STATISTIKA

Yenny Anggreini Sarumaha
Hetty Elfina
Rahmi Wahyuni
Suryanti
Novrianti
Ariyani Muljo
Retna Kristiana
Novita Aswan
Joni Wilson Sitopu
Laelatul Khikmah
Siti Ulfa Nabila
Hanna Hilyati Aulia

GETPRESS INDONESIA

STATISTIKA

Penulis :
Yenny Anggreini Sarumaha
Hetty Elfina
Rahmi Wahyuni
Suryanti
Novrianti
Ariyani Muljo
Retna Kristiana
Novita Aswan
Joni Wilson Sitopu
Laelatul Khikmah
Siti Ulfa Nabila
Hanna Hilyati Aulia

ISBN :
Editor : Ari Yanto., M.Pd
Resi Sevita,S.Pd
Penyunting: Yuliatri M.Hum.
Desain Sampul dan Tata Letak :
Penerbit : GETPRESS INDONESIA
Anggota IKAPI No. 033/SBA/2022

Redaksi :
Jl. Palarik RT 01 RW 06 Kelurahan Air Pacah
Kecamatan Koto Tangah Padang Sumatera Barat

website: www.getpress.co.id
email: adm.getpress@gmail.com

Cetakan pertama, November 2023

KATA PENGANTAR

Dengan mengucapkan puji syukur kehadirat Allah SWT, atas limpahan rahmat dan hidayahNya, maka Penulisan Buku dengan judul Statistika dapat diselesaikan dengan baik. Buku ini berisikan tentang konsep statistika, pemusatan data, penyebaran data, peluang, serta analisis korelasi dan regresi.

Buku ini masih banyak kekurangan dalam penyusunannya. Oleh karena itu, kami sangat mengharapkan kritik dan saran demi perbaikan dan kesempurnaan buku ini selanjutnya. Kami mengucapkan terima kasih kepada berbagai pihak yang telah membantu dalam proses penyelesaian Buku ini. Semoga Buku ini dapat menjadi sumber referensi dan literatur yang mudah dipahami.

Padang, November 2023

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB 1
PENGERTIAN STATISTIKA

Oleh Yenny Anggreini Sarumaha

1.1 Definisi

Statistika adalah ilmu menganalisis data di mana peluang umum berperan (Ewens and Brumberg, 2023). Statistika secara informal sudah ada berabad-abad yang lalu. Catatan awal korespondensi antara matematikawan Prancis, Pierre de Fermat dan Blaise Pascal pada tahun 1654 sering disebut-sebut sebagai contoh awal analisis probabilitas statistika. Statistika merupakan cabang dari matematika terapan yang melibatkan pengumpulan, deskripsi, analisis, yang hasilnya digunakan untuk membuat keputusan dan menyelesaikan masalah (Kahl, 2023). Sebagai bagian dari matematika, statistika berupaya membuat susunan dari beragam koleksi fakta dengan merangkum sifat-sifat dari kelompok besar (Carrol and Carrol, 2023).

Sebagian buku menyebutkan bahwa statistika adalah seni dan ilmu yang menitikberatkan pada pengembangan dan metode studi untuk mengumpulkan, menganalisa, menginterpretasi, dan menampilkan data empiris, biasanya dalam kumpulan besar data numerik (Blatchley, 2019), untuk menjawab pertanyaan investigasi (Hinton, 2004; Shayib, 2013; Agresti, Franklin and Klingenberg, 2023). Statistika dapat dikomunikasikan pada berbagai tingkat mulai dari deskriptor non-numerik (tingkat nominal) hingga numerik yang mengacu pada titik nol (tingkat rasio). Teknik statistik merangkum data dan menyatakan ulang data kompleks menjadi lebih sederhana sehingga informasi yang diperoleh dapat digunakan dan

ditindaklanjuti. Obyek dasar dari statistika matematika adalah model statistik yang memiliki data untuk diobservasi sehingga menghasilkan variabel random dengan nilai tertentu dan distribusinya diberikan oleh hasil pengukuran dalam bentuk perkiraan (Lauritzen, 2023). Variabel dalam statistika berbeda dari variabel dalam aljabar, dimana dalam statistika variabel merujuk pada karakteristik individu atau benda (Gould, Wong and Ryan, 2020).

Orang yang mempelajari statistika disebut sebagai ahli statistik atau *statistician.* Ahli statistika ini sangat fokus dalam menentukan bagaimana menarik simpulan yang reliabel tentang kelompok besar dan situasi umum dari tingkah laku dan karakteristik lain yang dapat diamati dari sampel. Sampel ini mewakili sebagian besar atau sejumlah contoh fenomena atau kondisi umum. Namun, pada kenyataannya, tidak hanya ahli statistik yang mempelajari statistika, berbagai kalangan mulai dari bisnis administrasi, ilmu informasi, psikologi, sosiologi, pariwisata, dan lain sebagainya juga mempelajari statistika. Dalam setiap lini kehidupan, surat kabar, lingkungan kerja, pengambilan keputusan, dalam kehidupan sehari-hari, statistika digunakan (Salvati *et al.*, 2021; Rumsey, 2023)

Statistika merupakan ilmu pada area atau bidang interdisipliner, yang berarti statistika digunakan dihampir semua disiplin ilmu, seperti sains, teknologi, teknik, obat-obatan, psikologi dan lain sebagainya. Tidak heran jika kita sering menemukan statitika diterapkan hampir di semua bidang ilmiah dan pertanyaan penelitian di berbagai bidang ilmiah yang memotivasi pengembangan metode dan teori baru dari statistik (Agresti, Franklin and Klingenberg, 2023). Penelitian sendiri bermakna proses investigasi sistematis yang dilakukan untuk memperkaya pengetahuan yang dapat membantu kita memahami sebanyak mungkin variabel yang kita bisa (Watt and Collins, 2023). Sebagian besar pertanyaan penelitian bersifat bivariat dan analis

data kemudian mengeksplorasi hubungan antara variabel atau pengaruh satu variabel terhadap variabel lainnya.

Data analisis selalu dimulai dengan menggunakan statistik dan plot untuk merangkum, mendeskripsikan, dan menampilkan data dari variabel tunggal atau analisis data univariat (Blaine, 2023). Karenanya variabel-variabel tersebut dikumpulkan kemudian dianalisis dengan bantuan beberapa alat, seperti SPPS, *Excel*, R, atau *Machine Learning Toolbox* (MATLAB) (Kronthaler, 2023; Mathworks, 2023; Salcedo and McCormick, 2023). Teori matematika yang berada di balik statistika sangat bergantung pada kalkulus diferensial dan integral, aljabar linier, dan teori probabilitas. Beberapa teknik pengambilan sampel dapat digunakan untuk menyusun data statistik, anatar lain pengambilan sampel acak sederhana, sistematik, bertingkat, dan cluster.

Pada praktiknya, statistika adalah ide bahwa kita dapat mempelajari sifat-sifat sekumpulan besar objek atau peristiwa (populasi) dengan mempelajari karakteristik sejumlah kecil objek atau peristiwa serupa (sampel) (Hinton, 2004). Pengumpulan data komprehensif tentang populasi dalam banyak kasus menghabiskan banyak biaya, sulit atau tidak mungkin dilakukan, sehingga, akan jauh lebih meringankan jika dimulai dengan sampel yang dapat diamati, yang bisa diperoleh dengan mudah atau terjangkau. Ahli statistika mengukur dan mengumpulkan data tentang individu atau elemen sampel, kemudian menganalisis data tersebut untuk menghasilkan statistik deskriptif. Kemudian, dengan menggunakan karakteristik yang telah diamati dari data sampel, yang disebut dengan statistik, untuk membuat simpulan atau dugaan tentang karakteristik populasi yang lebih luas yang tidak terukur atau tidak dapat diukur, yang dikenal sebagai parameter.

Dua ide dasar dalam bidang statistika adalah data (ketidakpastian) dan variasi (Gould, Wong and Ryan, 2020). Banyak kejadian atau situasi yang kita temukan di sains (atau secara umum dalam kehidupan) memiliki hasil yang

tidak pasti. Sebagai contoh, beberapa kasus mengenai ketidakpastian ini disebabkan jawaban dari pertanyaan belum dapat dipastikan (seperti apakah besok akan hujan?) atau dalam kasus lainnya, ketidakpastian disebabkan ketika hasil telah ditentukan namun kita tidak menyadarinya (seperti apakah kita lulus ujian tertentu). Statistika tidak hanya dipelajari dengan tujuan untuk memahami kemampuan teknis statistik, sebaliknya dengan mempelajari statistika kemampuan teknis dapat dikembangkan dan digunakan untuk mendiskusikan dan mengkomunikasikan ide-ide tentang fenomenta penting dengan tujuan utama mendiskusikan implikasi dari informasi untuk aksi sosial (Series, 2022).

Pendekatan statistik untuk menganalisis data dimulai dengan model probabilitas untuk meskripsikan proses generalisasi data, karenanya untuk mempelajari statistika, seseorang harus telebih dahulu belajar bahasa probabilitas (Zhu, 2023). Probabilitas adalah bahasa matematika yang digunakan untuk mendiskusikan kejadian yang tidak pasti. Selain itu, probabilitas juga merupakan penalaran statistik dasar yang digunakan untuk melihat kemungkinan berbagai hasil yang dapat terjadi (Agresti, Franklin and Klingenberg, 2023). Probabilitas adalah ukuran statistik yang umum ditemui yang membantu untuk menemukan kemungkinan hasil dari kejadian tertentu (Kahl, 2023). Setiap kegiatan pengukuran atau pengumpulan data berada di bawah sejumlah sumber variasi. Dalam hal ini, jika pengukuran yang sama diulangi, jawaban yang dihasilkan kemungkinan besar akan berubah. Dengan kata lain, teori probabilitas membuat deduksi atau implikasi, sedangkan statistika membuat induksi atau inferensi (Ewens and Brumberg, 2023). Setiap induksi atau inferensi yang dibuat selalu didasarkan pada data dan kalkulasi teori probabilitas yang berkorespondensi dengan data tersebut. Para ahli statistika berupaya memahami dan mengendalikan (jika

memungkinkan) sumber dalam situasi apa pun. Dengan memahami kemungkinan suatu peristiwa terjadi, peneliti dapat membuat prediksi lebih lanjut tentang masa depan kumpulan data, serta lebih memahami data yang dikumpulkan.

Statistika digunakan secara luas di berbagai aplikasi dan profesi, mulai dari lembaga pemerintah hingga investasi. Statistika dilakukan setiap kali data dikumpulkan dan dianalisis. Statistika memberikan informasi untuk mendidik bagaimana segala sesuatunya bekerja. Statistika sering ditemui saat melakukan penelitian, mengevauasi hasil, mengembangkan pemikiran kritis, dan membuat keputusan. Statistika dapat digunakan untuk menyelidiki hampir semua bidang studi untuk menginvestigasi mengapa sesuatu terjadi, kapan hal tertentu terjadi, dan apakah terulangnya hal tersebut dapat diprediksi.

Statistika adalah sebuah alat. Jika digunakan secara tidak hati-hati akan menyesatkan, dan jika digunakan tanpa tanggung jawab maka hasilnya membawa pada pemikiran irasional (Watt and Collins, 2023). Terdapat beberapa alasan pentingnya mempelajari statistika (Kronthaler, 2023). Pertama, sebagian besar pengetahuan di dunia didasarkan pada data dan analisis data. Argumen kedua adalah bahwa data kadang digunakan untuk membuat opini, seburuk-buruknya digunakan untuk memanipulasi individu. Terakhir, analisis data membantu kita membuat keputusan yang tepat.

1.2 Data

Statistika adalah ilmu tentang data (Gould, Wong and Ryan, 2020). Dalam statistika, hal yang sangat menentukan adalah analisis data (Kronthaler, 2023). Data diartikan sebagai bukti yang dikumpulkan secara sistematis (Gould, Wong and Ryan, 2020)untuk menyediakan informasi mengenai pertanyaan penelitian yang dimiliki (Watt and Collins, 2023). Data dapat diperoleh

menggunakan satu atau lebih cara, baik dalam observasi maupun desain eksperimen (Agresti, Franklin and Klingenberg, 2023; Kahl, 2023). Namun, tidak menutup kemungkinan, data diperoleh dari pengambilan acak atau eksperimen random.

1.2.1 Observasi

Cara pertama yang dapat digunakan untuk mengumpulkan data adalah dengan melakukan observasi. Dalam observasi, peneliti tidak melakukan intervensi dalam pengumpulan data yang akan diolah, sebalikanya peneliti mengambil data langsung pada saat kejadian berlangsung dan berusaha memastikan terdapatnya tren khusus atau hasil pada pengambilan data tersebut. Cara lain yang digunakan dalam pengambilan data melalui observasi adalah melalui penambangan data (*data mining*), dengan kata lain memperoleh informasi dari kumpulan data yang telah dilakukan sebelumnya, seperti data sejarah.

1.2.2 Desain eksperimen

Suatu penelitian atau studi yang secara acak menetapkan peserta sample ke tingkat variabel X disebut sebagai eksperimen (Blaine, 2023). Dalam desain ekperimen, peneliti dengan kesengajaan melakukan perubahan pada variabel yang dapat dikontrol dalam suatu sistem, skenario, atau proses, mengamati data yang diperoleh setelah perubahan dilakukan dan kemudian membuat simpulan tentang perubahan yang terjadi. Desain eksperimen memiliki peran yang cukup besar dalam sains, manufaktur, ilmu kesehatan, dan teknik karena membantu peneliti mengeliminasi faktor-faktor perancu atau pengecoh sehingga dapat menarik simpulan yang kuat. Bagian penting dari desain eksperimen adalah pengujian hipotesis. Hipotesis adalah sebuah ide tentang faktor atau proses yang diterima atau ditolak peneliti berdasarkan data (Shayib, 2013). Prosedur pengambilan keputusan terhadap

hipotesis ini disebut sebagai pengujian hipotesis. Pengujian hipotesis adalah salah satu cara untuk memperoleh data selama eksperimen, di mana prosedur ini memungkinkan peneliti untuk secara tepat menyeleksi berbagai faktor yang akan dibuktikan atau disangkah sebagai bagian dari pelaksanaan eksperimen.

1.3 Variabel

Terdapat banyak tipe data berbeda yang digunakan dalam statistika. Data diperoleh dari karakteristik tentang individu, obyek, atau kondisi atau keadaan (Carrol and Carrol, 2023). Data dalam statistika juga dikenal dengan nama variabel. Statistika sangat bergantung pada variabel. Secara definisi, variabel adalah sifat-sifat atau karakteristik dari suatu kejadian, objek atau orang yang memiliki nilai dan jumlah yang berbeda. Dalam desain ekperimen dan pengujian hipotesis, nilai-nilai tersebut dimanipulasi oleh peneliti sebagai bagain dari penelitian. Variabel adalah kumpulan data yang dapat dihitung, yang menandai suatu ciri atau atribut suatu benda. Misalkan pada sebuah mobil, variabelnya antara lain merek, model, tahun, jarak tempuh, warna, kondisi, dan lain sebagainya. Dengan menggabungkan variabel-variabel tersebut pada suatu kumpulan data, seperti warna mobil di tempat parkir tertentu, statistika memungkinkan kita memahami tren dan memperoleh hasil yang lebih baik.

Menurut Kahl (2023), terdapat enam jenis variabel berbeda (Kahl, 2023) di antaranya variabel independen, dependen, diskrit, kontinu, kualitatif, dan kuantitatif. Hubungan antara input dan output dapat dijelaskan dengan model matematis yang berhubungan dengan input, perancu atau pengecoh dan pengubah efek (sering disebut sebagai variabel independen dan dilambangkan dengan X) dengan output (sering disebut sebagai variabel dependen dan dilambangkan dengan Y) (Campbell and Jacques, 2023). Di beberpa sumber ditulis bahwa variabel yang dimanipulasi

oleh peneliti disebut sebagai variabel independen, dan efeknya pada variabel dependen diukur. Jumlah variabel independen setara dengan jumlah kondisi atau perlakukan dalam ekperimen yang dilakukan. Dalam pengukuran tekanan darah, misalnya, tekanan daran sebagai *Y* dan tingkat stres dan jenis kelamin sebagai variabel *X*. Di sini *X* tidak membedakan antara perancu dan kasualitas. Dalam hal ini ingin dilihat apakah stres merupakan prediktor tenanan darah yang baik jika jenis kelamin seseorang diketahui. Untuk menlakukannya, perlu diasumsikan bahwa jenis kelamin dan stres membentuk kombinasi untuk mempengaruhi tekanan darah.

