
   Panjang kelas didapat melalui hasil bagi antara rentang kelas dan interval kelas yaitu 71 : 7 = 10,14 ≈ 10 atau 11, maka diperoleh panjang kelas data diatas adalah 11
4. Menyusun Tabel Distribusi Frekuensi
   Setelah rentang, internal, dan panjang kelas, maka disusunlah tabel distribusi frekuensi, dimulai dari data terkecil yaitu 22.

Tabel 2.4. Tabel Distribusi Frekuensi dengan Turus

Kelas	Interval Kelas	Turus	Frekuensi
1	22-28	llll	4
2	29-35	lll	3
3	36-42	llll l	6
Kelas	**Interval Kelas**	**Turus**	**Frekuensi**
4	43-49	llll l	6
5	50-56	llll llll llll	15
6	57-63	llll llll l	11
7	64-70	llll l	6
8	71-77	lll	3
9	78-84	lll	3
10	85-91	ll	2
11	92-98	l	1
Total			60

5. Menghitung frekuensi data pada setiap interval kelas, dengan menggunakan turus agar proses pendataan frekuensi pada data yang disajikan jadi lebih mudah.
6. Penggunaan turus mempermudah pendataan frekuensi data dengan cara memberi tanda (X) pada setiap data yang telah dimasukkan kedalam tabel distribusi frekuensi, misalnya pada data no 1 yaitu 52, maka data 52 diberi tanda (X) kemudian masukkan pada kelas no 5 dengan interval kelas 50-56 menggunakan turus. Kemudian data berikutnya adalah 42, maka data 42 diberi tanda (X) kemudian masukkan pada kelas no 3 dengan interval kelas 36-42 menggunakan turus.hingga semua data tersilang dan masuk kedalam tabel distribusi frekuensi.

2.2 Grafik

Grafik merupakan cara penyajian data yang sangat sering digunakan para ahli statistika untuk menggambarkan suatu keadaan objek-objek tertentu secara visual. Grafik merupakan kelanjutan dari penyusunan tabel distribusi frekuensi karena grafik dibuat berdasarkan keadaan tabel distribusi frekuensi. Secara umum grafik

terbagi menjadi 3 (tiga) macam, yaitu grafik garis (*Polygon*) , grafik batang (*histogram*) dan grafik lingkaran (*Piechart*) serta dapat dikembangkan lagi kedalam bentuk-bentuk yang menarik dan mudah dilihat.

1. Grafik Batang (*Histogram*)

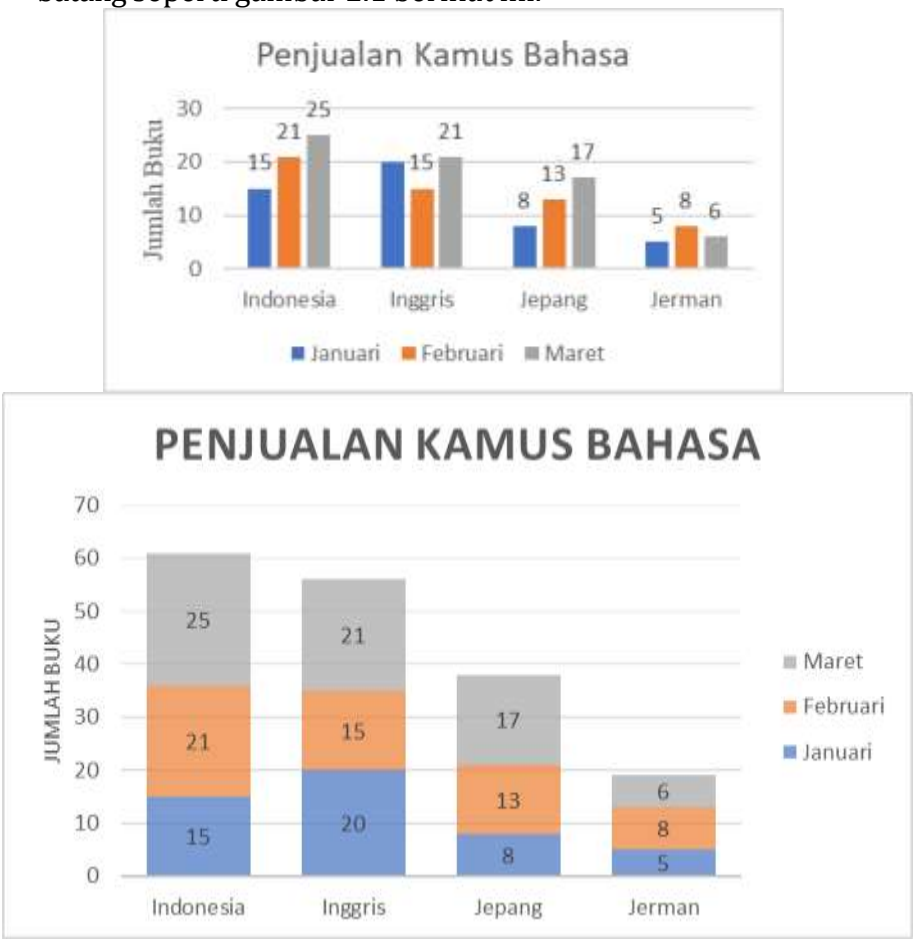
   Grafik batang merupakan cara penyajian data dalam bentuk batang-batang panjang segiempat yang berada pada sumbu x dan y. Penyajian diagram batang sudah mengalami banyak pengembangan dalam tampilan visualnya dari 2D menjadi 3D.

   Grafik batang pada gambar 2.2 menunjukkan data penjualan kamus bahasa dari bulan Januari hingga Maret di 4 (empat) negara yaitu Indonesia, Inggris, Jepang dan Jerman. Dalam gambar tersebut dapat dilihat grafik penurunan atau peningkatan penjualan kamus bahasa. Seperti halnya di Indonesia dan Jepang, penjualan kamus bahasa meningkat dari bulan Januari hingga Maret, sementara di Inggris mengalami penurunan dari Januari ke Februari serta di jerman, penjualan kamus bahasa juga mengalami penurunan dari februari ke maret, namun mengalami peningkatan dari januari ke februari. Maka dari itu penggunaan grafik sangat membantu dalam menyajikan data agar lebih mudah dilihat dan disimpulkan. Berikut data penjualan buku yang disajikan dalam bentuk tabel 2.5 kemudian diubah kedalam grafik batang.

Tabel 2.5. Tabel Penjualan Kamus

Negara penjual Kamus Bahasa	**Bulan Penjualan**		
	Januari	**Februari**	**Maret**
Indonesia	15	21	25
Inggris	20	15	21
Jepang	8	13	17
Jerman	5	8	6

Tabel Penjualan Kamus diatas kemudian diubah kedalam grafik batang seperti gambar 2.1 berikut ini.



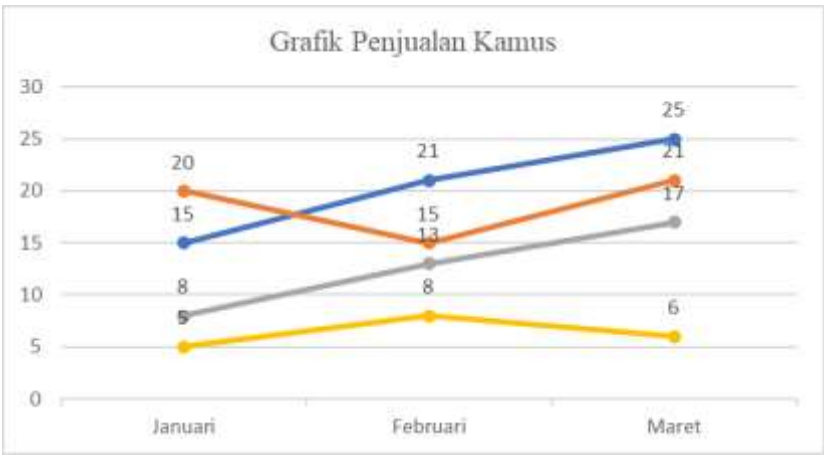
Gambar 2.1. Grafik penjualan kamus bahasa

Bentuk-bentuk penyajian data dalam grafik batang yang telah dikembangkan menjadi grafik batang balok yang beragam bentuk hingga menarik, grafik diatas dapat diubah menjadi berbagai bentuk lain seperti berikut :

Gambar 2.2. Macam-macam bentuk grafik batang (Histogram)

2. Grafik Garis (*Polygon*)

Grafik garis merupakan gambaran visual suatu kondisi data yang disajikan dalam bentuk garis yang naik turun, grafik garis sering digunakan untuk menunjukkan perubahan data dari masa ke masa. Grafik garis merupakan pengembangan dari grafik batang (*Histogram*), seperti contoh pada tabel 2.5 yang diubah menjadi grafik garis (*Polygon*) sebagai berikut.



Gambar 2.3. Grafik Garis Penjualan Kamus

3. Grafik Lingkaran (*Piechart*)

Grafik lingkaran atau *Piechart* merupakan cara menyajikan data dengan menampilkan visual berupa lingkaran yang digunakan sebagai pembanding antara kelompok-kelompok data, diagram lingkaran juga digunakan untuk menggambarkan sebuah persentase (%) dari total keseluruhan data, secara umum ada 3 bentuk grafik lingkaran yang sering dibuat kedalam soal-soal matematika, yaitu : 1) grafik lingkaran biasa, 2) grafik lingkaran derajat (^{0}), dan 3) grafik lingkaran persen (%).

Pada setiap bentuk lingkaran diatas, memiliki cara perhitungan yang berbeda-beda sesuai dengan pertanyaan yang diberikan. Berikut akan dibahas 3 bentuk grafik lingkaran tersebut.

a. Grafik Lingkaran Biasa

Grafik lingkaran biasa ditampilkan dalam bentuk angka yang menyatakan jumlah dari suatu data. Adapun cara membuat grafik lingkaran biasa menggunakan rumus sebagai berikut :

Langkah-langkah yang dilakukan dalam membuat grafik lingkaran biasa adalah :

1) Tentukan jumlah seluruh data yang disajikan dalam soal
2) Temukan besar sudut yang mewakili masing-masing data dengan rumus :

$$Jumlah\ Data = Total\ jumlah\ data - total\ data$$

$$Besar\ Sudut = \frac{Banyak\ data}{Total\ Data\ Keseluruhan}\ x\ 360^0$$

3) Tuliskan keterangan dalam setiap daerah lingkaran dengan jumlah masing-masing data.

Contoh 2.2 Dalam suatu kelas terdapat siswa-siswi yang menggemari satu jenis kegiatan ekstrakulikuler, setelah dikumpulkan data diperoleh 15 orang siswa gemar basket, 17 siswa gemar voli, 24 siswa gemar futsal dan 4 orang gemar pramuka. Sajikan dalam bentuk grafik lingkaran.

kegemaran	Jumlah siswa peminat
Basket	15
Voli	17
Futsal	24
Pramuka	4

Penjelasan :
Jumlah siswa dari tabel diatas adalah 60 siswa, maka besar masing- masing daerah adalah :

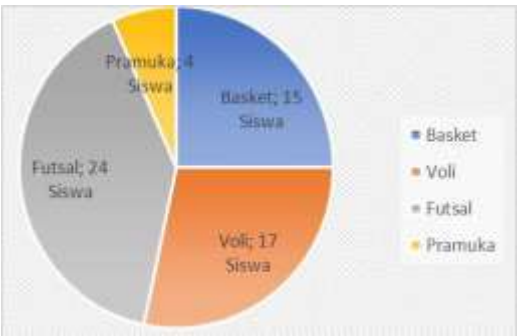
$\text{Basket} = \frac{15}{60} \times 360^0 = 90^\circ$

$\text{Voli} = \frac{17}{60} \times 360^0 = 102^\circ$

$\text{Futsal} = \frac{24}{60} \times 360^0 = 144^\circ$

$\text{Pramuka} = \frac{4}{60} \times 360^0 = 24^\circ$

Selanjutnya, besar masing-masing daerah tersebut dibuat kedalam bentuk grafik lingkaran dengan menggunakan busur derajat seperti berikut :



Gambar 2.4. Grafik Lingkaran Ekstrakulikuler Siswa

b. Grafik Lingkaran Derajat

Grafik lingkaran derajat menampilkan derajat busur atau besar sudut dalam grafiknya, bukan jumlah data. Cara menyusun grafik lingkaran derajat hampir sama dengan grafik lingkaran biasa, seperti contoh sebagai berikut :

Contoh 2.3 Survey kepada 100 orang di Aceh Barat terhadap makanan instan diperoleh data sebagai berikut, 45 orang menyukai Mie gelas, 35 orang menyukai Kentang Goreng dan 20 orang menyukai Roti Lapis. Buatlah Grafik Lingkaran Derajatnya.

Penjelasan :

Cara menyusun grafik lingkaran derajat sama dengan menyusun grafik lingkaran biasa, karena jumlah data sudah diketahui adalah 100, maka langkah selanjutnya adalah menemukan besar sudut setiap daerah, yaitu sebagai berikut :

$$\text{Mie Gelas} = \frac{45}{100} \times 360^0 = 162^\circ$$

$$\text{Kentang Goreng} = \frac{35}{100} \times 360^0$$

$$\text{Roti Lapis} = \frac{20}{100} \times 360^0 = 72^\circ$$

$$= 126^\circ$$

Setelah didapat besar sudut tiap-tiap daerah maka dapat disusun grafik lingkaran derajatnya dengan mengunakan busur derajat, perbedaannya dengan grafik lingkaran biasa adalah keterangan pada gambar lingkaran menggunakan derajat, seperti gambar 2.5 berikut

Gambar 2.5. Grafik Lingkaran derajat tentang data makanan

c. Grafik Lingkaran Persen

Grafik lingkaran persen adalah grafik lingkaran yang menggunakan data persen (%) dalam setiap keterangannya. Ada sedikit perbedaan antara grafik lingkaran biasa dan derajat dengan grafik lingkaran persen, yaitu pada langkah-langkah penyusunannya. Grafik lingkaran persen menggunakan rumus perubahan data menjadi persen (%) seperti hal berikut ini :

$$persen\ (\%) = \frac{Banyak\ data}{Total\ Data\ Keseluruhan}\ x\ 100\%$$

Untuk lebih jelasnya, perhatikan contoh 2.4 berikut ini.
Contoh 2.4

Hobi	Jumlah Siswa
Sepak bola	300
Basket	150
Voli	200
Bulu tangkis	250
Karate	100
Lainnya	200
Total	**1200**

Penjelasan :

Dari data diatas dapat dilihat sejumlah data hobi siswa dan jumlah siswa yang menyukai olahraga tersebut. Maka dari itu akan kita sajikan dalam bentuk grafik lingkaran persen sebagai berikut :

Jumlah data keseluruhan 1200 siswa, kemudian tentukan besar sudut tiap daerah.

$$\text{Sepak Bola} = \frac{300}{1200} \text{ x } 360^0 = 90^\circ$$
$$\text{Basket} = \frac{150}{1200} \text{ x } 360^0 = 45^\circ$$
$$\text{Voli} = \frac{200}{1200} \text{ x } 360^0 = 60^\circ$$
$$\text{Bulu Tangkis} = \frac{250}{1200} \text{ x } 360^0 = 75^\circ$$
$$\text{Karate} = \frac{100}{1200} \text{ x } 360^0 = 30^\circ$$
$$\text{lainnya} = \frac{200}{1200} \text{ x } 360^0 = 60^\circ$$

Kemudian setelah selesai mencari besar sudut tiap daerah, maka langkah selanjutnya adalah mengubah data menjadi bentuk persen.

$$\text{Sepak Bola} = \frac{300}{1200} \text{ x } 100\% = 25\%$$
$$\text{Basket} = \frac{150}{1200} \text{ x } 100\% = 12{,}5\%$$
$$\text{Voli} = \frac{200}{1200} \text{ x } 100\% = 16{,}7\%$$
$$\text{Bulu Tangkis} = \frac{250}{1200} \text{ x } 100\% = 20{,}83\%$$

$$\text{Karate} = \frac{100}{1200} \text{ x } 100\% = 8{,}3\%$$

$$\text{lainnya} = \frac{200}{1200} \text{ x } 100\% = 16{,}7\%$$

Maka didapatlah grafik lingkaran persen seperti berikut.

Gambar 2.6. grafik lingkaran persen tentang data hobi olahraga siswa

DAFTAR PUSTAKA

Fitri, A. 2011. Pengembangan Perangkat Pembelajaran Statistika Dasar Bermuatan Pendidikan Karakter Dengan Metode Problem Based Learning. *Jpp*, *1*(2).

Hadjar, I. 2014. Dasar-Dasar Statistik Untuk Ilmu Pendidikan, Sosial, & Humaniora. *Semarang: Pustaka Zaman*.

Hidayati, T. 2020. Statistika Dasar Panduan Bagi Dosen dan Mahasiswa.

Nugroho, S. 2008. *Dasar Dasar Metode Statistika*. Grasindo.

Rasul, A., & Sonda, R. 2022. *Statistika Pendidikan Matematika*.

CV Kreator Cerdas Indonesia.

Rohana, R., Hartono, Y., & Purwoko, P. 2009. Penggunaan Peta Konsep dalam Pembelajaran Statistika Dasar di Program Studi Pendidikan Matematika FKIP Universitas PGRI Palembang. *Jurnal Pendidikan Matematika*, *3*(2).

Winarsunu, T. 2017. *Statistik dalam penelitian psikologi dan pendidikan* (Vol. 1). UMMPress.

BAB 3
PENGUKURAN PEMUSATAN DATA

Oleh Rini Yunita

3.1 Pendahuluan

Dalam menyajikan data kita memerlukan suatu ukuran yang dapat mewakili sekumpulan data yang dapat menunjukkan karakteristik data disebut dengan ukuran pemusatan data.

3.2 Pengukuran Pemusatan Data

Pengukuran pemusatan data merupakan salah satu bentuk penyajian data yaitu: mean, median dan modus dan ukuran letak yaitu (kuartil, desil, dan persentil). Pengukuran pemusatan data untuk melihat seberapa besar kecenderungan data memusat pada nilai tertentu.

3.3 Rata-Rata atau Mean

Nilai Rata-rata atau mean merupakan hasil bagi jumlah keseluruhan data dengan banyaknya data. Nillai rata-rata dapat diperoleh dari data populasi maupun data sampel. Baik berupa data tunggal maupun data berkelompok. Populasi merupakan himpunan dari keseluruhan objek yang akan diteliti dan memiliki sifat yang sama, sedangkan sampel merupakan bagian dari populasi.

Untuk data yang berasal dari populasi terhingga yang berukuran N, maka nilai rata-rata atau mean nya dapat dihitung dengan menggunakan rumus berikut :

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$

Keterangan :
μ= Mean untuk data populasi
x_i= data ke-i
N= banyak data
Contoh

Diketahui data hasil pemeriksaaan % ketidakmurnian 10 kaleng ikan tuna sebagai berikut :

Tabel 3.1. Populasi hasil pemeriksaaan % ketidakmurnian ikan tuna

Kaleng	1	2	3	4	5	6	7	8	9	10
% Ketidak murnian	1.8	2.1	1.7	1.6	0.9	2.7	1.8	2.5	1.9	2.1

$$\mu = \frac{1.8 + 2.1 + 1.7 + 1.6 + 0.9 + 2.7 + 1.8 + 2.5 + 1.9 + 2.1}{10}$$

$$\mu = \frac{19.1}{10} = 1.91\%$$

Maka nilai rata-rata % ketidakmurnian dari 10 kaleng ikan tuna adalah sebesar 1.91%.

Sedangkan untuk data yang berasal dari sampel terhingga yang berukuran N, maka nilai rata-rata atau mean nya dapat dihitung dengan menggunakan rumus berikut :

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

Keterangan
$\bar{X}$= Mean untuk data sampel
x_i= data ke-i
N= banyak data

Contoh
Diketahui data hasil pemeriksaaan % ketidakmurnian 6 kaleng sampel ikan tuna secara acak dari 10 kaleng sebagai berikut :

Tabel 3.2. Sampel hasil pemeriksaaan % ketidakmurnian ikan tuna

kaleng	1	2	3	4	5	6
% ketidakmurnian	1.8	2.1	1.7	1.6	0.9	2.7

$$\bar{X} = \frac{1.8 + 2.1 + 1.7 + 1.6 + 0.9 + 2.7}{6}$$

$$\bar{X} = \frac{10.8}{6} = 1.8\%$$

3.3.1 Mean untuk Data Berkelompok

Untuk data berkelompok, nilai rata-rata atau mean dapat dihitung dengan rumus berikut :

$$\bar{X} = \frac{\sum_{i=1}^{n} f \cdot x_i}{\sum f}$$

Keterangan
$\bar{X}$= Mean untuk data berkelompok
x_i= data ke-i
f= frekuensi data
n= banyak data

Contoh

Diketahui nilai matematika dari 50 siswa kelas X adalah sebagai berikut :

9	8	8	7	6	7	8	8	9	10
7	7	7	6	6	8	8	9	9	9
6	6	7	7	7	7	8	8	8	8
9	9	9	9	9	6	6	6	6	8
7	7	8	8	8	7	7	7	6	10

Untuk mencari nilai rata-rata atau mean dari nilai matematika 50 siswa tersebut dapat dilakukan dengan cara berikut :

- Distribusikan nilai-nilai tersebut kedalam tabel distribusi frekuensi data tunggal.

Tabel 3.3. Distribusi Frekuensi Data Tunggal

Nilai Data	Frekuensi data f	$f.x$
6	10	60
7	14	98
8	14	112
9	10	90
10	2	20
	$\sum f$=50	$\sum f \, . \, x$ =380

Dengan menggunakan rumus mean untuk data berkelompok Diperoleh

$$\bar{X} = \frac{60+98+112+90+20}{10+14+14+10+2} \quad \bar{X} = \frac{380}{50} = 7.6$$

Maka rata-rata nilai matematika dari 50 siswa adalah 7.6.

Contoh

Diketahui data nilai statistika mahasiswa seperti dalam tabel distribusi frekuensi data berkelompok berikut, hitunglah berapa nilai rata-rata dari nilai statistika mahasiswa tersebut.

Tabel 3.4. Distribusi Frekuensi Data Berkelompok

Interval Nilai	Frekuensi (f)
51-60	8
61-70	12
71-80	17
81-90	9
91-100	4
	50

Jawab :

Dalam distribusi frekuensi data berkelompok kita butuh nilai X yang merupakan nilai tengah dari masing-masing kelas interval. Nilai X dapat dicari dengan rumus $X = \frac{X_1+X_2}{2}$. Yaitu interval nilai yang pertama nilai $X = \frac{60+51}{2} = 55.5$ dan seterusnya.

Tabel 3.5. Cara menghitung rata-rata atau Mean Nilai Statistika Mahasiswa

Interval Nilai	Nilai Tengah (x)	Frekuensi (f)	f.x
51-60	55,5	8	444
61-70	65.5	12	786
71-80	75.5	17	1283.5
81-90	85.5	9	769.5
91-100	95.5	4	382
		$\sum f$ =50	$\sum f\,x$ =3665

$$\bar{X} = \frac{444+786+1283.5+769.5+382}{50}$$

$$\bar{X} = \frac{3665}{50} = 73.3$$

3.3.2 Rata-rata atau Mean Menggunakan Angka Terkaan

Rata-rata hitung juga dapat dicari dengan menggunakan angka terkaan, dengan rumus:

$$\bar{X} = MT + \left(\frac{\sum fx'}{n}\right) i$$

Langkah-langkahnya :

1. Menerka letak mean (boleh disembarang interval) dan menjadikan titik tengahnya sebagai nilai mean terkaan (MT)
2. Memberi status deviasi/penyimpangan (x') pada setiap interval, dengan ketentuan letak interval MT akan diberi harga 0, kemudian interval dengan nilai diatas MT diberi tanda + (plus) mulai dari +1 dst, dan interval dengan nilai dibawah MT diberi tanda – (minus) mulai dari -1 dst
3. Mengalikan frekuensi dengan nilai deviasi
4. Menghitung nilai interval (i)

Tabel 3.6. Cara menghitung rata-rata atau Mean Dengan Angka Terkaan

Interval nilai	x	f	x'	fx'
5 – 11	8	11	-2	-22
12 – 18	15	20	-1	-20
19 – 25	22	19	0	0
26 – 32	29	8	1	8
33 – 39	36	2	2	4
Jumlah		60		-30

Maka nilai rata-rata dengan angka terkaan adalah sebagai berikut :

$$\bar{X} = MT + \left(\frac{\sum fx'}{n}\right) i$$

$$\bar{X} = 22 + \left(\frac{-30}{60}\right) 7$$

$$\bar{X} = 22 - 3.5 = 18.5$$

3.3.3 Rata-Rata Hitung Atau Mean Dengan Metode Coding

Rata-rata hitung juga dapat dicari dengan menggunakan metode koding dengan rumus:

$$\bar{X} = x_0 + \left(\frac{\sum f_i \cdot d_i}{n}\right)$$

Dimana:
x_0 = titik tengah terbesar pada kelas interval
d_i = $x_i - x_0$
x_i = nilai tengah kelas interval ke i
f_i = frekuensi kelas interval ke-i

Tabel 3.7. Cara menghitung rata-rata atau Mean Dengan Metode Coding

Interval nilai	x_i	f_i	d_i	$f_i \cdot x_i$
5 – 11	8	11	-7	-77
12 – 18	15	20	0	0
19 – 25	22	19	7	133
26 – 32	29	8	14	112
33 – 39	36	2	21	42
Jumlah		60		210

$$\bar{X} = 15 + \left(\frac{210}{60}\right)$$

$$\bar{X} = 15 + 3.5 = 18.5$$

3.3.4 Rata-Rata Hitung Atau Mean Untuk Data Berbobot

Jika nilai data x_1 , x_2 , x_3, ... x_n masing-masing mempunyai bobot w_1 , w_2 , w_3 .... w_n , maka rata-rata hitungnya ditentukan dengan menggunakan rumus :

$$\bar{X} = \frac{\sum wx}{\sum w}$$

Contoh

Diketahui daftar nilai seorang mahasiswa pada setiap mata kuliah beserta SKS dalam semester pertama seperti dalam tabel berikut. Tentukan IPK mahasiswa tersebut.

Tabel 3.8. Cara menghitung rata-rata atau Mean untuk data berbobot

No	Matakuliah	SKS/ Bobot (w)	Nilai	
			Huruf	Angka (x)
1	Pancasila	2	A	4,00
2	Pendidikan Agama Islam	2	A	4,00
3	Kalkulus	4	D	1,00
4	Peng.Teknologi Informasi	3	C	2,00
5	Algoritma &Pemrograman I	3	A	4,00
6	Logika Matematika	3	B	3,00
7	Statistika & Probabilitas	4	A	4,00
8	Bahasa Inggris	2	B	3,00

Gunakan Rumus $\bar{X} = \frac{\sum wx}{\sum w}$

Diperoleh

$$\bar{X} = \frac{2(4)+2(4)+4(1)+3(2)+3(4)+3(3)+4(4)+2(3)}{2+2+4+3+3+3+4+2}$$

$$\bar{X} = \frac{8+8+4+6+12+9+16+6}{23}$$

$$\bar{X} = \frac{69}{23} = 3{,}00$$

Maka nilai rata-rata IPK mahasiswa adalah 3,00

3.4 Rata-rata Letak (Median)

Nilai pengamatan yang terletak ditengah-tengah data hasil pengamatan yang telah diurutkan dari yang terkecil sampai yang terbesar atau sebaliknya.