Menurut Campbell dan Jacques (2023), data dapat dibagi ke dalam dua tipe, yaitu kualitatif dan kuantitatif (Campbell and Jacques, 2023). Pada sumber lain, disebutkan bahwa variabel dikelompokkan menjadi dua yaitu variabel numerik dan kategori (Gould, Wong and Ryan, 2020). Variabel kualitatif merupakan atribut spesifik yang seringkali bersifat non-numerik atau dapat dikatakan variabel ini tidak dapat dinyatakan dalam bentuk angka, biasanya dalam bentuk kategori. Jika merujuk pada contoh variabel mobil sebelumnya, contoh-contoh tersebut bersifat kualitatif. Contoh lain dari variabel kualitatif ini dalam statistika seperti jenis kelamin, warna rambut, kota asal, dan lainnya. Data kualitatif ini sering digunakan untuk menentukan persentase hasil yang terjadi untuk variabel kualitatif tertentu.

Variabel kualitatif juga merupakan variabel kategori (Campbell and Jacques, 2023; Carrol and Carrol, 2023). Analisis yang dilakukan dari data kualitatif ini seringkali tidak bergantung pada angka. Misalnya, dalm menentukan persentase jumlah ibu rumah tangga yang memiliki usaha kecil dengan menganalisis data kualitatif. Data kualitatif dapat dipresentasikan dalam bentuk diagram batang, *pie chart*, dan *pareto chart* (Shayib, 2013).

Variabel kuantitatif dipelajari secara numerik dan hanya memiliki kekurangan ketika variabel tersebut berada pada deskriptor non-numerik. Pada analisis kuantitatif, informasi yang diolah berakar pada angka (Watt and Collins, 2023). Pada contoh mobil sebelumnya, jarak tempuh yang dilalui mobil disebut sebagai variabel kualitatif, tapi angka 400 tidak memiliki nilai kecuali jika angka tersebut dikaitkan dengan total jarak tempuh. Dalam melaksanakan risetnya, peneliti menggunakan statistika untuk memperoleh informasi baru dari kelompok sampel kecil untuk meyakinkan ide tentang populasi yang lebih besar.

Variabel kuantitatif juga dapat dibagi menjadi dua kategori (Campbell and Jacques, 2023). Pertama, variabel diskrit, yang berada pada skala atau dalam rentang tertentu. Salah satu contoh yang dapat digunakan untuk menyatakan variabel diskrit ini adalah rentang usia pasien yang dipilih peneliti dalam penelitiannya. Misalkan pasien yang diinginkan adalah laki-laki yang berusia antara 35 dan 50 tahun. Usia setiap pasien dalam penelitian tersebut berada dalam skala diskrit dengan range antara 35 sampai 50 tahun. Maka setiap tahunnya, data usia yang dilaporkan seperti 35 atau 36 tahun, tapi tidak 36,5 tahun. Contoh lainnya bisa dilihat pada jumlah anak tiap keluarga atau jumlah anak yang terserang asma tiap bulannya.

Kedua adalah variabel kuantitatif berkelanjutan atau kontinu. Seperti waktu yang dibutuhkan untuk menjawab pertanyaan, bisa jadi berbeda, contohnya jawaban yang dihasilkan bisa berkisar antara 2,42 detik hingga 8,3761329 detik. Tidak ada langkah-langkah diskrit yang berkaitan dengan jenis data ini, oleh karena itu data tersebut dideskripsikan sebagai data kontinu. Untuk jenis data ini, akan lebih baik jika dilakukan pembatasan data dengan memotong nilainya pada titik tertentu, misalnya pada nilai tempat persepuluhan atau perseribu. Contoh

lainnya dapat diperoleh pada pengukuran tinggi, berat, atau tekanan darah.

Variabel kontinu sering dideskripsikan memiliki distribusi normal atau atau tidak normal (Campbell and Jacques, 2023). Memiliki distribusi normal berarti jika diplot sebuah histogram dari data yang dimiliki maka akan mengikuti kurva bentuk lonceng dengan nilai mean dan median yang cukup dekat. Distribusi tidak normal cenderung memiliki distribusi asimetris atau *skew* dan perbedaan signifikan antara mean dan median. Data kuantitatif dapat direpresentasikan dalam bentuk diagram titik, *stem and leaf display*, dan histogram (Shayib, 2013).

1.4 Statistika Deskriptif dan Inferensial

Dua bidang utama dari statistika adalah statistika deskriptif, yang menggambarkan sifat-sifat data sampel dan populasi, dan statistika inferensial, yang menggunakan sifat-sifat tersebut untuk menguji hipotesis dan menarik simpulan. Statistika deskriptif meliputi mean atau rata-rata, varian, *skewness* (kesimetrisan – gambaran distribusi data seperti miring ke kiri atau ke kanan), dan *kurtosis* (gambaran disribusi data seperti cenderung rata atau runcing). Sedangkan statistika inferensial meliputi analisis regresi linier, analisis varians (ANOVA), model logit/Probit, dan pengujian hipotesis.

1.4.1 Statistika Deskriptif

Statistika deskriptif sesuai namanya berperan menyediakan deskripsi pola dari sekumpulan data, agar data tersebut dapat diterima dan dijelaskan ke orang lain. Statistika deskriptif sebagian besar lebih berfokus pada kecenderungan umum, variabilitas, dan distribusi data sampel. Kecenderungan umum dapat diartikan sebagai perkiraan karakteristik, elemen tertentu dari suatu sampel atau populasi. Dalam hal ini, statistika deskriptif mencakup mean, median, dan modus. Variabilitas mengacu pada

sekumpulan statistik yang menunjukkan seberapa besar perbedaan antara karakteristik sampel atau populasi selama pengukuran dilakukan (Blaine, 2023; Watt and Collins, 2023). Pembahasan mengenai variabilitas dalam variabel numerik mengacu pada empat perhitungan statistik yaitu variansi, standar deviasi, *median absolute deviation* (MAD), dan *interquartile range* (IQR) (Blaine, 2023).

Distribusi mengacu pada keseluruhan "bentuk" data, yang dapat ditunjukkan atau dipresentasikan pada grafik seperti histogram atau diagram titik, dan mencakup sifat-sifat seperti fungsi distribusi, *skewness*, dan *kurtosis*. Statistik deskriptif juga dapat menggambarkan perbedaan antara karakteristik yang diamati dari elemen-elemen suatu kumpulan data. Hasil yang diperoleh dapat membantu kita memahami sifat-sifat kolektif dari elemen-elemen sampel data dan menjadi dasar untuk melakukan pengujian hipotesis dan membuat prediksi menggunakan statistik inferensial.

1.4.2 Statistika Inferensial

Statistika inferensial adalah sebuah alat yang digunakan para ahli statistik untuk menarik simpulan tentang karakteristik suatu populasi, yang diambil dari karakteristik suatu sampel, dan untuk menentukan seberapa yakin mereka terhadap reliabilitas simpulan tersebut. Berdasarkan ukuran dan distribusi sampel, ahli statistik dapat menghitung probabilitas statistik sampel tersebut di antaranya mengukur kecenderungan sentral, variabilitas, distribusi, dan hubungan antar karakteristik dalam suatu sampel data, memberikan gambaran akurat tentang parameter dari populasi yang merupakan sumber dari pengambilan sampel (Agresti, Franklin and Klingenberg, 2023).

Statistika inferensial digunakan untuk memperkirakan seberapa besar ketidakpastian dari

informasi yang telah kita kumpulkan, sehingga kita sadar seberapa banyak pengetahuan dari suatu populasi yang kita dapatkan. Seperti memperkirakan permintaan rata-rata suatu produk dengan mensurvei sampel kebiasaan membeli konsumen atau mencoba memprediksi kejadian di masa depan. Ini dapat berarti memproyeksikan keuntungan masa depan dari suatu sekuritas atau kelas aset berdasarkan keuntungan dalam periode tertentu (sampel). Terdapat dua jenis metode statistika inferensial yaitu estimasi parameter populasi dan pengujian hipotesis tentang nilai parameter.

Analisis regresi adalah teknik dalam statistik inferensial yang banyak digunakan untuk menentukan kekuatan dan sifat hubungan (korelasi) antara varabel terikat dan satu atau lebih variabel bebas (independen). Luaran dari model regresi sering kali dianalisis untuk mengetahui signifikansi statistiknya, yang mengacu pada klaim bahwa hasil temuan yang dihasilkan melalui pengujian atau eksperimen tidak mungkin terjadi secara acak atau kebetulan. Kemungkinan besar hal ini disebabkan oleh alasan spesifik atau khusus yang dapat dijelaskan oleh data.

1.5 Mean, Median, dan Modus

Distribusi frekuensi dan teknik grafis menghasilkan profil dasar dari kumpulan data yang dimiliki (Carrol and Carrol, 2023). Istilah mean, median, dan modus sering kali ditemui dalam pengolahan data. Ketiga perhitungan ini masuk ke dalam hasil pengolahan data umum yang dapat mendeskripsikan keumuman dari data atau kelompok sampel tertentu yang diperoleh.

Mean atau *average* adalah istilah yang digunakan untuk menyimpulkan nilai dari suatu kelompok dalam bentuk tunggal (Carrol and Carrol, 2023) dan lebih sering digunakan daripada median dan modus (Hinton, 2004). Mean adalah rata-rata dari seluruh nilai n, di mana n adalah

ukuran sampel (Blaine, 2023). Mean merupakan titik keseimbangan geometris dalam distribusi, atau nilai yang menyeimbangkan "berat" dari titik di atas dan di bawahnya. Mean melibatkan setiap nilai dalam perhitungannya sehingga dapat dipengaruhi oleh nilai yang sangat besar atau sangat kecil. Namun, potensi pengaruh nilai ekstrim tertentu dimoderasi oleh ukuran sampel. Semain besar n makan semakin kecil pengaruh nilai ekstrim terhadap mean. Dalam hal ini, mean akan berubah jika terdapat penambahan atau pengurangan sampel.

Median juga disebut sebagai nilai rata-rata yang didefinisikan dengan cara berbeda (Kronthaler, 2023). Median adalah nilai tengah (jika n ganjil) atau rata-rata dari dua nilai tengah (jika n genap) dari sekumpulan nilai x yang berurutan (Blaine, 2023) ketika kita menyusunnya dari nilai terendah hingga nilai tertinggi (Hinton, 2004). Median juga disebut sebagai kuartil kedua. Kebalikan dengan mean, median mendeskripsikan lokasi berdasarkan satu atau dua nilai, membuat median menjadi kurang memadai dibandingkan mean. Namun, median merupakan statistik yang lebih resisten dibandingkan dengan mean karena nilai yang sangat rendah ataupun sangat tinggi tidak akan mengubah nilai tengah.

Modus adalah nilai yang paling sering muncul (Kronthaler, 2023) atau batang terpanjang pada histogram (Hinton, 2004). Modus tidak dihitung secara statistik, modus diidentifikasi dari tabel frekuensi. Modus disebut kurang sesuai untuk merepresentasikan data karena hanya didasarkan pada nilai yang paling sering muncul dan perubahan kecil pada nilai sampel dapat mengakibatkan perubahan nilai modus (Blaine, 2023). Jika kumpulan data memiliki beberapa nilai dengan frekuansi tinggi yang sama, maka akan terdapat beberapa modus. Ketika terdapat beberapa nilai modus di mana tiap nilai berhubungan satu dengan lainnya, maka perlu adanya pertimbangan dalam analisis data.

1.6 Tingkat Pengukuran Statistik

Segala hal yang ada dan ada dengan jumlah tertentu maka akan dapat diukur. Pengukuran melibatkan kuantitas orang, obyek, atau kejadian dalam karakteristiknya. Ketika mengumpulkan informasi tentang orang, obyek, atau kejadian, informasi tersebut diubah ke dalam angka sehingga dapat diukur dan dideduksi tentang apa yang kita temui. Terdapat beberapa tingkat atau skala pengukuran dalam statistika (Shayib, 2013; Carrol and Carrol, 2023), di antaranya

1.6.1 Skala Pengukuran Nominal

Skala pengukuran pertama dalam hierarki adalah skala nominal. Istilah nominal berarti untuk diberikan nama. Sifat-sifat dari skala nominal yaitu kategori data eksklusif secara mutual dan tidak memiliki susunan logis. Pada skala ini, tidak ada nilai numerik atau kuantitatif, dan kualitas tidak diberi peringkat. Pengukuran pada skala nominal tidak hanya berupa label atau kategori sederhana dialihkan pada variabel lainnya. Angka atau nomor yang diberikan pada kategori-kategori tersebut tidak memiliki urutan, ranking, lebih tinggi dari, lebih rendah dari, lebih dari, dan kurang dari yang berkaitan. Dengan kata lain, skala pengukuran nominal merupakan fakta non numerik tentang variabel. Sebagai contoh Presiden Indonesia terpilih pada tahin 2019 adalah Ir. H. Joko Widodo.

1.6.2 Skala Pengukuran Ordinal

Tingkatan kedua dari skala dalam hierarki disebut sebagai ordinal, di mana terdapat ranking dan urutan dari suatu atribut dalam kategori yang berbeda. Hasil yang diperoleh melalui pengukuran ini dapat disusun secara berurutan, tapi seluruh data memiliki nilai atau bobot yang sama. Terdapat hubungan kualitatif di antara angka dalam skala ordinal. Berbeda dari skala nominal, angka dalam skala ordinal memiliki makna. Skala ordinal memberikan

informasi dan data yang lebih tepat dari pada skala nominal. Beberapa sifat dari skala ordinal adalah kategori data saling eksklusif, kategori data mempunyai urutan logis, dan kategori data diskalakan menurut jumlah karakteristik tertentu yang mereka miliki.

Meskipun bersifat numerik, skala ordinal, dalam statistika operasi pengurangan antar data yang dimiliki tidak dapat dilakukan karena dalam statistika hanya posisi dari titik data yang menjadi pertimbangan. Penggunaan skala ordinal yang paling sering digunakan adalah berkaitan dengan rating, pilihan, ranking, tujuan yang ingin dicapai, kepuasan, derajat kualitas, dan level persetujuan, seperti skala Likert yang sering digunakan dalam kuesioner. Skala ordinal sering dikategorikan sebagai statistika nonparametrik dan dibandingkan dengan total kelompok variabel. Contoh, buku terlaris pertama di dunia adalah *A tale of Two Cities* oleh Charles Dickens.

1.6.3 Skala Pengukuran Interval dan Rasio

Terdapat dua skala metrik dari pengukuran yaitu skala interval dan rasio. Karena sama-sama menggunakan metrik, pengukuran kedua skala ini memberikan data yang tepat. Hasil yang diperoleh dari pengukuran level interval ini dapat diatur secara berurutan dan perbedaan antara nilai-nilai dari data memiliki makna tersendiri. Dua titik data sering digunakan untuk membandingkan lama waktu yang dihabiskan atau perubahan kondisi dalam sekumpulan data. Dalam skala interval dan rasio, jarak antara angka 2 dan 3 dan antara 5 dan 6 tepat sama.

Satu-satunya perbedaan antara skala interval dan rasio adalah peran dari nol (0). Dalam skala interval, nol adalah buatan bukan sebenarnya. Sedangkan dalam skala rasio, peran dari nol benar nyata dan terdapat variabel yang tidak ada saat diukur. Pada skala interval seringkali tidak terdapat titik awal dari rentang data dan tidak memiliki nilai intrinsik nol yang berarti. Contoh pada April 2023

suhu tertinggi di Indonesia tercatat terjadi di Ciputat dengan suhu 37,2 derajart Celcius, sebelumnya suhu dengan nilai ini tercatat terjadi di Banten.

Sifat-sifat dari skala interval yaitu kategori data saling eksklusif, data kategori memiliki urutan logis yang pasti, kategori data diskalakan menurut jumlah karakteristik tertentu yang mereka miliki, perbedaan yang sama dalam karakteristik diwakili oleh perbedaan yang setara dalam angka yang ditetapkan pada kategori, dan titik nol (0) hanyalah titik lain pada skala.

Skala rasio mirip dengan skala interval dalam hal ekivalensi nilai antar angka. Perbedaannya adalah pada skala rasio, nol (0) berarti sesuatu yang penting. Artinya tidak ada skala apaun yang diukur. Jenis skala ini sering ditemukan pada ilmu sains dimana sebenarnya hanya terdapat ketiadaan, seperti berat, waktu, ketinggian, kalori, atau ketika nilai nol tersebut nyata bukan dibuat-buat. Contohnya ketika mengukur level nyeri atau sakit pada pasien pasca operasi (yang jelas tidak mungkin dilakukan), yang bisa digunakan dalam hal ini adalah skala rasio. Sifat-sifat skala rasio yaitu kategori data saling eksklusif, data kategori memiliki urutan logis yang pasti, kategori data diskalakan menurut jumlah karakteristik tertentu yang mereka miliki, perbedaan yang sama dalam karakteristik diwakili oleh perbedaan yang setara dalam angka yang ditetapkan pada kategori, dan titik nol (0) mencerminkan tidak adanya karakteristik tersebut.

1.7 Teknik Sampling

Sering kita temukan ketidakmampuan mengumpulkan informasi statistik dari seluruh titik data dalam suatu populasi. Statistika mengandalkan berbagai teknik pengambilan sampel yang berbeda dengan tujuan untuk menciptakan subkumpulan, dalam hal ini kita kenal dengan istilah sampel, yang berbeda untuk menciptakan sampel dari populasi yang representatif dan lebih mudah

dianalisis. Terdapat beberapa jenis pengambilan sampel utama dalam statistika, sebagai berikut

1.7.1 *Simple Random Sampling* (Sampel acak sederhana)

Pengambilan sampel acak sederhana mengharuskan setiap anggota populasi mempunyai peluang yang sama untuk dapat terpilih. Seluruh populasi digunakan sebagai dasar pengambilan sampel dan secara acak bisa memilih item sampel. Sebagai contoh, dari 100 orang dipilih 10 orang secara acak.

1.7.2 *Systemic Sampling* (Sampel sistematis)

Pengambilan sampel sistematis juga memerlukan sampel acak, namun, tekniknya sedikit dimodifikasi agar lebih mudah dilakukan. Sebuah nomor acak tunggal dihasilkan, dan individu kemudian akan dipilih pada interval reguler tertentu sampai urutan sampel selesai. Misalkan 100 orang berbaris dan diberi nomor. Individu ke-7 dipilih sebagai sampel, diikuti oleh setiap individu ke-9 berikutnya hingga 10 item sampel telah dipilih.

1.7.3 *Stratified Sampling* (Sampel bertingkat)

Pengambilan sampel bertingkat memerlukan kontrol yang lebih besar terhadap sampel. Populasi dibagi menjadi beberapa subkelompok berdasarkan karakteristik yang serupa. Kemudian dihitung berpa banyak orang dari setiap subkelompok yang akan mewakili seluruh populasi. Sebagai contoh dari 100 orang dikelompokkan berdasarkan jenis kelamin dan ras. Kemudian sampel dari masing-masing subkelompok diambil secara proporsional dengan keterwakilan subkelompok tersebut dalam populasi.

1.7.4 Cluster Sampling

Pengambilan sample klaster juga memerlukan subkelompok, namun setiap subkelompok harus mewakili

populasi. Seluruh subkelompok dipilih secara acak, bukan memilih individu dalam subkelompok secara acak.

DAFTAR PUSTAKA

Agresti, A., Franklin, C.A. and Klingenberg, B. (2023) *Statistics: The Art and Science of Learning from Data,5e, GE*. Harlow: Pearson Education Limited.

Blaine, B. (2023) *Introductory Applied Statistics*. Cham: Springer.

Blatchley, B. (2019) *Statistics in Context*, *The Mathematics Teacher*. New York: Oxford University Press Inc. doi:10.5951/mt.93.1.0054.

Campbell, M.J. and Jacques, R.M. (2023) *Statistics at Square Two*. Third. New Jersey: Wiley Blackwell.

Carrol, S.R. and Carrol, D.J. (2023) *Simplifying Statistics for Graduate Students*. Lanham: Rowman & Littlefield.

Ewens, W.J. and Brumberg, K. (2023) *Introductory Statistics for Data Analysis*. Cham: Springer.

Gould, R., Wong, R. and Ryan, C. (2020) *Introductory Statistics: Exploring the World Through Data*. Third, *Pearson*. Third. New Jersey: Pearson. doi:10.1080/00031305.2012.726904.

Hinton, P.R. (2004) *Statistics Explain*. Second. London: Routledge Taylor & Francis Group.

Kahl, A. (2023) *Introductory Statistics*. Pennsylvania: Bentham Books.

Kronthaler, F. (2023) *Statistics Applied With Excel*, *Statistics Applied With Excel*. Berlin: Springer-Verlag GmbH Germany, part of Springer Nature. doi:10.1007/978-3-662-64319-8.

Lauritzen, S. (2023) *Fundamentals of Mathematical Statistics*. Florida: Chapman and Hall/CRC.

Mathworks (2023) *Statistics and Machine Learning Toolbox ™ User ' s Guide*. Natick: MathWorks, Inc.

Rumsey, D. (2023) *Statistics All-in-One for Dummies*. New Jersey: John Wiley & Sons Inc.

Salcedo, J. and McCormick, K. (2023) *SPPS Statistics Workbook for Dummies*. New Jersey: John Wiley & SOns, Inc.

Salvati, N. *et al.* (2021) 'Studies in Theoretical and Applied Statistics', in *Springer Proceedings in Mathematics & Statistics Volume 406*. Pisa: Springer, p. 547.

Series, E. (2022) *Statistics for Empowerment and Social Engagement*. Edited by J. Ridgway. Cham: Springer.

Shayib, M.A. (2013) *Applied Statistics*. The Ebook COmpany: bookboon.com.

Watt, R. and Collins, E. (2023) *Statistics for Psychology*. Second. London: SAGE Publications, Inc.

Zhu, M. (2023) *Essential Statistics for Data Science*. Oxford: Oxford University Press.

BAB 2
MENGANALISIS UKURAN PEMUSATAN DATA

Oleh Hetty Elfina

2.1 Pendahuluan

Pada suatu hasil penelitian akan berkaitan dengan sekelompok data, yang berarti satu orang memiliki sekumpulan data atau sekelompok orang memiliki satu jenis data. Contohnya sekelompok mahasiswa yang memiliki nilai mata kuliah statistik atau sekelompok mahasiswa yang memiliki nilai mata kuliah statistik, kalkulus, dan aljabar linier.

Saat penelitian dilakukan, peneliti mendapatkan sekelompok data variabel dari sekelompok responden yang diteliti. Misalkan pada saat penelitian mengenai prestasi kerja karyawan di Perusahaan X. Prinsip dasar untuk kelompok yang akan dilakukan penelitian yaitu penjelasan atau pembahasan yang disajikan harus benar-benar mewakili seluruh karyawan di perusahaan X tersebut.

Terdapat beberapa cara untuk menjelaskan hasil akhir yang telah selesai diobservasi melalui data kuantitatif, selain penggunaan tabel dan gambar, dapat juga diberi penjelasan melalui teknik statistik seperti Mean, Median, dan Modus.

Mean, Median, dan Modus merupakan suatu teknik statistik dengan memberikan penjelasan kelompok berdasarkan pada gejala pusat (*tendency central*) dari kelompok tersebut. Tetapi dari ketiga macam teknik ini memiliki ukuran gejala pusat yang berbeda (Sugiono, 2011).

2.2 Ukuran Pemusatan Data Tunggal

a. Mean

Mean yaitu suatu cara memberikan penjelasan pada kelompok berdasarkan pada nilai rata-rata kelompok tersebut. Nilai rata-rata ini diperoleh dari menjumlahkan data setiap individu lalu dibagi dengan banyak individu yang ada di kelompok tersebut. Penjelasan diatas dapat dirumuskan seperti di bawah ini :

$$\bar{x} = \frac{\sum x}{n} \text{ atau } \bar{x} = \frac{\sum (f_i \,.\, x_i)}{\sum f_i}$$

Keterangan :

$\bar{x}$ = mean

$\sum x$ = jumlah data

n = banyak data

f_i = frekuensi untuk nilai x_i

x_i = nilai data

Contoh :

Duabelas karyawan di PT Maju memiliki penghasilan (dalam satuan ribu rupiah) adalah sebagai berikut

600, 400, 300, 500, 600, 750, 500, 550, 600, 350, 400, 700

Berdasarkan data di atas, maka rata-rata penghasilan dapat dihitung yaitu

$$\bar{x} = \frac{600 + 400 + 300 + 500 + 600 + 750 + 500 + 550 + 600 + 350 + 400 + 700}{12}$$

$$\bar{x} = \frac{6250}{12}$$

$$\bar{x} = 520{,}8$$

Jadi penghasilan rata-rata karyawan PT Maju adalah Rp 520.800,00

Sesuai pernyataan di atas bahwa untuk memberikan penjelasan mengenai keadaan kelompok berarti setiap pernyataan kualitatif atau kuantitatif yang diberikan terhadap kelompok tersebut menjadi perwakilan individu pada kelompok. Hal ini memiliki arti bahwa setiap pernyataan yang diberikan untuk kelompok tersebut tidak terjadi perbedaan yang signifikan untuk setiap individu.

Contoh :
Sepuluh penduduk di desa X memiliki penghasilan per bulan (dalam satuan ribu rupiah) sebagai berikut :

150, 250, 300, 600, 700, 800, 850, 900, 1200, 2000,

Penghasilan rata-rata dari sepuluh penduduk tersebut adalah :

$$\bar{x} = \frac{150 + 250 + 300 + 600 + 700 + 800 + 850 + 900 + 1200 + 2000}{10}$$

$$\bar{x} = \frac{7750}{10}$$

$$\bar{x} = 775$$

Rata-rata penghasilan penduduk desa X adalah Rp 775.000,00. Dari penghasilan rata-rata tersebut dapat dilihat bahwa rata-rata penghasilan kurang mewakilkan individu yang memiliki penghasilan Rp 300.000,00 ke bawah dan Rp 1.200.000,00 ke atas. Berdasarkan data ini diperoleh informasi bahwa terdapat jarak penghasilan yang signifikan. Pada kasus ini ada baiknya tidak menggunakan *mean* dalam menjelaskan kondisi penghasilan penduduk, namun digunakan *median* yang akan dipelajari di bagian selanjutnya (Sugiono, 2011).

b. Median

Median merupakan suatu teknik penjelasan kelompok berdasarkan pada nilai tengah kelompok data yang telah diurutkan dari yang terkecil hingga terbesar, atau dari yang terbesar hingga terkecil (Sugiono, 2011). Median membagi nilai hasil observasi sehingga 50% berada di bawah median dan 50% berada di atas median. Median dapat digunakan ketika skala

pengukuran data serendah-rendahnya ordinal, sehingga nilai-nilai hasil observasi dapat dibuat dalam bentuk ranking untuk memperoleh nilai yang berada di tengah (Somantri, 2006). Misalnya diketahui umur karyawan perusahaan *X*.

Tabel 2.1. Umur Karyawan Perusahaan *X*

Umur Karyawan (tahun)	Jumlah
23	2
54	3
35	1
42	2
60	2
27	4
48	1

Untuk mencari mediannya, data ini terlebih dahulu diurutkan mulai yang terkecil hingga terbesar

23, 23, 27, 27, 27, 27, 35, 42, 42, 48, 54, 54, 54, 60,60

Nilai tengah pada data di atas adalah urutan ke-8 yaitu 42. Banyak data yang diberikan di atas yaitu 15 yang merupakan bilangan ganjil. Jika diberikan contoh untuk data yang berjumlah genap, maka dapat dicari seperti contoh berikut.

Diketahui berat badan mahasiswa semester III:

55, 50, 57, 53, 60, 53, 60, 54

Untuk mencari median, data tersebut terlebih dahulu diurutkan dari yang terkecil hingga terbesar sehingga urutan data akan berubah menjadi

50, 53, 53, 54, 55, 57, 60, 60

Dari data di atas, mediannya berada di antara data ke 4 dan ke 5. Maka mediannya menjadi $Me = \frac{54+55}{2} = 54{,}5$. Dengan demikian, rata-rata berat badan mahasiswa semester III adalah 54,5kg.

Selain menggunakan cara manual seperti di atas, median juga dapat dicari dengan menggunakan rumus

- ✓ Untuk *n* ganjil : $Me = x_{\frac{n+1}{2}}$
- ✓ Untuk *n* genap : $Me = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$

c. Modus

Modus merupakan suatu teknik dalam menjelaskan keadaan kelompok yang didasarkan pada nilai yang popular atau nilai yang paling banyak muncul (Sugiono, 2011). Modus adalah nilai yang memiliki frekuensi terbesar pada suatu perkumpulan data. Modus digunakan untuk mengetahui tingkat banyaknya yang terjadi suatu peristiwa. Jika terdapat nilai dengan frekuensi tertinggi ada dua disebut *bimodal.* Jika ada tiga disebut *trimodal,* dan jika ada banyak disebut *multimodal*. Modus dapat digunakan pada semua skala pengukuran data mulai dari nominal hingga rasio.

Contoh data kualitatif :

- Kebanyakan masyarakat Indonesia menyukai kopi
- Pada umumnya Pegawai Negeri Sipil tidak tepat waktu masuk kerja
- Pada tahun 1970-an, para siswa masih banyak menggunakan sepeda untuk ke sekolah.

Contoh data kuantitatif :

Berdasarkan contoh umur karyawan perusahaan *X* pada tabel 1 dapat dilihat bahwa yang paling banyak muncul adalah umur 27 tahun. Umur ini muncul sebanyak 4 kali atau frekuensinya 4, yang berarti ada 4 orang yang berumur 27 tahun.

Dari ketiga teknik penjelasan di atas, masing-masing teknik memiliki keunggulan masing-masing. Modus digunakan saat peneliti ingin menjelaskan keadaan kelompok. Namun karena hanya memiliki data yang paling banyak mucul pada kelompok, maka teknik ini dianggap kurang teliti. Median digunakan saat terdapat data dengan rentang perbedaan yang signifikan pada suatu

kelompok. Mean digunakan jika pada kelompok terdapat kenaikan data yang merata. Bila terdapat keraguan pada peneliti ketika menggunakan berbagai teknik penjelasan kelompok ini, maka ada baiknya ketiga teknik digunakan bersama (Sugiono, 2011).

2.3 Ukuran Pemusatan Data Kelompok

Pada data tunggal, angka yang akan dianalisis ditulis satu per satu. Namun berbeda halnya dengan data kelompok. Data pada kelompok telah di susun dalam tabel distribusi frekuensi sehingga peneliti hanya mengetahui rentang data beserta frekuensi atau jumlahnya.

Contohnya yaitu pada suatu penelitian yang diadakan di sekolah *Y* diperoleh hasil tes kemampuan berhitung matematika siswa. Hasil tersebut telah disusun dalam tabel berikut :

Tabel 2.2 Hasil Tes Kemampuan Berhitung Matematika

Nilai	Frekuensi
31 – 40	6
41 – 50	15
51 – 60	12
61 – 70	10
71 – 80	8
81 – 90	9

Berdasarkan data di atas, carilah hasil mean, median, dan modusnya!

a. Mean

Untuk menghitung mean (rata-rata) berdasarkan data di atas, digunakan rumus sebagai berikut :

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

$\bar{x}$ $= rata - rata\ (mean)$

$\sum f_i$ $= jumlah\ data\ atau\ frekuensi$

$\sum f_i x_i$ $= jumlah\ antara\ perkalian\ frekuensi\ dan\ nilai\ tengah\ kelas$

Tabel 2. 3 Perhitungan Jumlah Data

Nilai	Frekuensi	x_i	$f_i x_i$
31 – 40	6	$\frac{31+40}{2} = \frac{71}{2} = 35{,}5$	213
41 – 50	15	45,5	682,5
51 – 60	12	55,5	666
61 – 70	10	65,5	655
71 – 80	8	75,5	604
81 – 90	9	85,5	769,5
	$\sum f_i = 60$		$\sum f_i x_i = 3590$

Mean = $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{3590}{60} = 59{,}83$

Jadi rata-rata nilai kemampuan berhitung matematika siswa di sekolah *Y* adalah 59,83.

b. Median

Median (nilai tengah) data dapat dihitung dengan menggunakan rumus :

$$Me = b + p\left(\frac{\frac{1}{2}n - F}{f}\right)$$

$Me = median$

$b \quad = batas\ bawah\ kelas\ median$

$p \quad = panjang\ kelas/interval$

$n \quad = banyak\ data$

$F \quad = frekuensi\ kumulatif\ sebelum\ kelas\ median$
$f \quad = frekuensi\ kelas\ median$