3.4.1 Median untuk data tidak berkelompok

Untuk *n* data ganjil

Untuk sebarang konstanta $k = \frac{n-1}{2}$, maka median =X_{k+1}

Untuk *n* data genap

Untuk sebarang konstanta$k = \frac{n}{2}$, maka median =$\frac{1}{2}(X_k + X_{k+1})$

Contoh

Tentukan median untuk data berikut :

70, 60, 80, 60, 90, 50, 40, 70, 60

Jawab:

Urutkan data dari yang terkecil sampai yang terbesar yaitu:

40, 50, 60, 60, 60, 70, 70, 80, 90

Karena jumlah data ada 9 (jumlah data ganjil)

$k = \frac{9-1}{2}$=4, maka median =X_{4+1}

median merupakan data ke -5 yaitu 60

3.4.2 Median untuk data kelompok (dalam distribusi frekuensi)

$$Med = B_b + \left(\frac{\frac{1}{2}n - f_{kb}}{f_d} \right) i$$

Med : Median

B_b :Batas bawah dari interval yang mengandung nilai median

f_{kb} : frekuensi kumulatif interval yang nilainya dibawah /sebelum interval yang mengandung nilai median

f_d : frekuensi interval yang mengandung nilai median

i : lebar interval

n : jumlah total frekuensi

Contoh :

Diketahui data interval nilai seperti pada tabel berikut :

Tabel 3.9. Data Interval nilai untuk data berkelompok

Interval nilai	f
3 – 7 8 – 12 13 – 17 18 – 22 23 – 27 28 – 32	3 6 3 4 2 5
Jumlah	23

Untuk menghitung median pada data berkelompok, tentukan terlebih dahulu nilai frekuensi kumulatif dari masing-masing kelas interval data.

Tabel 3.10. Cara menghitung Median pada data Bekelompok

Interval nilai	f	fk
3 – 7 8 – 12 13 – 17 18 – 22 23 – 27 28 – 32	3 6 3 4 2 5	3 9 12 16 18 23
Jumlah	23	

Langkah-langkah:

1. Tentukan besar nilai ½ n
2. Temukan letak nilai ½ n pada fk
3. Temukan batas bawah dari interval yang mengandung nilai median (B_b)
4. Temukan fkb, yaitu fk yang terletak dibawah/sebelum interval yang mengandung nilai median
5. Temukan fd, yaitu frekuensi interval yg mengandung nilai median
6. Tentukan lebar interval (i)

Diperoleh:

7. ½ n = ½ 23 = 11,5
8. Interval kelas 13-17
9. Bb = 12,5
10. fkb = 9
11. fd = 3
12. i = 5

sehingga nilai Median dapat dihitung dengan menggunakan rumus

$$Med = 12.5 + \left(\frac{11.5 - 9}{3}\right)5$$

$$Med = 16.67$$

3.5 Modus

Nilai modus digunakan untuk menyatakan gejala yang paling sering terjadi atau yang paling banyak muncul

Modus suatu kelompok data tidak selalu ada, namun suatu kelompok data juga bisa juga mempunyai lebih dari satu modus

Contoh :

- Kelompok data 3, 4, 4, 5, 6, 8, 8, 8, 9 mempunyai satu modus yaitu 8
- Kelompok data 3, 4, 4, 5, 6, 8, 8, 9, 10 mempunyai dua modus yaitu 4 dan 8

3.5.1 Modus untuk data berkelompok (distribusi frekuensi)

Dapat ditentukan dengan menggunakan rumus :

$$Mod = B_b + \left(\frac{b_1}{b_1 + b_2}\right)c$$

Mod : Modus

B_b : Batas bawah dari interval yang mengandung modus

b_1 : Selisih antar frekuensi kelas modus dengan frekuensi kelas dibawah/sebelum kelas modus

b_2 : Selisih antar frekuensi kelas modus dengan frekuensi kelas diatas/sesudah kelas modus

c : Lebar kelas

Contoh :

tentukan nilai modus dari data berikut:

Tabel 3.11. Data Interval nilai untuk data berkelompok

Interval Nilai	Frekuensi
112-120	4
121-129	5
130-138	8
139-147	12
148-156	5
157-165	4
166-174	2
Jumlah	40

Jawab

Pertama tentukan kelas modus yaitu 12 terletak pada interval nilai 139-147.

Tabel 3.12. Cara untuk menghitung modus pada data berkelompok

Interval Nilai	Frekuensi
112-120	4
121-129	5
130-138	8
139-147	12
148-156	5
157-165	4
166-174	2
Jumlah	40

Bb = 138,5

b1 = 12-8 = 4

b2 = 12-5 = 7

c = 9

maka nilai modus dapat dihitung dengan menggunakan rumus:

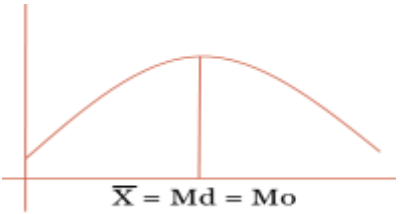
$$Mod = 138.5 + \left(\frac{4}{4+7}\right)9$$

$$Mod = 141{,}7$$

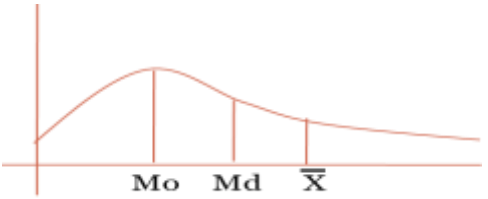
3.6 Hubungan antara mean, median dan modus

Hubungan antara mean, median dan modus ditentukan oleh bentuk dari kesimetrisan kurva distribusi data. Ada tiga kemungkinan kesimetrisan kurva, yaitu :

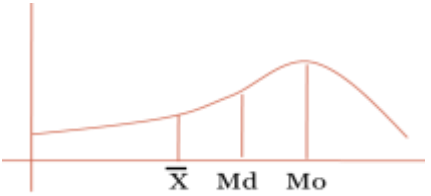
1. Jika nilai mean, median dan modus berdekatan (hampir sama) satu sama lain, maka kurva dari data tersebut akan mendekati simetri.
   Kurva simetris
   $\bar{X} = Md = Mo$



2. Jika nilai modus lebih kecil dari median dan median lebih kecil dari mean, maka kurva dari data tersebut akan miring ke kanan
   Kurva Miring ke Kanan
   $Mo < Md < \bar{X}$



3. Jika nilai mean lebih kecil dari median dan median lebih kecil dari modus, maka kurva dari data tersebut akan miring kekiri
   Kurva Miring Ke kiri
   $\bar{X} < Md < Mo$



3.7 Ukuran Letak

Merupakan sebuah ukuran yang menunjukkan pada bagian mana sebuah data terletak. Ukuran letak terdiri dari kuartil, desil dan persentil.

3.7.1 Kuartil

Kuartil adalah ukuran letak yang membagi 4 bagian yang sama. Dimana setiap bagian nialinya sebesar 25 % dari seluruh data, yaitu K1 sampai 25% data, K2 sampai 50% dan K3 sampai 75%.

1. **Untuk data tidak berkelompok**

   kuartil dapat dicari dengan rumus:

$$K_i = nilai\ ke\ \frac{i(n+1)}{4}$$

$$\text{dengan } i = 1,2,3, dan\ n = jumlah\ data$$

   Contoh

   Diberikan data upah bulanan karyawan (dalam ribuan rupiah) berikut ini , Tentukanlah nilai kuartil 1, 2 dan 3.

   30, 40,70, 80, 70, 40, 80, 40, 90, 100, 80, 30, 80, 120

   Jawab :

   data yang telah diurutkan adalah sebagai berikut :

   30, 30, 40, 40, 40, 70, 70, 80, 80, 80, 80, 90, 100, 120

   Maka nilai K1, K2 dan K3 adalah:

   K_1 = nilai ke - 1 (14+1) / 4 = 15/4 = nilai ke - 3 ¾

   $K_1 = x_3 + ¾ (x_4 - x_3) = 40 + ¾ (40 – 40) = 40$

   K_2 = nilai ke - 2 (14+1) / 4 = 30/4 = nilai ke – 7 ½

   $K_2 = x_7 + ½ (x_8 - x_7) = 70 + ½ (80 – 70) = 75$

   K_3 = nilai ke - 3 (14+1) / 4 = 45/4 = nilai ke – 11 ¼

   $K_3 = x_{11} + ¼ (x_{12} - x_{11}) = 80 + ¼ (90 – 80) = 82,5$

2. **Untuk data berkelompok**

$$K_i = B_b + c\left(\frac{\frac{in}{4} - f_{kb}}{fd}\right)$$

   K_i = Kuartil ke i

   n = Jumlah data

   B_b = Batas bawah pada interval yang mengandung kuartil

   f_{kb} = frekuensi kumulatif interval sebelum fk interval yang mengandung kuartil

f_d = frekuensi interval yang mengadung kuartil
c = lebar interval

Tabel 3.13. Cara untuk Menghitung K1, K2 dan K3 pada Data Berkelompok

Interval nilai	f	fk
3 – 7	3	3
8 – 12	6	9
13 – 17	3	12
18 – 22	4	16
23 – 27	2	18
28 – 32	5	23
Jumlah	23	

Untuk mencari nilai K_1
- → ¼ N = 23/4 = 5,75
- → Bb = 7,5
- → fkb = 3
- → fb = 6

Maka

$$K_1 = 7{,}5 + \left(\frac{5{,}75 - 3}{6}\right)5 = 9{,}79$$

Untuk mencari nilai K_2
- →1/2 N = 23/2 = 11,5
- → Bb = 12,5
- → fkb = 9
- → fb = 3

$$K_2 = 12{,}5 + \left(\frac{11{,}5 - 9}{3}\right)5 = 16{,}67$$

Untuk mencari nilai K_3
→3/4 N = 69/4 = 17,25
- → Bb = 22,5
- → fkb = 16
- → fb = 2

$\rightarrow\ K_3 = 22{,}5 + \left(\frac{17{,}25-16}{2}\right)5 = 25{,}62$

3.7.2 Desil

Desil adalah ukuran letak yang membagi data menjadi sepuluh bagian/kategori, yang masing-masing sebesar sepuluh persen.

1. **Untuk data tidak berkelompok**

   Desil dapat dicari dengan rumus:

$$D_i = nilai\ ke\ \frac{i(n+1)}{10}$$

   dengan $i = 1,2,3,\ldots,9\ dan\ n = jumlah\ data$

   Contoh:

   Tentukanlah nilai D1 dan D5 dari data upah bulanan karyawan (dalam ribuan rupiah) berikut ini :

   30, 40,70, 80, 70, 40, 80, 40, 90, 100, 80, 30, 80, 120

   Jawab :

   data yang telah diurutkan adalah sebagai berikut :

   30, 30, 40, 40, 40, 70, 70, 80, 80, 80, 80, 90, 100, 120

   Sehingga nilai D1, D5 adalah:

   D_1 = nilai ke - 1 (14+1) / 10 = 15/10 = nilai ke – 1½

   $D_1 = x_1 + ½ (x_3 - x_2) = 30 + ½ (40 – 30) = 35$

   D_5 = nilai ke - 5 (14+1) / 10 = 75/10 = nilai ke – 7 ½

   $D_5 = x_7 + ½ (x_8 - x_7) = 70 + ½ (80 – 70) = 75$

2. **Untuk data berkelompok**

$$D_i = L_0 + c\left(\frac{\frac{in}{10} - F}{f}\right)$$

   D_i = Desil ke -i

   i = 1, 2, 3...,9

   L_0 = Batas bawah kelas desil

   c = Lebar kelas

   F = jumlah frekuensi semua kelas sebelum kelas desil

   f = frekuensi

Tabel 3.14. Cara Untuk Menghitung D3 dan D9 pada Data Berkelompok

Interval nilai	f	fk
3 – 7	3	3
8 – 12	6	9
13 – 17	3	12
18 – 22	4	16
23 – 27	2	18
28 – 32	5	23
Jumlah	23	

Untuk mencari nilai D_3

→ 3/10 n = 3/10 x 23 = 6,9

→ L_0 = 7,5

→ F = 3

→ f = 6

maka

$$D_3 = 7{,}5 + 5\left(\frac{6{,}9 - 3}{6}\right) = 10{,}75$$

Untuk mencari D_3

→ 9/10 n = 9/10 x 23 = 20,7

→ L_0 = 27,5

→ F = 18

→ f = 5

$$D_3 = 27{,}5 + 5\left(\frac{20{,}7 - 18}{5}\right) = 30{,}2$$

3.7.3 Persentil

Persentil adalah ukuran letak yang membagi data menjadi seratus bagian/kategori, yang masing-masing sebesar satu persen.

1. **Untuk data tidak berkelompok**

   Persentil dapat dicari dengan rumus:

   $P_i = nilai\ ke\ \frac{i(n+1)}{100}$

   dengan $i = 1,2,3, \ldots ,99\ dan\ n = jumlah\ data$

Contoh:
Tentukanlah nilai P40 dan P60 dari data upah bulanan karyawan (dalam ribuan rupiah) berikut ini :
30, 40,70, 80, 70, 40, 80, 40, 90, 100, 80, 30, 80, 120
Jawab :
data yang telah diurutkan adalahsebagai berikut :
30, 30, 40, 40, 40, 70, 70, 80, 80, 80, 80, 90, 100, 120
Sehingga nilai P40, P60 adalah:
P40 = nilai ke - 40(14+1) / 100 = 600/100 = nilai ke –6
P40 = 40
P60 = nilai ke - 60 (14+1) / 100 = 900/10 = nilai ke – 9
P60 = 90

2. Untuk data berkelompok

$$P_i = L_0 + c\left(\frac{\frac{in}{100} - F}{f}\right)$$

P_i = Persentil ke -i
i = 1, 2, 3...,9
L_0 = Batas bawah kelas persentil
c = Lebar kelas
F = jumlah frekuensi semua kelas sebelum kelas persentil
f = frekuensi

Tabel 3.15. Cara untuk Menghitung P60 dan P75 pada Data Berkelompok

Interval nilai	f	fk
3 – 7	3	3
8 – 12	6	9
13 – 17	3	12
18 – 22	4	16
23 – 27	2	18
28 – 32	5	23
Jumlah	23	

Untuk mencari P60

→ 60/100 n = 60/100 x 23 = 13,8

→ $L_0 = 17,5$

→ F = 12

→ f = 4

$$P_{60} = 17,5 + 5\left(\frac{13,8 - 12}{4}\right) = 19,75$$

Untuk mencari P75

→ 75/100 N = 75/100 x 23 = 17,25

→ $L_0 = 22,5$

→ F = 16

→ f = 2

→ $P_{75} = 22,5 + 5\left(\frac{17,25-16}{2}\right) = 25,62$

3.8 Penutup

Ukuran pemusatan data merupakan nilai yang mencerminkan karakteristik sekumpulan data yang letaknya cenderung ditengah kelompok data. Nilai mean, median dan modus mempengaruhi terhadap bentuk kurva data, apakah simetris, miring kekiri atau miring ke kanan.

DAFTAR PUSTAKA

Ananda, Rusydi, M.F. 2018. *Statistik Pendidikan Teori dan Praktik dalam Pendidikan*. 1st edn. Edited by S. Saleh. Medan: CV Widia Puspita.

Asari, Andi, D. 2023. *Pengantar Statistika*. 1st edn. Edited by A. Asari. Solok Sumatera Barat: Mafy Media Literasi Indonesia anggota IKAPI.

Hamid, R.S. and Patra, I.K. 2019. *Pengantar Statistika untuk Bisnis dan Ekonomi Konsep Dasar dan Aplikasi SPSS Versi 25*. 1st edn. Edited by Khaeruman. Banten: AA Rizky.

Hamzah, Lies Maria. Imam Awaludin, E.M. 2016. *Pengantar Statistika Ekonomi*. 1st edn. Bandar Lampung: Anugrah Utama Rahaja (AURA).

Mundir. 2012. *Statistik Pendidikan Pengantar Analisis Data untuk Penelitian Skripsi dan Tesis*. 1st edn. Edited by M. Hisbiyatul Hasanah. Jember: STAIN Jember Press.

Nuryadi *et al.* 2017. *Buku Ajar Dasar-dasar Statistik Penelitian*. 1st edn, *Sibuku Media*. 1st edn. Yogyakarta: Gramasurya.

Walpole, R.E. 1992. *Pengantar Statistika*. 3rd edn. jakarta: Gramedia Pustka Utama.

BAB 4
UKURAN PENYEBARAN DATA

Oleh Enos Lolang

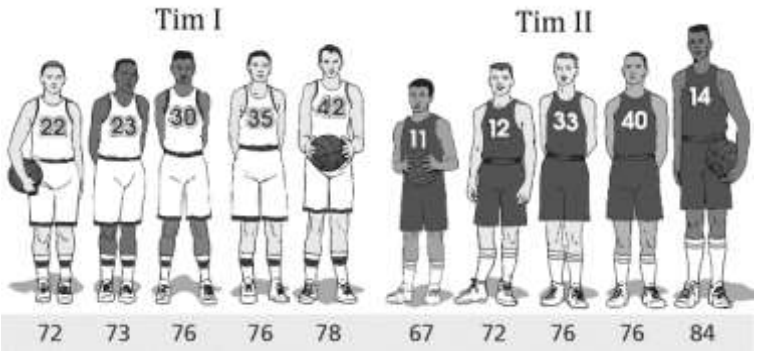
4.1 Pendahuluan

Ukuran penyebaran atau dispersi atau variabilitas data, adalah statistik yang digunakan untuk mengetahui sejauh mana data tersebar atau menyimpang dari nilai pusat seperti rata-rata atau median. Menurut Winarsunu (2009), ukuran variabilitas adalah suatu ukuran tentang derajat penyebaran nilai-nilai variabel dari suatu tendensi sentral dalam suatu distribusi data. Dua kelompok distribusi data dapat memiliki nilai tendensi sentral yang sama, tetapi derajat penyebarannya bisa sangat berbeda. Ukuran ini membantu seseorang memahami keragaman atau variasi kumpulan data.

Pada dasarnya, variabilitas adalah inti dari statistik. Variasi selalu ada dalam kumpulan data, terlepas dari karakteristik yang diukur, karena tidak semua individu akan memiliki nilai yang tepat sama untuk setiap karakteristik yang diukur. Tanpa ukuran variabilitas, seseorang tidak dapat membandingkan dua kumpulan data secara efektif.

Bagaimana jika dua kumpulan data memiliki nilai rata-rata dan median yang sama? Apakah itu berarti bahwa data tersebut semuanya sama? Tidak sama sekali. Misalnya, kumpulan data 199, 200, 201, dan 0, 200, 400 keduanya memiliki rata-rata yang sama, yaitu 200, dan median yang sama, yang juga 200. Namun kedua kumpulan data tersebut memiliki penyebaran yang sangat berbeda. Kumpulan data pertama memiliki penyebaran yang sangat kecil dibandingkan dengan yang kedua.

Ukuran variabilitas juga adalah suatu nilai yang menggambarkan penyebaran dari kumpulan data tentang nilai pusatnya (McCune, 2010). Interpretasi ukuran tendensi sentral (kecenderungan pemusatan) untuk kumpulan data akan lebih baik jika variabilitas nilai sentral diketahui.



Gambar 4.1. Perbedaan tinggi badan dua tim pemain basket (Sumber: Weiss, 2016)

Gambar 4.1 memperlihatkan dua tim pemain basket yang memiliki ukuran rata-rata tinggi badan yang sama, berdasarkan ukuran pemusatan data. Rata-rata tinggi badan kedua tim adalah 75 incii, median 76 inci, dan modus 76 inci. Meskipun demikian, variasi data ukuran tinggi badan kedua kelompok tersebut jelas-jelas berbeda. Tinggi badan tim II lebih bervariasi dibandingkan dengan tinggi badan tim I. Untuk menggambarkan perbedaan tersebut secara kuantitatif, digunakan ukuran deskriptif yang menyatakan ukuran besarnya variasi atau sebaran dalam kumpulan data yang dimaksud. Ukuran variasi data yang paling sering digunakan dalam statistik adalah rentang dan standar deviasi (simpangan baku) (Weiss, 2016).

4.2 Ukuran Penyebaran Data

Ukuran penyebaran data, merupakan suatu konsep penting dalam statistik yang digunakan untuk mengukur sejauh mana data tersebar di sekitar nilai rata-rata. Ukuran penyebaran membantu kita memahami keragaman data dan bagaimana titik data individual berbeda satu sama lain. Kuantitas ukuran penyebaran data yang akan diuraikan pada bab ini adalah rentang, varians, simpangan rata-rata, standar deviasi, dan kuartil.

Salah satu ukuran penyebaran yang paling sederhana adalah rentang. Rentang adalah selisih antara nilai tertinggi dan terendah dalam kumpulan data. Ini memberikan gambaran kasar tentang

sebaran data, tetapi rentan terhadap data pencilan (*outlier*), yaitu data yang jauh dari sebagian besar data lainnya.

Varians (keragaman) adalah ukuran lain yang digunakan mengukur sebaran data dengan membandingkan variabilitas data terhadap nilai rata-rata. Ini adalah akar kuadrat dari varians, dan seperti simpangan baku, rentan terhadap pencilan. Standar deviasi adalah variasi yang paling sederhana dan merupakan akar kuadrat dari nilai harapan kuadrat perbedaan antara setiap titik data dan rata-rata.

Simpangan rata-rata adalah alternatif untuk simpangan baku. Ini mengukur rata-rata jarak antara setiap titik data dan rata-rata. Namun, simpangan rata-rata tidak sekuat simpangan baku dalam mendeteksi pencilan karena tidak mempertimbangkan perbedaan mutlak.

Standar deviasi adalah ukuran penyebaran data yang umum digunakan. Kuantitas ini mengukur sejauh mana data menyebar di sekitar rata-rata. Semakin besar simpangan baku, semakin besar sebaran data. Dalam perhitungannya, setiap data diberi bobot berdasarkan sejauh mana ia berbeda dari rata-rata, sehingga data ekstrim memiliki pengaruh yang signifikan.

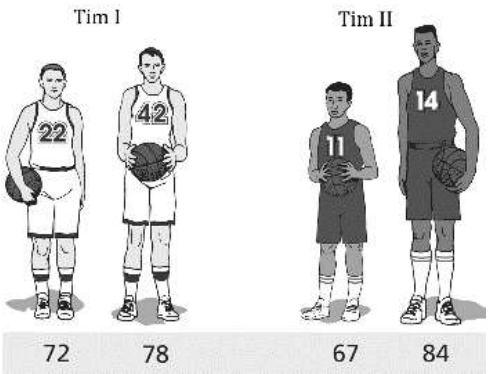
Simpangan kuartil adalah ukuran penyebaran yang lebih tahan terhadap pencilan. Ini berdasarkan pada perbedaan antara kuartil atas (nilai tengah tertinggi 25% data) dan kuartil bawah (nilai tengah terendah 25% data). Simpangan kuartil memberikan gambaran tentang sebaran sebagian besar data, mengabaikan pencilan yang mungkin ada.

Dalam analisis statistik, pemilihan ukuran penyebaran data yang tepat sangat tergantung pada sifat data dan tujuan analisis. Ukuran penyebaran data membantu kita memahami apakah data cenderung berkumpul di sekitar rata-rata atau apakah mereka sangat bervariasi. Ini adalah komponen penting dalam pengambilan keputusan, pemodelan statistik, dan penelitian ilmiah, karena membantu kita mengidentifikasi tren, anomali, dan pola dalam data kita. Sebagai contoh, jika kita ingin memahami sebaran gaji karyawan dalam sebuah perusahaan, kita akan menggunakan ukuran penyebaran data untuk melihat sejauh mana gaji berbeda satu sama lain, dan ini dapat memengaruhi kebijakan kompensasi

dan perencanaan sumber daya manusia. Dengan demikian, ukuran penyebaran data adalah alat yang penting dalam analisis statistik yang membantu kita menganalisis karakteristik dari data yang ada.

4.2.1 Rentang

Rentang adalah selisih antara nilai maksimum dan minimum dalam kumpulan data yang sama. Dengan kata lain rentang adalah data terbesar dengan data terkecil. Ini merupakan ukuran penyebaran data yang paling sederhana. Namun, rentang sangat ditentukan oleh nilai-nilai ekstrem dan mungkin tidak memberikan gambaran yang akurat jika terdapat pencilan dalam data. Perbedaan ukuran tinggi badan Tim I dan Tim II dapat terlihat dengan jelas apabila langsung dipilih pemain yang tinggi badannya paling rendah dan paling tinggi (Gambar 4.2).



Gambar 4.2. Rentang tinggi badan (Sumber: Weiss, 2016)

Rentang data dihitung dengan langkah-langkah berikut:

1. Menentukan nilai minimum (min). Nilai minimum adalah nilai terendah dalam kumpulan data. Nilai minimum merupakan titik awal perhitungan rentang.
2. Menentukan nilai maksimum (maks). Nilai maksimum adalah nilai tertinggi dalam kumpulan data. Ini adalah titik akhir perhitungan rentang.
3. Menentukan rentang. Rentang data dihitung dengan mengurangkan nilai minimum dari nilai maksimum.

Rentang = Maks – Min

Sebagai contoh, misalkan terdapat sekelompok data:

10, 15, 18, 22, 25, 30, 35

Maka rentang data tersebut dapat dihitung sebagai berikut.

1. Nilai minimum (min) adalah 10.
2. Nilai maksimum (maks) adalah 35.
3. Rentang = maks - min = 35 - 10 = 25
   Jadi, rentang data dari kumpulan data ini adalah 25.

Pada Gambar 4.2 dapat dilihat bahwa untuk Tim I, rentang data tinggi badan anggota tim adalah 78 – 72 = 6 inci, sedangkan untuk tim II adalah 84 – 67 = 17 inci. Rentang data mudah dihitung tetapi kuantitas ini hanya melibatkan dua data yaitu data terbesar dan data terkecil. Sifat ini menimbulkan keraguan karena tidak terdapat perbedaan antara kumpulan 100 data dengan dua data yang kebetulan menyatakan data terkecil dan terbesar dari kumpulan 100 data tadi. Selai itu, rentang sangat sensitif terhadap pencilan. Jika terdapat nilai ekstrim, rentang akan sangat dipengaruhi oleh nilai tersebut dan mungkin tidak mencerminkan sebaran yang sebenarnya. Rentang juga tidak memberikan informasi tentang sebaran data di antara nilai-nilai minimum dan maksimum, serta tidak memperhitungkan sebaran nilai-nilai dalam kumpulan data. Ketika data mengandung pencilan atau perbedaan yang signifikan antara nilai-nilai ekstrim, rentang data dapat sangat besar, meskipun sebaran data sebenarnya lebih terkonsentrasi. Ini bisa menyesatkan dalam analisis statistik karena rentang data tidak memberikan gambaran yang baik tentang sebaran sebagian besar data. Karena alasan-alasan ini, dua ukuran variasi lainnya yaitu standar deviasi dan rentang antarkuartil lebih dipertimbangkan dari pada rentang.

Dalam analisis statistik sering juga digunakan ukuran penyebaran lain seperti rentang antarkuartil (*interquartile range*), simpangan baku (*standard deviation*), atau simpangan baku sampel

(*sample standard deviation*) untuk memberikan gambaran yang lebih komprehensif tentang sebaran data.

Rentang antarkuartil, misalnya, mengukur sebaran data di antara kuartil pertama (Q1) dan kuartil ketiga (Q3) dan lebih tahan terhadap pengaruh pencilan dibandingkan dengan rentang data. Simpangan baku mengukur sebaran data sehubungan dengan rata-rata, dan ini memberikan gambaran yang lebih komprehensif tentang sebaran data.

Jadi, tingkat ketelitian dalam mengukur penyebaran data dapat ditingkatkan dengan menggunakan ukuran yang lebih canggih seperti rentang antarkuartil atau simpangan baku, yang memberikan informasi yang lebih akurat tentang sebaran data secara keseluruhan, tidak terlalu dipengaruhi nilai-nilai ekstrim.

4.2.2 Varians

Varians adalah rata-rata dari kuadrat selisih antara masing-masing data dengan nilai rata-rata. Ini memberikan gambaran tentang seberapa jauh data tersebar dari nilai rata-ratanya. Varians yang tinggi menunjukkan penyebaran yang besar, sedangkan varians rendah menunjukkan penyebaran yang kecil.

Varians data adalah salah satu konsep penting dalam statistika yang digunakan untuk mengukur sejauh mana data tersebar atau bervariasi. Varians sering digunakan dalam analisis statistik untuk mengevaluasi sejauh mana data dapat diandalkan atau memiliki konsistensi. Seperti yang dinyatakan oleh (Montgomery, Peck dan Vining, 2012), "Varians adalah ukuran penting yang digunakan dalam analisis statistik untuk mengidentifikasi sejauh mana data berbeda dari nilai rata-rata."