Sebelum menentukan nilai median, terlebih dahulu dihitung letak kelas median yaitu $\frac{1}{2}n = \frac{1}{2}(60) = 30$. Maka letak kelas median ada di data ke 60 yaitu pada interval 51 – 60. Pada interval ini memiliki batas bawah (b) yaitu 51 – 0,5 = 50,5. Panjang kelas/interval (p) yaitu 10 dan frekuensi pada interval 51 – 60 adalah 12 (panjang interval dapat dilihat pada tabel). Nilai F nya yaitu 6 + 15 = 21. Sehingga nilai median menjadi

$$Me = b + p\left(\frac{\frac{1}{2}n - F}{f}\right)$$
$$Me = 50{,}5 + 10\left(\frac{30 - 21}{12}\right)$$
$$Me = 50{,}5 + 10\left(\frac{9}{12}\right)$$
$$Me = 50{,}5 + 7{,}5$$
$$Me = 58$$

c. Modus

Untuk mencari nilai modus dari data yang telah dibuat dalam tabel distribusi frekuensi, maka digunakan rumus :

$$Mo = b + p\left(\frac{d_1}{d_1 + d_2}\right)$$

$Mo = modus$
$b \quad = batas\ bawah\ kelas\ modus$
$p \quad = panjang\ kelas$

d_1 = *selisih frekuensi kelas modus dengan kelas sebelumnya*

d_2 = *selisih frekuensi kelas modus dengan kelas setelahnya*

Sebelum menghitung nilai modus, maka diperhatikan terlebih dahulu frekuensi mana yang paling tinggi. Dari tabel di atas, frekuensi tertinggi yaitu 15 di interval 41 – 50. Batas bawah kelas modus (b) yaitu 41 – 0,5 = 40,5. Panjang kelas (p) yaitu 10, selisih frekuensi kelas modus dengan kelas sebelumnya (d_1) yaitu 15 – 6 = 9 dan selisih frekuensi kelas modus dengan kelas setelahnya (d_2) yaitu 15 – 12 = 3. Sehingga hasil perhitungan modus menjadi

$$Mo = b + p\left(\frac{d_1}{d_1 + d_2}\right)$$

$$Mo = 40{,}5 + 10\left(\frac{9}{9+3}\right)$$

$$Mo = 40{,}5 + 10\left(\frac{9}{12}\right)$$

$$Mo = 40{,}5 + 7{,}5$$

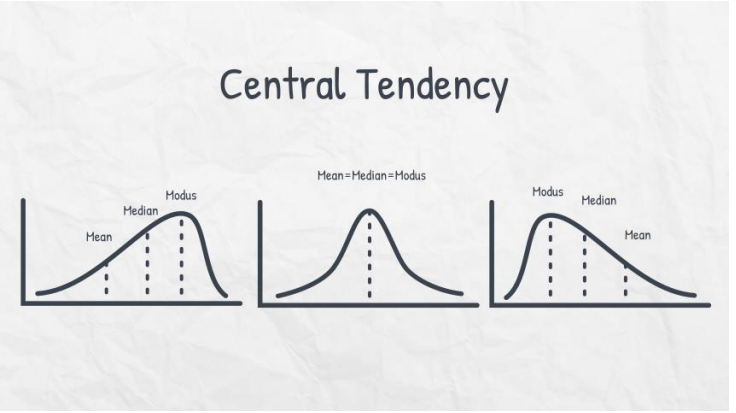
$$Mo = 48$$

2.4 Hubungan Mean, Median, dan Modus

Hubungan antara mean, median, dan modus dari suatu distribusi frekuensi yaitu :

- Jika nilai mean, nilai median, dan nilai modus sama besar ($\bar{x} = Me = Mo$) artinya nilai mean, median, dan modus terletak di satu titik dari kurva distribusi frekuensi, dan kurva berbentuk simetris (*symmetrical curve*)
- Jika nilai mean lebih besar dari nilai median dan nilai modus ($\bar{x} > Me > Mo$) artinya nilai mean terletak di sebelah kanan kurva distribusi frekuensi, selanjutnya median di tengah dan modus di kiri. Kurva berbentuk tidak simetris dan menceng ke sebelah kanan (*skewed right*)

- Jika nilai mean lebih kecil dari nilai median dan nilai modus ($\bar{x} < Me < Mo$) artinya nilai mean terletak di sebelah kiri kurva distribusi frekuensi, selanjutnya median di tengah dan modus di kanan. Kurva berbentuk tidak simetris dan menceng ke sebelah kiri (*skewed left*) (Somantri, 2006).



Gambar 2.1 Hubungan Mean, Median, dan Modus

(Sumber : https://www.rachmatwahid.com/2020/02/mengenal-central-tendency-mean-median.html)

DAFTAR PUSTAKA

Somantri, Muhidin. 2006. *Aplikasi Statistika Dalam Penelitian*. Bandung : Pustaka Setia

Sugiono. 2011. *Statistika untuk Penelitian.* Bandung : Alfabeta

BAB 3
UKURAN PENYEBARAN DATA

Oleh Rahmi Wahyuni

3.1 Pendahuluan

Sejarah ukuran penyebaran data berkaitan erat dengan perkembangan ilmu statistik dan analisis data. Pada Abad ke-19 Sebelum Masehi, Konsep dasar tentang ukuran penyebaran data sudah ada sejak zaman kuno. Bahkan dalam karya Aristoteles, ia membahas tentang variabilitas dalam pengukuran dan perbedaan antara berbagai pengukuk. Awal Abad ke-20 konsep-konsep statistik modern mulai berkembang. Tokoh seperti Karl Pearson dan Francis Galton berkontribusi pada statistik dan menganalisis variasi data. Mereka memperkenalkan konsep-konsep awal tentang deviasi dari nilai rata-rata, yang pada akhirnya berkembang menjadi konsep simpangan baku.

Seiring dengan perkembangan komputasi dan statistik matematis, ukuran penyebaran data mulai didefinisikan dengan lebih rinci. Pada tahun 1940-an dan 1950-an, penggunaan teknik-teknik matematika seperti analisis varians dan regresi semakin berkembang, dan ukuran-ukuran seperti varians dan deviasi baku menjadi lebih umum digunakan. Dalam beberapa dekade terakhir, penggunaan komputer dan perangkat lunak statistik telah mempermudah perhitungan dan analisis ukuran penyebaran data. Sekarang, statistisi dan analis data dapat dengan mudah menghitung ukuran-ukuran ini untuk set data yang besar dan kompleks.

Selain ukuran-ukuran tradisional seperti simpangan baku dan varians, ada juga pengembangan ukuran alternatif yang mungkin lebih sesuai dalam beberapa konteks. Sebagai contoh, simpangan rata-rata mutlak digunakan sebagai alternatif yang lebih tahan terhadap outlier dalam beberapa kasus.

Ukuran penyebaran data digunakan dalam berbagai disiplin ilmu, termasuk statistik, ilmu sosial, ilmu alam, ekonomi, dan bisnis. Mereka digunakan untuk menganalisis data, membuat prediksi, mengidentifikasi anomali, dan membuat keputusan. Seiring perkembangan ilmu statistik dan teknologi, penggunaan dan pemahaman tentang ukuran penyebaran data terus berkembang. Hari ini, ini adalah konsep dasar dalam analisis data modern dan menjadi bagian integral dari banyak disiplin ilmu yang bergantung pada pengumpulan dan interpretasi data.

Pada bab ini akan diuraikan mengenai ukuran penyebaran. Selain dapat mengetahui lebih lanjut tentang bentuk suatu distribusi bilangan, dapat juga mengetahui skor mentah, skor terolah, dan beberapa skor baku seperti stanin, skor huruf, skor-z, dan skor-T serta dapat mengubah skor mentah menjadi skor terolah dan skor baku (Ruseffendi, 1993).

Berdasarkan besar kecilnya penyebaran kelompok data dibagi menjadi 2 yaitu:

- Kelompok data homogen
  Penyebaran relatif kecil jika seluruh data sama, maka disebut kelompok data homogen 100%.
- Kelompok data heterogen
  Penyebaran data relatif besar.

Tujuan penyajian materi dalam bab ini sebagai berikut:

1. Mengetahui ukuran penyebaran antara lain jangkauan atau range, deviasi rata-rata, variansi, deviasi baku, sebaran atau jangkauan antar kuartil, sebaran semi antar kuartil, dan sebaran keempat serta artinya masing-masing.
2. Dapat mengetahui kelebihan ukuran penyebaran tertentu terhadap ukuran penyebaran lainnya.
3. Dapat menghitung ukuran-ukuran penyebaran sekelompok bilangan baik data tersusun maupun data yang tidak tersusun.

4. Dapat menghitung deviasi baku dan variansi sekumpulan bilangan dengan menggunakan program kalkulator.
5. Dapat memahami keunggulan peringkat persen dari peringkat, serta
6. Dapat memahami keunggulan skor-T dan skor-Z dan sebaliknya.

Rata-rata, median, dan kebanyakan ukuran terdensi sentral lainnya merupakan sebuah bilangan yang dapat menggambarkan keadaan sekumpulan bilangan. Untuk mengetahui lebih lanjut sekumpulan bilangan itu kita memerlukan ukuran lain, antara lain ialah ukuran penyebaran data.

Dispersi atau ukuran penyebaran data adalah suatu ukuran yang menyatakan seberapa besar penyimpangan nilai-nilai data dari nilai-nilai pusatnya atau ukuran yang menyatakan seberapa banyak nilai-nilai data yang bervariasi atau berbeda dengan nilai pusatnya. Dalam membandingkan dua atau lebih kumpulan data, penggunaan ukuran pusat data saja tidak akan memberikan hasil yang baik, mungkin juga dapat memberikan hasil yang menyesatkan (Kustituanto & Badrudin, 1994). Ukuran penyebaran pada dasarnya adalah pelengkap dari ukuran nilai pusat dalam menggambarkan sekumpulan data. Dengan adanya ukuran penyebaran maka penggambaran sekumpulan data akan menjadi lebih jelas. Ukuran penyebaran membantu mengetahui sejauh mana suatu nilai menyebar dari nilai tengahnya, apakah semakin kecil atau semakin besar. Untuk jelasnya, perhatikan contoh berikut:

Nilai X adalah hasil dari tes masuk ke Perguruan Tinggi Negeri sebanyak 10 orang, dan Y adalah hasil tes masuk ke Perguruan Tinggi Swasta juga sebanyak 10 orang.

X: 72 72 69 69 68 67 67 67 65 65

Y: 97 89 86 74 61 60 60 53 53 47

Kedua kelompok bilangan itu rata-ratanya sama yaitu 68. Tetapi jika kita perhatikan selanjutnya, nilai-nilai di X ada disekitar

rata-rata yaitu 68. Sedangkan beberapa nilai di kelompok Y jauh bedanya dengan 68. Jadi dapat disimpulkan untuk mengetahui keadaan sekumpulan bilangan, kita tidak cukup hanya dengan mengetahui ukuran tendensi sentralnya saja. Namun, ukuran penyebarannya pun perlu diketahui.

3.2 Jenis-Jenis Ukuran Penyebaran Data

Beberapa jenis ukuran penyebaran adalah sebagai berikut:

3.2.1 Range, Inter-kuartir dan Deviasi Kuartil

Range

Jangkauan sering kita sebut Range atau Rentang, jangkauan disimbolkan jangkauan dengan huruf R. jangkauan berbagai 2 macam; data jangkauan tunggal dan data jangkauan kelompok. Salah satu cara yang paling sederhana untuk mengukur penyebaran tentang sekumpulan bilangan ialah range. Range adalah selisih atau beda antara pengukuran bilangan terbesar dengan nilai terkecil. Ini memberikan gambaran kasar tentang sebaran data. Kelebihan Range ini bisa digunakan untuk mengukur gambaran mengenai penyebaran data dengan waktu yang cukup singkat. Namun menurut Alvin dalam (Marhawati et al., 2022) range juga memiliki kekurangan range tidak tepat untuk menentukan suatu perhitungan apalagi untuk data yang kompleks.

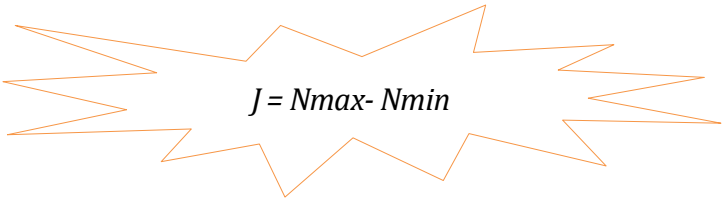
Contoh sederhana dari penggunaan "range" (rentang) sebagai ukuran penyebaran data adalah sebagai berikut:

1. Range untuk data tunggal

Misalkan Anda memiliki data berikut ini yang mencatat jumlah poin yang dicetak oleh tim basket Kalandra dalam 10 pertandingan berturut-turut:

[10, 15, 18, 20, 25, 28, 30, 35, 40, 45]

Untuk menghitung rentang (range) dari data ini, Anda cukup mengurangkan nilai terkecil dari nilai terbesar dalam set data tersebut:



Keterangan:

J : Daerah Jangkauan

N : Nilai terbesar dari serangkaian data

N : Nilai terbesar dari serangkaian data

Dalam kasus ini: [Rentang = 45 - 10 = 35]

Jadi, dalam kasus di atas adalah berarti poin yang dicetak oleh tim basket Kalandra adalah 35. Ini berarti data poin yang di cetak tersebut mencakup poin dari terbanyak 45 hingga poin terendah 10. Rentang adalah salah satu ukuran penyebaran yang paling sederhana dan memberikan gambaran kasar tentang seberapa tersebar atau bervariasinya data dalam set tersebut. Semakin besar rentangnya, semakin tersebar data tersebut.

2. Range untuk data Kelompok

Berikut ini data mengenai nilai 30 peserta ujian mata pelajaran Fisika di SMA Kalandra

Tabel 3.1

TABEL	FREKUENSI
40-49	4
50-59	6
60-69	2
70-79	10
80-89	4
90-99	4

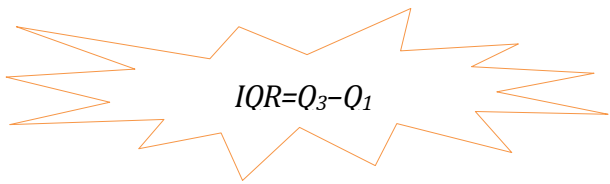
Range nilai 30 peserta ujian mata pelajaran Fisika dengan menggunakan Rumus:

Rentang = Nilai Terbesar – Nilai Terkecil

Rentang = 99-40=59

Inter-Kuartil

Inter-kuartil, atau interquartile range (IQR), adalah ukuran penyebaran statistik yang mendeskripsikan rentang di mana sebagian besar data berada. IQR dihitung sebagai perbedaan antara kuartil ketiga (Q3) dan kuartil pertama (Q1). Ini membantu kita dalam mengidentifikasi seberapa tersebar data kita dan dapat memberikan informasi tentang variabilitas data serta mengidentifikasi outlie. IQR dihitung dengan cara mengurangkan kuartil pertama (Q1) dari kuartil ketiga (Q3):



Keterangan:
Kuartil pertama (Q1): nilai median dari separuh data pertama
Kuartil kedua (Q2): nilai median dari seluruh data
Kuartil ketiga (Q3): nilai median dari separuh data terakhir

Deviasi Kuartil

Deviasi kuartil, sering juga disebut sebagai "semi-interquartile range", adalah ukuran penyebaran yang menggambarkan seberapa jauh data tersebar di sekitar median. Deviasi kuartil dihitung dengan mengambil setengah dari rentang interkuartil (IQR).

Rumus untuk deviasi kuartil adalah:

$$Deviasi\ Kuartil = \frac{IQR}{2}$$

Atau

$$Deviasi\ Kuartil\ \frac{Q_3 - Q_1}{2}$$

Keterangan:
Q_3 adalah kuartil ketiga.
Q_1 adalah kuartil pertama.

Deviasi kuartil memberikan gambaran tentang seberapa jauh data tersebar di sekitar median dan sering digunakan ketika distribusi tidak simetris. Seperti IQR, deviasi kuartil juga memiliki keuntungan karena tidak mudah dipengaruhi oleh outlier atau nilai-nilai ekstrem dalam dataset.

Berikut adalah contoh untuk menghitung interkuartil (IQR) dan deviasi kuartil:

Nilai = {1, 2, 3, 4, 5, 6, 7, 8, 9}

1) Menentukan kuartil pertama (Q1) dan kuartil ketiga (Q3).
   - Q1 adalah median dari separuh data pertama. Dalam hal ini, data separuh pertama adalah {1, 2, 3, 4} dan median-nya adalah (2+3)/2 = 2,5
   - Q3 adalah median dari separuh data kedua. Dalam hal ini, data separuh kedua adalah {6, 7, 8, 9} dan median-nya adalah (7+8)/2 = 7,5
2) Menghitung IQR dengan menggunakan rumus:
   IQR = Q3 - Q1
   IQR= 7.5 - 2.5 = 5
3) Menghitung deviasi kuartil dengan menggunakan rumus:

$$Deviasi\ Kuartil = \frac{IQR}{2}$$

$$Deviasi\ Kuartil = \frac{5}{2} = 2{,}5$$

Dalam contoh ini, IQR dan deviasi kuartil memberikan ukuran tentang seberapa jauh data kita tersebar di sekitar median nilai tersebut.

3.2.2 Variansi

Variansi (Variance) adalah salah satu ukuran statistik yang digunakan untuk mengukur seberapa jauh titik data individual dalam satu set data tersebar dari nilai tengah atau rata-rata dari set data tersebut. Varian merupakan metode untuk mengukur sebaran data dengan menghitung rata-rata kuadrat perbedaan antara setiap titik data dan nilai rata-rata. Terdapat dua jenis variansi yang didasarkan atas perolehan sumber data penelitian. Jenis pertama yaitu varian untuk data populasi disimbolkan σ^2 dan untuk data sampel disimbolkan dengan s^2.

1. Variansi Data tunggal atau nilai f yang sama berdasarkan populasi

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

   Variansi Data tunggal atau nilai f yang sama berdasarkan sampel

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

2. Variansi Data Berkelompok atau nilai f yang sama berdasarkan populasi

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

Variansi Data tidak Berkelompok atau nilai f yang sama berdasarkan populasi

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

3.2.3 Standar Deviasi

Standar Deviasi atau simpangan baku adalah ukuran statistik yang digunakan untuk mengukur sebaran atau dispersi data dalam suatu himpunan (Santosa, 2018). Ini memberikan informasi tentang sejauh mana data tersebar dari nilai rata-rata atau mean. Standar deviasi dihitung sebagai akar kuadrat dari varian data. Ini adalah ukuran penyebaran yang lebih intuitif dan umum digunakan daripada variansi. Semakin besar simpangan baku, maka semakin besar variasi dalam data.

1. Standar Deviasi Data tunggal

   Rumus standar Deviasi Data tunggal berdasarkan populasi :

$$\sigma_x = \sqrt{\frac{\sum(X_{1-}\mu_x)^2}{N}}$$

Keterangan :

σ_x : Standar Deviasi populasi
x_1 : Data ke-i dari variabel acak X
μ_x : Rata-rata populasi
N : Ukuran Populasi

Rumus Standar Deviasi Data tunggal berdasarkan sampel :

$$s = \sqrt{\frac{\sum(X_{1-}\mu_x)^2}{N-1}}$$

Keterangan :

s : Standar Deviasi sampel
x_1 : Data ke-i dari variabel acak X
μ_x : Rata-rata populasi
N : Ukuran Populasi

2. Standar Deviasi Data Berkelompok

Rumus Standar Deviasi Data Berkelompok berdasarkan populasi :

$$\sigma =. \sqrt{\left[\frac{\sum U_1^2 . f_1 - \left(\frac{\sum U_1 . f_1}{N}\right)^2}{N}\right]} i$$

Keterangan :

σ : Simpangan baku populasi
i : Interval kelas
U_i : Kode U pada kelas ke-i
f_1 : Frekuensi kelas ke-i
N : Ukuran populasi

Rumus Standar Deviasi Data Berkelompok berdasarkan sampel :

$$x = i. \sqrt{\left[\frac{\sum U_1^2 . f_1 - \left(\frac{\sum U_i . f_1}{N}\right)^2}{n-1}\right]}$$

Keterangan :
s : Simpangan baku populasi
i : *I*nterval kelas
U_i : Kode U pada kelas ke-i
f_1 : Frekuensi kelas ke-i
n : Ukuran sampel

Koefisien Variasi

Koefisien varian (CV) adalah ukuran relatif dari variasi atau penyebaran data (Ruseffendi, 1993). Ini adalah rasio antara deviasi standar (σ) dan rata-rata (mean, μ) dari kumpulan data.

Rumus Koefisien Variasi untuk populasi:

$$CV = \left(\frac{\sigma}{\mu}\right) X\ 100\%$$

Keterangan:

CV : koefisien varian

σ : deviasi standar populasi

μ : rata-rata populasi

Rumus Koefisien Variasi untuk sampel:

$$CV = \frac{s}{x} x\ 100\%$$

Keterangan:

S: Simpangan baku atau deviasi standar

X: Rata-rata sampel

Untuk mencari koefisien variansi, pertama-tama kita perlu menemukan rata-rata dan deviasi standar dari data. Berikut adalah langkah-langkahnya:

Mari kita gunakan contoh berikut:

Data: (10, 20, 30, 40, 50)

Langkah 1: Temukan rata-rata:

$$x = \frac{10 + 20 + 30 + 40 + 50}{5} = \frac{150}{5} = 30$$

Langkah 2: Temukan deviasi standar. Pertama, temukan varians:

$$Varians = \frac{1}{5} = ((10 - 30)^2 + (20 - 30)^2 + (30-$$

$$30)^2+(40-30)^2+(50-30)^2)$$

$$=\frac{1}{5}(400+100+0+100+400)$$

$$= \frac{1}{5}(1000)$$

$$=200$$

Kemudian temukan deviasi standar:

$$s = \sqrt{200}$$

$$s = 14{,}14$$

Langkah 3: Temukan koefisien variansi:

$$CV= (\frac{14{,}14}{30})x100\%$$

$$=47{,}13\%$$

Jadi, koefisien variansi untuk data ini adalah sekitar (47.13%).

DAFTAR PUSTAKA

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

Marhawati, M., Mahmud, R., Nurdiana, S. P., Sri Astuty, S. E., STrKes, P., Fahradina, N., La One, S. T., Faelasofi, M. T. R., Widyasari, T., & Mawardati, R. (2022). Statistika Terapan. *Penerbit Tahta Media Group*.

Ruseffendi, E. T. (1993). Statistika Dasar untuk Penelitian Pendidikan. In *Jakarta: Departemen Pendidikan dan Kebudayaan*.

Santosa. (2018). *Statistika Hospitalitas: edisi Revisi*.

BAB 4
KAIDAH PENCACAHAN

Oleh Suryanti

4.1 Pengertian Kaidah Pencacahan

Kaidah pencacahan berguna untuk membantu kita memperkirakan jumlah kejadian atau rangkaian dengan cara yang lebih sistematis. Kaidah pencacahan dapat digunakan dalam banyak kasus mulai dari masalah kombinatorik, probabilitas hingga statistik.

Artinya kaidah matematika adalah salah satu cabang matematika yang membahas tentang cara menghitung banyaknya susunan atau kombinasi suatu sesuatu tanpa semua kemungkinan susunanya. Bayangkan, jika disuruh membuka kunci dengan 10 ribu kemungkinan tentu akan memakan waktu lama bukan? Nah, disinilah materi dari kaidah pencacahan membantu untuk mendapatkan peluang dalam waktu singkat.

Aturan pencacahan juga dapat dipahami sebagai aturan penghitungan yang menentukan berapa banyak peristiwa atau objek tertentu yang akan muncul. Disebut pencacahan karena dihasilkan dari hasil bilangan bulat. Ada tiga kaidah pencacahan, yaitu kaidah pengisian tempat yang tersedia, kaidah permutasi, dan kaidah kombinasi.

Contoh 1
Dalam perombaan renang 200 m terdapat 5 orang anak yang sampai di babak final yaitu P(Putri), Q(Queensa), R(Rani), S(Santi), dan T(Tina). Panitia menyediakan tiga hadiah bagi tiga peserta tercepat. Ada berapa susunan pemenang yang akan muncul?
Penyelesaian:

Pemenang pertama, kedua dan ketiga yang kemungkinan akan muncul, dapat disusun sebagai berikut:

PQR. PQS, PQT, PRS, PRT, PRQ, PST, PSQ, PSR, PTQ, PTR, PTS, QRS, QRT,QRP, QST, QSP, QSR, QTP, QTR,QTS, QPR,QPS, QPT, RST, RSP, RSQ, RTP, RTQ, RTS, RPQ,RPS, RPT, RQS, RQT, RQP, STP, STQ, STR, SPQ, SPR, SPT, SQR, SQT, SQP, SRT, SRP, SRQ, TPQ, TPR, TPS, TQR, TQS, TQP, TRS, TRP, TRQ, TSP, TSQ, TSR.

Langkah-langkah yang digunakan untuk menyusun pemenang sebagai berikut

Langkah 1:

Terdapat 5 peserta lomba yang semuaya dapat keluar sebagai peringkat pertama.

Langkah 2:

Apabila 1 orang telah sampai garis finis, maka masih akan tersisa 4 orang peserta lomba yang dapat menduduki peringkat ke dua.

Langkah 3:

Apabila terdapat 2 orang yang menduduki peringkat pertama dan kedua, maka akan tersisa 3 orang peserta yang dapat menduduki peringkat ketiga.

Jadi, banyaknya susunan pemenanga yang mungkin terjadi adalah $5 \times 4 \times 3 = 60$ cara.

Contoh 2

Anindya memiliki 5 buah baju, 3 buah celana dan 2 buah sandal. Ada berapa cara Anindya berpakaian lengkap?

Penyelesaian:

Anindya bisa memilih baju dengan 5 cara, sedangkan celana ada 3 cara memilih, disusul 2 cara memilih sandal.

Jadi, banyaknya cara agar Anindya dapat berpakaian dengan lengkap adalah $5 \times 3 \times 2 = 30$ cara.

Berdasarkan uraian diatas, dapat disimpulkan bahwa

n_1 : Banyaknya cara mengisi tempat pertama

n_2 : Banyaknya cara mengisi tempat yang kedua, setelah tempat pertama terisi

n_3 : Banyaknya cara mengisi tempat ketiga, setelah tempat pertama dan kedua telah terisi

:
:
n_k : Banyak cara mengisi tempat ke k, setelah tempat pertama, kedua, ketiga sampai tempat ke (k – 1) terisi.

Jadi, dapat disimpulkan bahwa cara mengisi tempat sampai dengan tempat ke k adalah $n_1 \times n_2 \times n_3 \times \ldots \times n_k$.

Cara untuk mengisi tempat yang tersedia seperti yang dijabarkan diatas disebut dengan aturan perkalian dalam pencacahan.

4.2 Faktorial

Hasil kali bilangan bulat positif yang dimulai dari bilangan bulat ke 1 sampai ke n disebut faktorial atau dalam matematika dilambangkan dengan !.

$1! = 1$
$2! = 2 . 1$
$3! = 3 . 2 . 1$
$4! = 4 . 3 . 2 . 1$
$5! = 5 . 4 . 3 . 2 . 1$

Berdasarkan uraian diatas dapat dituliskan rumus dari faktorial adalah

$$n! = n(n-1)(n-2) \ldots 3.2.1$$

Contoh:

1. 2! = 1.2 = 2.1 = 2
2. 5! = 1.2.3.4.5 = 5.4.3.2.1 = 120
3. 6! = 6.5.4.3.2.1 = 6.5!
4. $\frac{7!}{6!} = \frac{7.6!}{6!} = 7$
5. $\frac{8!}{6!} = \frac{8.7.6!}{6!} = 8.7 = 56$

4.3 Permutasi

Penyusunan n objek dalam urutan tertentu disebut permutasi. Permutasi juga dapat dipahami sebagai suatu proses pencatatan yang memperhatikan urutan atau susunan. Misalnya

kita mengetahui himpunan P = {a, b, c, d}. Jika anggota-anggota himpunan P tersusun berpasangan, maka terdapat himpunan yang terdiri dari 12 anggota, disebut {ab, ac, ad, ba, bc, bd, ca, cb, cd, da, db, dc}. Banyaknya anggota suatu himpunan juga dapat dihitung dengan menggunakan aturan permutasi, seperti:
Jika n objek berbeda disusun dalam r objek, banyaknya susunan dapat ditentukan dengan rumus:

$$P(n,r) = \frac{n!}{(n-r)!}$$

Posisi pertama pada n objek dapat dipilih dengan n cara berbeda, posisi kedua pada permutasi dapat dipilih dengan n – 1 cara, dan posisi ketiga pada permutasi dapat dipilih dengan n – 2 cara. Demikian pula kita mengetahui bahwa posisi ke-r (posisi terakhir) pada permutasi r suatu benda dapat dipilih dengan n – (r – 1) cara atau n – (n – 1) = n – r + 1 cara.

Teorema 1
P (n, r) = n (n – 1) (n – 2) … (n – r + 1)
Atau

$$P\,n,r = \frac{n!}{(n-r)!}, r \leq n$$

Membuktikan $\left(n(n-1)(n-2)\ldots(n-r+1)\right) = \frac{n!}{(n-r)!}$ adalah sebagai berikut

$$n(n-1)(n-2)\ldots(n-r+1)$$
$$= \frac{n(n-1)(n-2)\ldots(n-r+1).(n-r)!}{(n-r)!} = \frac{n!}{(n-r)!}$$

Contoh 1:
Suatu sekolah akan mengadakan pemilihan ketua OSIS, wakil ketua, bendahara dan sekertaris dari 12 peserta yang akan mengikuti pemilihan. Berapa banyak cara untuk menentukan pilihan?

Penyelesaian:
Dari 12 orang peserta akan dipilih 4 orang
Sehingga $n = 12, r = 4$
$P(12,4) = \frac{12!}{(12-4)!} = \frac{12.11.10.9.8!}{8!} = 12.11.10.9 = 11.800$ cara.

Contoh 2:
Misalkan ada empat orang anak akan membeli tiket nonton bioskop. Ada berapa urutan yang dapat terbentuk jikan penjaga tiket akan melayani mereka secara berurutan satu per satu?
Penyelesaian:
Misalkan ke empat anak tersebut terdiri dari K, L, M dan N.
Urutan yang dapat dibentuk yaitu
KLMN, KLNM, KMNL, KMLN, KNLM, KNML, LMNK, LMKN, LNKM, LNMK, LKMN, LKNM, MNKL, MNLK, MKLN, MKNL, MLNK, MLKN, NKLM, NKML, NLMK, NLKM, NMKL, NMLK.
Sehingga banyaknya urutan penjaga tiket untuk melayani ke empat anak tersebut adalah
$P(4,4) = \frac{4!}{(4-4)!} = \frac{4!}{0!} = 4! = 4.3.2.1 = 24$ urutan.

Dari contoh di atas dapat disimpulkan:

$$P(n,n) = \frac{n!}{(n-n)!} = \frac{n!}{0!} = \frac{n!}{1} = n!$$

Teorema Akibat
Ada n! permutasi dari n objek atau:

$$P\ (n,n)\ =\ n!$$

4.3.1 Permutasi dengan Beberapa Unsur yang Sama

Terkadang kita ingin mengetahui banyaknya perubahan suatu objek yang berbeda dan beberapa diantaranya ada yang sama. Untuk alasan ini, teorema berikut dapat digunakan.

Teorema 2

Banyaknya permutasi dari n objek yang terdiri atas n_1 objek sama, n_2 objek sama, ... ,n_r objek sama adalah:

$$\frac{n!}{n_1!.n_2!\ldots n_r!}$$

Contoh:

Hitung jumlah perubahan permutasi yang dapat dilakukan pada setiap huruf pada setiap kata berikut.

1. SEKOLAH
2. MAKASSAR
3. STATISTIK

Penyelesaian:

1. Kata "SEKOLAH" terdiri dari 7 huruf yang berbeda. Sehingga banyaknya permutasi yang dapat terbentuk dari kata "SEKOLAH" $= 7! = 7.6.5.4.3.2.1 = 5040.$
2. Kata "MAKASSAR" terdiri dari 8 huruf. Dimana terdapat 2 huruf yang sama yaitu A sebanyak 3 buah dan huruf S sebanyak 2 buah. Sehingga banyaknya permutasi yang dapat terbentuk dari kata "MAKASSAR" adalah
   $\frac{8!}{3!.2!} = \frac{8.7.6.5.4.3!}{3!.2.1} = \frac{6720}{2} = 3360.$
3. Kata "STATISTIK" terdiri dari 9 huruf. Dimana terdapat 3 huruf yang sama yaitu huruf S sebanyak 2 buah, huruf T sebanyak 3 buah dan huruf I sebanyak 2 buah. Sehingga banayaknya permutasi yang dapat terbentuk dari kata "STATISTIK" adalah
   $\frac{9!}{2!.3!.2!} = \frac{9.8.7.6.5.4.3!}{2.1.2.1.3!} = \frac{60480}{4} = 15120.$

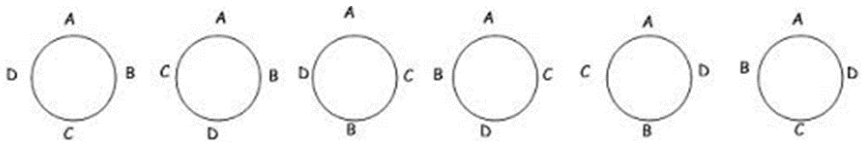
4.3.2 Permutasi Siklis (Sirkuler)

Permutasi siklik adalah permutasi yang terjadi bila unsur-unsur diletakkan melingkar menurut arah putarannya.