Varians data adalah salah satu konsep penting dalam statistika yang digunakan untuk mengukur sebaran atau variasi dari sekumpulan data. Konsep ini pertama kali diperkenalkan oleh ilmuwan Inggris bernama Francis Galton pada tahun 1869 yang menyatakan bahwa "Varians adalah ukuran untuk menunjukkan sejauh mana letak sekelompok data dari nilai rata-ratanya."

Varians memiliki beberapa kegunaan yang penting dalam analisis statistik. Salah satunya adalah dalam pengukuran risiko dan kerentanan dalam keuangan. Sebagai contoh, dalam pasar saham,

varians harga saham dapat digunakan untuk mengevaluasi tingkat fluktuasi atau risiko investasi. Hal ini juga dapat diterapkan dalam ilmu sosial, seperti dalam penelitian tentang tingkat variabilitas pendapatan dalam suatu populasi.
Rumus varians adalah sebagai berikut:

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

Dalam rumus tersebut, σ^2 adalah varians, n adalah jumlah data, x_i adalah nilai data ke-i, dan $\bar{x}$ adalah nilai rata-rata dari seluruh data. Hasil perhitungan varians akan memberikan kita gambaran tentang sejauh mana data tersebar atau bervariasi dalam kumpulan data tersebut.

Pengukuran varians data dapat memberikan wawasan tentang keragaman dan ketidakpastian dalam kumpulan data. Menurut Gunst dan Mason (1980), "Variabilitas data dapat diidentifikasi dengan mengukur varians. Semakin besar nilai varians, semakin besar keragaman dalam dataset." Ini berarti bahwa varians data yang tinggi menunjukkan bahwa data memiliki variasi yang besar antara titik-titiknya.

Varians data biasanya dihitung dengan menggunakan rumus yang melibatkan selisih kuadrat antara setiap titik data dan nilai rata-rata, kemudian dijumlahkan dan dibagi dengan jumlah total titik data. Seperti yang dijelaskan oleh Taylor (2022), "Rumus varians yang umum digunakan adalah rumus varians sampel, yang mengukur variabilitas dalam sampel data. Ini diperoleh dengan mengkuadratkan selisih antara setiap titik data dan rata-rata, menjumlahkannya, dan kemudian membaginya dengan jumlah pengamatan minus satu."

Varians data juga berhubungan erat dengan konsep standar deviasi. Standar deviasi adalah akar kuadrat dari varians, dan sering digunakan untuk memberikan gambaran yang lebih mudah dipahami tentang sebaran data. Menurut Gupta dan Kapoor (1989), "Standar deviasi adalah ukuran statistik yang sering digunakan untuk mengukur sebaran data."

Pemahaman tentang varians data dapat membantu dalam pengambilan keputusan yang lebih baik dalam berbagai bidang,

termasuk ilmu pengetahuan, bisnis, dan teknik. Dengan mengukur sebaran data, kita dapat mengidentifikasi tren, anomali, dan pola yang mungkin tidak terlihat jika hanya melihat nilai rata-rata. Dalam statistik, varians data adalah alat penting untuk memahami keragaman dan konsistensi dalam dataset (Montgomery, Peck dan Vining, 2012).

Dalam penelitian ilmiah, penggunaan varians data dapat membantu dalam menentukan apakah hasil eksperimen signifikan atau tidak. Varians yang rendah menunjukkan bahwa data eksperimen cenderung konsisten, sementara varians tinggi menunjukkan tingkat keragaman yang tinggi. Oleh karena itu, pemahaman yang kuat tentang varians data adalah kunci dalam mengevaluasi hasil penelitian dan membuat kesimpulan yang valid (Taylor, 2022).

Variabilitas dalam data juga penting dalam analisis bisnis. Varians data dapat membantu manajer dan analis bisnis dalam mengidentifikasi tren dan fluktuasi dalam kinerja bisnis. Dengan memahami varians data, perusahaan dapat mengambil tindakan yang sesuai untuk meningkatkan efisiensi dan produktivitas (Gunst dan Mason, 1980).

Jadi dapat disimpulkan bahwa varians adalah ukuran yang penting dalam statistika yang digunakan untuk mengukur sebaran dan variasi dalam dataset. Varians membantu kita memahami sejauh mana data berbeda dari nilai rata-rata dan memberikan wawasan tentang keragaman dalam data. Dalam berbagai konteks, pemahaman yang kuat tentang varians data sangat penting untuk pengambilan keputusan yang lebih baik dan analisis yang lebih akurat.

4.2.3 Simpangan Rata-rata

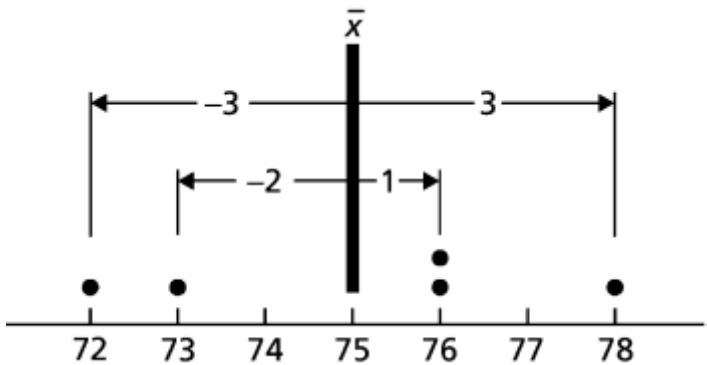
Tinggi badan dari kelima pemain Tim I dalam Gambar 1 masing-masing adalah 72, 73, 76, 76, dan 78. Nilai rata-rata dari data ini adalah

$$\bar{x} = \frac{\sum x_i}{n} = \frac{75+73+76+76+78}{5} = \frac{375}{5} = 75 \text{ inci}$$

Untuk menghitung simpangan dari nilai rata-rata x_i, maka masing-masing nilai tersebut diselisihkan dengan nilai rata-rata $(x_i - \bar{x})$. Sebagai contoh, simpangan tinggi badan 72 inci dari nilai rata-rata adalah $x_i - \bar{x} = 72 - 75 = -3$. Simpangan dari nilai rata-rata untuk kelima pengamatan ditunjukkan pada tabel berikut.

Tabel 4.1. Simpangan dari nilai rata-rata

Tinggi badan, x	Deviasi dari nilai rata-rata $x - \bar{x}$
72	-3
73	-2
76	1
76	1
78	3



Gambar 4.3. Data pengamatan dan simpangan dari nilai rata-rata

4.2.4 Standar Deviasi

Standar deviasi atau simpangan baku (*Standard Deviation*) adalah akar kuadrat dari varians. Ini sering digunakan karena satuan pengukuran standar deviasi sesuai dengan satuan pengukuran data asli. Semakin besar standar deviasi, semakin besar penyebaran data.

Berbeda dengan rentang, standar deviasi melibatkan semua data pengamatan. Ini adalah ukuran variasi yang lebih dipertimbangkan ketika mean digunakan sebagai ukuran pemusatan. Secara kasar, standar deviasi mengukur variasi

pengamatan dari nilai rata-rata. Untuk kumpulan data dengan jumlah variasi yang besar, rata-rata pengamatan akan jauh dari rata-rata sehingga standar deviasi menjadi lebih besar. Untuk kumpulan data dengan jumlah variasi yang kecil, pengamatan akan rata-rata mendekati rata-rata; sehingga standar deviasi akan kecil. Semakin besar variasi suatu data, semakin besar pula standar deviasi kumpulan data tersebut. Hal ini menunjukkan bahwa standar deviasi memenuhi kriteria dasar dari ukuran suatu variasi. Dalam banyak masalah statistik, standar deviasi pada umumnya digunakan untuk mengukur variasi. Namun demikian, standar deviasi memiliki kelemahan. Salah satu kelemahannya adalah bahwa nilai standar deviasi tidak resisten karena nilainya sangat dipengaruhi oleh data hasil pengamatan yang terlalu menyimpang.

Rumus untuk standar deviasi data sampel dan data populasi sedikit berbeda. Pada bagian ini, kita berkonsentrasi pada standar deviasi sampel. Langkah pertama dalam menghitung standar deviasi sampel adalah menentukan simpangan dari rata-rata, yaitu seberapa jauh setiap data pengamatan dari nilai rata-rata. Standar deviasi untuk variabel x dilambangkan dengan s_x.

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

dimana n adalah ukuran sampel dan $\bar{x}$ adalah rata-rata sampel. Rumus ini dinamakan *rumus definisi*.

Tinggi badan dari kelima anggota Tim II masing-masing adalah 67, 72, 76, 76, dan 84. Standar deviasi tinggi badan Tim II dihitung mengikuti tiga langkah sebagai berikut.

Langkah 1: menghitung nilai rata-rata sampel $(\bar{x})$.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{67+72+76+76+84}{5} = \frac{375}{5} = 75 \text{ inci.}$$

Langkah 2: menyusun tabel untuk deviasi kuadrat $\sum(x_i - \bar{x})^2$

Tabel 4.2. Menghitung deviasi kuadrat tinggi badan Tim II

x	$x - \bar{x}$	$(x - \bar{x})^2$
67	-8	64
72	-3	9
76	1	1
76	1	1
84	9	81
Jumlah		156

Langkah 3. menggunakan rumus definisi.

$s_x = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{156}{5-1}} = \sqrt{39} = 6{,}2$ inci.

Dengan demikian dapat disimpulkan bahwa rata-rata variasi tinggi badan anggota Tim II dari tinggi badan rata-rata 75 adalah sekitar 6,2 inci.

Langkah kedua dalam menghitung standar deviasi sampel adalah menentukan ukuran deviasi total dari nilai rata-rata untuk masing-masing pengamatan. Meskipun kuantitas $x_i - \bar{x}$ menyatakan deviasi dari nilai rata-rata, menjumlahkan semua nilanya untuk memperoleh deviasi total tidak memiliki makna karena hasil penjumlahannya, $\sum(x_i - \bar{x})$, selalu menghasilkan nilai nol. Agar nilai penjumlahannya tidak sama dengan nol, maka masing-masing selisihnya dikuadratkan. Jumlah deviasi kuadrat tersebut dinamakan jumlah deviasi kuadrat yang menghasilkan ukuran deviasi total untuk semua pengamatan.

Jumlah kuadrat simpangan $(x - \bar{x})$ tinggi badan Tim I dapat dilihat pada Tabel 4.3. Jumlah kuadrat simpangan adalah 24.

Tabel 4.3. Kuadrat simpangan tinggi badan Tim I

Tinggi badan (inci)	Simpangan dari nilai rata-rata $(x - \bar{x})$	Kuadrat simpangan $(x - \bar{x})^2$
72	-3	9
73	-2	4
76	1	1
76	1	1
78	3	9
Jumlah		24

Langkah ketiga dalam menghitung standar deviasi sampel adalah menghitung nilai rata-rata dari kuadrat simpangan yang dilakukan dengan cara membagi jumlah kuadrat simpangan dengan $n-1$, bukannya membagi dengan jumlah sampel (n). Jika kita membagi dengan n, varians sampel yang didapatkan adalah nilai rata-rata kuadrat simpangan. Meskipun membagi dengan n merupakan langkah yang benar dalam menentukan rata-rata, varians sampel dihitung dengan membagi $n-1$ karena alasan berikut ini. Salah satu tujuan utama menghitung varians sampel adalah untuk memperkirakan varians populasi. Pembagian dengan n cenderung memperkirakan varians populasi terlalu rendah. Oleh karena itu pembagian dengan $n-1$ secara rata-rata akan menghasilkan nilai yang lebih akurat.

Nilai yang diperoleh dinamakan varians sampel, dinyatakan dengan s_x^2 atau singkatnya s^2.

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

Standar deviasi diperoleh dari akar s^2.

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

Tinggi badan dari kelima anggota Tim II masing-masing adalah 67, 72, 76, 76, dan 84. Dari data ini dapat dihitung standar deviasi data tinggi badan dengan langkah-langkah sebagai berikut.

Langkah 1. Menghitung nilai rata-rata tinggi badan.
$\bar{x} = \frac{\sum x_i}{n} = \frac{67+72+76+76+84}{5} = \frac{375}{5} = 75$ inci.

Langkah 2. Menghitung jumlah deviasi kuadrat $\sum(x_i - \bar{x})^2$.
Nilai ini sudah dihitung pada Tabel 2. Kolom ketiga menunjukkan bahwa jumlah kuadrat deviasi adalah 156.

Langkah 3. Menghitung standar deviasi
Diketahui bahwa $n = 5$ dan $\sum(x_i - \bar{x})^2 = 156$. Dengan nilai-nilai ini dapat dihitung standar deviasi data tinggi badan anggota Tim II.

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{156}{5-1}} = \sqrt{39} = 6{,}2$$

Dengan demikian, tinggi badan Tim II rata-rata menyimpang sebesar 6,2 inci dari ukuran tinggi badan rata-rata 75 inci.

Berdasarkan contoh-contoh yang telah ditunjukkan, terlihat dengan jelas perbedaan nilai standar deviasi tinggi badan dari Tim I dan Tim II, masing-masing adalah 2,4 dan 6,2. Jadi dapat disimpulkan bahwa tinggi badan Tim II lebih bervariasi dibandingkan dengan tinggi badan Tim I.

Alternatif lain untuk mencari nilai standar deviasi adalah menggunakan *rumus hitung*,

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

dengan n adalah ukuran sampel. Rumus hitung menghasilkan nilai standar devisi yang sama dengan rumus definisi, meskipun rumus hitung biasanya lebih cepat memeberikan hasil dibandingkan dengan rumus definisi. Mungkin ada perbedaan kecil dalam pembulatan hasil akhir dan juga mengurangi kemungkinan salah hitung, tetapi tidak akan mempengaruhi nilai standar deviasi yang dimaksud.

Sebelum menjelaskan lebih lanjut mengenai rumus hitung, mari kita menyelidiki perbedaan $\sum x_i^2$ dengan $(\sum x_i)^2$. Pernyataan $\sum x_i^2$ menyatakan jumlah kuadrat data. Untuk mencari nilai tersebut, masing-masing data dikuadratkan kemudian masing-masing kuadrat dijumlahkan. Dengan kata lain, $\sum x_i^2 = x_1^2 + x_2^2 + x_3^2 + x_4^2 + \cdots$. Sedangkan pernyataan $(\sum x_i)^2$adalah kuadrat jumlah data. Kuantitas ini dinyatakan dalam bentuk $(\sum x_i)^2 = (x_1 + x_2 + x_3 + x_4 + \cdots)^2$. Dari kedua rumus di atas, standar deviasi data tinggi badan Tim II dapat dihitung dengan terlebih dulu menentukan $\sum x_i^2$ dan $(\sum x_i)^2$. Nilai ini ditunjukkan pada Tabel 4.4.

Tabel 4.4. Nilai $\sum x_i$ dan $\left(\sum x_i^2\right)$ data tinggi badan Tim II

x_i	x_i^2
67	4489
72	5184
76	5776
76	5776
84	7056
375	28281

Jadi standar deviasi data tinggi badan Tim II adalah

$$s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{28281 - \frac{(375)^2}{5}}{5-1}}$$

$$= \sqrt{\frac{28281 - 28125}{4}} = \sqrt{\frac{156}{4}} = \sqrt{39} = 6{,}2$$

Terlihat bahwa hasil yang diperoleh dari perhitungan menggunakan rumus hitung sama dengan nilai yang diperoleh dari perhitungan yang menggunakan rumus definisi.

Seperti diketahui, standar deviasi adalah ukuran variasi data. Semakin bervariasi suatu kumpulan data, semakin besar standar deviasinya. Pada Tabel 4.5 diperlihatkan dua kelompok data, masing-masing terdiri atas 10 pengukuran. Kelompok data II memiliki variasi lebih besar dari kelompok data I.

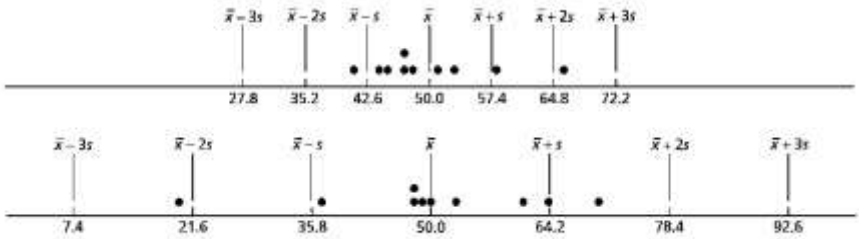
Tabel 4.5. Perbedaan variasi data Kelompok I dan Kelompok II

Kelompok I	41	44	45	47	47	48	51	53	58	66
Kelompok II	20	37	48	48	49	50	53	61	64	70

Nilai rata-rata dan standar deviasi kedua kelompok data pada Tabel 4.5. Mahasiswa dapat menghitung dan mengetahui bahwa nilai rata-ratanya sama $(\bar{x} = 50)$ tetapi standar deviasinya berbeda. Standar deviasi data kelompok I adalah 7,4 sedangkan kelompok II adalah 14,2.

Untuk membandingkan variasi kedua kelompok data secara visual, perhatikan Gambar 4.4. Dengan skala yang sama dan standar

deviasi berbeda, terlihat bahwa simpangannya berbeda diukur dari nilai rata-rata. Hal ini disebabkan karena perbedaan standar deviasi.



Gambar 4.4. Visualisasi variasi data kelompok I dan kelompok II (Sumber: Weiss, 2016)

Posisi horisontal yang ditandai dengan $\bar{x} + 2s$ menyatakan suatu nilai dua kali standar deviasi di sebelah kanan nilai rata-rata, dan posisi yang ditandai dengan $\bar{x} - 3s$ menyatakan nilai tiga kali standar deviasi di sebelah kiri nilai rata-rata. Dalam hal ini, $\bar{x} + 2s = 50{,}0 + 14{,}8 = 64{,}8$ dan $\bar{x} - 3s = 50{,}0 - 22{,}2 = 27{,}8$. Hampir semua data pengukuran statistik terletak dalam batas tiga kali standar deviasi diukur ke kiri dan kanan nilai rata-rata. Ini merupakan salah satu sifat standar deviasi yang penting. Selain itu beberapa sifat penting dari standar deviasi adalah:

1. Standar deviasi tidak pernah bisa bernilai negatif.
2. Nilai terkecil yang mungkin untuk standar deviasi adalah 0 ketika kumpulan data terdiri atas angka yang sama.
3. Standar deviasi sangat dipengaruhi oleh data pencilan karena jarak dari rata-rata dipengaruhi oleh pencilan.
4. Standar deviasi memiliki satuan yang sama dengan data asli, sedangkan varians dalam satuan kuadrat.
5. Koefisien variasi (*Coefficient of Variation*). Koefisien variasi adalah rasio antara standar deviasi dengan nilai rata-rata. Koefisien ini berguna untuk membandingkan sejauh mana data tersebar secara relatif terhadap rata-rata.

$$K_v = \frac{Standar\ Deviasi}{Rata - Rata} \times 100\%$$

Dapat dilihat dari rumus tersebut bahwa semakin besar standar deviasi, akan semakin besar pula koefisien variasi.

4.2.5 Kuartil dan Rentang Antarkuartil

Kuartil dan rentang antarkuartil adalah dua konsep penting dalam statistik deskriptif yang digunakan untuk mengukur sebaran data. Kuartil adalah titik pemisah data ke dalam empat kelompok yang memiliki jumlah data yang sama. Kuartil pertama (Q1) adalah nilai tengah dari data yang lebih rendah dari median, kuartil kedua (Q2) adalah median itu sendiri, dan kuartil ketiga (Q3) adalah nilai tengah dari data yang lebih tinggi dari median. Kuartil dapat memberikan gambaran tentang sebaran data dan membantu mengidentifikasi pencilan.

Rentang antarkuartil, juga dikenal sebagai IQR (*Interquartile Range*), adalah perbedaan antara kuartil ketiga (Q3) dan kuartil pertama (Q1). Ukuran ini lebih meyakinkan daripada rentang karena tidak terpengaruh oleh pencilan. IQR sering digunakan dalam analisis boxplot untuk mengidentifikasi data yang ekstrim. Semakin besar IQR, semakin besar sebaran data di dalam "kotak" pada boxplot.

Misalnya, jika sebuah penelitian mengamati tinggi badan siswa dalam sebuah sekolah, IQR dapat memberikan informasi tentang sebaran tinggi badan siswa di sekolah tersebut. Penjelasan ini didukung oleh penelitian terbaru yang diterbitkan dalam "Statistics in Medicine" oleh Johnson, R. B. (2021), yang mendiskusikan penggunaan kuartil dan IQR dalam analisis data tinggi badan siswa.

Kuartil dan IQR juga dapat membantu mengidentifikasi apakah data memiliki distribusi yang simetris atau tidak. Jika Q1, Q2, dan Q3 memiliki posisi yang hampir sama, data cenderung simetris. Namun, jika salah satu kuartil berada jauh dari yang lain, data cenderung asimetris. Referensi yang dapat digunakan untuk memahami konsep ini lebih dalam adalah buku "Statistics" karya Robert S. Witte dan John S. Witte (2020).

Dalam praktiknya, kuartil dan IQR digunakan dalam berbagai disiplin ilmu, termasuk ilmu sosial, ekonomi, dan ilmu alam, untuk menganalisis dan memahami data. Mereka adalah alat

yang berguna dalam statistik deskriptif untuk menggambarkan sebaran data secara singkat dan efektif.

Selain itu, kuartil dan IQR juga berguna dalam analisis data ekstensif dan digunakan sebagai dasar untuk pengambilan keputusan di berbagai bidang seperti bisnis, ilmu kedokteran, dan ilmu lingkungan. Mereka membantu dalam pemahaman pola data, mengidentifikasi anomali, dan memungkinkan pengambilan keputusan yang lebih informasional. Referensi tambahan tentang aplikasi kuartil dan IQR dapat ditemukan dalam artikel "Practical Statistics for Data Scientists" karya Andrew Bruce dan Peter Bruce (2021).

Dalam kesimpulan, kuartil dan IQR adalah alat statistik yang penting dalam mengukur sebaran data. Mereka membantu dalam pemahaman pola data, identifikasi pencilan, dan memberikan wawasan tentang distribusi data. IQR khususnya berguna karena mampu mengatasi masalah pencilan. Dengan pemahaman yang baik tentang konsep ini, analisis data menjadi lebih informatif dan dapat mendukung pengambilan keputusan yang lebih baik dalam berbagai bidang. Kuartil dan rentang antarkuartil dapat ditentukan dengan mengikuti langkah-langkah berikut.

1. Urutkan Data
   Langkah pertama adalah mengurutkan data dari yang terkecil hingga yang terbesar atau dari data terbesar hingga terkecil. Kuartil dan rentang antarkuartil lebih mudah ditentukan pada kelompok data yang berurutan dibandingkan dengan yang tidak.

2. Tentukan Kuartil Ketiga (Q3)
   Kuartil ketiga (Q3) adalah nilai yang membagi 25% data tertinggi. Untuk menentukan Q3, digunakan rumus
   $$Q_3 = \left(\frac{3n+1}{4}\right)$$
   di mana n adalah jumlah data dalam kumpulan data. Jika hasil dari rumus ini adalah bilangan bulat, Q_3 adalah nilai di posisi tersebut dalam data yang sudah diurutkan. Jika hasilnya bukan bilangan bulat, kita dapat mencari dua angka

yang berada di posisi tersebut kemudian mengambil rata-ratanya.

3. Menentukan Kuartil Pertama (Q_1)
   Kuartil pertama (Q_1) adalah nilai yang membagi 25% data terendah. Untuk menemukan (Q_1), digunakan rumus yang sama dengan (Q_3).
   $$Q_1 = \frac{(n+1)}{4}$$
   Seperti sebelumnya, jika hasilnya adalah bilangan bulat, Q_1 adalah nilai di posisi tersebut dalam data yang sudah diurutkan. Jika hasilnya bukan bilangan bulat, carilah dua angka yang berada di posisi tersebut dan hitung nilai rata-ratanya.

4. Hitung Rentang Antarkuartil (IQR)
   Rentang antarkuartil (IQR) adalah selisih antara Q_3 dan Q_1, yaitu $Q_3 - Q_1$. IQR memberikan informasi tentang sebaran data di sekitar nilai tengah.

5. Identifikasi Data Pencilan
   Dalam beberapa analisis statistik, kita mungkin ingin mengidentifikasi data pencilan. Pencilan adalah data yang terletak di luar rentang IQR. Data-data ini dapat dianggap sebagai data yang ekstrem atau tidak biasa dalam distribusi.

   Dengan langkah-langkah di atas, kita dapat menentukan kuartil dan rentang antarkuartil dalam kumpulan data. Kuartil dan rentang antarkuartil adalah kuantitas yang berguna untuk menggambarkan sebaran data dan mengidentifikasi keragaman dalam data. Selain itu kuartil dan rentang antarkuartil juga membantu dalam mengidentifikasi data ekstrem yang dapat memengaruhi analisis statistik.

Memahami ukuran penyebaran data sangat penting dalam analisis statistik untuk membantu mengidentifikasi pola, pencilan, dan karakteristik distribusi data yang sedang dianalisis. Pemilihan ukuran penyebaran yang tepat tergantung

pada sifat data dan tujuan analisis statistik yang sedang dilakukan.

4.3 Indeks Variasi Kualitatif dan Indeks Segregasi

Indeks variasi kualitatif (*index of qualitative variation*, IQV) adalah suatu ukuran sebaran yang bermanfaat untuk berbagai variabel kategori (dikotomi atau poliotomi nominal, atau variabel ordinal berkelompok). Rentang nilai IQV adalah dari 0 sampai 1. IQV = 1 menunjukkan bahwa data tidak menyebar apabila data hanya terdiri atas satu kategori. IQV = 1 merupakan sebaran maksimum data ketika semua data terdistribusi secara merata dalam semua kategori. IQV dihitung dengan rumus sebagai berikut:

$$IQV = \frac{k(N^2 - \sum f^2)}{N^2(k-1)}$$

dimana k adalah banyaknya kategori dalam variabel (yang mencakup semua kategori meskipun beberapa kategori memiliki frekuensi nol), N adalah total kasus, dan f adalah banyaknya kasus pada kategori tertentu. Langkah pertama dalam rumus di atas menunjukkan bahwa yang harus dilakukan adalah menetapkan jumlah kasus di dalam setiap kategori kemudian dikuadratkan, dan akhirnya menjumlahkannya untuk semua k kategori.

Misalkan kita memiliki daftar frekuensi status perkawinan seperti ditunjukkan pada Tabel 5. Dari tabel diketahui bahwa banyaknya kategori k adalah 5, ukuran sampel $N = 1498$, $N^2 = 1498^2 = 2.244.004$, dan jumlah kuadrat frekuensi, masing-masing kategori:

f_{kawin^2}	882^2	784.996
$f_{janda/duda^2}$	132^2	17.424
f_{cerai^2}	132^2	26.896
f_{pisah^2}	49^2	2.401
$f_{tidak\ kawin^2}$	267^2	71.289
Jumlah		903.006

Tabel 4.6. Tabel frekuensi kategori status perkawinan

Kategori	f	%	% valid	% kumulatif
Kawin	886	59,1	59,1	59,1
Janda/duda	132	8,8	8,8	68,0
Cerai	164	10,9	10,9	78,9
Pisah	49	3,3	3,3	82,2
Tidak kawin	267	17,8	17,8	100
Total	1498	99,9	100	

Dengan nilai-nilai ini, dapat dihitung IQV:

$$IQV = \frac{5(2.244.004 - 903.006)}{(2.244.004)4} = 0{,}747$$

Perhitungannya membosankan serta melibatkan angka-angka besar, tetapi sebenarnya cukup sederhana. Hasil akhirnya (0,747) menyatakan bahwa distribusi ini memiliki penyebaran yang cukup besar. Kita dapat memverifikasi ini dengan melihat persentasenya. Kategori paling banyak (yaitu kawin) meliputi hampir 60 persen data. Kategori lainnya memiliki jumlah data yang jauh lebih sedikit. Jadi ada cukup banyak ketidaksetaraan, keragaman, atau penyebaran dalam status perkawinan.

Indeks segregasi (atau indeks perbedaan) merupakan pengembangan dari ide dasar yang sama dengan IQV. Indeks segregasi digunakan untuk menggambarkan seberapa berbedanya distribusi kasus-kasus dalam suatu populasi dengan distribusi yang terjadi dalam populasi lainnya. Salah satu contoh penerapannya dalam penelitian bidang sosiologi, misalnya untuk menggambarkan seberapa besar perbedaan distribusi orang kulit hitam dengan orang kulit putih dalam suatu wilayah tertentu. Tentu saja hal sama dapat dilakukan untuk mengukur perbedaan dari berbagai jenis kelompok.