Banyaknya n unsur yang berbeda dalam permutasi siklis dapat dirumuskan:

$$Ps = (n - 1)!$$

Contoh:
Berapa banyak cara yang dapat disusun jika A,B,C,D akan disusun secara melingkar?
Penyelesaian:
Jika A, B, C dan D akan disusun dalam lingkaran secara satu per sata maka:



Jadi, banyaknya sususna yang terbentukada 6 cara.
Ketika menggunakan rumus dari permutasi siklis maka:

$$P_s = (4 - 1)! = 3! = 3 \times 2 \times 1 = 6 \text{ cara}$$

4.4 Kombinasi

Misalkan kita mempunyai himpunan n objek. Kombinasir objek dari n objek merupakan pemilihan r objek dari n objek tanpa memperhatikan urutannya. Jadi ab dan ba dianggap sama.
Misalkan kita mempunyai kumpulan n objek. Kombinasi r objek dari
Banyak kombinasi r objek dari n objek dapat dirumuskan:

$$C\ (n,r) = \frac{n!}{(n-r)!.\,r!}, r \leq n$$

Contoh 1:
Ada lima orang dalam satu ruangan yang masih belum saling mengenal. Ketika mereka ingin saling mengenal dengan saling berjabat tangan satu kali. Berapa banyak jabat tangan yang akan terjadi?
Penyelesaian:

Misalkan nama-nama kelima orang tersebut adalah P, Q, R, S, T
Kemungkinan yang terjadi

1. Satu kejadian, misalkan PQ
2. Ubah urutan menjadi QR
3. PQ artinya P menyalami Q, berarti QP artinya Q menyalami P.
4. jika P menyalami Q maka otomatis Q juga menyalami P, sehingga $PQ = QP \Longrightarrow Kombinasi$

Jadi, banyak cara jabat tangan yang akan terjadi adalah

$C(5,2) = \frac{5!}{(5-2)1.2!} = \frac{5!}{3!.2!} = \frac{5.4.3!}{2.1.3!} = \frac{20}{2} = 10$ cara

Contoh 2:
Suatu klub matematika di SMA "Pelita Kasih" mempunyai anggota sebanyak 15 orang siswa yang terdiri dari 10 orang perempuan dan 5 orang laki-laki. Berapa banyak cara yang dapat dipilih jika klub tersebut akan mengikuti lomba Olimpiade Matematika yang terdiri dari 3 orang perempuan dan 2 orang laki-laki?
Penyelesaian:

- Untuk siswa perempuan $n = 10\ dan\ r = 3$, maka
  $C(10,3) = \frac{10!}{(10-3)!.3!} = \frac{10}{7!.3!} = \frac{10.9.8.7!}{3.2.1.7!} = \frac{720}{6} = 120$ cara
- Untuk siswa laki-laki $n = 5\ dan\ r = 2$, maka
  $C(5,2) = \frac{5!}{(5-2)!.2!} = \frac{5!}{3!.2!} = \frac{5.4.3!}{2.1.3!} = \frac{20}{2} = 10$ cara

Jadi, banyaknya cara yang dapat dipilih untuk mengikuti lomba adalah $120 \times 10 = 1200$ cara

Teorema 3
C (n, n – r) = C (n, r)
Bukti:

$$C(n, n-r) = \frac{n!}{(n-r)!\ \big(n-(n-r)\big)!}$$
$$= \frac{n!}{r!(n-r)!}$$

Terbukti: C (n, n – r) = C (n, r)

Contoh:

$$C(5,3) = \frac{5!}{3!.(5-2)!} = \frac{5!}{3!.2!} = 10$$

$$C(5,2) = \frac{5!}{2!.(5-2)!} = \frac{5!}{2!.3!} = 10$$

Teorema 4

$C(n+1,r) = C(n,r-1) + C(n,r)$

Bukti:

$$C(n+1,r) = \frac{(n-1)!}{r!\,(n+1-r)}$$

$$C(n,r-1) = \frac{n!}{(r-1)!\,(n-(r-1))} = \frac{n!}{(r-1)!\,(n-r+1)!}$$

$$C(n,r) = \frac{n!}{r!\,(n-r)!}$$

$$C(n,r-1) + C(n,r) = \frac{n!}{r!\,(n-r+1)!} + \frac{n!}{r!\,(n-r)!}$$

$$= \frac{n!}{(r-1)!\,(n-r+1)(n-r)!} + \frac{n!}{r(r-1)!\,(n-r)!}$$

$$= \frac{n!\,r + n!\,(n-r+1)}{r(r-1)!\,(n-r+1)(n-r!)}$$

$$= \frac{n!\,r + n!\,n - n!\,r + n!}{r!\,(n-r+1)!}$$

$$= \frac{n!\,n + n!}{r!\,(n-r+1)!}$$

$$= \frac{(n+1).n!}{r!\,(n-r+1)!}$$

$$= \frac{(n+1).n!}{r!\,(n+1-r)!}$$

$$= C(n+1,r)$$

Terbukti: $C(n+1,r) = C(n,r-1) + C(n,r)$

Contoh:

$$C(5,4) = \frac{5!}{4!\,(5-4)!} = \frac{5!}{4!\,.1!} = 5$$

$$C(5,4-1) = C(5,3) = \frac{5!}{3!\,(5-3)!} = \frac{5!}{3!\,.2!} = 10$$

$$C(5+1,4) = C(6,4) = \frac{6!}{4!\,(6-4)!} = \frac{6!}{4!.\,2!} = 15$$

Jadi, $C(5,4) + C(5,3) = 5 + 10$

$= 15$

$= C(6,4)$

DAFTAR PUSTAKA

Kusriani. 2004. *Peluang*. Bagian Pengembangan Kurikulum Direktorat Pendidikan Menengah Kejuruan Jenderal Pendidikan Dasar dan Menengah Departemen Pendidikan Nasional.

Sudjana. 2002. *Metoda Statistika. Edisi 6*. Bandung: Tarsito.

_____. *Modul Kaidah Pencacahan dan Peluang.* Matematika Wajib Kelas XII MIPA/IPS.

BAB 5
PELUANG KEJADIAN

Oleh Novrianti

5.1 Pendahuluan

Peluang merupakan kajian secara matematis mengenai suatu hal yang sifatnya tidak pasti. Dalam hal menerka dan mengira dikarenakan suatu pengharapan sesuatu hal yang diinginkan terjadi, atau bahkan justru sebaliknya, supaya suatu hal tersebut tidak terjadi. Perjudian merupakan salah satu lapangan yang memakai ilmu peluang dan kajian peluang secara luas diterapkan. Hal ini bukanlah rahasia umum bahwasanya ilmu peluang kemungkinan besar pengembangannya adalah dilatar belakangi oleh gilanya atau maraknya perjudian di suatu negara tertentu. Disisi lain, kajian peluang digunakan pula dalam dunia politik, bisnis, dan kajian – kajian ilmiah yang bersifat ramalan (Walpole, 1995).

5.2 Ruang contoh dan himpunan bagiannya

Di dalam mempelajari ilmu peluan, ada istilah – istilah yang akan sering digunakan, baik dalam teori maupun dalam menyelesaiakan persoalan. Diantaranya adalah percobaan, ruang contoh, kejadian, titik contoh, dll. Masing – masing kata ini, tentu memiliki makna yang berbeda. Perhatikan definisi dan penggunaan kata – kata tersebut dalam setiap pernyataannya.

Percobaan adalah suatu kegiatan atau proses yang dilaksanakan / dilakukan untuk membangkitkan data. Disamping itu, hal yang akan terjadi dalam pelaksanaan percobaan tersebut dinamakan dengan kejadian. Sebagai contoh sederhana adalah percobaan pelemparan sekeping uang logam (koin). Kejadian yang mungkin adalah munculnya angka, atau munculnya gambar pada permukaan koin bagian atas. Pada dasarnya, tidak akan mungkin

ada kejadian lain, selain munculnya angka atau gambar, pada percobaan tersebut.

Ruang contoh adalah himpunan semua kemungkinan hasil (kejadian) suatu percobaan, dan dilambangkan dengan huruf S (Walpole, 1995).

Kejadian merupakan himpunan bagian dari ruang contoh, dan penamaan dari suatu kejadian biasanya dilambangkan dengan huruf kapital. **Kejadian** merupakan suatu peristiwa yang terjadi baik secara sengaja ataupun tidak disengaja dan merupakan sesuatu yang tidak dapat dikontrol penomenanya. Sebagai contoh selain dari pelemparan sekeping uang logam di atas adalah kejadian munculnya mata dadu ganjil pada pelemparan suatu dadu setimbang yang bersisi enam. Pelemparan dadu dapat dilakukan oleh siapa saja, akan tetapi mata dadu yang akan muncul di atas permukaan, setelah dadu tersebut berhenti, tidak ada yang bisa mengontrolnya. Keinginan suatu keadaan munculnya mata dadu ganjil, merupakan hal yang dikatakan akan terjadi, jika mata dadu yang berada di paling atas (permukaan), yang muncul dari pelemparan dadu tadi adalah mata dadu 1, 3, atau 5. Sementara kemungkinan yang lain, yakni munculnya mata dadu genap yakni 2, 4, dan 6. Mata dadu genap ini merupakan pelengkap dari keseluruhan sisi dadu, sehingga semua kejadian yang mungkin terjadi pada pelemparan suatu dadu setimbang yang bersisi enam adalah munculnya mata dadu 1, 2, 3, 4, 5, dan 6. Seluruh item ini merupakan titik contoh yang disatukan pada himpunan yang dinamakan dengan ruang contoh dari kejadian pelemparan suatu dadu setimbang bersisi enam.

5.2.1 Jenis ruang contoh

Jenis ruang contoh digolongkan menjadi dua. Pembagian ini didasarkan banyaknya titik contoh dari ruang contoh itu sendiri. Oleh karena itu, ruang contoh digolongkan menjadi ruang contoh yang berhingga, dan ruang contoh yang tidak berhingga.

A. Ruang contoh berhingga

Ruang contoh berhingga adalah ruang contoh yang banyak titik contohnya ada batasannya dan dapat dihitung / dicacah.

Contoh 1: Pada pelemparan sebuah dadu setimbang bersisi enam, dan sekeping uang logam, akan dimiliki titik – titik contoh yang dihimpun dalam ruang contoh S, yaitu

$S = \{1A, 1G\ 2A, 2G, 3A, 3G, 4A, 4G, 5A, 5G, 6A, 6G\}$.

Banyaknya titik contoh dari kejadian ini adalah $n(S) = 12$.

B. Ruang contoh tak hingga

Ruang contoh tak hingga adalah ruang contoh yang tidak memiliki batasan dari contohnya, dan / atau tidak dapat dihitung banyak titik contohnya.

Contoh 2: Garis bilangan merupakan salah satu contoh dari suatu objek yang tidak memiliki batasan, baik dari segi nilainya ataupun dari segi pencacahannya. Ruang contoh pada garis bilangan dapat ditulis ke dalam himpunan ruang contoh S, yaitu $S = \{x | x \geq 0, x \in \mathbb{R}\}$, sehingga ruang contoh ini tidak bisa dihitung, atau ruang contoh ini memiliki titik contoh yang tidak berhingga. Banyaknya titik contoh dari himpunan S adalah $n(S) = \infty$.

5.2.2 Jenis kejadian

Kejadian, yang merupakan himpunan bagian dari ruang contoh, digolongkan menjadi tiga jenis. Hal ini didasarkan pada banyaknya kejadian yang diinginkan dalam suatu percobaan, yakni kejadian sederhana, kejadian majemuk, dan kejadian nol.

a. Kejadian sederhana

Kejadian sederhana adalah kejadian yang hanya terdiri dari satu kejadian dalam suatu percobaan

Contoh 3: Kejadian munculnya mata dadu ganjil pada pelemparan sebuah dadu setimbang bersisi enam.

b. Kedian majemuk

Kejadian majemuk adalah kejadian yang banyaknya lebih dari satu kejadian dalam suatu percobaan.

Contoh 4: Kejadian munculnya mata dadu ganjil dan prima pada pelemparan sebuah dadu setimbang bersisi enam.

c. kejadian nol

Kejadian nol adalah kejadian dimana banyak anggota yang diinginkan tidaklah mungkin terjadi. Atau dengan kata lain banyaknya anggota dari kejadian ini adalah nol.

Contoh 5: Kejadian munculnya gambar pada pelemparan sebuah dadu setimbang berisisi enam.

Sementara itu, operasi yang mungkin terjadi dalam beberapa kejadian pada suatu percobaan adalah operasi irisan, gabungan, komplemen, dan selisih.

Supaya lebih memahami perbedaan antara percobaan, kejadian, ruang contoh, titik contoh, dan banyaknya anggota dari suatu kejadian atau ruang contoh, perhatikanlah rincian dalam pernyataan di bawah ini:

Pada pelemparan sekeping uang logam (koin), informasi – informasi yang dimiliki adalah sebagai berikut:

- Percobaan: Pelemparan sekeping uang logam.
- Kejadian: Munculnya angka / gambar (pilih salah satu), pada pelemparan sekeping uang logam tersebut.
- Ruang contoh : $S = \{A, G\}$.
- Titik contoh: $\{A\}$, dan $\{G\}$.
- $n(S) = 2$.

5.3 Menghitung titik contoh

Dalam mendata setiap kemungkinan yang terjadi pada suatu percobaan, diharuskan untuk dilakukan dengan sangat teliti. Hal ini dikarenakan apapun bentuk yang diperoleh, akan mempengaruhi proses selanjutnya, hingga pada akhirnya akan diperoleh nilai peluang dari suatu kejadian yang diinginkan. Adapun metode yang dapat digunakan dalam melakukan perhitugan terhadap titik contoh

pada suatu percobaan adalah metode pencacahan, penjumlahan, perkalian, permutasi, dan kombinasi.

5.3.1 Metode pencacahan

Jika terdapat suatu percobaan yang akan diamati, maka untuk mendapatkan banyak titik contoh dari keseluruhan kejadian adalah dengan mencacah setiap keadaan yang mungkin terjadi di dalam percobaan tersebut. Jika terdapat lebih dari satu objek yang akan diamati, maka dilakukan pemasangan / pencocokan setiap kejadian yang mungkin pada objek yang satu ke objek yang lainnya. Setelah itu, dilakukan pencacahan terhadap semua kemungkinan kejadian yang akan terjadi, dan hasilnya di daftarkan kedalam himpunan ruang contoh *S*.

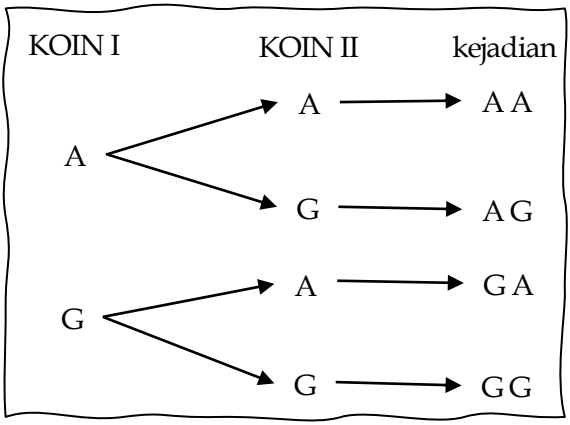
Jika akan diamati satu objek saja, maka adalah suatu hal yang mudah untuk mendaftarkan seluruh kemungkinan yang akan terjadi. Akan tetapi, untuk mendaftarkan anggota kedalam ruang contoh pada percobaan yang memiliki lebih dari satu objek pengamatan, dapat menggunakan diagram papan catur, atau dengan menggambarkan diagram akar dari masing – masing objek, yang kemudian saling dipasangkan.

Agar lebih memahaminya, perhatikanlah pencacahan objek pada percobaan pelemparan dua keping uang logam dengan menggunakan papan catur berikut ini !

Table 5.1 Papan catur pada pelemparan dua keping uang logam

Koin I \ Koin II	A	G
A	AA	AG
G	GA	GG

Dengan demikian, ruang contoh dari pelempalaran dua keping uang logam (koin) ini adalah $S = \{AA, AG, GA, GG\}$.
Metode lain untuk mendapatkan titik contoh dari percobaan adalah dengan menggunakan diagram akar seperti gambar di bawah ini.



Gambar 5.1 Diagram akar pelemparan dua keping uang logam (koin)

Tentu hasil yang diperoleh adalah sama dengan ketika menggunakan metode papan catur, yakni $S = \{AA, AG, GA, GG\}$.

Akan tetapi, jika percobaan dilakukan untuk lebih dari dua objek, metode diagram akar tentu akan lebih mudah untuk digunakan daripada metode papan catur yang akan bertransformasi menjadi tabel multi dimensi.

Dengan menggunakan kedua jenis pencacahan di atas, diperoleh hasil yang sama, bahwasanya banyaknya titik contoh dari pelemparan dua keping uang logam adalah $n(S) = 4$.

5.3.2 Metode penjumlahan

Bila suatu himpunan ruang contoh S terbagi ke dalam bagian $S_1, S_2, \ldots, S_n$ yang tidak saling tumpang tindih, maka banyaknya titik contoh dari himpunan ruang sampel S akan sama dengan jumlah dari semua unsur yang ada dalam setiap himpunan $S_1, S_2, \ldots, S_n$.

Sebagai contoh, Jika terjadi pelemparan sekeping uang logam atau pelemparan sebuah dadu setimbang bersisi enam, maka titik contoh yang mungkin muncul dari kejadian tersebut adalah muculnya sisi angka atau gambar pada uang logam, atau munculnya mata dadu 1, 2, 3, 4, 5, atau 6, sehingga ruang

contohnya akan dituliskan sebagai $S = \{A, G, 1, 2, 3, 4, 5, 6\}$. Ini bearti banyaknya titik contoh dari S adalah $n(S) = n(A) + n(B) = 2 + 6 = 8$, dengan A adalah kejadian pada pelemparan sekeping uang logam, dan B adalah kejadian pada pelemparan sebuah dadu setimbang bersisi enam.

5.3.3 Metode perkalian

Sebagaimana namanya, dalam menggunakan metode atau aturan perkalian, proses yang akan digunakan adalah dengan mengalikan semua kemungkinan yang terjadi dalam setiap objek percobaan. Metode ini dikenal juga dengan nama kaidah penggandaan umum (Walpole, 1995), yakni suatu operasi yang dapat dilakukan secara bersama dengan operasi lainnya yang kemudian dipasangkan secara berurutan. Jadi untuk dapat menggunakan aturan ini, terlebih dahulu kita harus mendata seluruh kemungkinan – kemungkinan yang terjadi pada setiap objek, secara detail.

Contoh 6: Andi memiliki 3 warna baju kaos, yakni hitam, biru, dan merah. 3 jenis celana yakni celana pendek, celana panjang, dan celana ¾. Jika Andi memiliki 2 jenis topi yakni *bucket* dan metal, dan 2 jenis sandal yakni sandal jepit, dan sandal gunung, maka berapa banyak cara memasangkan pakaian yang dapat dibuat oleh Andi dengan semua jenis item yang dimilikinya tersebut!

Penyelesaian:

Jika telah memahami aturan perkalian dengan baik, maka persoalan ini sangat mudah untuk diselesaikan. Dalam hal ini, kita hanya butuh untuk mendata semua kemungkinan yang terjadi pada setiap item jenis nya. Adapun pendataan yang kita peroleh adalah sebagai berikut:

- ✓ Baju kaos = hitam, biru, merah. (**3** kemungkinan)
- ✓ Celana = celana pendek, celana panjang, celana ¾. (**3** kemungkinan).
- ✓ Topi = bucket, metal. (**2** kemungkinan)
- ✓ Sandal = jepit, gunung. (**2** kemungkinan)

Maka banyaknya cara dalam memasangkan pakaian yang dapat dikenakan Andi adalah $3 \times 3 \times 2 \times 2 = 36$ cara berpakaian.

5.3.4 Metode permutasi

Untuk memahami penggunaan metode permutasi, perhatikan suatu proses yang menjadi kata kunci dalam penyelesaian persolan yang akan ditemui. Apakah persoalan tersebut mempermasalahkan urutan dari objek – objek yang akan disusun, atau tidak. Aturan permutasi hanya akan berfokus kepada penyusunan objek – objek yang memperhatikan urutan dalam penyusunannya. Hal ini bearti tidaklah sama antara AB dengan BA. Jika ada dua kejadian dilabeli dengan AB dan BA, hal ini bermakna ada dua kejadian yang terjadi, yakni posisi A di kanan B, atau posisi A berada di kiri B.

Ada lima jenis kejadian yang harus diselesaikan dengan menggunakan aturan permutasi.

1. Menyusun n objek ke dalam n tempat, diselesaikan dengan menggunakan permutasi $= \boldsymbol{n}!$.
   Gambaran sederhana dalam menggunakan permutasi jenis ini adalah untuk menyusun n orang untuk didudukkan pada n kursi yang berjajar. Tentu pada tempat pertama semua orang memiliki hak yang sama untuk mendapat kesempatan mendudukinya. Setelah diperoleh satu orang duduk ditempat pertama, maka objek sisanya akan beralih kekursi kedua, dan seterusnya hingga semua objek duduk di kursi masing – masing. Dalam prosesnya, hal ini menunjukkan bahwa banyaknya susunan yang terjadi menurut posisi tempat duduk berbanjar tersebut adalah:
   $$\boldsymbol{P} = n \times (n-1) \times (n-2) \times (n-3) \ldots 1 = \boldsymbol{n}!.$$

   Contoh 7: Untuk menempatkan 5 orang putra dan 3 orang putri untuk duduk pada 8 kursi yang berjajar, maka mereka dapat disusun dengan susunan sebanyak 8! = 40320 cara.

2. Menyusun dari n objek menjadi r objek, tanpa adanya kejadian berulang, diselesaikan menggunakan permutasi $\boldsymbol{P_r^n = \frac{n!}{(n-r)!}}$.
   Untuk dapat memahami bentuk persoalan yang harus diselesaikan dengan menggunakan rumusan ini, haruslah teliti terhadap pernyataan yang diberikan pada kalimat soal. Haruslah pernyataan pada soal memiliki makna bahwasanya objek yang disusun adalah berbeda, atau tidak boleh berulang, baik secara tersirat ataupun tidak.

   Contoh 8: Banyaknya cara untuk menyusun perangkat kelas yang terdiri dari Ketua, Sekretaris, dan Bendahara (tidak boleh rangkap jabatan) di kelas III B yang terdiri dari 20 orang siswa/i adalah ... cara.
   Penyelesaian:
   Karena akan dilakukan penyusunan perangkat kelas yang terdiri dari Ketua, Sekretaris, dan bendahara, hal ini bearti kejadian yang terjadi adalah memperhatikan urutan, dan keterangan tidak boleh rangkap jabatan menunjukkan bahwasanya kejadian terpilihnya A sebagai ketua, tidak boleh dibarengi dengan A menjadi sekretaris atau bendahara. Hal ini bermakna kejadian tidak boleh berulang. Maka banyak cara untuk menyusun 20 orang ke dalam perangkat kelas Ketua, sekretaris, dan bendahara adalah sebanyak $P_3^{20} = \frac{20!}{(20-3)!} = \frac{20.\ 19.18.17!}{17!} = 6840$ cara.

3. Menyusun dari n objek menjadi r objek, boleh dengan adanya pengulangan, diselesaikan menggunakan permutasi $\boldsymbol{P = n^r}$.
   Perbedaan persoalan dengan menggunakan metode ini dan metode pada poin 2 adalah terletak pada kondisi pengulangan kejadian. Pada kasus ini, kejadian yang terjadi boleh berulang.

Contoh 9: Dari kumpulan angka 2, 3, 4, dan 5, berapa banyakkah bilangan ratusan yang dapat dibentuk?
Penyelesaian:
Berdasarkan kalimat pada soalan tersebut, tidak disebutkan informasi pengulangan boleh terjadi atau tidak. Maka dalam hal ini, tersirat makna bahwasanya bilangan yang terbentuk adalah bebas, asalkan bilangan tersebut merupakan bilangan ratusan. Oleh karena itu, mungkin saja yang terbentuk adalah bilangan yang terdiri dari angka berulang, seperti 232, 233, atau bahkan bisa saja 555 dan banyak susunan lainnya yang memuat bilangan berulang.
Maka banyaknya cara untuk membuat bilangan ratusan dari bilangan 2, 3, 4, dan 5 adalah $P = 4^3 = 64$ cara.

4. Menyusun objek dalam keadaan melingkar diselesaikan dengan menggunakan $\boldsymbol{P = (n - 1)!}$.
   Penyusunan objek – objek dalam keadaan melingkar, merupakan suatu jenis khusus yang mungkin tidak akan ditemukan pada jenis perhitungan lainnya. Dalam menyelesaikan perhitungan penyusunan obek – objek dalam keadaan melingkar, perlu ditetapkan satu objek yang menjadi patokan, sehingga tidak terjadi pengulangan dalam penyusunan. Dikarenakan faktor inilah, untuk mendapatkan banyak cara, banyak objek harus dikurangi satu.

   Contoh 10: Untuk menyusun 5 orang yang akan duduk di meja bundar, maka dapat disusun kedalam P = (5 – 1)! = 4! = 24 cara.

5. Menyusun n objek ke dalam n tempat, dengan adanya objek yang berulang, diselesaikan dengan menggunakan permutasi

$$\boldsymbol{P = \frac{n!}{n_1!\, n_2! \ldots n_k!}}.$$

Penyusunan objek yang didalamnya ditemukan beberapa objek yang sejenis, tentu akan mempengaruhi banyak susunan yang terjadi. Hal ini dikarenakan objek yang sejenis, yang akan dihitung, tentu tidak dapat dibedakan dalam penyusunannya berdasarkan urutan, sehingga penyusunan harus dibagi sebanyak objek yang sejenis tersebut. Disamping itu, objek – objek yang sejenis, tentu dapat disusun pula dalam kelompok mereka, sehingga proses pembagiannya tidak hanya pada jumlah item mereka, melainkan dengan menggunakan operator faktorial pula.

Contoh 11: Huruf – huruf dari kata MATEMATIKA, akan disusun sedemikian rupa walaupun tidak memiliki makna. Jika semua huruf harus digunakan, banyaknya susunan yang dapat dibuat adalah sebanyak ... cara.
Penyelesaian: Langkah pertama yang harus dilakukan untuk menyelesaikan persoalan ini adalah mengelompokkan huruf – huruf yang sejenis terlebih dahulu, yakni huruf M ada 2, A ada 3, T ada 2.
Maka banyaknya susunan adalah $P = \frac{10!}{2!\,3!\,2!} = 151200$ cara.

5.3.5 Metode kombinasi

Dalam menyelesaikan permasalahan yang ditemukan pada suatu kejadian yang tidak mempertimbangkan faktor urutan dari objek – objek yang diinginkan, maka akan diselesaikan dengan menggunakan rumusan kombinasi. Adapun rumusan yang digunakan untuk menyusun r objek dari n objek, dengan tidak memperhatikan atau memperhitungkan urutan yang terjadi pada kejadian yang muncul adalah $C_r^n = \frac{n!}{(n-r)!r!}$.

Contoh 12: Banyak cara untuk memilih 3 orang mahasiswa yang akan dikirim ke luar daerah untuk mengikuti program kegiatan

pengabdian kepada masyarakat, dari 20 orang mahasiswa di kelas A, adalah sebanyak ... cara.

Penyelesaian: Pemilihan tiga orang mahasiswa dari 20 orang mahasiswa untuk pergi mengikuti suatu kegiatan bermakna bahwasanya urutan tidaklah diperhatikan. A, B, dan C yang pergi mengikuti pengabdian, adalah sama saja maknanya dengan B, A, dan C yang pergi mengikuti kegiatan pengabdian. Maka banyaknya cara dalam memilih mahasiswa tersebut adalah $C_3^{20} = \frac{20!}{(20-3)!3!} = 1140$ cara.

Setelah mengetahui masing – masing metode untuk mendapatkan banyaknya titik contoh dari suatu percobaan, maka pelajarilah dengan seksama bagaimana caranya untuk dapat menggunakan metode tersebut secara bersamaan untuk suatu soal tertentu. Adakalanya suatu persoalan membutuhkan lebih dari satu atau dua metode untuk dapat menghitung banyak titik contoh dari percobaan yang dipersoalkan.

5.4 Teori Peluang

Rangkaian pembahasan yang telah dikaji di atas akan diterapkan dalam perhitungan nilai peluang dari suatu kejadian yang diinginkan. Berikut ini akan dibahas tentang pengertian dari peluang, jenis – jenis peluang berdasarkan kejadian, dan soal – soal latihan serta pembahasannya.

5.4.1 Pengertian peluang

Peluang atau probabilitas adalah perbandingan nilai dari suatu kejadian yang diinginkan terjadi terhadap seluruh kemungkinan yang akan terjadi atau ruang contohnya. Secara umum peluang dari suatu kejadian A dibandingkan dengan seluruh kejadian (S) yang mungkin dituliskan dalam bentuk

$$P(A) = \frac{n(A)}{n(S)}, \qquad \dots (1)$$

dimana $P(A)$ merupakan nilai peluang dari kejadian A, $n(A)$ merupakan banyaknya titik contoh dari kejadian A , dan $n(S)$

merupakan banyaknya semua titik contoh dalam suatu percobaan yang mungkin terjadi.

Contoh 13: Sekeping uang logam setimbang dilemparkan sebanyak dua kali. Berapa peluang sekurang – kurangnya sisi gambar muncul satu kali?

Penyelesaian: Ruang contoh dari percobaan pelemparan sekeping uang logam yang dilakukan sebanyak dua kali adalah $S = \{AA, AG, GA, GG\}$. Jika B menyatakan kejadian munculnya sisi gambar sekurang – kurangnya satu kali pada pelemparan sekeping uang logam yang dilemparkan sebanyak dua kali, maka titik contoh dari kejadian B adalah *AG, GA*, dan *GG*. Dengan menggunakan persamaan (1), maka

$$P(B) = \frac{n(B)}{n(S)} = \frac{3}{4}.$$

Peluang memiliki nilai hanya berada pada kisaran 0 hingga 1. Jika mendapatkan nilai kurang dari nol, atau lebih dari satu, hal ini sudah dapat dipastikan bahwa pada perhitungan yang dilakukan terdapat suatu kekeliruan, dan perlu diperiksa kembali. Peluang suatu kejadian yang bernilai nol bermakna bahwa kejadian tersebut mustahil terjadi, dan jika peluang kejadian bernilai satu, maka kejadian tersebut dikatakan suatu hal yang pasti terjadi.

Definisi Peluang suatu kejadian (Walpole, 1995): Peluang suatu kejadian A adalah jumlah peluang semua titik contoh dalam A, dengan demikian

$$0 \leq P(A) \leq 1, \quad p(\emptyset) = 0, \quad P(S) = 1.$$

Untuk memahami definisi ini, mari dengan seksama perhatikan persoalan pada contoh 13. Masing – masing titik contoh pada sebuah koin setimbang yang dilemparkan sebanyak dua kali memiliki peluang $\frac{1}{4}$. Berdasarkan Definisi peluang suatu kejadian, maka masing – masing titik contoh pada kejadian B akan memiliki nilai peluang $\frac{1}{4}$. Oleh karena itu, peluang dari kejadian B adalah

jumlah peluang semua titik contoh pada kejadian B tersebut, yakni $P(B) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$.

Proses ini tentu memberikan nilai yang sama dengan penyelesaian pada contoh 13 di atas, yakni $P(B) = \frac{3}{4}$.

Nilai peluang dari suatu ruang contoh adalah pasti bernilai satu. Hal ini dikarenakan titik contoh dalam ruang contoh merupakan pendataan dari keseluruhan kejadian yang mungkin terjadi, sehingga $n(A) = n(S)$.

Komplemen dari suatu kejadian A merupakan lawan dari kejadian A itu sendiri. Hal ini mengakibatkan total nilai peluang dari suatu kejadian dan komplemennya akan bernilai 1. Dikarenakan kejadian tersebut merupakan dua hal yang saling bertentangan, maka jika kedua kejadian tersebut disatukan, gabungan dari kedua kejadian ini akan menjadi ruang contoh. Sebagai akibatnya, peluang dari komplemen kejadian A adalah sama dengan 1 dikurangi dengan peluang kejadian A tersebut, dituliskan dengan $P(A^c) = 1 - P(A)$.

Contoh 14: Jika diketahui peluang munculnya mata dadu genap pada pelemparan sebuah dadu yang tidak setimbang dan bersisi enam adalah $\frac{1}{4}$, berapakah peluang munculnya mata dadu ganjil pada pelemparan dadu tersebut?

Penyelesaian: Diketahui dadu yang dilemparkan adalah dadu yang tidak setimbang, sehingga tidak diketahui peluang muncul setiap titik contoh dengan baik. Akan tetapi, kita mengetahui bahwasanya pada pelemparan sebuah dadu bersisi enam, pastilah yang akan muncul adalah mata dadu genap, atau mata dadu ganjil. Bearti, untuk mendapatkan peluang munculnya mata dadu ganjil, kita dapat menggunakan rumusan komplemen kejadian. Jika A adalah peluang munculnya mata dadu genap pada pelemparan sebuah dadu tak setimbang yang bersisi enam, maka A^c adalah kejadian munculnya mata dadu ganjil pada pelemparan dadu tersebut, sehingga $P(A^c) = 1 - P(A) = 1 - \frac{1}{4} = \frac{3}{4}$.

5.4.2 Soal – soal Latihan

Selesaikanlah persoalan di bawah ini dengan cermat dan tepat!

1. Bila dilakukan suatu percobaan untuk mengambil satu huruf alfabet, secara acak di dalam suatu kotak. Hitunglah peluang terambilnya:
    a. Huruf vokal
    b. Huruf konsonan
    c. Angka

2. Pada pelemparan dua buah dadu setimbang bersisi enam, tentukanlah peluang kejadian:
    a. Jumlah mata dadu adalah 8.
    b. Jumlah mata dadu lebih dari 8.
    c. Jumlah mata dadu adalah bilangan prima.