Perhatikan segregasi antara laki-laki dan wanita dalam hal perbedaan pekerjaan (Tabel 4.7). Secara langsung, terlihat ada perbedaan distribusi laki-laki dan wanita dalam semua kategori pekerjaan. Untuk memastikan dugaan ini, digunakan indeks segregasi.

$$I = \frac{1}{2}\sum_{i=1}^{n} |M_i - F_i|$$

Perhatikan bahwa dalam rumus di atas, kita memperkurangkan proporsi wanita dalam kategori pekerjaan dengan proporsi laki-laki dalam kategori pekerjaan yang sama. Kita akan menjumlahkan nilai mutlak selisih tersebut kemudian dibagi dua.

Tabel 4.7. Distribusi Jenis Kelamin dan Pekerjaan

	Sopir	Disainer	Barista	Total
Wanita	2	40	60	102
Laki-laki	30	4	10	44
Total	32	44	70	146

Proporsi laki-laki dan wanita adalah
$M_i \equiv \frac{m_i}{m}$ dan $F_i \equiv \frac{f_i}{f}$
Hal ini berarti jumlah laki-laki yang bekerja dalam kategori pekerjaan tertentu akan dibagi dengan jumlah laki-laki dalam sampel. Demikian juga untuk wanita.

$$s = 0{,}5\left\{\sum_{i=1}^{3}\left|\frac{m_i}{m} - \frac{f_i}{f}\right|\right\}$$
$$= 0{,}5\left\{\left|\frac{m_1}{m} - \frac{f_1}{f}\right| + \left|\frac{m_2}{m} - \frac{f_2}{f}\right| + \left|\frac{m_3}{m} - \frac{f_3}{f}\right|\right\}$$
$$= 0{,}5\left\{\left|\frac{30}{44} - \frac{2}{102}\right| + \left|\frac{4}{44} - \frac{40}{102}\right| + \left|\frac{10}{144} - \frac{60}{102}\right|\right\}$$
$$= 0{,}5\{|0{,}6818 - 0{,}0196| + |0{,}0909 - 0{,}3922| + |0{,}2273 - 0{,}5882|\}$$
$$= 0{,}5(0{,}6622 + 0{,}3013 + 0{,}3609)$$
$$= 0{,}5(1{,}3244)$$
$$= 0{,}6622 = 66{,}22\%.$$

Hal ini berarti bahwa untuk memenuhi proporsi yang sama antara laki-laki dan wanita, kita harus memindahkan 66,22 persen orang yang jenis kelamin yang sama ke kelompok kategori pekerjaan yang lain. Jadi, sebagai contoh, misalkan kita

memindahkan 66,22% wanita ke perkerjaan lainnya maka kita harus memindahkan 67,5 orang wanita ke kategori pekerjaan yang dimaksud. Kita menambahkan 67 wanita ke pekerjaan sebagai supir, mengurangi 31 orang wanita dari desainer dan 36 orang wanita dari barista, seperti terlihat pada Tabel 8. Proporsi 66,22% wanita dari 102 orang berjumlah sekitar 67, yang merupakan jumlah wanita yang akan dipindahkan dari profesinya supaya laki-laki dan wanita memiliki proporsi pekerjaan yang sama.

Tabel 4.8. Mengatur jumlah wanita untuk memperoleh proporsi yang sama

	Sopir	Disainer	Barista	Total
Wanita	2+67=69	40-31=9	60-36=24	102
Laki-laki	30	4	10	44
Total	99	13	34	146

Tabel 4.9. Selisih proporsi sebagai ukuran penyebaran

	Rendah	Sedang	Tinggi	Total
Aktual	0,25	0,50	0,25	1,00
Teoritis	0,33	0,33	0,33	1,00
Selisih absolut	0,08	0,17	0,08	0,33

Jika kita menambahkan selisih data wanita (69+24=134), kita akan membagi dua jumlah tersebut (134/2=67) karena kalau tidak, jumlah wanita akan terhitung dua kali (satu kali ketika dipindahkan dari profesinya, dan kedua ketika ditambahkan ke profesi yang baru).

Jadi, jika $S = 0$, artinya tidak ada orang dalam kelompoknya (jenis kelamin, ras, dan lain-lain) yang harus dipindahkan dari pekerjaannya agar proporsinya sama. Jika $S = 1$, maka semua orang yang berada pada suatu kelompok harus dipindahkan ke kelompok lainnya untuk memenuhi proporsi yang sama.

Salah satu kelemahan dari metode ini adalah bahwa proporsi yang sama dapat dicapai dengan cara memindahkan proporsi sesuai indeks, dengan ketentuan profesi yang ada juga harus diubah untuk mencapai proporsi yang diinginkan. Perlu diperhatikan bahwa pada saat kita memindahkan wanita, maka

akan diperoleh jumlah orang yang lebih banyak pada profesi sopir (profesi desainer berkurang).

Variasi dari indeks segregasi menyatakan perbandingan distribusi proporsi suatu kelompok kategori X dengan suatu distribusi teoritis dari distribusi yang sama, sebagaimana diperlihatkan pada Tabel 4.9.

Jumlah selisih absolut dari proporsi (0,08+0,17+0,08) adalah 0,33. Setengah dari jumlah ini adalah 0,165. Jadi, 16,5% dari semua data yang sebenarnya harus dipindahkan ke kategori lainnya untuk membuat distribusi yang sama dengan penyebaran atau kesetaraan sempurna.

DAFTAR PUSTAKA

Gunst, R.F. dan Mason, R.L. 1980. *Regression Analysis and Its Application: A Data-Oriented Approach*. New York: Taylor & Francis.

Gupta, S.C. dan Kapoor, V.K. 1989. *Fundamentals of Mathematical Statistics*. Sultan Chand \& Sons. Tersedia pada: https://books.google.co.id/books?id=k_HKwQEACAAJ.

McCune, S. 2010. *Practice Makes Perfect Statistics*. New York: McGraw Hill.

Montgomery, D.C., Peck, E.A. dan Vining, G.G. 2012. *Introduction to Linear Regression Analysis*. 5th ed. New Jersey: John Wiley & Sons, Inc.

Taylor, J.R. 2022. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books (G - Reference,Information and Interdisciplinary Subjects Series). Tersedia pada: https://books.google.co.id/books?id=htEUzwEACAAJ.

Weiss, N.A. 2016. *Introductory Statistics*. 10 ed. Boston: Pearson.

Winarsunu, T. 2009. *Statistik Dalam Penelitian Psikologi dan Penelitian*. Malang: Malang: UMM Press.

BAB 5
DISTRIBUSI PELUANG DISKRIT

Oleh Ratu Sarah Fauziah Iskandar

5.1 Pendahuluan

Variabel acak memiliki peran krusial sebagai parameter dalam distribusi probabilitas. Variabel acak merupakan suatu entitas di mana nilai-nilainya ditetapkan berdasarkan hasil eksperimen. Umumnya, variabel acak diwakili oleh huruf besar, seperti X, dan nilai-nilai terkaitnya diwakili oleh huruf kecil, seperti x. Sebagai contoh, dalam situasi melempar dua koin, variabel Y digunakan untuk menunjukkan jumlah kejadian yang diamati, dengan nilai-nilai yang dapat berupa y = 0, 1, atau 2. Setiap nilai yang berbeda dari variabel acak memiliki probabilitas yang terkait, yang umumnya disebut sebagai distribusi probabilitas. Dua jenis probabilitas yang dapat dibedakan adalah distribusi peluang diskrit dan distribusi peluang kontinu. Distribusi peluang diskrit mengacu pada probabilitas terjadinya masing-masing nilai variabel acak yang bersifat diskrit. Variabel acak diskrit mengacu pada variabel acak yang memiliki nilai yang dapat dihitung. Setiap nilai potensial dari fungsi variabel acak diskrit selalu memiliki nilai yang tidak identik dengan nol. Dalam pembahasan probabilitas, distribusi probabilitas diskrit dicirikan sebagai pencacahan atau pengaturan dimana variabel acak mengambil setiap nilai dengan probabilitas tertentu. Variabel-variabel diskrit ini mempunyai kuantitas nilai yang dapat dibayangkan atau jumlah nilai yang dapat dihitung yang jumlahnya tidak terbatas. Istilah "dihitung" menyiratkan bahwa variabel acak dapat dikuantifikasi secara numerik dengan menggunakan nomor urut yaitu 1, 2, 3, dan seterusnya. Misalnya, jumlah rumah yang ada salam suatu *cluster* sebagai contoh variabel diskrit, karena variabel tersebut dapat dihitung.

5.2 Distribus Peluang Diskrit

Beberapa Distribusi Peluang Diskrit, yaitu sebagai berikut.

5.2.1 Distribusi Bernouli

Distribusi Bernoulli berasal dari eksperimen Bernoulli, yang terdiri dari dua hasil yaitu Sukses dan Gagal. Distribusi Bernoulli merupakan distribusi random yang melibatkan dua kejadian yang bersifat saling melengkapi, seperti sukses-gagal, ya-tidak, baik-buruk, dan sebagainya. Sebagai contoh, pelemparan satu koin dapat dijadikan contoh, di mana terdapat dua kemungkinan hasil dari satu lemparan, yakni Angka dan Gambar.

Jika kita menganggap munculnya Angka sebagai kejadian "Sukses" dengan peluang p, dan munculnya Gambar sebagai kejadian "Gagal" dengan peluang 1-p, selanjutnya variabel acak XXX yang terkait dengan eksperimen ini akan mendapatkan nilai 1 dengan peluang p jika kejadian "Sukses" terjadi, dan mendapatkan nilai 0 jika kejadian "Gagal" atau dapat dinyatakan dengan peluang $1 - p$. Jadi, variabel acak XXX dapat dikategorikan sebagai variabel acak dengan distribusi Bernoulli.

Contoh:

Seorang calon ibu pasti menghadapi dua potensi hasil saat melahirkan. Pertama, ia dapat melahirkan dengan sukses, dan kedua, ia mungkin mengalami komplikasi yang mengakibatkan kelahiran yang tidak selamat. Tentu saja, harapan setiap ibu adalah dapat melahirkan dengan selamat. Jika seorang ibu berhasil melahirkan tanpa masalah, itu dianggap sebagai sukses sesuai dengan harapannya. Sebaliknya, jika proses kelahiran tidak berjalan dengan selamat, dianggap sebagai kegagalan karena tidak sesuai dengan harapannya.

Berikut adalah karakteristik Distribusi Bernoulli:

1. Eksperimen atau kejadian hanya dijalankan sekali.
2. Setiap eksperimen atau kejadian hanya menghasilkan dua kemungkinan hasil, yakni sukses ataupun gagal.
3. Peluang keberhasilan, dilambangkan dengan p, dan peluang kegagalan, diwakili oleh q, dihitung sebagai $q = 1 - p$.

Adapun rumus Distribusi Bernoulli sebagai berikut.

Atau

$$P_B\,(x,p) = \begin{cases} p & x = 1 \\ (1-p=q) & x = 0 \\ 0 & x \neq atau\ 1 \end{cases}$$

$$P_B\,(x,p)\,P^x\,(1-p)^{1-x}\ ;= 0,1\ ;\ 0 \leq p \leq 1$$

Contoh :

Satu koin dilempar sekali, dengan catatan bahwa hasilnya dapat berupa muka "M" atau belakang "B". Berapa probabilitas kemunculan muka?

Penyelesaian

Peluang munculnya muka “M” (p) = 0.5

Maka

$q = 1 - p = 1 - 0.5 = 0.5$

Jadi,

$f(x) = p^x q^{1-x}$

$f(1) = (0.5)^1(0.5)^{1-1}$

$f(1) = (0.5)(1)$

$f(1) = 0.5$

5.2.2 Distribusi Binomial

Eksperimen yang dilakukan secara berulang dapat menghasilkan dua hasil yang mungkin. Misalnya, melakukan lemparan koin berulang-ulang pasti akan menghasilkan gambar atau nilai numerik. Dalam skenario khusus ini, kita dapat menetapkan kesuksesan ketika sebuah gambar terwujud dan kegagalan ketika sebuah angka muncul. Ilustrasi lain dapat kita temukan pada fenomena kelahiran. Setiap kejadian kelahiran bayi perempuan bisa dianggap sukses, sedangkan kelahiran bayi laki-laki bisa dianggap gagal. Demikian pula konsep ini dapat diterapkan pada proses produksi lampu pijar. Setiap bohlam yang diperiksa dapat diklasifikasikan sebagai rusak (memerlukan perbaikan lebih lanjut) atau sehat (untuk dikirim untuk diproses lebih lanjut di departemen penjualan). Contoh lain yang lebih rumit melibatkan pengambilan empat kartu dari setumpuk kartu bridge dan

menentukan keberhasilan atau kegagalan berdasarkan keberadaan kartu hitam. Setiap kali sebuah kartu diambil, hasilnya tidak bergantung pada undian sebelumnya, sehingga memastikan kemungkinan sukses yang konstan di seluruh percobaan berturut-turut. Eksperimen yang menunjukkan ciri-ciri tersebut di atas disebut eksperimen binomial.

Karakteristik eksperimen binomial adalah sebagai berikut.

1. Eksperimen diulang sebanyak n kali.
2. Setiap percobaan menghasilkan hasil yang dapat digolongkan sebagai gagal atau sukses.
3. Peluang keberhasilan p konstan dari satu percobaan ke percobaan lainnya.
4. Percobaan yang diulang bersifat saling bebas.

Suatu variabel acak X mempunyai distribusi binomial jika (untuk suatu $0 \leq p \leq 1$)

$$P(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} ; x = 0,1,2 \\ 0 ; lainnya \end{cases}$$

Dengan :
n = jumlah percobaan
x = jumlah kejadian
p = peluang keberhasilan

Distribusi binomial merupakan jumlah variabel acak *x* yang berdistribusi Bernoulli.

Kriteria Distribusi Binomial:

1. Jumlah percobaan merupakan bilangan bulat. Contoh melambungkan koin 5 kali.
2. Setiap eksperimen mempunyai dua hasil (outcome). Contoh: sukses atau gagal dan tinggi atau pendek.
3. Peluang sukses sama di setiap ekperimen. Contoh: Jika pada sebuah dadu, yang diharapkan adalah keluar mata 3 (tiga), maka dikatakan peluang sukses adalah $\frac{1}{6}$, sedangkan

peluang gagal adalah $\frac{5}{6}$. Untuk itu peluang sukses dilambangkan p, sedangkan peluang gagal adalah $(1-p)$ atau biasa juga dilambangkan q, di mana $q = 1-p$.

Contoh :

Probabilitas bahwa seorang bayi tidak mendapatkan imunisasi campak adalah 0,2. Di suatu hari di Tempat vaksin Sehat, terdapat 4 bayi. Peluang bahwa dari keempat bayi tersebut, tiga belum diimunisasi campak adalah...

Penyelesaian :
Diketahui :
Peluang tidak imunisasi (p) =0.2
Peluang imunisasi (q) = $1 - 0.2 = 0.8$
Banyak subjek (n) = 4
Belum diimunisasi $(x) = 3$

Ditanyakan :
Peluang bayi yang belum diimunisasi $P(x)$?
Jawab

$$P(x) = \binom{n}{x} p^x q^{n-x}$$
$$P(x) = \binom{4}{3} (0.2)^3 (0.8)^{4-3}$$
$$P(x) = \frac{4!}{4!3!} (0.008)(0.8)$$
$$P(x) = 0.025$$

5.2.3 Distribusi Poisson

Distribusi Poisson, disebut [pwasɔ̃], merupakan distribusi probabilitas diskrit yang menggambarkan kemungkinan jumlah peristiwa yang terjadi dalam suatu periode waktu tertentu, dengan catatan bahwa rata-rata kejadian diketahui dan kejadian-kejadian itu saling bebas dalam interval waktu sejak kejadian terakhir. (Distribusi Poisson juga dapat diterapkan pada jumlah kejadian dalam interval tertentu, seperti jarak, luas, atau volume.)

Suatu eksperimen dianggap sebagai eksperimen Poisson jika eksperimen tersebut menghasilkan hasil berupa jumlah

keberhasilan yang terjadi selama periode waktu tertentu atau dalam wilayah tertentu. Durasi waktu dapat berkisar dari detik hingga tahun, mencakup interval seperti detik, menit, jam, hari, minggu, bulan, atau tahun. Demikian pula, daerah yang ditunjuk dapat terdiri dari garis, luas, sisi, atau benda. Contoh ilustrasi interval waktu antara lain jumlah kendaraan roda empat yang melintasi Jalan ABC dalam satu jam atau jumlah bayi baru lahir yang dirawat di rumah sakit per bulan. Suatu eksperimen yang menghasilkan variabel acak X yang mewakili jumlah hasil selama interval waktu tertentu atau dalam "area" atau "luas" tertentu disebut sebagai eksperimen Poisson. Karakteristik distribusi Poisson meliputi:

1. Keberhasilan dalam satu interval tidak dipengaruhi oleh interval lain.
2. Peluang terjadinya satu percobaan dalam suatu selang waktu kecil (yang jarang terjadi).
3. Peluang terjadi lebih dari satu keberhasilan dalam selang waktu singkat tersebut dapat diabaikan

Suatu variabel acak x mempunyai distribusi poisson bila (untuk suatu $\mu > 0$, disebut parameter distribusi)

$$\boldsymbol{p(x)} = \begin{cases} \dfrac{\boldsymbol{\mu^x e^{-\mu}}}{\boldsymbol{x!}}; \boldsymbol{x = 0, 1, 2, \ldots} \\ \boldsymbol{0; lainnya} \end{cases}$$

Dalam hal ini, μ menyatakan rata-rata keberhasilan percobaan
Dengan,
e = 2,71828
μ = rata – rata sukses = n . p
x = Banyaknya sukses sampel atau variabel random diskrit (1,2,3, . . , x)
n = Jumlah populasi
P = probabilitas berhasil

Contoh :

1. Dua ratus penumpang telah melakukan pemesanan tiket untuk sebuah penerbangan internasional. Jika probabilitas penumpang yang sudah memesan tiket tidak datang adalah

0,007, berapakah probabilitasnya bahwa dua penumpang tidak akan datang?
Penyelesaian:
Diketahui:
$n = 200$
$p = 0{,}007$
$x = 2$
$\mu = n.p = 200.0{,}007 = 1{,}4$
Jawab:
$p(x) = \frac{\mu^x . e^{-\mu}}{x\,!}$
$= \frac{(1{,}4)^2 (2{,}71\ldots)^{-1{,}4}}{2!}$
$= \frac{(1{,}96)(0{,}247)}{2}$
$= 0{,}2421$

2. Diketahui probabilitas terjadinya reaksi syok selama imunisasi dengan vaksin meningitis adalah 0,0005. Jika di suatu kota terdapat 4000 orang yang menjalani vaksinasi, hitunglah probabilitas terjadinya syok pada tepat tiga orang.

   Jawab
   Diketahui : n = 4000 p = 0,0005
   μ = n.p
   = 4000.0,0005
   = 2
   $p(x) = \frac{e^{-\mu}.\ \mu^x}{x!}$
   $p(x = 3) = \frac{2{,}71828^{-2}.\ 2^3}{3!}$
   $= 0{,}18404\ atau\ 18{,}04\%$

5.2.4 Distribusi Geometrik

Distribusi geometrik adalah distribusi probabilitas yang menggambarkan keberhasilan pertama kali dalam rangkaian kegagalan. Dengan kata lain, dalam situasi di mana upaya-upaya independen dan berulang menghasilkan keberhasilan dengan

peluang p, dan kegagalan dengan peluang q = 1 - p, distribusi ini menunjukkan probabilitas variabel acak yang menyatakan jumlah upaya yang diperlukan hingga mencapai keberhasilan pertama.

Contohnya, pertimbangkan situasi ujian kenaikan tingkat di mana seseorang akan terus mengikuti ujian berulang kali sampai mereka berhasil lulus. Namun, setelah berhasil lulus sekali, prosesnya dianggap selesai.

Secara konsep, distribusi geometri mencerminkan suatu percobaan acak yang melibatkan,

1. Dilakukan berulang-ulang
2. Kejadian antar ulangnya saling bebas
3. Probabilitas keberhasilan pada setiap percobaan sama, yaitu p, dengan $0 < P < 1$
4. Percobaan ini diulang sampai diperoleh satu keberhasilan.

Distribusi geometrik merupakan varian khusus dari distribusi binomial negatif yang melibatkan n percobaan dan berakhir saat keberhasilan pertama ditemukan. Suatu variable acak X mempunyai distribusi Geometrik jika (untuk suatu $0 \leq p \leq 1$)

$$p\,(X = x = p(x) = \begin{cases} p(1-p)^{x-1} \\ 0\,;\, lainnya \end{cases} ; x = 1$$

Contoh :

Pengujian hasil pengelasan logam mencakup proses pengelasan hingga terjadi patahan. Dalam jenis pengelasan tertentu, patahan terjadi karena logam itu sendiri sebanyak 80%, dan 20% disebabkan oleh penyinaran pada pengelasan. Beberapa hasil pengelasan diuji, dan kita ingin menghitung peluang menemukan patahan pertama pada pengujian ketiga.

Jawab :

Dengan menggunakan distribusi geometrik dengan x = 3, p =0,2 dan q=1- 0,2 = 0,8.

Maka diperoleh :

$g\ (x : p) \quad = pqX - 1$

$g\ (3 : 0.2) = 0.2(0.8)2 \quad = 0.128$

LATIHAN

1. Sebuah pengiriman berisi 7 pesawat televisi, di mana 2 di antaranya rusak. Sebuah hotel membeli 3 pesawat televisi dari pengiriman tersebut secara acak. Jika X menyatakan jumlah pesawat televisi rusak yang dibeli oleh hotel, sampaikan hasilnya dalam bentuk distribusi peluang.
2. Sebuah koin dilempar sebanyak 3 kali. Jika kita menyatakan X sebagai jumlah kemunculan muka, apakah X merupakan suatu variabel acak? Dan tentukan ruang sampel dari X.
3. Dari sebuah kantong yang memuat 5 kelereng hitam dan 3 kelereng hijau, dilakukan pengambilan 4 kelereng secara acak satu per satu dengan pengembalian setiap kali sebelum pengambilan berikutnya. Tentukan distribusi peluang untuk jumlah kelereng hijau yang berhasil diambil.
4. Selama jam-jam sibuk, seringkali sulit untuk mendapatkan layanan telepon melalui jaringan. Misalkan diketahui bahwa probabilitas berhasil mendapatkan koneksi pada jam sibuk adalah 5%. Berapakah probabilitasnya bahwa, dalam jam sibuk, diperlukan 5 percobaan panggilan agar berhasil terkoneksi?
5. Misalkan 4% dari total uang di ATM merupakan uang palsu. Tentukan probabilitasnya terdapat 2 lembar uang palsu dari 100 lembar uang yang diambil secara acak dari ATM tersebut.
6. Sebuah studi yang dilakukan di Universitas Jakarta dan Institut Kesehatan Nasional menginvestigasi pandangan masyarakat terhadap obat penenang. Temuan dari penelitian ini menunjukkan bahwa sekitar 70% dari penduduk meyakini bahwa obat penenang tidak mengobati kondisi apapun; sebaliknya, obat tersebut hanya menyembunyikan gejala penyakit sebenarnya. Berdasarkan penelitian ini, berapakah peluang bahwa paling tidak 3 dari 5 penderita yang dipilih secara acak akan memiliki pandangan serupa?

DAFTAR PUSTAKA

BAB 6
VARIABEL ACAK KONTINU

Oleh Retno Andriyani

Variabel acak kontinu adalah kumpulan titik penghubungnya yang membentuk garis luruspada suatu garis interval atau untuk garis lurus pada garis A. Nilainya berupa pecahan atau bilangan bulat. variabel acak kontinu membuat sesuatu garis lurus jika ditampilkan dalam satu garis interval, yang akan menjadi sebuah rangkaian dari titik kontinu.

6.1 Probabilitas Variabel Acak Interval Berbentuk Kontinu

Dalil

Jika X adalah variabel variabelacak dan a dan b adalah dua bilangan real konstan dengan a < b, maka:

$$P(a \leq x \leq b) = a \leq x \leq b = a$$

Contoh Pertanyaan

1. Misalkan S ruang sampel mata uang yg di tos tiga kali . " Banyak wajah " atau "X banyak wajah " M“ adalah wajah dan B adalah belakang. seimbang sisi wajah (M) dan belakang (B). tentukan itu fungsi X dalam Uang X:S R
   a. Ambil Sebuah ruang peluang dari (S, 2s, P). Periksa apakah X merupakan variabel acak pada S. Apakah X kontinu atau diskrit?
   b. Jika Sx range X dan Ωx = 2sx yang dirumuskan (Sx , Ωx , dan P x) Sebuah ruangan peluang jika Sx rentang X dan bidang Ωx = 2sx kejadian di Sx . Px ({12})

Solusi:

a. S {BBB, BBM, BMB, MBB, BMM, MBM, MMB, MMM} S adalah terpisah . Seragam, dengan N (S) = 8 dan P(s) = 8 1,, s € S

Dalam hal ini adalah X(BBB) = X(BBB) = 0
X (BBM) = X (BMB) = X (MBB) = 1
X (BMM) = X (MBM) = X (MMB) = 2 dan
X(MMM) = 3

Dapat didemonstrasikan Bagaimana asosiasi X: S → R mendefinisikan fungsi A secara pasti . Menurut ke definisi 1, maka X adalah variabel acak pada S, dan Rx = Sx = rentang X = {0, 1, 2, 3 , } . Karena S, terhitung , X adalah diskrit variabel acak .

b. Karena N (Sx = 4, maka N (Ωx) = N (2sx) = 24 = 16, dan Ωx = { , Sx , {0}, {.....{3}, {0, 1}, ..., {1, 2, 3, }}. Sehingga _ dengan definisi 2, fungsi X benar-benar variabel acak . Untuk contoh , A = {1, 2} € Ωx , maka Px (A) = P (X4 (A)) = P {BBM, BMB, MBB, BMM, MBM, MMB} = 4 3

2. Variabelacak X: SR adalah didefinisikan dengan range Sx = {x | 0 ≤ x ≤ fungsi f: R→ R.di defenisikan:
f(x) = ...
   a. Tunjukkan , itu fa fkp dari X
   b. Hitung P [- 1 < x ≤ 1]
   c. Hitung P [X = Saya]

Penyelesaian :

a. Karena Sx = [0,2] adalah kontinu, maka X adalah variabel acak kontinu . f (x) 1 – ½ x ≥ 0, untuk 0 ≤ x ≤ 2, dan f (x) = 0, untuk x < 0 atau x > 2. Artinya bahwa f (x) ≥ 0, ∀ x ∈ R, dan karena x ∈ R, dan karena ∫_R▒ 〖 f(x)dx=∫_(-x)^0▒ 〖 0 dx+∫_0^2▒ 〖 (1-1/2 x)dx+∫_2^0▒ 〖 0 dx=x-1 /4 〖 x 〗 ^2 〗 〗 〗 〗 maka f adalah A fkp dari X

b. P [- 1 < x 1] = P[- 1 < x 1] = ∫_4^1 ▒ 〖 f(x)dx=∫_1^0▒ 〖 0 dx+∫_0^21▒ 〖 (1-1/2x)dx=x-1/(4x^2)=3/4 〗 〗 〗
c. P (X = 1) = 0

6.2 Distribusi Fungsi Kumulatif Kontinyu

Jika X adalah variabel kontinu acak, misalnya, distribusi fungsi kumulatif dari X akan berbentuk sebagai berikut :

$$F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(t)\, dt$$

Dengan f(t) adalah fungsi densitas dari X pada t dimana f(t) mewakili k densitas fungsi dari X pada t. Distribusi fungsi nilai untuk variabel acak kontinu seringkali berbentuk konstanta dan fungsi. Grafik distribusi fungsi Ada banyak pilihan untuk variabel acak kontine, termasuk: garis yang berimpit dengan kurva dan sumbu datar; garis berimpit dengan sumbu datar, garis lurus, dan garis sejajar dengan sumbu datar; atau garis berimpit dengan sumbu datar, garis melengkung, dan garis sejajar dengan sumbu datar.

Notasi untuk variabel fungsi distribusi acak kontinu perlu penulisan. Untuk fungsi distribusi digunakan huruf F besar sebagai notasinya; Namun, huruf besar G dapat juga digunakan H, atau K, diikuti dengan nilai variabelacak, dan lebih baik disesuaikan dengan fungsi notasi densitasnya.

1. Notasi fungsi distribusi fungsi densitas variabel acak Y harus G(y) jika ditampilkan dengan g(y).
2. jika ditampilkan dengan f(y), maka F(y) harus digunakan
3. jika ditampilkan dengan h(y), maka H(y) harus digunakan.
4. Distribusi variabel acak Y harus mempunyai fungsi notasi K(y) jika densitas fungsinya adalah k(y).

Sebelumnya telah dijelaskan bahwa peluang fungsi atau densitas fungsi dapat digunakan untuk menghitung peluang dari suatu variabel acak yang memiliki nilai interval. Selain itu, peluang nilai berdasarkan distribusi fungsi dapat diperoleh, baik yang

bersifat diskrit maupun kontinu. Masalah ini dapat diselesaikan dengan menggunakan rumus:

$$P(a \leq X \leq b) = F(b) - F(a)$$

dimana a < b dan a dan b adalah dua bilanhan real.

Rumusnya di bawah ini Bisa menjadi digunakan menentukan peluang variabel acak:

$$P(X = b) = FX(b) - FX(b-)$$

Kami Bisa mengetahui distribusi suatu fungsi jika peluangnya atau densitas variabel acak diketahui. D sisi lain, distribusi fungsi dapat ditemukan jika fungsi peluang atau fungsi densitas variabel acak tidak diketahui:

Contoh soal:

1. Misalnya variabel acak X memiliki FMP sebagai berikut :

$$f(x) = \begin{cases} \frac{x}{6} ; x = 1,2,3 \\ 0 ; x \text{ yang lainnya} \end{cases}$$

Tentukan fungsi distribusi kumulatif dari X

Solusi:

a. Untuk $-\infty < X < 1$

F(X) = 0

b. Untuk $1 \leq X < 2$

F(1) = P($-\infty < X < 1$) + P(X=1)

F(1) = 0 +1/6 = 1/6

c. Untuk $2 \leq X < 3$

F(2) = P($-\infty < X < 1$) + P(X=1) + P($1 < X < 2$) + P(X=2)

F(2) = 0 + 1/6 + 0 + 2/6 = 3/6

d. Untuk $P(3 \leq X < \infty)$

F(3) = P($-\infty < X < 1$)+P(X=1)+P($1 < X < 2$) + P(X=2) + P($2 < X < 3$) + $P(X = 3)$

F(3) = 0 + 1/6 + 0 + 2/6 + 0 + 3/6 = 1

2. Misalnya variabel acak X memiliki FKP berikut:

$$f(x) = \begin{cases} \frac{3x^2}{8}; 0 < x < 2 \\ 0 ; x\ yg\ lain \end{cases}$$

Tentukan fungsi distribusi kumulatif dari X!

Solusi:

1) $-\infty < X < 0$

   $F(X) = 0$

2) $0 \leq X < 2$

   $F(X) = \int_{-\infty}^{0} f(t)dt + \int_{0}^{x} f(t)dt$

   $F(X) = \int_{-\infty}^{0} 0\ dt + \int_{0}^{x} \frac{3t^2}{8}\ dt = [\frac{1}{8}t^3]_0^x = \frac{1}{8}x^3$

3) $2 \leq X < \infty$

   $F(X) = P(X \leq x)$

   $F(X) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{0} f(t)dt + \int_{0}^{2} f(t)dt + \int_{2}^{x} f(t)dt$

   $F(X) = 0+1+0 = 1$

Sehingga:

$$F(X) = \begin{cases} 0 ; -\infty < X < 0 \\ \frac{1}{8}x^3; 0 \leq X < 2 \\ 1 ; 2 \leq X < \infty \end{cases}$$

6.3 Peluang Distribusi Kontinu

Distribusi peluang kontinu adalah variabel acak yang menerima setiap nilai pada skala kontinu. Jika ada banyak titik sampel dalam satu ruang sampel, itu disebut ruang sampel kontinu. Distribusikontinu berlaku jika itu fungsi f(x) adalah fungsi densitas variabel peluang acak kontinu X yang ditentukan merupakan semua bilangan real R. $f(x) \geq 0\ untuk\ semua\ x \in R$

1. $\int_{\infty}^{\infty} f(x)dx = 1$
2. $p(a < X < b) = \int_{\infty}^{\infty} f(x)dx$

Distribusi Normal atau " Distribusi Gaussian"

Distribusi normal adalah distribusi yang luas dan konsisten yang digunakan dalam berbagai konteks. Karena distribusi ini berkesinambungan, hal ini memerlukan terus menerus pengukuran

dari variabel, seperti tinggi badan, berat badan, IQ, angka tetesan hujan curah , ujian hasil , dll.

Definisi

Fungsi densitas normal _ secara acak _ variabel x telah diatur sebelumnya nilai rata -rata dan penyimpangan σ. Jika

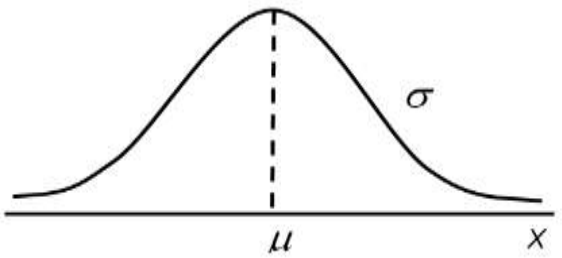
$$f\ (x) = \frac{1}{\sigma\sqrt{2x}} e^{-1/2(\frac{x-\mu}{\sigma})^2}$$

N $(x,\ \mu,\sigma^2)$ merupakan notasi penulisan untuk bilangan variabel acak dengan distribusi normal, yang mana variabel acak x berdistribusi normal dengan rata - rata dari μ dan varians dari σ^2.

Parameter:

1. E (x) =σ (rata-rata)
2. Var (x) = σ^2 (varians)
3. M x (t) = e ($\frac{\mu t+\sigma^2 t^2}{2}$); t ϵ R (Fungsi generator momen)

Berikut penjelasan dari kurva distribusi normal



Distribusi Normal Standar

Definisi

Sebuah variabel x memiliki fungsi densitas normal bsku dengan parameter 0 dan standar deviasi 1. Jika:

$$f\ (x) = \frac{1}{\sqrt{2x}} e^{-1/2(x)^2}$$

Parameter:

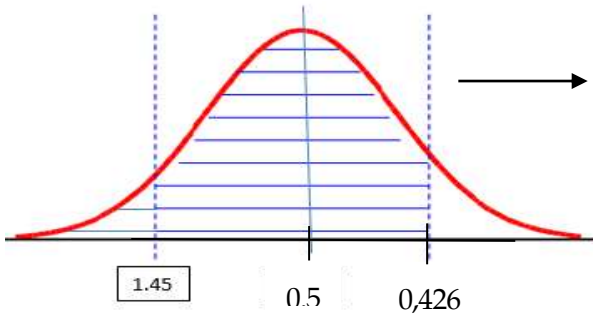
1. E (x) =$\mu = 0$
2. Var (x) = $\sigma^2 = 1$
3. M x (t) = e ($\frac{1}{2}t^2$); t ϵ R

Contoh **Soal**

Tentukan peluang suatu variabel acak Z yang berdistribusi normal tipikal mempunyai tanda.:

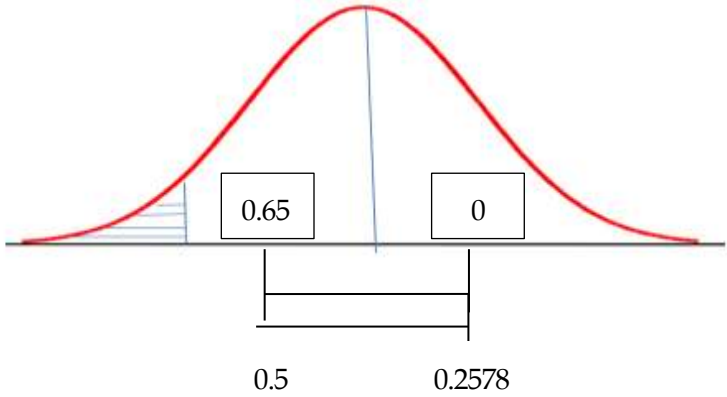
1. Kurang dari 1,45

   Menjawab



Nilai 0.4265 adalah nilai dari z = 0 ke z = 1.45 didapat dari tabel normal standar (tabel Z)

2. Kurang dari – 0,65



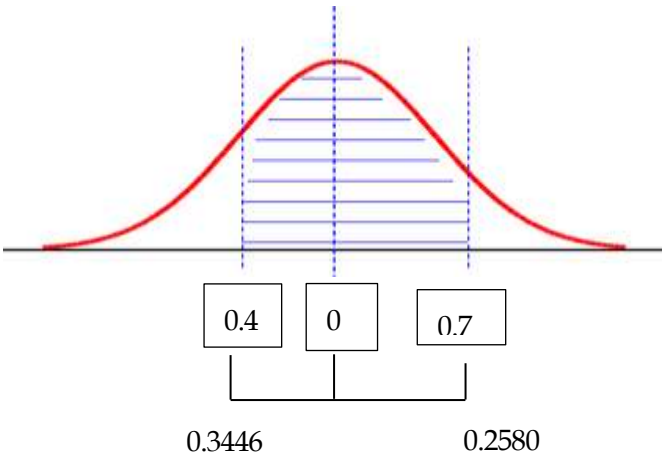
Dengan menggunakan tabel Z diperoleh :

LD = 0,5 – 0,2578

= 0,2422

Jadi luas areanya tidak cukup - 0,65 adalah 0,2422

3. Antara – 0,4 dan 0,7



Dengan menggunakan tabel Z diperoleh :

LD = 0,3446 + 0,2580

= 0,6026

Jadi luas area antara -0,40 dan 0,7 adalah 0,6026

LATIHAN

1. X ~N (65.36). Hitunglah ;
   a. P(X=36)
   b. P(X > 70)
   c. P(X <50)
   d. P(55 < x < 72)
   e. P (|X- μ|≤9 **)**
2. Hitunglah luas daerah di bawah _kurva berbatasI _ antara z = -1,97 dan z = 0,86!
3. Aki mobil biasanya berumur 3,0 tahun dengan standar deviasi 0,5 tahun. Dengan asumsi usia berdistribusi normal, ada peluangumur baterai kurang dari 2,3 tahun.
4. Koin dilempar lima ratus kali. Tentukan peluangjumlah kepala No akan melebihi 250, Sebanyak
   a. lebih dari 10
   b. Lebih dari 30

5. Pabrik pipa air memproduksi pipa dengan ukuran panjang 6m. setelah diukur dengan teliti ternyata panjang pipa yang dihasilkan disana rata-rata 599,5 cm. Dengan simpangan baku σ = 0,5 cm. Pipa terpanjang adalah 601 cm. Jika diambil secara acak satu pipa maka Berapa Banyak Pipa yang Mungkin : _
   a. Memiliki panjang tidak cukup dari 599 cm
   b. Memiliki panjang lebih dari 601 cm
6. Perusahaan listrik Amir memproduksi bola lampu yang bertahan seumur hidup berdistribusi normal dengan rata-rata 800 jam dan standar deviasi 40 jam, coba hitung peluang sebuah bola lampu dapat menyala antara 778 dan 834 jam.

DAFTAR PUSTAKA

BAB 7
TEOREMA PROBABILITAS

Oleh Ukta Indra Nyuswantoro

7.1 Konsep dasar probabilitas

Probabilitas adalah cabang yang sangat mendasar dalam bidang matematika yang secara khusus memusatkan perhatiannya pada analisis kemungkinan suatu kejadian terjadi. Di inti konsep dasar probabilitas ini terdapat fokus pada pengukuran peluang atau kecenderungan suatu peristiwa. Dalam setiap situasi, peluang dapat dilihat sebagai suatu angka yang mewakili sejauh mana kejadian tersebut mungkin terjadi. Dengan kata lain, probabilitas memberikan kerangka kerja untuk mengukur ketidakpastian. Konsep ini bukan hanya sekadar teori matematika, tetapi juga menjadi dasar penting bagi berbagai disiplin ilmu seperti statistik, kecerdasan buatan, ekonomi, dan berbagai bidang lainnya (Ross, 2014).

Penting untuk memahami bahwa probabilitas erat kaitannya dengan ketidakpastian. Kita menggunakan konsep ini untuk menggambarkan dan memahami situasi di mana hasil yang pasti tidak dapat diprediksi dengan pasti. Misalnya, dalam konteks pelemparan dadu, probabilitas memberikan alat untuk memprediksi kemungkinan hasil tertentu, seperti angka mana yang kemungkinan besar akan muncul. Dengan demikian, probabilitas tidak hanya menyediakan kerangka kerja matematis, tetapi juga memberikan wawasan yang sangat berguna dalam menghadapi situasi ketidakpastian dalam berbagai konteks kehidupan (Dewadi et al., 2023).

Contoh konkret lainnya dapat ditemukan dalam teori permainan, di mana pemahaman probabilitas memainkan peran kunci dalam merancang strategi yang efektif. Dalam kehidupan sehari-hari, ketika kita membuat keputusan berdasarkan informasi terbatas, probabilitas membantu kita mengevaluasi risiko dan memperhitungkan kemungkinan hasil yang berbeda. Dengan

demikian, konsep dasar probabilitas memiliki dampak signifikan dalam membentuk cara kita memahami dan berinteraksi dengan dunia di sekitar kita.

Pengenalan terhadap konsep dasar probabilitas membawa kita ke inti dari analisis kemungkinan dan ketidakpastian. Probabilitas memberikan fondasi yang kokoh dalam memahami dan memodelkan situasi di mana hasil tidak dapat diprediksi dengan pasti. Dengan aplikasi yang meluas dari ilmu statistik hingga kecerdasan buatan, probabilitas bukan hanya menjadi alat matematis, tetapi juga menjadi landasan utama untuk pengambilan keputusan di berbagai bidang kehidupan (Alfaris, Dewadi, et al., 2022).

7.2 Sejarah dan perkembangan teorema probabilitas

Sejarah teorema probabilitas mencerminkan perjalanan perkembangan konsep ini melalui kontribusi penting tokoh-tokoh berpengaruh. Salah satu pionir dalam pengembangan probabilitas adalah Blaise Pascal, yang hidup pada abad ke-17. Bersama Fermat, Pascal memulai diskusi awal tentang teori probabilitas, khususnya dalam konteks perjudian. Kontribusi ini kemudian menjadi dasar bagi perkembangan lebih lanjut dalam abad ke-18 oleh Pierre-Simon Laplace. Laplace mengembangkan teori probabilitas dengan menggabungkan prinsip-prinsip matematika dengan pemikiran filosofis (Agustianti et al., 2022).

Pada abad ke-20, teorema probabilitas mengalami kemajuan signifikan. Andrey Kolmogorov, seorang ahli matematika Rusia, menyusun dasar-dasar matematika modern untuk probabilitas dalam karyanya "Grundbegriffe der Wahrscheinlichkeitsrechnung" (1933). Kontribusi ini memberikan kerangka kerja aksiomatik yang lebih ketat untuk probabilitas. Selain itu, teorema limit pusat, yang merupakan konsep kunci dalam statistika, diperkenalkan oleh matematikawan seperti Sir Francis Galton dan George Udny Yule pada awal abad ke-20. Teorema limit pusat memainkan peran vital dalam memahami distribusi probabilitas dari hasil pengamatan acak yang besar (Hacking, 2006).

Sejarah teorema probabilitas mencerminkan perjalanan yang menarik dari konsep sederhana hingga menjadi landasan matematika yang kokoh. Kontribusi tokoh-tokoh seperti Pascal, Laplace, Kolmogorov, Galton, dan Yule menciptakan kerangka kerja teorema probabilitas yang kita kenal saat ini. Dengan dasar matematika yang semakin kuat, teorema probabilitas terus berkembang, memainkan peran kunci dalam pemahaman risiko, prediksi, dan pengambilan keputusan di berbagai disiplin ilmu (Alfaris, Sarumaha, *et al.,* 2022).

7.3 Pentingnya probabilitas dalam berbagai bidang

Probabilitas memegang peran krusial dalam berbagai bidang, memberikan landasan untuk pengambilan keputusan yang informasional dan merencanakan strategi. Dalam ranah statistik, probabilitas berfungsi sebagai alat utama untuk membuat inferensi tentang populasi berdasarkan sampel data yang terbatas. Penggunaan probabilitas dalam analisis statistik memungkinkan peneliti dan analis untuk mengukur ketidakpastian dan mengambil keputusan yang didukung oleh dasar matematis yang kuat. Referensi kepada probabilitas dalam kecerdasan buatan menyoroti peran kunci algoritma probabilitas dalam mendukung pembelajaran mesin dan pengambilan keputusan di dunia digital (Alfaris, 2023).

Di sektor keuangan, probabilitas memiliki dampak penting dalam manajemen risiko investasi. Melalui model probabilitas, para profesional keuangan dapat mengevaluasi potensi keuntungan dan risiko yang terlibat dalam suatu investasi. Bidang ilmu sosial, kedokteran, dan teknik juga tidak terkecuali, karena mereka mengandalkan probabilitas untuk melakukan analisis dan prediksi. Dalam ilmu sosial, probabilitas digunakan untuk merumuskan model matematika yang dapat menjelaskan perilaku sosial. Di kedokteran, probabilitas berperan dalam diagnosis dan perencanaan perawatan, sementara di bidang teknik, probabilitas digunakan untuk mengevaluasi keandalan sistem dan merancang solusi yang dapat diandalkan (Ross, 2014).

Pentingnya pemahaman probabilitas tidak hanya terletak pada aplikasinya dalam keilmuan tertentu, tetapi juga pada kontribusinya dalam membentuk dasar untuk pengambilan

keputusan secara umum. Sebagai contoh, dalam merencanakan strategi bisnis, pengusaha sering mengandalkan probabilitas untuk mengidentifikasi peluang dan mengukur risiko. Dengan demikian, pemahaman mendalam terhadap konsep dasar probabilitas menjadi esensial dalam menyikapi kompleksitas dan ketidakpastian dalam berbagai bidang kehidupan (Alfaris, Gustian, *et al.*, 2022).

7.4 Ruang sampel dan kejadian

Ruang Sampel dan Kejadian adalah konsep yang sangat fundamental dalam teorema probabilitas, menyediakan landasan matematis untuk memodelkan dan mengukur peluang berbagai hasil dalam suatu eksperimen acak. Ruang Sampel (sample space) merupakan himpunan semua hasil yang mungkin terjadi dalam suatu eksperimen. Sebagai contoh sederhana, jika kita melempar dadu, ruang sampelnya adalah himpunan {1, 2, 3, 4, 5, 6}, mencakup semua kemungkinan hasil dari lemparan dadu tersebut. Ruang sampel memberikan gambaran komprehensif tentang variasi hasil yang mungkin terjadi (Subakti et al., 2022).

Kejadian (event), di sisi lain, adalah sub-himpunan dari ruang sampel yang mewakili hasil atau serangkaian hasil yang memenuhi suatu kondisi atau kriteria tertentu. Sebagai contoh, kejadian "munculnya angka genap" dalam pelemparan dadu dapat direpresentasikan oleh himpunan {2, 4, 6}. Dalam pemodelan probabilistik, kejadian memungkinkan peneliti untuk fokus pada hasil-hasil tertentu yang relevan dengan tujuan analisis atau prediksi.

Pentingnya ruang sampel dan kejadian menjadi jelas ketika kita mengukur probabilitas. Probabilitas (P) dari suatu kejadian A dapat dihitung dengan membagi jumlah kejadian yang diinginkan dengan jumlah total kejadian dalam ruang sampel. Matematis, ini dapat dirumuskan sebagai P(A) = (Jumlah kejadian A) / (Jumlah ruang sampel). Probabilitas berkisar antara 0 dan 1, di mana nilai 0 menunjukkan kejadian tersebut tidak mungkin terjadi, sementara nilai 1 menunjukkan kejadian tersebut pasti terjadi.

7.5 Aksioma probabilitas

Aksioma probabilitas adalah seperangkat aturan dasar yang memberikan fondasi matematis bagi konsep probabilitas dalam teorema probabilitas. Tiga aksioma utama membentuk landasan yang kritis untuk memastikan konsistensi dan validitas dalam perhitungan probabilitas. Pertama, aksioma non-negativitas menyatakan bahwa probabilitas suatu kejadian tidak dapat lebih kecil dari nol, yakni $P(A) \geq 0$ untuk setiap kejadian A. Kedua, aksioma normalisasi memastikan bahwa total probabilitas dari semua kejadian dalam ruang sampel sama dengan 1, dinyatakan sebagai $\sum P(A) = 1$, di mana $\sum$ melibatkan semua kejadian yang mungkin (Jaynes, 2003). Ketiga, aturan penjumlahan probabilitas menyatakan bahwa probabilitas dari suatu kejadian gabungan dari dua kejadian saling eksklusif dapat dihitung dengan menjumlahkan probabilitas masing-masing, yaitu $P(A \cup B) = P(A) + P(B)$.

Aksioma non-negativitas memastikan bahwa probabilitas selalu berada dalam rentang antara 0 dan 1, mencerminkan kemungkinan terjadinya suatu kejadian. Aturan normalisasi menjamin bahwa total probabilitas dalam ruang sampel setara dengan kepastian, mendukung interpretasi probabilitas sebagai ukuran ketidakpastian atau keyakinan dalam suatu konteks eksperimen. Selain itu, aturan penjumlahan probabilitas memberikan kerangka kerja matematis yang kokoh untuk menghitung probabilitas dari berbagai skenario eksperimen acak (Octavianti et al., 2022).

Penerapan aksioma probabilitas sangat luas dalam berbagai bidang, termasuk statistik, teori permainan, dan kecerdasan buatan. Dalam analisis statistik, aksioma ini menjadi dasar untuk pengembangan model probabilistik yang mendukung pengambilan keputusan dan inferensi. Di teori permainan, aksioma probabilitas membantu merinci strategi optimal dalam situasi ketidakpastian. Dengan memahami dan menerapkan aksioma probabilitas, ilmuwan dapat memanfaatkan alat matematis ini untuk memahami dan meramalkan ketidakpastian dalam berbagai konteks ilmiah dan praktis (Muttaqin et al., 2023).

7.6 Variabel Acak dan Distribusi Probabilitas

Variabel acak dan distribusi probabilitas adalah konsep-konsep kunci dalam teorema probabilitas yang memungkinkan para peneliti untuk memodelkan dan menganalisis ketidakpastian serta variasi dalam berbagai fenomena. Memahami kedua konsep ini membuka pintu untuk pemahaman yang lebih mendalam tentang peluang dan perilaku acak, dengan aplikasi yang meluas dari statistik hingga kecerdasan buatan.

Variabel acak adalah suatu kejadian atau nilai yang dapat memiliki berbagai hasil dalam suatu eksperimen atau situasi tertentu. Dalam matematika, variabel acak dapat berupa diskrit atau kontinu. Variabel acak diskrit memiliki sejumlah nilai yang terbatas dan terpisah, sementara variabel acak kontinu dapat mengambil nilai dalam suatu rentang yang kontinu. Sebagai contoh, pelemparan dadu dapat dianggap sebagai variabel acak diskrit, sementara suhu udara pada suatu waktu tertentu dapat dianggap sebagai variabel acak kontinu.

Distribusi probabilitas menggambarkan sejauh mana nilai-nilai yang mungkin dari suatu variabel acak mungkin terjadi. Dalam kasus variabel acak diskrit, distribusi probabilitas diwakili oleh fungsi massa probabilitas (probability mass function/PMF). Contoh dari variabel acak diskrit adalah distribusi Poisson, yang sering digunakan untuk menggambarkan jumlah kejadian langka dalam suatu interval waktu tertentu. Di sisi lain, untuk variabel acak kontinu, kita menggunakan fungsi densitas probabilitas (probability density function/PDF), yang menggambarkan peluang nilai tertentu jatuh dalam suatu rentang. Distribusi normal adalah contoh klasik dari distribusi probabilitas kontinu (Casella & Berger, 2002).

Fungsi massa probabilitas (PMF) memberikan probabilitas bahwa variabel acak diskrit akan mengambil nilai tertentu. Sebagai contoh, PMF dapat memberikan probabilitas bahwa dadu yang dilempar akan menunjukkan angka tertentu. Sebaliknya, fungsi densitas probabilitas (PDF) memberikan peluang bahwa variabel acak kontinu akan jatuh dalam suatu rentang nilai. Misalnya, PDF dapat memberikan peluang bahwa suhu pada suatu waktu akan berada dalam suatu interval tertentu.

Harapan matematik (*expected value*) dan varians adalah ukuran-ukuran yang memberikan informasi tentang pusat dan sebaran distribusi probabilitas. Harapan matematik adalah nilai rata-rata dari suatu variabel acak dan dihitung dengan menjumlahkan produk nilai-nilai dengan probabilitas masing-masing nilai. Varians mengukur seberapa jauh nilai-nilai tersebut tersebar dari rata-rata. Kedua konsep ini memberikan pemahaman yang lebih dalam tentang karakteristik distribusi probabilitas.

Dalam penelitian probabilitas, pemahaman konsep variabel acak dan distribusi probabilitas memiliki aplikasi luas. Banyak model matematika dan statistik yang bergantung pada konsep ini untuk menggambarkan variasi dan ketidakpastian dalam data. Sebagai contoh, dalam analisis risiko keuangan, variabel acak dan distribusi probabilitas dapat digunakan untuk meramalkan variasi harga saham atau nilai mata uang.

7.7 Teorema Probabilitas Kondisional dan Independensi

Teorema probabilitas kondisional dan independensi merupakan aspek penting dalam teorema probabilitas, memungkinkan peneliti dan ilmuwan untuk memodelkan dan menganalisis hubungan antara kejadian-kejadian dalam suatu konteks. Pemahaman mendalam terhadap probabilitas kondisional, teorema Bayes, dan independensi memberikan dasar yang kuat untuk menganalisis situasi ketika kejadian bergantung pada kejadian lainnya (Grinstead & Snell, 1997).

1. Probabilitas Kondisional:
   Probabilitas kondisional adalah probabilitas suatu kejadian yang terjadi, diketahui bahwa kejadian lain telah terjadi. Secara matematis, probabilitas kondisional dari kejadian A, diberikan kejadian B, dilambangkan sebagai P(A | B). P(B) adalah probabilitas kejadian B. Pemahaman probabilitas kondisional memungkinkan analisis yang lebih mendalam tentang bagaimana informasi tambahan dapat memengaruhi estimasi peluang.

2. Teorema Bayes:
   Teorema Bayes adalah alat yang sangat kuat dalam menganalisis probabilitas kondisional, memungkinkan kita untuk membalikkan informasi dan menghitung probabilitas kejadian awal berdasarkan informasi baru. Teorema Bayes dirumuskan sebagai berikut:

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

   Teorema ini memberikan cara sistematis untuk memperbarui keyakinan kita tentang suatu kejadian setelah mendapatkan informasi tambahan. Dalam konteks medis, teorema Bayes dapat digunakan untuk menghitung probabilitas penyakit berdasarkan hasil tes diagnostik.
3. Independensi Kejadian:
   Kejadian A dan B dikatakan independen jika terjadi $P(A \cap B) = P(A).P(B)$. Dengan kata lain, kejadian A tidak memengaruhi kejadian B dan sebaliknya. Probabilitas kondisional dalam konteks kejadian independen menjadi $P(A/B)=P(A)$ dan $P(B/A) = P(B)$. Memahami independensi sangat penting dalam analisis probabilitas, karena memberikan indikasi bahwa dua kejadian beroperasi secara terpisah satu sama lain.
4. Aplikasi Teorema Probabilitas Kondisional dalam Kehidupan Nyata:
   Aplikasi teorema probabilitas kondisional dapat ditemukan dalam berbagai aspek kehidupan sehari-hari. Misalnya, dalam industri keuangan, probabilitas kondisional dapat digunakan untuk menghitung probabilitas gagal bayar kredit pelanggan berdasarkan faktor-faktor tertentu seperti kondisi ekonomi. Pemahaman ini memberikan landasan untuk pengelolaan risiko yang lebih efektif.
   Pemahaman yang mendalam tentang probabilitas kondisional dan teorema Bayes juga memainkan peran vital dalam kecerdasan buatan, khususnya dalam pengembangan sistem pembelajaran mesin. Algoritma pembelajaran mesin sering kali

bergantung pada probabilitas kondisional untuk membuat prediksi dan pengambilan keputusan yang lebih baik.

Sebagai contoh, dalam pengolahan bahasa alami, probabilitas kondisional dapat digunakan untuk menentukan arti suatu kata berdasarkan konteks kalimat. Pemahaman yang baik tentang teorema Bayes juga dapat diterapkan dalam pengenalan pola, seperti pengenalan wajah dalam kecerdasan buatan.

Penting untuk dicatat bahwa independensi kejadian, meskipun sering diinginkan, tidak selalu terpenuhi dalam konteks kehidupan nyata. Oleh karena itu, peneliti perlu mempertimbangkan dengan cermat apakah dapat dianggap bahwa dua kejadian independen dalam suatu analisis.

Dalam kesimpulan, teorema probabilitas kondisional dan independensi memberikan landasan matematis untuk analisis probabilistik yang mendalam. Pemahaman konsep-konsep ini membuka pintu untuk pengembangan model yang lebih canggih, analisis data yang lebih akurat, dan aplikasi dalam berbagai disiplin ilmu. Dengan memanfaatkan prinsip-prinsip probabilitas kondisional dan teorema Bayes, para peneliti dapat membawa pemahaman kita tentang ketidakpastian dan hubungan antar kejadian ke tingkat yang lebih tinggi, memainkan peran penting dalam pengambilan keputusan dan prediksi dalam dunia yang penuh dengan ketidakpastian ini.

7.8 Variabel Acak Bersama dan Distribusi Gabungan

Variabel acak bersama dan distribusi gabungan adalah konsep-konsep kunci dalam teori probabilitas dan statistika yang memungkinkan kita untuk memodelkan hubungan antara dua atau lebih variabel acak. Dalam konteks ini, mari kita bahas masing-masing konsep dengan lebih rinci.

1. Pengertian Variabel Acak Bersama:
   Variabel acak bersama merujuk pada dua atau lebih variabel acak yang terkait satu sama lain dalam suatu eksperimen atau situasi tertentu. Dua variabel acak bersama umumnya disimbolkan sebagai X dan Y, dan dapat merepresentasikan

berbagai situasi, seperti hasil pelemparan dua dadu atau suhu dan kelembaban udara pada suatu waktu tertentu. Pemahaman tentang hubungan antar variabel acak bersama memungkinkan kita untuk menganalisis dan memodelkan situasi yang lebih kompleks.

2. Fungsi Distribusi Gabungan:
   Fungsi distribusi gabungan, atau joint distribution function, menggambarkan probabilitas bersama dari dua atau lebih variabel acak. Fungsi ini umumnya dilambangkan sebagai $F_{XY}(x,y)$ dan memberikan probabilitas bahwa X kurang dari atau sama dengan x dan Y kurang dari atau sama dengan y. Fungsi distribusi gabungan mencakup seluruh rentang nilai yang mungkin untuk kedua variabel acak bersama.
3. Marginal dan Kondisional Distribusi:
   Marginal Distribusi: Marginal distribution dari satu variabel acak dalam variabel acak bersama adalah distribusi probabilitas tunggal dari variabel tersebut, tanpa memperhatikan nilai variabel lainnya. Misalnya, marginal distribution dari X disimbolkan sebagai $F_{XY}(x)$ dan diperoleh dengan menjumlahkan atau mengintegrasikan probabilitas bersama $F_{XY}(x,y)$ untuk semua nilai y.
4. Ketergantungan dan Korelasi:
   Ketergantungan antara variabel acak bersama merujuk pada hubungan statistik atau probabilitas di antara mereka. Ketergantungan dapat bersifat positif, di mana peningkatan nilai satu variabel berkaitan dengan peningkatan nilai variabel lainnya, atau negatif, di mana peningkatan nilai satu variabel berkaitan dengan penurunan nilai variabel lainnya. Ketergantungan dapat dievaluasi dengan menggunakan kondisional distribusi.
   Korelasi adalah ukuran numerik dari hubungan linier antara dua variabel acak. Koefisien korelasi Pearson, sering disimbolkan sebagai ρ, berkisar dari -1 hingga 1. Nilai positif menunjukkan korelasi positif, nilai negatif menunjukkan korelasi negatif, dan nilai nol menunjukkan ketiadaan korelasi linier. Korelasi memberikan gambaran tentang sejauh mana

perubahan dalam satu variabel berkaitan dengan perubahan dalam variabel lainnya.

Dalam pengembangan model statistika dan analisis data, pemahaman tentang ketergantungan dan korelasi antar variabel acak bersama sangat penting. Ini membantu para peneliti dan analis untuk membuat prediksi yang lebih akurat dan menyelidiki sejauh mana informasi dari satu variabel dapat digunakan untuk memprediksi atau menjelaskan variasi dalam variabel lainnya.

Dengan merinci konsep-konsep tersebut, kita dapat lebih memahami dinamika variabel acak bersama dan distribusi gabungan. Penerapan konsep ini melibatkan perhitungan matematis yang cermat dan analisis statistika yang teliti, yang dapat membantu kita menggambarkan dan memahami kompleksitas hubungan antar variabel acak dalam berbagai konteks ilmiah dan praktis. Konsep-konsep ini merupakan fondasi bagi pengembangan model statistika yang kuat dan pemahaman yang lebih baik tentang variasi dan hubungan dalam data.

7.9 Hukum Besar Bilangan dan Teorema Batas Pusat

Hukum Besar Bilangan (*Law of Large Numbers*) dan Teorema Batas Pusat (*Central Limit Theorem*) adalah dua konsep mendasar dalam statistika yang membentuk landasan bagi banyak metode inferensi statistika dan teori estimasi. Mari kita eksplorasi masing-masing konsep secara rinci.

1. Hukum Besar Bilangan: Lemah dan Kuat:
   Hukum Besar Bilangan Lemah (LLN): Hukum Besar Bilangan Lemah menyatakan bahwa rata-rata aritmatika dari sejumlah besar pengamatan acak akan mendekati nilai harapan dari distribusi tersebut. Secara matematis, untuk setiap μ adalah nilai harapan dari distribusi.
   Hukum Besar Bilangan Kuat (LLN): Hukum Besar Bilangan Kuat lebih kuat daripada versi lemahnya. Hukum ini menyatakan bahwa rata-rata empiris dari sejumlah besar pengamatan acak akan hampir pasti konvergen ke nilai harapan distribusi.

2. Teorema Batas Pusat: Versi Lindeberg-Lévy, Lyapunov, dan Cramér:
   Teorema Batas Pusat (CLT): Teorema Batas Pusat menyatakan bahwa distribusi rata-rata dari sejumlah besar pengamatan acak, tanpa memperhatikan bentuk distribusi awal, akan mendekati distribusi normal saat ukuran sampel *n* meningkat.
   Versi Lyapunov: Menggeneralisasi CLT, Versi Lyapunov memperbolehkan variabel acak tidak identik dan tidak independen, dengan syarat moment ketiga terbatas.
   Versi Cramér: Memberikan batasan lebih ketat untuk kondisi-kondisi di mana distribusi normal muncul dalam pengulangan independen dari variabel acak.
3. Penerapan dalam Statistik dan Teori Estimasi:
   Estimasi Parameter: Hukum Besar Bilangan dan Teorema Batas Pusat memberikan landasan untuk metode-metode estimasi parameter dalam statistika.
   Pengujian Hipotesis: Dalam pengujian hipotesis statistik, terutama ketika sampel besar, distribusi normal sering digunakan sebagai dasar untuk menghitung nilai-nilai kritis dan p-value.

Penerapan dalam Statistika Non-parametrik: Meskipun Hukum Besar Bilangan dan Teorema Batas Pusat berasal dari konteks statistika parametrik, konsep-konsep ini juga memainkan peran penting dalam statistika non-parametrik. Mereka memberikan dasar untuk distribusi sampling dari berbagai statistik, memungkinkan kita untuk membuat inferensi tanpa harus mengasumsikan bentuk tertentu dari populasi (Wackerly et al., 2008).

Pemahaman mendalam tentang Hukum Besar Bilangan dan Teorema Batas Pusat menjadi kunci dalam analisis statistika dan inferensi. Dengan memahami bahwa rata-rata dari sejumlah besar pengamatan acak mendekati distribusi normal, para statistikawan dan peneliti dapat membuat inferensi yang lebih kuat dan membangun dasar yang kokoh untuk pengambilan keputusan berdasarkan data empiris. Konsep ini memiliki dampak yang luas

dalam berbagai disiplin ilmu dan aplikasi praktis dalam dunia statistika.

7.10 Aplikasi Teorema Probabilitas dalam Bidang Lain

Teorema probabilitas memiliki dampak yang signifikan dalam berbagai bidang, membuka pintu untuk analisis statistik yang mendalam, pengembangan algoritma kecerdasan buatan, pengelolaan risiko dalam ekonomi dan keuangan, serta berbagai aplikasi praktis. Mari kita eksplorasi penerapan probabilitas dalam beberapa bidang kunci (Hastie et al., 2009).

1. Probabilitas dalam Statistik dan Analisis Data:
   Probabilitas memainkan peran krusial dalam statistik dan analisis data. Dalam statistik inferensial, probabilitas digunakan untuk membuat inferensi tentang populasi berdasarkan sampel data yang terbatas. Konsep seperti interval kepercayaan dan uji hipotesis bergantung pada teorema probabilitas untuk memberikan dasar matematis bagi keputusan statistik. Misalnya, dalam uji hipotesis, nilai p-nilai dihitung berdasarkan probabilitas distribusi sampel yang diharapkan (Bishop, 2006).
2. Penerapan Probabilitas dalam Ilmu Komputer dan Kecerdasan Buatan:
   Algoritma Probabilistik: Probabilitas digunakan dalam pengembangan algoritma probabilistik dalam ilmu komputer. Algoritma ini memungkinkan model untuk membuat keputusan berdasarkan probabilitas, seperti algoritma Naive Bayes dalam klasifikasi teks atau algoritma Markov Chain Monte Carlo dalam pemodelan statistik.
   Pemelajaran Mesin: Dalam kecerdasan buatan, pemahaman probabilitas membentuk dasar pembelajaran mesin. Algoritma pembelajaran mesin seperti regresi logistik dan mesin vektor dukungan mengandalkan konsep probabilitas untuk membuat prediksi dan mengukur ketidakpastian model.
3. Probabilitas dalam Ekonomi dan Keuangan:
   Manajemen Risiko: Probabilitas memainkan peran kunci dalam manajemen risiko dalam ekonomi dan keuangan. Dalam pengambilan keputusan investasi, analisis probabilitas

membantu dalam mengevaluasi risiko dan pengembalian investasi. Model seperti Black-Scholes dalam opsi saham juga bergantung pada konsep probabilitas.

Analisis Statistik Ekonomi: Dalam ekonomi, analisis probabilitas digunakan untuk memodelkan perilaku konsumen, perkiraan pertumbuhan ekonomi, dan membuat keputusan kebijakan ekonomi berdasarkan proyeksi probabilitas berbagai skenario.

4. Studi Kasus dan Contoh-contoh Praktis:

   Diagnosa Medis: Dalam dunia medis, probabilitas digunakan dalam proses diagnosa. Algoritma probabilitas dapat membantu dokter dalam menentukan kemungkinan diagnosis berdasarkan gejala dan hasil tes.

   Jaringan Sosial: Dalam analisis jaringan sosial, probabilitas dapat digunakan untuk memodelkan hubungan antar individu. Misalnya, model perambatan informasi atau penyebaran virus dalam suatu populasi.

   Analisis Penjualan: Dalam bisnis dan pemasaran, probabilitas dapat digunakan untuk memprediksi perilaku konsumen, mempersonalisasi pengalaman pelanggan, dan mengoptimalkan strategi penjualan.

Penerapan probabilitas dalam berbagai bidang ini menunjukkan keberagaman dan fleksibilitas konsep ini dalam mengatasi tantangan di dunia nyata. Dengan memahami probabilitas, para profesional dapat membuat keputusan yang lebih informasional, mengoptimalkan strategi mereka, dan memahami kompleksitas dari berbagai situasi. Seiring dengan kemajuan teknologi dan metode analisis, penerapan probabilitas terus berkembang, membuka pintu untuk inovasi dan pemahaman yang lebih baik dalam berbagai disiplin ilmu.

DAFTAR PUSTAKA

Agustianti, R., Nuryami, Fajriah, N. A., Nasruddin, Nay, F. A., Mahmud, R., Kumanireng, L. B., Y, W. N., Faelasofi, R., Prasetyo, A., Alfaris, L., Anim, Asmin, L. O., Nanang, & Sari, M. E. 2022. *Filsafat Pendidikan Matematika*. PT. Global Eksekutif Teknologi.

Alfaris, L. 2023. DERET BERHINGGA DAN TAK HINGGA. In A. Yanto (Ed.), *Aljabar Elementer* (pp. 167–173). PT. Global Eksekutif Teknologi.

Alfaris, L., Dewadi, F. M., Munim, A., Taba, H. T., Khasanah, Maing, C. M. M., Susano, A., & Rukhmana, T. 2022. *Matriks & Ruang Vektor*. Cendikia Mulia Mandiri.

Alfaris, L., Gustian, D., Setyorini, R., Romli, I., Putri, A. Y. P., Herjuna, S. A. S., Syamsiyah, N., Yuniansyah, Aziza, N., Muhammad, A. C., Umar, N., & Wali, M. 2022. *Riset Operasi*. Indie Press.

Alfaris, L., Sarumaha, Y. A., Sitopu, J. W., Agustianti, R., Rizqi, V., Yenni, Jamaludin, Nurvicalesti, N., Murtako, A., Akbar, N., & Abidin, Z. 2022. *Logika dan Struktur Diskrit*. PT. Global Eksekutif Teknologi.

Bishop, C. M. 2006. *Pattern Recognition and Machine Learning*. Springer.

Casella, G., & Berger, R. L. 2002. *Statistical Inference*. Duxbury Press.

Dewadi, F. M., Milasari, L. A., A., H., Wibowo, C., Abdi, Alfaris, L., Saputra, A. A., & Gobel, F. F. 2023. *Desain Penelitian Bidang Teknik*. Get Press Indonesia.

Grinstead, C. M., & Snell, J. L. 1997. *Introduction to Probability*. American Mathematical Society.

Hacking, I. 2006. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*. Cambridge University Press.

Hastie, T., Tibshirani, R., & Friedman, J. 2009. *The Elements of Statistical Learning*. Springer.

Jaynes, E. T. 2003. *Probability Theory: The Logic of Science*. Cambridge University Press.

Muttaqin, Alfaris, L., Muhammad, A. C., Arafah, M., Limbong, A., Suryani, Abdal, N. M., Fairuzabadi, M., Pungus, S. R., Nurahman, A., Siagian, R. C., & Hazriani. 2023. *Representasi Pengetahuan* (A. Karim (ed.)). Yayasan Kita Menulis.

Octavianti, C. T., Mulyawati, I., Setiawan, J., Sitopu, J. W., Alfaris, L., Ratuanik, M., Hasanah, N., Agustianti, R., Pangestika, R. R., & Wulandari, Y. O. 2022. *Konsep Dasar Matematika*. PT. Global Eksekutif Teknologi.

Ross, S. 2014. *Introduction to Probability and Statistics for Engineers and Scientists*. Academic Press.

Subakti, H., Romli, I., Syamsiyah, N., Herianto, A. A. B. H., Alfaris, L., Hasin, M. K., Hadi, A., Farida, F., Rismayani, R., Setiawan, A., & Khadafi, R. K. H. S. 2022. *Artificial Intelligence.* Media Sains Indonesia.

Wackerly, D., Mendenhall, W., & Scheaffer, R. L. 2008. *Mathematical Statistics with Applications*. Cengage Learning.

BAB 8
RATA-RATA DAN VARIANS DARI VARIABEL ACAK

Oleh Siti Aminah

8.1 Pendahuluan

Menurut katadata media *network* (2022), waktu rata-rata yang digunakan masyarakat Indonesia dalam penggunaan media sosial adalah 197 menit atau 3,2 hari perjam. Apa artinya? Apa itu rata-rata? Apakah ini berarti setiap hari seseorang menggunakan media sosial 197 menit? Atau apakah ini berarti setiap orang menggunakan media sosial 197 menit sehari? Bagaimana waktu yang dihabiskan oleh orang yang berbeda berbeda satu sama lain? Ini adalah *mean* dan variabilitasnya adalah varians dalam probabilitas dan statistik. Nilai rata-rata ini akan memberi tahu Anda tentang waktu yang paling dekat tentang penggunaan media sosial dalam sehari. Anda dapat dengan mudah melihat perbedaan setiap data yang diperoleh katadata media network dari nilai rata-rata ini. Perbedaan nilai ini menunjukkan variabilitas nilai yang mungkin dari variabel acak.

8.2 *Mean* Variabel Acak

Misalkan variable acak X , maka mean X dapat dinotasikan dengan μ_x dibaca *miu x*. Mean variabel acak ini dikenal dengan ekspektasi matematis atau nilai harapan atau rata-rata yang dinotasikan dengan E[X]. Menurut (Walpole and All, 2002). Definisi dari E[X] dijelaskan jika X diskrit dan X kontinu sebagai berikut.

Definisi *Mean* Variabel Acak

Misalkan X adalah variabel acak dari distribusi peluang $f(x)$, rata-rata atau nilai harapan dari X adalah

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

untuk X diskrit dan

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

untuk X kontinu

Dapat kita lihat bahwa perhitungan rata-rata variabel acak menggunakan distribusi peluang. Bedanya, pada perhitungan rata-rata sampel menggunakan data. Namun, dapat disimpulkan bahwa perhitungan rata-rata sampel dan variabel acak merepresentasikan nilai pusat dari distribusi data. Adapun generalisasi defisini rata-rata variabel acak adalah sebagai berikut(Walpole and All, 2002)

Teorema *Mean* Variabel Acak

Misalkan X adalah variabel acak dari distribusi peluang $f(x)$. Rata-rata atau nilai harapan dari variabel acak $g(X)$ adalah

$$\mu_{g(X)} = E[g(X)] = \sum_{x} g(x) \cdot f(x)$$

untuk X diskrit dan

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

untuk X kontinu

Dalam perhitungan nantinya, kita akan dipermudah dengan menggunakan sifat-sifat dari rata-rata variabel acak yang telah disebutkan pada teorema berikut(*Properties of Expected Value*, no date).

Teorema Sifat-Sifat *Mean* Variabel Acak

Misalkan X adalah variabel acak, maka berlaku sifat-sifat berikut

1. c adalah konstanta dan $X = c$, maka $E[X] = E[c] = c$
2. c adalah konstanta, maka $E[cX] = c \cdot E[X]$
3. $c_1\ dan\ c_2$ merupakan konstanta, maka $E[\,c_1X + c_2] = c_1 \cdot E[X] + c_2$

8.2.1 Contoh *Mean* Variabel Acak Diskrit

Contoh 1 : Ekspektasi dari sebuah dadu(Rosen, 2012)

Misalkan X adalah angka yang muncul pada pelemparan sebuah dadu. Apakah nilai yang diharapkan dari X?

Solusi:

X adalah nilai 1, 2, 3, 4, 5, atau 6. Setiap nilai tersebut mempunyai probabilitas 1/6. Maka nilai yang diharapkan

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{21}{6} = \frac{7}{2}$$

Contoh 2:

Sebuah koin dilempar tiga kali. Misalkan S adalah ruang sampel dari delapan kemungkinan hasil, dan misalkan X adalah variabel acak yang menetapkan banyaknya hasil percobaan setiap kemungkinan kejadian. Berapa nilai X yang diharapkan?

Solusi:

Nilai X untuk delapan kemungkinan hasil ketika sebuah uang logam dilempar tiga kali. Karena pelemparannya independen, maka probabilitas setiap nilai X adalah 1/8. Sehingga

$$\begin{aligned} E(X) = \frac{1}{8} [&X(AAA) + X(AAG) + X(AGA) + X(GAA) + X(GAG) \\ &+ X(AGG) + X(GGG) \\ &= \frac{1}{8}(3 + 2 + 2 + 2 + 1 + 1 + 1 + 0) = \frac{12}{8} \end{aligned}$$

Jadi, nilai harapan pelemparan koin sebanyak 3 kali adalah 3/2.

Contoh 3:

Jika X adalah variabel acak diskrit dari nilai fungsi peluang yang diberikan dalam tabel berikut.

X	1	2	3	4
$P(X = x)$	0,4	0,1	0,3	0,2

Tentukan $E(2x)$

Solusi:

Mean dari X dari fungsi peluang yang diberikan pada tabel adalah

$$E[X] = \sum_{x=1}^{4} 2x \cdot p(X = x)$$

$$= 1(0{,}4) + 2(0{,}1) + 3(0{,}3) + 4(0{,}2) = 2{,}3$$

Mean dari $2X$ dari fungsi peluang yang diberikan pada tabel adalah

$$E[2X] = \sum_{x=1}^{4} 2x \cdot p(X = x)$$

$$= 2\sum_{x=1}^{4} x \cdot p(X = x)$$

$$= 2(2{,}3) = 4{,}6$$

8.2.2 Contoh *Mean* Variabel Acak Kontinu

Contoh 5:

Fungsi kepadatan peluang dari X didefinisikan pada fungsi berikut

$$f(x) = \begin{cases} 2(1-x), untuk & 0 < x < 1 \\ 0, & untuk \quad x\ lainnya \end{cases}$$

Nilai dari $E[X]$ adalah ….

Solusi:

X pada soal ini adalah variabel acak kontinu karena nilai x ada didalam interval selang tertentu. Sesuai dengan definisi *mean* dari variabel acak kontinu diperoleh

$$E[X] = \int_0^1 x \cdot f(x)\, dx$$

$$= \int_0^1 x \cdot 2(1-x)\, dx$$

$$= 2\int_0^1 (x - x^2)\, dx$$

$$= 2\left[\frac{1}{2}x^2 - \frac{1}{3}x^3\right]_0^1$$

$$= 2\left[0 - \left(\frac{1}{2}(1)^2 - \frac{1}{3}(1)^3\right)\right]$$

$$= 2 \cdot \frac{1}{6} = \frac{1}{3}$$

Jadi nilai $E[X] = \frac{1}{3}$

Contoh 6:
Fungsi kepadatan peluang dari X didefiniskan sebagai berikut

$$f(x) = \begin{cases} \frac{x^2}{3}, -1 < x < 2 \\ 0, \qquad x\ lainnya \end{cases}$$

Nilai ekspektasi dari $g(X) = 4X + 3$ adalah ....
Solusi:

$$E[g(X)] = E[4X + 3]$$
$$= \int_{-1}^{2} (4x + 3) \cdot \frac{x^2}{3} dx$$
$$= \frac{1}{3} \int_{-1}^{2} (4x^3 + 3x^2)\, dx$$
$$= \frac{1}{3} [x^4 + x^3]_{-1}^{2}$$
$$= \frac{1}{3} \left((2^4 - (-1)^4) + (2^3 - (-1)^3) \right) = 8$$

Jadi, ekspektasi $g(X) = 4X + 3$ sama dengan 8.

8.3 Varians Variabel Acak

Varians ini merepresentasikan ukuran variasi nilai-nilai X terhadap rata-ratanya. Varians digunakan untuk menggambarkan penyebaran nilai di sekitar nilai rata-rata. Contohnya sebagai wisatawan kita akan berkunjung ke suatu negara, maka kita memerlukan persiapan baju yang sesuai dengan suhu di negara tersebut. Kita akan kesulitan menggunakan rata-rata suhu harian di negara tersebut, karena perubahan suhu di pagi, siang hingga malam pasti berbeda. Sehingga kita membutuhkan data suhu minimal dan maksimal pada negara tersebut. Jika selisih minimal dan maksimal suhu terlalu tinggi, makan wisatawan harus mempersiapkan pakaian yang dibawa jika suhu fluktuatif. Namun jika selisih suhu minimal dan maksimal rendah, maka tidak akan terjadi perubahan suhu yang signifikan. Dari sini wisatawan bisa menentukan jenis baju yang sesuai dan dibutuhkan saat di negara tersebut. Definisi varians variabel acak ada;ah sebagai berikut(Walpole and All, 2002)

Definisi Varians Variabel Acak

Jika X adalah variabel acak, maka varians dari X dinyatakan $var(X)$ dapat diperoleh melalui formula berikut

$$var\ (X) = E[(X-\mu)^2] = \sum_x (x-\mu)^2 \cdot f(x)$$

untuk X adalah distribusi diskrit

$$var\ (X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx$$

untuk X adalah distribusi kontinu

Dengan perhitungan yang lebih sederhana, varians dapat ditentukan dengan menggunakan teorema berikut (Walpole and All, 2002).

Teorema Varians Variabel Acak

Jika X adalah variabel acak dan $E[X]$ merupakan ekspektasi dari X. Varians dari X dapat diperoleh dengan $var\ (X) = E[X^2] - E[X]^2$

Dalam perhitungan nantinya, kita akan dipermudah dengan menggunakan sifat-sifat dari varians variabel acak yang telah disebutkan pada teorema berikut

Teorema Sifat-Sifat Varians Variabel Acak

Jika X adalah variabel acak

1. $X = c$ adalah konstanta, maka $Var\ (X) = Var\ (c) = 0$
2. c adalah konstanta, maka $Var(cX) = c^2 Var(X)$
3. c_1 dan c_2 adalah konstanta, maka $Var(c_1 X + c_2) = {c_1}^2 Var(X)$

8.3.1 Contoh Varians Variabel Acak Diskrit

Contoh 7 :

Berapakah varians dari X, dimana X adalah bilangan yang muncul pada pelemparan sebuah dadu?

Solusi:

Kita mempunyai $Var(X) = E(X^2) - E(X)^2$. $E(X)$ pada contoh 1 adalah $7/2$. Peluang muncul setiap mata dadu adalah $1/6$, sehingga

$$E(X^2) = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 = \frac{91}{6}$$

$$Var(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

8.3.2 Contoh Varians Variabel Acak Kontinu

Contoh 8:

Jika fungsi kepadatan peluang X didefinisikan sebagai berikut

$$f(x) = \begin{cases} 2(1-x), 0 < x < 1 \\ 0, \qquad\quad x\ lainnya \end{cases}$$

Nilai dari $Var[X]$ adalah ….

Solusi:

Nilai $E[X]$ telah didapatkan pada contoh 5.

$$E[X^2] = \int_0^1 x^2 \cdot f(x)\, dx$$

$$= \int_0^1 x^2 \cdot 2(1-x)\, dx$$

$$= 2\int_0^1 (x^2 - x^3)\, dx$$

$$= 2\left[\frac{1}{3}x^3 - \frac{1}{4}x^4\right]_0^1$$

$$= 2\left[\frac{1}{3} - \frac{1}{4}\right] = \frac{1}{6}$$

Maka

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

$$= \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{18}$$

DAFTAR PUSTAKA

Properties of Expected Value (no date). Available at: https://hyperskill.org/learn/step/16751#properties-of-expected-value.

Rosen, K. H. 2012. *Discrete Mathematics and Its Applications*. Seventh Ed. New York: McGraw-Hill.

Walpole, R. E. and All, E. 2002. *Probability & Statistics for Engineers & Scientists*. Boston: Prentice Hall.

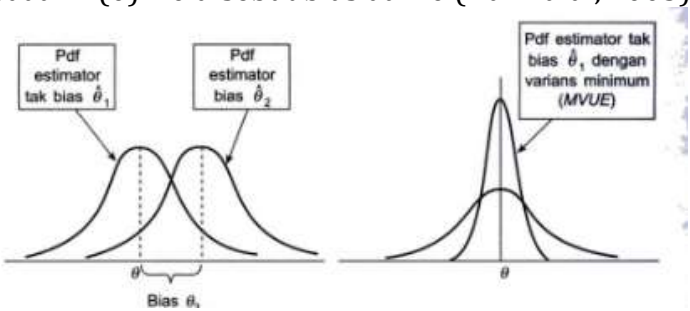
BAB 9
ESTIMASI DAN INTERVAL KEPERCAYAAN

Oleh Ridha Yuniara

9.1 Estimasi

Terdapat perbedaan makna yang harus dipahami terlebih dahulu dalam memaknai kata *estimate,* estimator, dan estimasi. *Estimate* (hasil estimasi) merupakan Nilai spesifik dari suatu statistik seperti nilai mean sampel, persentase sampel, dan varians sampel, yang dijadikan sebagai contoh untuk mengetahui nilai yang sebenarnya pada populasi. Sedangkan estimator adalah setiap statistik (mean sampel, varians sampel, dll) yang dimanfaatkan untuk mengestimasi suatu parameter. Sebagai contoh, mean sampel ($\bar{X}$) dapat mengestimasi mean populasi μ_x (Harinaldi, 2005). Ciri-ciri estimator yang baik adalah:

1. Tidak bias, artinya hasil yang diperoleh mengandung nilai parameter yang diestimasi, Secara matematis, dapat dinyatakan bahwa jika sebuah estimator $\hat{\theta}$ ialah estimator tak-bias dari parameter θ maka $E(\hat{\theta}) = \theta$ untuk semua nilai θ yang mungkin. Jika $\hat{\theta}$ bukan estimator tak-bias, maka perbedaan $E(\hat{\theta}) - \theta$ disebut bias dari $\hat{\theta}$ (Harinaldi, 2005).



Gambar 9.1. Estimator bias dan tak-bias

Sumber: Buku *Prinsip-prinsip Statistik untuk Teknik dan Sains,* karangan Harinaldi, 2005 jakarta: Erlangga.

2. Efisien, artinya hasil estimasi menggunakan nilai tersebut sudah memuat nilai parameternya walaupun nilai berada pada rentang yang kecil saja. Standar deviasi kecil, sehingga kemungkinan mendekati nilai parameter lebih besar.
3. Konsisten, artinya keakuratan estimasi nilai parameter meningkat bila jumlah sampel lebih besar. Terlepas dari ukuran sampel, rentang tersebut akan berisi nilai parameter yang sedang diestimasi.

Adapun estimasi merupakan semua proses dalam menggunakan estimator untuk mendapatkan *estimate* dari suatu parameter (Harinaldi, 2005). Kita dapat melakukan dua jenis estimasi dalam menduga nilai parameter, yaitu sebagai berikut:

1. Estimasi titik (*point estimation*) berisi perhitungan angka tunggal untuk memperkirakan nilai parameter suatu populasi (Lind, dkk, 2007, 383). Estimasi titik didapat dengan menentukan statistik yang tepat dan menghitung nilainya dari sampel. Statistik yang dipilih disebut estimator titik (*point estimator*) dan Estimasi titik yang dapat digunakan untuk memperkirakan parameter populasi adalah rata-rata sampel terhadap rata-rata populasi, proporsi sampel terhadap proporsi populasi, varians, serta jumlah variabel tertentu yang termasuk dalam sampel untuk memperkirakan jumlah variabel tersebut dalam populasi.

   teori estimasi dapat dimanfaatkan untuk memprediksi berapa banyak pengunjung atau meramalkan prospek (kemungkinan Penyembuhan) suatu penyakit.

   Contoh (1)

   Sebuah penelitian dengan mengambil sampe sebanyak 210 lansia di Kabupaten Aceh Tengah diperoleh bahwa Hb rata-rata 9,4%. Jika kita menduga kadar Hb lansia di Kabupaten Aceh Tengah dengan estimasi titik, maka dapat dikatakan bahwa kadar Hb lansia di Kabupaten Aceh Tengah adalah 9,4%.

   Contoh (2)

Untuk memperkirakan rata-rata tinggi badan mahasiswa Universitas Gajah Putih, maka diambil sebanyak 30 orang dengan hasil sebagai berikut.

165	155	171	153	164	163	162	161
154	162	159	162	155	166	168	171
153	155	157	163	161	160	156	154
152	172	150	151	167	154		

Kemudian diperoleh estimasi titik terhadap tinggi badan mahasiswa Universitas Gajah Putih yaitu X = 159,8. Tinggi badan 159,8 cm adalah titik estimasi terhadap tinggi badan mahasiswa Universitas Gajah Putih.

Adapun kelemahan dari estimasi titik adalah kita tidak dapat mengetahui kekuatan estimasi kita, serta besarnya kemungkinan estimasi tersebut bernilai salah. Kelemahan estimasi titik ini dapat dihilangkan dengan menggunakan estimasi selang.

2. Estimasi selang (Interval estimation), berisi dua bilangan di antara posisi suatu parameter populasi yang dijadikan sebagai estimasi dari parameter tersebut (murray, 2007). Dalam estimasi interval ini, sampel-sampel yang diambil dari suatu populasi akan berdistribusi (normal) sekitar μ dan memiliki simpangan baku. Sehingga dapat kita tentukan batas minimum dan maksimum dari letak nilai μ.

9.2 Interval Kepercayaan

Misalkan μ_s dan σ_s masing-masing diketahui sebagai mean dan deviasi standar (error stand dari distribusi sampling suatu statistik *S*. Kemudian jika distribusi sampling *S* mendekati normal (yang terbukti benar untuk banyak statistik di mana ukuran sampel $N \geq 30$), maka dapat diharap kita akan menemukan sebuah statistik sampel aktual *S* yang berada di dalam interval $\mu_s - \sigma_s$ sampai dengan $\mu_s + \sigma_s$, $\mu_s - 2\sigma_s$ sampai dengan $\mu_s + 2\sigma_s$, atau $\mu_s - 3\sigma_s$ sampai dengan $\mu_s + 3\sigma_s$ masing-masing sekitar 68,27%, 95,45%, dan 99,73% dari waktu yang kita habiskan (Spiegel dkk. 2007).

Kita juga dapat berharap (kita *percaya*), dengan cara yang sama, akan menemukan μ_s dalam interval $S - \sigma_s$ sampai dengan $S +$

σ_s, $S - 2\sigma_s$ sampai dengan $S + 2\sigma_s$, atau $S - 3\sigma_s$ sampai dengan $S + 3\sigma_s$ masing-masing sekitar 68,27%, 95,45%, dan 99,73% dari waktu yang ada. Atas dasar ini dapat disebutkan bahwa masing-masing interval ini sebagai *interval kepercayaan* (*confidence interval*) 68,27%, 95,45%, dan 99,73% untuk mengestimasi μ_s. Bilangan-bilangan di kedua ujung dari interval-interval ini ($S - \sigma_s$ sampai dengan $S \pm \sigma_s$, $S \pm 2\sigma_s$ $S \pm 3\sigma_s$) masing-masing diketahui sebagai *batas kepercayaan* (*confidence limit*), atau *batas fidusial* (*fiducial limit*) 68,27%, 95,45%, dan 99,73% (Spiegel dkk. 2007). Dengan kata lain, Interval kepercayaan merupakan rentang nilai yang dipakai dalam statistik inferensial untuk mengetahu sejauh mana kita bisa yakin bahwa parameter populasi (seperti rata-rata, proporsi, dan varians) berada dalam rentang tertentu (Arraniri dkk, 2023).

Cara yang sama untuk $S \pm 1{,}96\sigma_s$ dan $S \pm 2{,}58\sigma_s$ adalah batas-batas kepercayaan 95% dan 99% (atau 0,95 dan 0,99) untuk *S.* Persentase kepercayaan sering disebut juga sebagai *tingkat kepercayaan* (*confidence level*). *Koefisien kepercayaan,* atau *nilai kritis*, disimbolkan oleh z_c. Dari tingkat kepercayaan tersebut, dapat dicari koefisien kepercayaan, dan sebaliknya. Nilai-nilai z_c yang berkorespondensi dengan berbagai tingkat kepercayaan dapat dilihat dengan contoh pada tabel 9.1 berikut (Spiegel dkk. 2007).

Tabel 9.1. Nilai-nilai z_c yang berkorespondensi dengan berbagai tingkat kepercayaan

Tingkat Kepercayaan	z_c
99,73%	3,00
99%	2,58
98%	2,33
96%	2,05
95,45%	2,00
95%	1,96
90%	1,645
80%	1,28
68,27%	1,00
50%	0,6745

Sumber: Buku *Schaum's Outlines Teori dan Soal-soal Statistik,* karangan Murray R.S dan Larry J.S, 2007

Perlu diketahui bahwa semakin tinggi tingkat kepercayaan yang dipilih, semakin lebar juga interval kepercayaannya. Hal ini mengindikasikan bahwa kita lebih yakin tetapi kurang presisi dalam mengestimasi parameter populasi. Sebaliknya, semakin rendah tingkat kepercayaan yang dipilih, semakin sempit juga interval kepercayaannya dan kita semakin kurang yakin. Tetapi estimasi parameternya menjadi lebih presisi (Arraniri dkk, 2023).

Contoh :

Hasil penelitian menunjukkan bahwa diyakini 90 persen bahwa tingkat pendapatan masyarakat kelas menengah di Surabaya berkisar antara $500 hingga 1000$ per bulan. Dari pernyataan tersebut, yang dimaksud dengan tingkat kepercayaan adalah 90 persen dan interva kepercayaan adalah *range* pendapatannya, yaitu $(2000-5000) per bulan.

Namun seringnya tingkat kepercayaan ini tidak dinyatakan secara eksplisit, tetapi dinyatakan dengan suatu nilai dengan simbol α, yang berarti peluang dugaan parameter tidak memenuhi interval. Sehingga α = 1 – tingkat kepercayaan. Jika tingkat kepercayaannya sebesar 95 persen, maka α = 0.05.

Interval Kepercayaan untuk Mean

Misalkan statistik S adalah mean sampel $\bar{X}$, maka batas-batas kepercayaan 95% dan 99% untuk mengestimasi mean populasi μ masing-masing dirumuskan sebagai $\bar{X} \pm 1{,}96\sigma_{\bar{X}}$ dan $\bar{X} \pm 2{,}58\sigma_{\bar{X}}$. Bentuk umum batas kepercayaannya dapat ditulis sebagai $\bar{X} \pm z_c\sigma_{\bar{X}}$, dimana z_c bergantung pada tingkat kepercayaan yang diinginkan, seperti pada contoh tabel 11.1. Adapun batas kepercayaan untuk mean populasi dinyatakan sebagai,

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}}$$

Jika samplingnya berasal dari sebuah populasi tak berhingga atau sampling dengan pengembalian dari sebuah populasi berhingga. Kemudian dapat dinyatakan sebagai,

$$\bar{X} \pm z_c \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}}$$

Jika samplingnya adalah sampling tanpa pengembalian dari sebuah populasi dengan ukuran berhingga N_p (Spiegel dkk. 2007).

Deviasi standar populasi σ umumnya tidak diketahui nilainya sehingga untuk memperoleh batas-batas kepercayaan di atas, maka digunakan estimasi sampel $\hat{s}$ atau *s.* Penggunaan estimasi sampel sebagai pengganti deviasi standar populasi dalam menghitung batas kepercayaan dapat memberikan hasil yang baik jika $N \geq 30$. Sedangkan untuk N < 30, aproksimasinya akan buruk sehingga harus digunakan teori sampling kecil (Spiegel dkk. 2007).

Interval Kepercayaan untuk Proporsi

Misalkan terdapat sampel berukuran *N* yang diambil dari sebuah populasi binomial dan didalamnya terdapat statistik *S* yang merupakan proporsi dari "keberhasilan-keberhasilan". Maka *p* menyatakan proporsi keberhasilannya (probabilitas keberhasilan) dan batas-batas kepercayaan untuk *p* diberikan oleh $P \pm z_c\sigma_P$, dimana *P* adalah proporsi dari keberhasilan-keberhasilan dalam sampel berukuran *N.* Batas-batas kepercayaan juga dapat dirumuskan sebagai berikut,

$$P \pm z_c\sqrt{\frac{pq}{N}} = P \pm z_c\sqrt{\frac{p(1-p)}{N}}$$

Jika samplingnya berasal dari sebuah populasi tak-berhingga atau sampling dengan pengembalian dari sebuah populasi berhingga, dan dapat dinyatakan dengan:

$$P \pm z_c\sqrt{\frac{pq}{N}}\sqrt{\frac{N_p-N}{N_p-1}}$$

Jika samplingnya adalah sampling tanpa pengembalian dari sebuah populasi dengan ukuran berhingga N_p (Spiegel dkk. 2007).

Interval Kepercayaan untuk Selisih dan Jumlah

Misalkan S_1 dan S_2 adalah dua statistik sampel yang terdistribusi mendekati normal, maka batas-batas kepercayaan untuk selisih antara parameter-parameter populasi yang berkorespondensi dengan S_1 dan S_2 dinyatakan sebagai berikut,

$$S_1 - S_2 \pm z_c\sigma_{S_1-S_2} = S_1 - S_2 \pm z_c\sqrt{{\sigma_{S_1}}^2 + {\sigma_{S_2}}^2}$$

Sedangkan batas-batas kepercayaan untuk jumlah parameter-parameter populasi adalah,

$$S_1 + S_2 \pm z_c\sigma_{S_1+S_2} = S_1 + S_2 \pm z_c\sqrt{{\sigma_{S_1}}^2 + {\sigma_{S_2}}^2}$$

dengan syarat-syarat sampelnya saling bebas (Spiegel dkk. 2007).

Contoh (1)

Batas-batas kepercayaan untuk selisih dua mean populasi dimana populasinya tak berhingga dinyatakan sebagai berikut,

$$\overline{X_1} - \overline{X_2} \pm z_c\sigma_{\overline{X_1}-\overline{X_2}} = \overline{X_1} - \overline{X_2} \pm z_c\sqrt{\frac{{\sigma_1}^2}{N_1} + \frac{{\sigma_2}^2}{N_2}}$$

di mana $\overline{X_1}$, σ_1, N_1 dan $\overline{X_2}$, σ_2, N_2 masing-masing adalah mean, deviasi standar dan ukuran dari kedua sampel yang diambil dari populasi-populasi tersebut (Spiegel dkk. 2007).

Contoh (2)

Batas-batas kepercayaan untuk selisih dua proporsi populasi dimana populasi-populasinya tak berhingga dinyatakan sebagai berikut,

$$P_1 - P_2 \pm z_c\sigma_{P1-P2} = P_1 - P_2 \pm z_c\sqrt{\frac{P_1(1-P_1)}{N_1} + \frac{P_2(1-P_1)}{N_2}}$$

di mana P_1 dan P_2 adalah proporsi sampel, N_1 dan N_2 adalah ukuran dari kedua sampel yang diambil dari populasi, dan p_1 dan p_2 adalah proporsi di dalam kedua populasi tersebut (diestimasi oleh P_1 dan P_2) (Spiegel dkk. 2007).

Interval Kepercayaan untuk Deviasi Standar

Batas-batas kepercayaan untuk deviasi standar σ dari sebuah populasi yang berdistribusi normal seperti yang diestimasi dari sebuah sampel dengan deviasi standar *s* adalah,

$$s \pm z_c\sigma_s = s \pm z_c\frac{\sigma}{\sqrt{2N}}$$

(Spiegel dkk. 2007).

Error yang Mungkin

Parameter populasi yang berkorespondensi dengan statistik S dengan batas kepercayaan 50% dinyatakan oleh $S \pm 0{,}674\sigma_s$. Kuantitas $0{,}674\sigma_s$ dikenal sebagai *error yang mungkin* (*probable error*) dari estimasi (Spiegel dkk. 2007).

DAFTAR PUSTAKA

Arraniri, I., Sabtohadi, J., Suhartawan, B., Sarie, F., Abubakar, R., Maryani, S., Hikmah., Rahayu, B., Posmaningsih, D,A,A., & Korosando, F. 2023. *Pengantar Statistika.* Batam: Yayasan Cendikia Mulia Mandiri.

Lind, D.A., Marchal, W.G., & Wathen, S.A. 2007. *Teknik-teknik Statistika dalam Bisnis dan Ekonomi Menggunakan Kelompok Data Global*. Jakarta: Salemba empat.

Harinaldi. 2005. *Prinsip-prinsip Statistik untuk Teknik dan Sains.* Jakarta: Erlangga.

Spiegel, M.R. & Stephens, L.J. 2007. *Schaum's Outlines Teori dan Soal-soal Statistik*. Jakarta: Erlangga.

BAB 10
REGRESI DAN KORELASI SEDERHANA

Oleh Kusnaeni

10.1 Pendahuluan

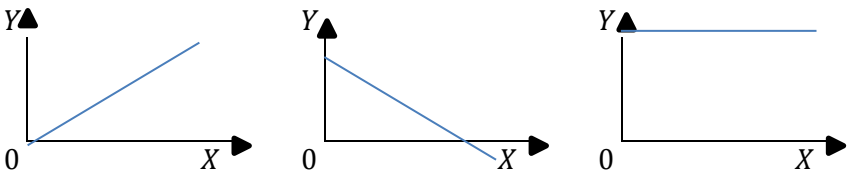
Dalam ranah statistik, dua konsep yang digunakan untuk mengevaluasi hubungan dua variabel, yaitu regresi dan korelasi. Regresi dan korelasi merupakan instrumen yang efektif dalam menilai seberapa erat hubungan dua variabel tersebut. Pentingnya regresi dan korelasi dalam analisis statistik tidak dapat diabaikan karena keduanya menyediakan kerangka kerja yang kuat untuk memahami dan mengukur hubungan antara variabel-variabel. Regresi, memungkinkan kita untuk memodelkan dan memprediksi sejauh mana satu atau lebih variabel bebas dapat mempengaruhi variabel terikat. Hal ini memberikan gambaran tentang kekuatan hubungan, dan juga memfasilitasi estimasi nilai yang mungkin untuk variabel terikat berdasarkan nilai variabel bebas. Sementara itu, korelasi membantu kita mengukur sejauh mana dua variabel bergerak bersama-sama tanpa memerhatikan arah hubungannya. Dengan kata lain, korelasi membantu kita menilai sejauh mana perubahan dalam satu variabel terkait dengan perubahan dalam variabel lainnya.

Setelah mempelajari bab ini, diyakini dapat mengetahui dan memahami terkait regresi dan korelasi yang sederhana. Kemudian menambah wawasan yang berharga tentang hubungan linier atau non-linier antara variabel-variabel tersebut. Selanjutnya, memahami cara mengkonstruksi persamaan regresi linear dan taksiran nilai variabel terikat bersesuaian nilai variabel yang ditentukan, menghitung koefisien regresi dan korelasi serta menginterpretasi nilainya.

10.2 Pengetian Regresi

Definisi regresi adalah metode statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen yang juga dikenal sebagai variabel prediktor dan satu variabel dependen yang juga dikenal sebagai variabel respons. Tujuannya adalah untuk menilai dan menjelaskan sejauh mana variabel independen dapat memberikan pengaruh terhadap variabel dependen. Regresi digunakan untuk menganalisis, memodelkan, dan memprediksi hubungan antara variabel-variabel (Moore, McCabe and Craig, 2014).

Sifat keterkaitan antara variabel bebas (X) dan variabel terikat (Y) dapat mengambil bentuk positif, negatif, atau tidak menunjukkan adanya keterkaitan. Keterkaitan positif disebut sebagai hubungan searah, yang berarti peningkatan nilai X berakibat nilai Y juga meningkat, atau sebaliknya jika nilai X menurun, maka nilai Y juga menurun. Sebaliknya, keterkaitan negatif disebut sebagai hubungan berlawanan arah, di mana peningkatan nilai X berakibat pada penurunan nilai Y, dan sebaliknya. Tidak adanya keterkaitan menunjukkan bahwa perubahan dalam nilai X tidak mempengaruhi nilai Y; dengan kata lain, nilai Y tetap stabil meskipun nilai X berubah. Grafik yang mencerminkan ketiga jenis sifat hubungan antara dua variabel pada gambar di bawah ini. (Wirawan, 2016).



a) Hubungan Positif b) Hubungan Negatif c) Tak ada Hubungan

Gambar 10.1. Jenis hubungan antara variabel *X* dan *Y*
(Sumber: Penulis, 2023. Diolah)

Sebagai contoh untuk ketiga jenis hubungan antara dua variabel, dapat disajikan sebagai berikut. Sebagai contoh untuk hubungan positif, terdapat keterkaitan antara pengeluaran biaya iklan dan jumlah penjualan, juga antara harga suatu barang dan

penawaran. Sebagai contoh untuk hubungan negatif, dapat disorot hubungan antara jumlah peserta KB dan tingkat kelahiran, atau antara suku bunga dan investasi. Untuk contoh tanpa adanya hubungan, dapat diilustrasikan melalui kecepatan kendaraan dan tingkat kecerdasan siswa di sebuah kelas. Untuk mengkaji hubungan-hubungan linier antara variabel bebas dan variabel terikat, regresi linier memiliki dua bentuk, yakni regresi linier sederhana dan regresi linier berganda.

10.3 Regresi Linier Sederhana

Analisis Regresi Linier sederhana adalah analisis yang digunakan untuk melihat hubungan fungsional variabel bebas dan variabel terikat yang dinyatakan dalam persamaan matematik dan garis serta memiliki hubungan yang searah(Kemendikbud, 2023). Seperti namanya, regresi linier sederhana adalah sebuah metode sederhana untuk memprediksi nilai variabel dependen X. Secara matematis persamaa linier sederhana dapat dilipada persamaan

$$Y \approx b_0 + b_1 X$$

dengan keterangan

1. Y : *Dependent variable* adalah akibat dari suatu sebab, misalnya pengaruh rangking murid terhadap lama waktu berlajarnya
2. X : *Independent variable* adalah hal yang diasumsikan menjadi sebab yang nantinya mempengaruhi nilai Y
3. b_1 : *coefficient* adalah suatu unit / proporsi yang menunjukkan seberapa besar perubahan dalam Y untuk setiap perubahan satu unit dalam X.
4. b_0 : *intercept* adalah nilai konstanta pada regresi linier sederhana

10.3.1 Metode Kuadrat Terkecil

Regresi adalah sebuah instrumen pengukuran yang digunakan untuk menilai apakah terdapat atau tidak adanya hubungan antara variabel-variabel tersebut. Persamaan garis regresi merujuk pada rumus yang digunakan untuk menentukan garis regresi dalam konteks data pada diagram pencar (scatter plot).

Scatter plot merupakan alat yang bermanfaat dalam analisis regresi untuk secara visual menggambarkan hubungan antara variabel independen (variabel prediktor) dan variabel dependen (variabel respons). Sebelum menetapkan persamaan garis regresi, disarankan untuk membuat scatter plot terlebih dahulu. Dengan langkah ini, persamaan regresi yang dipilih dapat lebih akurat. Pendekatan ini bertujuan untuk mengurangi sebanyak mungkin kesalahan (error) antara data aktual dan estimasi. Scatter plot memberikan dua manfaat utama, yaitu menunjukkan apakah terdapat hubungan yang signifikan antara dua variabel dan membantu menentukan jenis persamaan yang mencerminkan hubungan di antara keduanya.

Metode Kuadrat Terkecil diterapkan dalam menentukan persamaan garis regresi dengan memilih satu garis linier dari berbagai garis yang mungkin dibentuk dari data yang ada. Garis yang dipilih adalah yang memiliki kesalahan paling kecil antara nilai sebenarnya dan nilai prediksi. Prinsip dasar dalam pemilihan garis regresi ini adalah memilih garis yang memiliki jumlah kuadrat deviasi nilai observasi Y terhadap nilai Y prediksinya yang minimal, sehingga garis regresi yang dihasilkan dianggap optimal.

Persamaan regresi estimasi yang baik secara umum

$$\hat{Y} = b_0 + b_1 X, \tag{1}$$

dengan nilai koefisien b_0 dan b_1 dapat dihitung(Yuliara, 2016),

$$b_1 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}, \tag{2}$$

$$b_0 = \frac{\sum Y_i}{n} - \frac{b_1 \sum X_i}{n}. \tag{3}$$

Keterangan

b_0 = titik potong terhadapa sumbu Y bila $X = 0$

b_1 = slope garis, yaitu perubahan nilai Y akibat perubahan 1 unit X

$\hat{Y}$ = taksiran nilai Y

X = variabel bebas

Y = variabel terikat

n = banyaknya pasangan data.

10.3.2 Interpretasi terhadap nilai koefisien regresi

Tanda positif atau negatif koefisien regresi menunjukkan arah hubungan variabel bebas X terhadap variabel terikat Y. Interpretasi terhadap nilai koefisien regresi, b_0 adalah sebagai berikut :

1. Jika nilai b_0 bertanda positif hal ini mengindikasikan bahwa ketika variabel bebas X bertambah 1 unit, maka variabel terikat Y cenderung meningkat rata-rata sebesar nilai b_0 dan sebaliknya.
2. Jika nilai b_0 bertanda negatif hal ini menandakan jika nilai variabel bebas X bertambah 1 unit, variabel terikat Y cenderung berkurang rata-rata sebesar nilai b_0 dan kebalikannya.

10.3.3 Syarat dan Prosedur Regresi Linier sederhana

Sebelum melakukan regresi linier sederhana, maka data pengamatan harus memenuhi syarat (Ghozali, 2018), yaitu :

1. Jumlah sampel yang dipakai harus sama,
2. Terdapat satu variabel bebas,
3. Nilai residual berdistribusi normal,
4. Adanya korelasi linier antara variabel bebas dan terikat,
5. Tidak terkena heterokedastisitasm
6. Tidak terkena autokorelasi.

Setelah memenuhi syarat tersebut, maka terdapat 8 langkah yang harus dilakukan dalam menganalisis regresi linier sederhana, yaitu(Sitopu and dkk, 2023),

1. Menetapkan tujuan dari regresi linier sederhana.
2. Mengidentifikasi variabel terikat dan variabel bebas.
3. Mengumpulkan/kolektif data (dalam bentuk tabel).
4. Menghitung nilai X^2, Y^2 dan total dari masing-masing.
5. Menghitung slope b_0 dan intercept b_1.
6. Menyusun model persamaan regresi.
7. Memprediksi variabel terikat dan variabel bebas
8. Menetapkan taraf signifikansi dan melakukan uji T.

Sebagai ilustrasi untuk mempermudah pemahaman diberi contoh dibawah ini.

Contoh :

Sejumlah data terdiri dari enam pasangan informasi tentang jumlah pendapatan dan pengeluaran bulanan (dalam juta rupiah) dari enam karyawan di sebuah perusahaan swasta di sektor industri.

Tabel 10.1. Deskripsi besar pendapatan dan berat badan

No	Karyawan	Pendapatan (juta rupiah)	Konsumsi (juta rupiah)
1	Marie	8	7
2	Yudha	12	9
3	Teti	16	12
4	Ridwan	20	14
5	Aqita	24	13
6.	Sinta	26	15

Sumber: Penulis, 2023. Diolah

Berdasarkan data tersebut,

1. Tentukan persamaan regresi.
2. Interpretasi nilai koefisien regresi.
3. Jika seorang karyawan yang pendapatannya Rp 10 juta. Tentukan nilai taksiran berdasarkan hasil regresi

Penyelesaian

1. Berdasarkan prosedur regresi linier sederhana,
   a. Menetapkan tujuan dari regresi linier sederhana yaitu untuk mengetahui apakah besarnya pendapat dapat mempengaruhi besarnya konsumsi karyawan swasta yang bergerak dalam bidang industri.
   b. Mengidentifikasi variabel bebas dan variabel terikat:
      Variabel bebas (X) = jumlah pendapatan
      Variabel terika (Y) = jumlah konsumsi
   c. Mengumpulkan/kolektif data (susun bentuk tabel) dan menghitung nilai untuk X^2, Y^2 dan total dari masing-masing

Tabel 10.2. Perhitungan Unsur-unsur Persamaan Regresi

Data	X	**Y**	XY	X^2	Y^2
1	8	7	56	64	49
2	12	9	108	108	81
3	16	12	192	256	144
4	20	14	280	400	196
5	24	13	312	576	169
6	26	15	390	676	225
Jumlah	106	70	1338	2116	864

Sumber: Penulis, 2023. Diolah

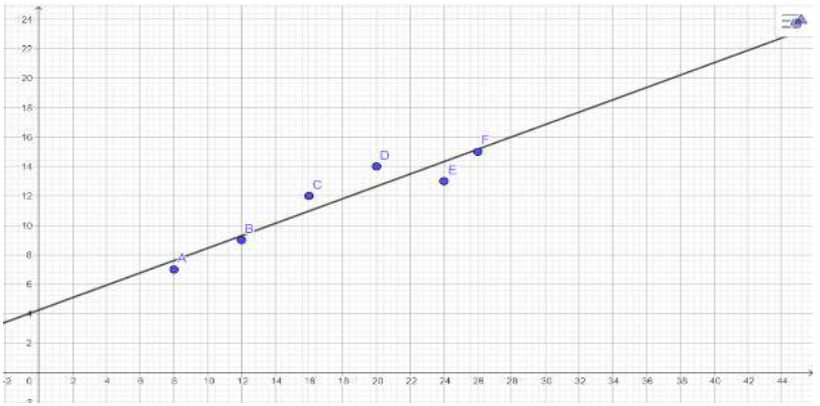
d. Menghitung niali b_0 dan b_1

$$b_1 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$
$$= \frac{6\,(1338) - (106)(70)}{6(2116) - (106)^2}$$
$$= \frac{608}{1460}$$
$$= 0{,}42,$$
$$b_0 = \frac{\sum Y_i}{n} - \frac{b_1 \sum X_i}{n}$$
$$= \frac{70}{6} - \frac{0{,}42(106)}{6}$$
$$= 11{,}67 - 7{,}42$$
$$= 4{,}25,$$

Jadi persamaan regresinya :

$$\hat{Y} = 4{,}25 + 0{,}42X.$$

e. Menyusun model persamaan regresi



Gambar 10.2. Garis regresi hubungan antara variabel X dan Y
(Sumber : Penulis, 2023. Diolah)

f. Memprediksi variabel bebes dan terikat

Analisis korelasi yang hasilnya merupakan koefisien korelasi, maka hasil koefisien korelasi yang didapatkan yaitu

$$r = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{[n\sum X_i^2 - (\sum X_I)^2][n\sum Y_i^2 - (\sum Y_I)^2]}}$$

$$r = \frac{6(1338) - (106)(70)}{\sqrt{[6(2116) - (106)^2][6(864) - (70)^2]}}$$

$$r = \frac{608}{\sqrt{(1460)(284)}} =$$

$$r = \frac{608}{643{,}92} = 0{,}95$$

Dari hasil diatas menunjukkan bahwa hubungan sangat kuat (95%)

Antara variabel bebas dengan variabel terikat. Ini berarti besar pendapatan sangat mempengaruhi besar konsumsi karyawan kelas industri. Sedangkan untuk koefisien determinasi menujukkan bahwa variabel bebas mampu menjelaskan variabel terikat sebesar $r^2 = 0{,}90$ atau 90%.

g. Menetapkan tarif siginifikan dan melakukan uji t
Diketahui $r^2 = 0{,}9$, jumlah data n = 6
Hipotesis yang diajukan
H_0: $\beta = 0$; variabel X tidak berdampak signifikan terhadap Y
H_1: $\beta \neq 0$; variabel X berdampak signifikan terhadap Y
Tingkat signifikansi = 5%
Nilai t_{hit}, yaitu

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}95\sqrt{10-2}}{\sqrt{1-0.90}} = 8{,}497$$

Sehingga $t_{hit} = 8{,}497$, derajat kebebasan $\alpha = n - k = 6 - 2 = 4$. Gunakan tabel uji t dengan tingkat signifikansi 5%, $\alpha = 6$ maka diperoleh $t_{tab} = 2{,}776$. Sehinggah nilai $t_{hit} > t_{tab}$, hal ini menunjukkan pengaruh yang signifikan atara variabel bebas dan terikat.

2. Interpretasi terhadap nilai koefisien regresi, $b_1 = 0{,}42$.Nilai `$b_1 = 0{,}42$, Secara statistik, jika terdapat hubungan regresi antara pendapatan dan konsumsi dengan koefisien regresi sebesar 0,42, dapat disimpulkan bahwa setiap kenaikan pendapatan satu juta rupiah berkorelasi dengan peningkatan rata-rata konsumsi sebesar 0,42 juta rupiah. Sebaliknya, jika terjadi penurunan pendapatan satu juta rupiah, konsumsi akan mengalami penurunan rata-rata sebesar 0,42 juta rupiah. Ini menunjukkan arah dan besar pengaruh antara kedua variabel tersebut dalam konteks hubungan linier yang dijelaskan oleh model regresi.

3. Taksiran nilai konsumsi karyawan yang berpendapatan 10 juta dengan persamaan regresi yang diperoleh yaitu $\hat{Y} = 4{,}25 + 0{,}42X$, akan ditaksir nilai Y untuk $X = 10$, sebagai berikut untuk $X = 10 \rightarrow \hat{Y} = 4{,}25 + 0{,}42X$

$$\hat{Y} = 4{,}25 + 0{,}42(10) = 4{,}25 + 4{,}2 = 8{,}45$$

Jadi konsumsi seorang karyawan yang berpendapatan sebesar 10 juta ditaksir sebesar 8,45 juta.

10.4 Korelasi

Analisis regresi bertujuan untuk meneliti pola hubungan antara dua variabel, baik itu berbentuk linier atau tak linier. Fokus utama adalah memahami pengaruh variabel bebas X terhadap variabel terikat Y serta memberikan estimasi nilai berdasarkan nilai variabel X. Sementara itu, analisis korelasi bertujuan untuk mengevaluasi sejauh mana kedekatan hubungan antara variabel bebas X dan variabel terikat$Y$, tanpa mempertimbangkan apakah hubungan tersebut berbentuk linier atau tak linier. Korelasi sederhana adalah teknik statistik yang digunakan untuk menilai bentuk hubungan antara variabel bebas dan variabel terikat, sekaligus mengukur kekuatan hubungan tersebut dengan memperhatikan aspek hubungan timbal balik (dua arah). (Kemendikbud, 2023).

10.4.1 Koefisien Korelasi dan Koefisien Determinasi

Analisis korelasi untuk mengevaluasi sejauh mana hubungan antara dua variabel, tanpa mempertimbangkan variabel yang mempengaruhi atau dipengaruhi. Koefisien korelasi r merupakan indeks yang digunakan untuk mengukur tingkat kedekatan suatu hubungan antar variabel. Menurut Supranto (2008), persamaan untuk menghitung koefisien korelasi dapat dinyatakan sebagai berikut (J. Supranto, 2008). (J.Supranto, 2008)

$$r = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{[n\sum X_i^2 - (\sum X_I)^2][n\sum Y_i^2 - (\sum Y_I)^2]}} \quad (4)$$

Keterangan
X : Variabel prediktor
Y : Variabel respon
$i = 1,2, \ldots.$
n =jumlah data

Koefisien korelasi dapat mengambil nilai positif atau negatif dan rentang nilai koefisien korelasi berkisar antara -1 hingga 1. Korelasi negatif terlihat melalui koefisien korelasi yang bernilai negatif, sementara korelasi positif terindikasikan oleh nilai koefisien korelasi yang positif.

Koefisien determinasi adalah ukuran kontribusi variabel bebas X terhapad perubahan variabel Y. Koefisien determinasi dapat diketahui dengan hasil kuadrat koefisien korelasi r^2.

10.4.2 Interpretasi terhadap nilai koefisien korelasi

Untuk mengevaluasi kekuatan atau kelemahan derajat hubungan antara variabel X dan Y, secara sederhana, dapat merujuk pada pedoman yang disajikan oleh Elifson, dkk. (1990)., dalam Tabel 10.3.

Tabel 10.3. Klasifikasi nilai koefisien korelasi

Nilai koefisien korelasi (r) (positif/negatif)	**Interpretasi**
0,01-0,30	Hubungan variabel X dan variabel Y mempunyai korelasi yang lemah
0,31-0,70	Hubungan variabel X dan variabel Y mempunyai korelasi yang sedang
0,71-0,99	Hubungan variabel X dan variabel Y mempunyai korelasi yang kuat atau tinggi
1	Hubungan variabel X dan variabel Y mempunyai korelasi sempurna

(Sumber : Elifson, 1990. Diolah)

10.4.3 Klasifikasi Koefisien Korelasi

Adapun koefisien korelasi dibedakan menjadi :

1. Koefisien korelasi Pearson Correlation :
   Koefisien korelasi Pearson Correlation digunakan untuk mengukur hubungan dua variabel yang berbentuk data rasio maupun interval. Metode terdiri dari:
   a. Metode Least Square
      Koefisien korelasi least square dinyatakan sebagaimana persamaan sebagai berikut,

$$r = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{[n\sum X_i^2 - (\sum X_I)^2][n\sum Y_i^2 - (\sum Y_I)^2]}} \quad (5)$$

   b. Metode Product Moment Correlation

Data yang berbentuk skala ordinal dan bersifat non-parametrik statistik (Universitas Negeri Yogyakarta, 2022)

$$r = \frac{\sum X_i Y_i}{\sqrt{\sum X_i^2 \sum Y_i^2}} \quad (6)$$

Keterangan :
r = koefisien korelasi
x = deviasi rata-rata variabel X
y = deviasi rata-rata variabel Y.

2. Koefisien Korelasi Rank Spearman:
Indeks yang digunakan untuk mengukur keeratan hubungan diantara dua variabel yang berupa data ordinal. Biasanya isimbolkan dengan 'r_s'. Berikut formulanya :

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

Keterangan
r_s = koefisien korelasi rank spearman
d = selisih dalam ranking
n = banyaknya pasangan rank.

3. Koefisien Determinasi
Koefisien ini digunakan untuk menjelasakan seberapa besar dampak suatu nilai variabel bebas pada naik/turunnya nilai variabel terikat. Formulanya adalah :

$$r^2 = (r)^2 \times 100\%$$

Keterangan:
r = Koefisien korelasi.
Sebagai ilustrasi untuk mempermudah pemahaman diberi contoh dibawah ini.

Contoh korelasi Pearson Correlation:
Perusahaan beras mencatat produksi (Y) beras dalam satuan ton pada tahun 2023 dan upah tenaga kerja (X) dalam juta rupiah yang dilampirkan sebagai berikut:

Tabel 10.4. Deskripsi Upah Pegawai dan Produksi beras perusahaan

Pegawai	Upah Pegawai (juta rupiah)	Produksi beras (satuan ton)
A	2	4
B	4	5
C	6	8
D	7	10
E	8	12
F	10	16
G	11	16
H	12	20
I	13	21
J	14	24

Sumber: Penulis, 2023. Diolah

Dengan menganggap data tersebut sampel acak

1. tentukanlah persamaan regresi dengan metode kuadrat terkecil,.
2. koefisien garis regresi yang diperoleh, mohon diinterpretasikan
3. Hubungan koefisien korelasi dan berikan interpretasi
4. Hubungan koefisien determinasi dan berikan interpretasi

Penyelesaian:

1. Menghitung variabel yang akan digunakan untuk metode kuadrat terkecil.
   a. menetapkan tujuan dari regresi linier sederhana yaitu untuk mengetahui apakah kenaikan upah gaji dapat mempengaruhi produksi perusahaan beras
   b. Mengidentifikasi variabel bebas dan variabel terikat
      Variabel bebas (X) = jumlah produksi perusahaan
      Variabel terika (Y) = jumlah konsumsi
   c. Mengumpulkan/kolektif data (susun bentuk tabel) dan menghitung nilai untuk X^2, Y^2 dan total dari masing-masing

Tabel 10.5. Perhitungan Unsur-unsur Persamaan Regresi

Pegawai	X_i	Y_i	X_i^2	Y_i^2	X_iY_i
A	2	4	4	16	8
B	4	5	16	20	25
C	6	8	36	64	48
D	7	10	49	100	70
E	8	12	64	144	96
F	10	16	100	256	160
G	11	16	121	256	176
H	12	20	144	400	240
I	13	21	169	441	273
J	14	24	196	576	336
Jumah	87	138	899	2346	1449

Sumber: Penulis, 2023. Diolah

Banyak data $n = 10$

d. Menghitung nilai b_0 dan b_1

$$b_1 = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2}$$

$$= \frac{10\ (1449) - (87)(138)}{6(899) - (87)^2}$$

$$= \frac{14490 - 12006}{8990 - 7569} = \frac{2484}{1421} = 1{,}75$$

$$b_0 = \frac{\sum Y_i}{n} - \frac{b_1 \sum X_i}{n}$$

$$b_0 = \frac{138}{10} - \frac{1{,}75(87)}{10} = 13{,}8 - 15{,}23 = -1{,}43$$

Jadi persamaan regresinya :

$$\hat{Y} = -1{,}43 + 1{,}75X.$$

2. Interpretasi koefisien persamaan regresi. Nilai $b_1 = 1{,}75$, dapat dinterpretasikan jika upah tenaga kerja naik 1 juta maka produksi beras meningkat 1,75ton atau jika upah

tenaga kerja turun 1 juta rupiah maka produksi akan berkurang sebanyak 1,75 ton.

3. Hubungan koefisien korelasi dan berikan interpretasi, Analisis korelasi yang hasilnya merupakan koefisien korelasi, maka hasil koefisien korelasi yang didapatkan yaitu

$$r = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{[n\sum X_i^2 - (\sum X_I)^2][n\sum Y_i^2 - (\sum Y_I)^2]}}$$

$$r = \frac{10(1449) - (87)(138)}{\sqrt{[10(899) - (87)^2][10(2346) - (138)^2]}}$$

$$r = \frac{2484}{\sqrt{(1421) * (4416)}}$$

$$r = \frac{2484}{2505,02} = 0,99$$

Dari hasil diatas menunjukkan bahwa hubungan sangat kuat (99%). Antara variabel bebas dengan variabel terikat. Ini berarti besar UPAH gaji sangat mempengaruhi besar produk beras karyawan kelas industri.

4. Hubungan koefisien determinas dan berikan interpretasi
Koefisien determinasi

$$r^2 = (r)^2 = (0,99)^2 = 0,98$$

Nilai koefisien determinasi $r^2 = 0,98$ dapat dinterpretasikan bahwa 98% variasi produksi dipengaruhi oleh upah tenaga kerja dan 2% di pengaruhi faktor lain.

Contoh Koefisien Korelasi Rank Spearman

Berikut adalah data mengenai pengalaman kerja (dalam tahun) dan jumlah penjualan (juta rupiah) dari 10 orang sales untuk sebuah peru sahaan untuk HP.

Tabel 10.6. Deskripsi pengalaman kerja sales dan hasil penjualan

Salesman	Pengalaman	Hasil penjualan
A	6	50
B	5	33
C	3	41
D	8	61
E	9	35
F	4	32
G	10	57
H	7	30
I	1	45
J	2	22

Sumber: Penulis, 2023. Diolah

Berdasarkan data yang dikumpulkan, tentukan koefisien korelasi peringkatnya.

Penyelesaian

1. Mengidentifikasi variabel
   Variabel bebas (X) = waktu pengalaman kerja
   Variabel terika (Y) = hasil penjualan
2. Melakukan perangkingan terhadap variabel.

Tabel 10.7. Perhitungan Unsur-unsur Koefisien Korelasi

Sales	Pengalaman kerja (X)	Rank X	Penjualan Y	Rank Y	Selisih rank d_i	d_i^2
A	6	5	50	3	2	4
B	5	6	33	7	-1	1
C	3	8	41	5	3	9
D	8	3	61	1	2	4
E	9	2	35	6	-4	16
F	4	7	32	8	-1	1
G	10	1	57	2	-1	1
H	7	4	30	9	-5	25
I	1	10	45	4	6	36
J	2	9	22	10	-1	1
Jumlah						96

Sumber: Penulis, 2023. Diolah

Dari table 13.7, dapat diketahui bahwa $\sum {d_i}^2 = 98$
Sehingga dengan menggunakan rumus koefisien spearman, maka didapat

$$r_s = 1 - \frac{6\sum d^2}{n(n^2-1)}$$
$$r_s = 1 - \frac{6(96)}{10(10^2-1)}$$
$$r_s = 1 - \frac{588}{990}$$
$$r_s = 1 - 0{,}59$$
$$r_s = 0{,}41$$

Koefisien korelasi peringkatnya yang diperoleh sebesar 0,41.

Soal-soal Latihan

1. Pakar ekonomi meriset hubungan inflasi dan peredaran jumlah uang. Berikut data mengenai tingkat inflasi dan peredaran jumlah uang disuatu negara selama 10 tahun (2012 -2021).

Tabel 10.8. Deskripsi antara inflasi dan jumlah uang yang beredar

Tahun	Inflasi	Jumlah uang beredar (Triliun rupiah)
2012	10	164,2
2013	12,5	170,0
2014	12,6	171,2
2015	13,7	177,5
2016	14,1	166,4
2017	15,0	169,1
2018	14,4	165,9
2019	16,9	170,8
2020	14,4	168,9
2021	17,6	173,6

Sumber: Penulis, 2023. Diolah

Berdasarkan data tersebut,

a. Tentukan persamaan regresi dari kasus diatas
b. Tentukan koefisien regresi dan berikan interpretasi
c. Tentukan koefisien korelasi dan determinasi, serta masing-masing interpretasi

2. Dua orang pengunjung warkop yaitu Bambang dan Andi Bau sedang memberikan penilaian untuk rasa 12 jenis kopi yang ada di di warkop tersebut. Kopi yang terenak memiliki skor terbesar yaitu 12 dan yang tidak enak diberi skor 1. Berikut hasil perankingan yang diperoleh:

Tabel 10.9. Deskripsi peringkat jenis kopi berdasarkan rasanya

Kopi	Rank Bambang	Rank Sentosa
AA	4	1
BB	2	10
CC	7	5
DD	1	3
EE	9	6
FF	5	7
GG	8	9
HH	10	4
II	3	2
JJ	6	8

Sumber: Penulis, 2023. Diolah

Berdasarkan data diatas, hitunglah koefisien korelasi peringkatnya.

DAFTAR PUSTAKA

J.Supranto. 2008. *Statistik Teori dan Aplikasi*. Jakarta: Erlangga.

Kemendikbud. 2023. *Korelasi Linier Sederhana dan Regresi Linier*. Available at: https://lmsspada.kemdikbud.go.id/pluginfile.php/651977/mod_resource/content/2/BAB 5. Korelasi Linear Sederhana dan Regresi Linear.pdf .

Moore, D.S., McCabe, G.P. and Craig, B.A. 2014. *Introduction to the Practice of Statistics*. W.H Freeman.

Sitopu, J.W. and dkk. 2023. *UNTUK EKONOMI DAN BISNIS : di Lengkapi dengan Program IBM SPSS*. Padang: GPress.

Universitas Negeri Yogyakarta. 2022. *Analisis Korelasi dan Regresi*. Available at: https://pendidikan-akuntansi.fe.uny.ac.id/.https://pendidikan%0Aakuntansi.fe.uny.ac.id/sites/pendidikan akuntansi.fe.uny.ac.id/files/Korelasi dan Regresi.pdf%0A.

Wirawan, N. 2016. '67$7,67,.$ (.2120', *Statistika Deskritptif*, p. 330.

Yuliara, M. 2016. *Modul Regresi Linier Sederhana*. Available at: https://simdos.unud.ac.id/.https://simdos.unud.ac.id/uploads/file_pendidikan_1_dir/3218126438990fa0771ddb555f70be42.pdf%0A%0A.

BIODATA PENULIS

Raden Sri Ayu Ramadhana, S.Pd .I, M.Pd.
Dosen Program Studi Pendidikan Matematika
Fakultas Keguruan dan Ilmu Pendidikan

Raden Sri Ayu Ramadhana lahir di Kwala Begumit, tanggal 15 Februari 1992. Menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika Institut Agama Islam Negeri Sumatera Utara (IAIN SU) dan melanjutkan S2 pada Jurusan Pendidikan Matematika di Universitas Negeri Medan. Pada tahun 2014-2019 sebagai Tenaga Pengajar di Sekolah Tingkat Menengah dan Madrasah Aliyah Negeri di Langkat kemudian pada tahun 2019 sampai saat ini merupakan dosen tetap pada Program Studi Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan (FKIP) Universitas Al Washliyah Labuhanbatu.

BIODATA PENULIS

Arief Aulia Rahman, S.Pd., M.Pd.,CISHR
Dosen Program Studi Pendidikan Matematika FKIP
Universitas Muhammadiyah Sumatera Utara

Lahir di langsa, 11 Oktober 1991, anak kedua dari bapak Drs. Ahmad As'adi dan ibu Dra. Aminah Sulaiman. Telah menyelesaikan sekolah dasar, sekolah menengah pertama dan sekolah menengah atas di kota Langsa, Provinsi Nanggroe Aceh Darussalam. Setelah itu melanjutkan pendidikan pada program sarjana pendidikan Matematika di Universitas Muhammadiyah Sumatera Utara (UMSU) selama 3 tahun 8 bulan, pada tahun 2014 mengambil program master pendidikan matematika di Universitas Negeri Medan (UNIMED) selama 1 tahun 5 bulan. Pernah menjadi guru bidang Matematika di MAN 2 Model Medan dan Sekarang aktif sebagai dosen pendidikan matematika di Universitas Muhammadiyah Sumatera Utara (UMSU) dua kali mendapat penghargaan pada tahun 2017 dalam rangka pengabdian kepada masyarakat di Aceh Tamiang dan peringkat 3 dosen berprestasi tingkat LLDIKTI XIII Aceh Tahun 2020.

BIODATA PENULIS

Rini Yunita, S.Si, M.Sc
Dosen LLDIKTI Wilayah X
Pada Sekolah Tinggi Teknologi Payakumbuh

Penulis lahir di Padang pada tanggal 16 Juni 1984. Penulis saat ini mengajar pada Program Sudi Teknik Komputer dan Teknik Sipil pada Sekolah Tinggi Teknologi Payakumbuh. Penulis menyelesaikan pendidikan S1 Jurusan Matematika di Universitas Negeri Padang kemudian melanjutkan pendidikan S2 Jurusan Matematika di Universitas Gajah Mada.

Penulis merupakan dosen aktif di Sekolah Tinggi Teknologi Payakumbuh. Penulis mengampu mata kuliah Matematika Dasar, Aljabar Linear, Statistika Dasar, Matematika I dan II, dan Matematika Teknik.

BIODATA PENULIS

Enos Lolang, S.Si., M.Pd.
Dosen Program Studi Pendidikan Fisika
Fakultas Keguruan dan Ilmu Pendiidkan
Universitas Kristen Indonesia Toraja
Email: enos@ukitoraja.ac.id

Enos Lolang, Lahir di Makale, 11 Mei 1969. Penulis menempuh Pendidikan Sekolah Menengah Atas di SMA Negeri 1 Makale dan lulus pada tahun 1989. Penulis menyelesaikan pendidikan program sarjana di Universitas Hasanuddin pada tahun 1995 kemudian melanjutkan pendidikan magister di Universitas Negeri Malang dan menyelesaikann studi pada tahun 2013.

Penulis bertugas sebagai dosen program studi Pendidikan Matematika di Universitas Kristen Indonesia Toraja sejak 2000. Selanjutnya tahun 2017 ditetapkan sebagai dosen program studi Pendidikan Fisika sampai saat ini. Penulis telah menulis beberapa buku ajar dan buku referensi, khususnya dalam bidang pendidikan matematika. Tiga buku ajar untuk mahasiswa program studi pendidikan matematika yaitu Persamaan Diferensial (2011), Aljabar Abstrak (2015), dan Matematika Diskrit (2017). Selain itu penulis juga berpartisipasi dalam menyusun beberapa *book chapter*, antara lain Metode Penelitian Berbagai Bidang Keilmuan, Dasar Matematika, Kalkulus I (Diferensial), dan Aplikasi SPSS Untuk Analisis Data Penelitian Kesehatan, serta Statistitik dan Probabilitas pada tahun 2023. Penulis mulai mengembangkan kemampuan menulis setelah mengikuti Sertifikasi Penulisan Buku Non-Fiksi

yang diselenggarakan oleh Badan Nasional Sertifikasi Profesi pada tahun 2019, dan Pelatihan Asesor Kompetensi dalam bidang Personil Teknologi Informasi dan Asesor Kompetensi yang diselenggarakan oleh Lembaga Sertifikasi Personil PT. Pilar Pendidikan Pelatihan Indonesia pada tahun 2022.

BIODATA PENULIS

Ratu Sarah Fauziah Iskandar, M.PMat.
Dosen Program Studi Pendidikan Matematika
Universitas Muhammadiyah Tangerang

Penulis lahir di Bandung, 17 April 1989. Saat ini penulis tercatat sebagai mahasiswa Program Doktoral Pendidikan Matematika di Universitas Pendidikan Indonesia. Selain menjadi mahasiswa, penulispun bergabung dalam komunitas pendidikan dan aktif mengikuti kegiatan relawan di Tangerang dan mengajar di Program Studi Pendidikan matematika Universitas Muhammadiyah Tangerang sejak tahun 2014.

BIODATA PENULIS

Retno Andriyani , M.Pd
Dosen Program Studi Pendidikan Matematika
Fakultas Keguruan Dan Ilmu Pendidikan
Universitas Muhammadiyah Tangerang

Penulis lahir di Penual tanggal 11 Nopember 1990 Penulis adalah dosen tetap pada Program Studi Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan, Universitas Muhammadiyah Tangerang. Menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika Universitas Bung Hatta tahum 2007 dan melanjutkan S2 pada Jurusan pendidikan matematika Universitas Negeri Malang 2012.

BIODATA PENULIS

Ukta Indra Nyuswantoro, S.T.
Sr. Structure Engineer di PT Asiatek Energi Mitratama

Penulis menyelesaikan Pendidikan di SMAN 1 Geger Madiun, selanjutnya menempuh pendidikan S1 pada Jurusan Teknik Kelautan, Institut Teknologi Sepuluh Nopember Surabaya. Penulis mempunyai pengalaman selama 10 tahun sebagai structure engineer di beberapa perusahaan di Indonesia dan memiliki ketertarikan di bidang Mekanika Klasik dan Analisis Numerik. Buku yang pernah ditulis "Pengantar Matematika Lubang Hitam" dan "Representasi Pengetahuan".

BIODATA PENULIS

Siti Aminah S.Si., S.Pd., M.Pd.
Dosen Program Studi Informatika
Sekolah Tinggi Informatika & Komputer Indonesia

Penulis lahir di Malang tanggal 6 November 1989. Penulis adalah dosen tetap pada Program Studi Informatika Sekolah Tinggi Informatika & Komputer Indonesia. Menyelesaikan pendidikan S1 dengan gelar ganda pada Jurusan Matematika dan Pendidikan Matematika di Universitas Negeri Malang. Kemudian melanjutkan S2 pada Jurusan Pendidikan Matematika di Universitas Negeri Malang. Penulis menekuni bidang menulis sejak tahun 2020. Buku pertama penulis bersama teman-teman menulis buku berjudul "*Membaca Arah Zaman Mozaik Pemikiran Kuliah Kreatif di Era new Normal*". Pada tahun 2022 penulis bersama teman-teman menulis kembali buku berjudul "*Metodologi Penelitian Ilmiah dalam Disiplin Ilmu Sistem Informasi*".

BIODATA PENULIS

Ridha Yuniara, S.Pd., M.Pd.
Dosen Program Studi Ekonomi Pembangunan
Fakultas Ekonomi dan Bisnis Universitas Gajah Putih

Penulis lahir di Takengon tanggal 11 Juni 1993. Penulis adalah Dosen Tetap di Program Studi Ekonomi Pembangunan Fakultas Ekonomi dan Bisnis, Universitas Gajah Putih. Menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika dan melanjutkan S2 pada Jurusan yang sama di Universitas Syiah Kuala.

BIODATA PENULIS

Kusnaeni, S.Si., M.Si.
Dosen Program Studi Matematika
Jurusan Sain Universitas B.J. Habibie

Penulis lahir di Makassar tanggal 16 Januari 1995. Penulis adalah dosen Program Studi Matematika Institut Teknologi Bacharuddin Jusuf Habibie. Menyelesaikan pendidikan S1 pada Jurusan Matemtika Universitas Hasanuddin dan melanjutkan S2 pada Jurusan Statistika dan Sains Data Institut Pertanian Bogor. Penulis menekuni penelitian-penelitian pada bidang sains data.