3. Pada suatu perguruan tinggi swasta di Provinsi A, akan diadakan kegiatan EXPO tahunan yang bertema "Kembangkan UMKM bergengsi di era society 5.0". Dalam kegiatan EXPO tersebut akan diadakan perlombaan dengan 10 kategori pilihan pemerintah daerah untuk dijagokan bagi provinsi A, dan ajang ini boleh diikuti oleh UMKM manapun di provinsi A. Jika pemerintah menginginkan 3 juara dari masing – masing kategori pilihan yang diurutkan menjadi juara I, II, dan III, maka dari 25 UMKM yang mengikuti kegiatan, berapa banyak kemungkinan susunan pemenang dari kegiatan tersebut?

4. Pak Tanto memiliki 5 keping uang logam. Ketika sedang bermain dengan anaknya, Budi, Pak Tanto melemparkan uang logam tersebut secara bersamaan, dan menyerukan kepada Budi untuk menebak permukaan yang menghadap ke atas dari masing – masing uang logam tersebut. Budi menebak ada 2 sisi angka, dan 3 sisi gambar yang menghadap ke atas.
    a. Berapakah peluang Budi BENAR?
    b. Berapa pula peluang Budi SALAH?

5. Dari angka 0, 1, 2, 3, 4, dan 5 akan dibentuk bilangan ribuan yang berbeda, sehingga terbentuklah bilangan yang kurang dari 4000, tetapi lebih dari 1000. Ada berapa banyakkah bilangan yang dapat dibentuk dengan ketentuan tersebut?

Penyelesaian:

1. Diketahui alfabet yaitu A, B, C, ... , Z. Keseluruhan terdiri dari 26 huruf, bearti $n(S) = 26$. Selanjutnya huruf – huruf ini akan dibagi sesuai dengan kejadian yang diinginkan.
    a. Huruf vokal terdiri dari 5 huruf, yaitu A, I, U, E, dan O. Misalkan kejadian terpilihnya huruf vokal dilambang-kan dengan kejadian A, maka $n(A) = 5$. Maka peluang terpilihnya huruf vokal adalah $P(A) = 5/26$.
    b. Huruf konsonan adalah huruf pada alfabet selain huruf vokal, bearti sebanyak 21 huruf. Jika kejadian terpilihnya huruf konsonan dilambangkan dengan kejadian B, maka $P(B) = 21/26$.
    c. Di dalam alfabet, tidaklah ada angka, sehingga isi kejadian munculnya angka adalah 0 kejadian. Jika kejadian terpilihnya angka dilambangkan dengan C, maka $P(C) = 0/26 = 0$.

2. Pada pelemparan dua buah dadu setimbang bersisi enam, memiliki ruang sampel $S = \{11, 12, 13, \ldots, 64, 65, 66\}$, dengan $n(S) = 36$. Jika masing – masing titik contoh dijumlah, maka akan memiliki nilai berkisar 2 – 12.
    a. Kejadian munculnya jumlah mata dadu sama dengan delapan adalah $A = \{26, 35, 44, 53, 62\}, n(A) = 5$. Maka $P(A) = 5 / 36$.
    b. Jika B adalah kejadian munculnya mata dadu yang berjumlah lebih dari delapan, maka $B = \{36, 45, 46, 54, 55, 56, 63, 64, 65, 66\}, n(B) = 10$. Maka $P(B) = 10/36$.

c. Pada pelemparan dua buah dadu, jumlah mata dadu prima yang mungkin muncul adalah 2, 3, 5, 7, dan 11. Jika C adalah kejadian munculnya jumlah mata dadu yang

prima yakni 2, 3, 5, 7, dan 11. Maka titik contoh dari $C = \{11, 12, 21, 14, 23, 32, 41, 16, 25, 34, 43, 52, 61, 56, 65\}, n(C) = 15$. Maka $P(C) = 15/36$.

3. Banyak kemungkinan susunan pemenang adalah $(P_3^{25})^{10}$.

4. $n(S)$ dari pelemparan 5 keping uang logam adalah $n(S) = 2^5$. Tebakan bahwasanya yang menghadap ke atas adalah 2 sisi angka, dan 3 sisi gambar dapat dihitung dengan menggunakan aturan kombinasi, yakni $C_2^5 . C_3^3$
    a. Jika kejadian A adalah peluang Budi benar yakni ketika kelima uang logam tersebut menunjukkan 2 sisi angka dan 3 sisi gambar, maka peluang Andi benar adalah:
    $P(A) = \frac{C_2^5 . C_3^3}{2^5} = \frac{10}{32}$.
    b. Kejadian B adalah komplemen dari kejadian A, sehingga $P(B) = 22/32$.

5. Perhatikan angka – angka yang diberikan, yakni angka 0, 1, 2, 3, 4, dan 5 . Bilangan berbeda yang dibuat lebih dari 1000 tetapi kurang dari 4000, dapat ditentukan dengan aturan pengisian tempat.
    a. Bilangan pada tempat pertama, boleh diisi oleh angka 1, 2, dan 3. Ini bearti banyaknya **3** kemungkinan.
    b. Pada tempat kedua, boleh diisi oleh sisa dari angka pada poin a (2 angka yang tidak terpakai), dan juga boleh diisi 0, 4, dan 5. Ini bearti banyaknya ada **5** kemungkinan angka.
    c. Ditempat ketiga, dapat dipakai sisa angka dari poin b, bearti ada **4** kemungkinan.
    d. Dan ditempat keempat dapat dipakai angka sisa dari poin c, bearti ada **3** kemungkinan.

Oleh karena itu, banyaknya susunan bilangan ribuan berbeda yang lebih dari 1000 tetapi kurang dari 4000 adalah sebanyak $3 \times 5 \times 4 \times 3 = \mathbf{180}$ cara.

DAFTAR PUSTAKA

Walpole, R., Myers, R., & Myers, S. (1998). *Probability & Statistics for Engineers & Scientists - Instructor's solution manual*.

Walpole, R. (1995). *Pengantar Statistika Edisi ke – 3*. Gramedia Pustaka Utama, Jakarta.

BAB 6
Distribusi Binomial, Distribusi Poisson, Distribusi Geometrik

Oleh Ariyani Muljo

6.1 Peluang Bersyarat

DEFENISI. Peluang bersyarat K jika J sudah diketahui, diungkapkan dengan P (K/J), menggunakan rumus

$P(J/K) = \frac{P(J \cap K)}{P(J)}$ bila $P(J) > 0$

Misalkan, pertimbangkan situasi berikut di mana kita memiliki ruang sampel yang mewakili populasi orang dewasa yang telah menyelesaikan pendidikan menengah atas (SMA) di sebuah kota kecil. Populasi ini dapat dikelompokkan berdasarkan jenis kelamin dan status pekerjaan mereka.

	Bekerja	Tak Bekerja	Jumlah
Lelaki	460	40	500
Wanita	140	260	400
Jumlah	600	300	900

Kota ini sedang berupaya untuk mengembangkan sektor pariwisata, dan mereka berencana untuk memilih secara acak seseorang dari populasi ini untuk menjadi duta pariwisata kota tersebut, yang akan bertugas mempromosikan daerah mereka ke seluruh negeri. Dalam konteks ini, kami tertarik untuk menyelidiki dua peristiwa berikut:

P = Seseorang yang terpilih adalah seorang lelaki.

U = Seseorang yang terpilih adalah seseorang yang sedang bekerja saat ini.

Dengan mempersempit ruang sampel U, kita dapatkan
U (P/U) = $\frac{460}{600} = \frac{23}{30}$

Jika kita ingin mengukur jumlah elemen dalam himpunan A, kita dapat menggunakan notasi matematis n(A).
U(P/U) = $\frac{n(U \cap P)}{n(U)} = \frac{n(U \cap P)/n(S)}{n(U)/n(s)} = \frac{U(U \cap P)}{U(U)}$
U(U∩P) dan P(U) diperoleh dari ruang sampel asal S: Untuk memverifikasi hasil ini, perlu dicatat bahwa....
U(U) = $\frac{600}{900} = \frac{2}{3}$
Dan
U(U∩P) = $\frac{460}{900} = \frac{23}{45}$
Jadi
U(P/U) = $\frac{\frac{23}{45}}{\frac{2}{3}} = \frac{23}{30}$
Contoh.
Probabilitas penerbangan tepat waktu dapat diungkapkan sebagai berikut: P(B) = 0,83, sedangkan probabilitas tiba tepat waktu adalah P(S) = 0,82, dan probabilitas berangkat dan tiba tepat waktu adalah P(B∩S) = 0,78. Tugas kita adalah untuk menentukan peluang bahwa pesawat...
Jawab.
A probabilitas pesawat tiba tepat waktu ketika keberangkatannya tepat waktu.

$$P(S/B) = \frac{P(B \cap S)}{P(B)}$$
$$= \frac{0,78}{0,83} = 0,94$$

6.2 Peubah Acak

Dalam statistika, kita berfokus pada pengambilan kesimpulan mengenai populasi dan karakteristik populasi tersebut.

Percobaan dalam statistika menghasilkan data yang memiliki unsur ketidakpastian. Salah satu contoh eksperimen statistik adalah pengujian sejumlah komponen elektronik, yang merupakan proses mengamati hasil yang dapat bervariasi. Dalam konteks ini, sangat penting untuk menetapkan nilai numerik sebagai acuan untuk mengevaluasi hasilnya. Sebagai ilustrasi, dalam situasi di mana kita menguji tiga komponen elektronik, kita perlu secara rinci mencatat setiap kemungkinan hasil yang mungkin terjadi.
J = {LLL, LLK, LKL, KLL, LKK, KLK, KLK, KKL, KKK}

Dalam situasi di mana "baik" direpresentasikan oleh L dan "cacat" direpresentasikan oleh K, tujuan kita adalah untuk mengukur jumlah produk cacat. Oleh karena itu, setiap titik dalam ruang uji akan diberi nilai numerik 0, 1, 2, atau 3. Secara dasarnya, angka-angka ini adalah variabel acak yang mencerminkan hasil dari percobaan. Nilai-nilai ini dapat dianggap sebagai representasi dari variabel acak Z, yang menggambarkan jumlah produk cacat ketika tiga komponen elektronik diuji.

DEFINISI. Peubah acak adalah fungsi yang memberikan nilai bilangan nyata untuk setiap elemen dalam ruang sampel

Biasanya, peubah acak menggunakan huruf besar, misalnya Z, sedangkan nilai yang dimiliki oleh peubah tersebut direpresentasikan dengan huruf kecil, seperti Z. Dalam situasi sebelumnya yang berkaitan dengan suku cadang elektronik, peubah acak Z akan memiliki nilai 2 untuk setiap bagian dalam kumpulan tersebut.
E = {KKL, KLK, LKK}

Dari ruang sampel J, kita dapat menyatakan bahwa nilai X mewakili suatu peristiwa yang merupakan subbagian dari himpunan percobaan ini.

Misalkan: Dua manik diambil secara kontinu tanpa mengembalikannya dari kantung yang berisi 4 manik kuning dan 3

manik abu. Jika kita mengacu pada Y sebagai variabel acak yang menggambarkan jumlah manik merah yang diambil, maka kita dapat mencatat beragam kemungkinan nilai yang dapat dimiliki oleh Y

Ruang Sampel	y
KK	2
KA	1
AK	1
AA	0

Dari contoh di atas, ruang sampel terdiri dari jumlah anggota yang terbatas. Namun, ketika satu dadu dilempar hingga angka 5 muncul, kita akan memiliki ruang sampel yang terdiri dari deretan anggota yang tak terbatas.

J = {T, YT, YYT, YYYT, . . . }

Dalam konteks ini, kita menggunakan T untuk menyatakan kejadian munculnya angka 5, dan Y untuk menyatakan kejadian ketika bukan angka 5 yang muncul. Meskipun dalam percobaan ini terdapat sejumlah tak terbatas unsur, kita tetap dapat mengidentifikasi mereka dengan menggunakan bilangan bulat, seperti unsur pertama, unsur kedua, unsur ketiga, dan seterusnya, sehingga mereka dapat dihitung secara berurutan.

DEFINISI. Dalam kasus di mana sebuah ruang sampel terdiri dari sejumlah elemen yang terbatas atau serangkaian anggota yang dapat dihitung dalam bilangan bulat, maka ruang sampel tersebut digolongkan sebagai ruang sampel diskrit.

Waktu Anda mengkaji hasil eksperimen statistika, terkadang mereka dapat berkaitan dengan nilai yang tak terbatas atau tak terhitung. Sebagai ilustrasi, dalam penelitian tentang sejauh mana mobil merek tertentu dapat pergi dengan menggunakan 5 liter bensin di jalan tertentu, jarak yang dapat ditempuh dianggap sebagai variabel yang diukur dengan tingkat ketelitian tertentu.

Dalam situasi semacam ini, ada banyak kemungkinan jarak

yang dapat dicapai dalam ruang sampel, yang tak terbatas, sehingga tak dapat diidentifikasi sebagai bilangan bulat. Hal yang sama berlaku jika kita ingin mencatat berapa lama waktu yang dibutuhkan dalam suatu reaksi kimia. Oleh karena itu, dapat disimpulkan bahwa ruang sampel tidak selalu memiliki karakteristik yang bersifat diskrit.

DEFENISI. Ketika ruang sampel memiliki elemen tak terbatas dan setara dalam jumlah dengan titik pada segmen garis tertentu, maka ruang sampel tersebut dikategorikan sebagai ruang sampel kontinu..

6.3 Distribusi peluang diskret

Sebuah peubah acak diskret mengambil setiap nilai dengan probabilitas tertentu. Sebagai contoh, dalam situasi melempar sebuah koin tiga kali, peubah acak X yang mewakili jumlah muka yang muncul, memiliki nilai 2 dengan probabilitas $\frac{3}{8}$. Ini disebabkan oleh tiga dari delapan hasil yang mungkin menghasilkan dua muka dan satu belakang.

Dalam kasus kejadian sederhana yang diberi bobot pada contoh dengan Pak Ali, Badu, dan Cokro, peluang bahwa tidak ada petani yang menerima kembali topinya dengan benar, yaitu peluang bahwa C mendapatkan nilai 0 adalah $\frac{1}{3}$. Nilai-nilai yang mungkin untuk C dan peluangnya diberikan oleh rumus yang diberikan dalam situasi tersebut.

C	0	1	3
P(C = c)	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

Harap diperhatikan bahwa peubah acak C mencakup semua kemungkinan nilai yang mungkin, sehingga total probabilitasnya sama dengan 1.

Seringkali lebih praktis untuk menyatakan semua probabilitas dari suatu peubah acak X dalam sebuah rumus. Rumus semacam ini biasanya menggambarkan fungsi nilai numerik x yang akan dinotasikan sebagai f(x) = P(X = x); contohnya, f(3) = P(X = 3). Sekumpulan pasangan terurut (x, f(x)) ini disebut sebagai fungsi peluang atau sebaran probabilitas dari peubah acak diskret X.

DEFINISI. Ketika kita memiliki himpunan pasangan terurut (x, f(x)), ini dapat digunakan untuk mewakili suatu fungsi peluang, fungsi massa peluang, atau sebaran peluang dari peubah acak diskret X. Ini berlaku untuk setiap nilai potensial x.

1. $f(x) \geq 0$
2. $\sum_x f(x) = 0$
3. $P(X = x) = f(x)$

Misalkan.

Dalam suatu pengiriman, terdapat 8 komputer PC yang semua identik. Dalam 8 komputer ini, 3 di antaranya rusak. Sekolah ingin membeli 2 komputer secara acak dari pengiriman tersebut, dan kami ingin menentukan sebaran peluang untuk jumlah komputer rusak yang akan mereka beli.

Jawab.

Jika kita menggunakan X untuk mewakili jumlah komputer yang rusak yang dibeli oleh sekolah, X bisa memiliki nilai 0, 1, atau 2.. Maka , ...

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{5}{2}}{\binom{8}{2}} = \frac{10}{28}$$

$$f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{5}{1}}{\binom{8}{2}} = \frac{15}{28}$$

$$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{5}{0}}{\binom{8}{2}} = \frac{2}{28}$$

Maka, kita dapat menentukan sebaran probabilitas dari X.

X	0	1	2
f(x)	$\frac{10}{28}$	$\frac{15}{28}$	$\frac{2}{28}$

Misalkan.

Apabila 50% dari mobil yang akan dijual oleh agen adalah bermesin diesel, kita ingin menemukan rumus sebaran probabilitas untuk jumlah mobil dengan mesin diesel dari 4 mobil berikutnya yang akan dijual oleh agen tersebut.

Jawab.

Karena peluang menjual mobil bermesin diesel atau bensin adalah 0,5, setiap dari 2^4 kemungkinan hasil mempunyai peluang yang sama. Jadi, pembagi untuk semua peluang, termasuk fungsi peluang, adalah 16. Untuk menentukan berapa banyak cara menjual 3 mobil bermesin diesel, kita perlu mempertimbangkan berapa banyak cara membagi 4 hasil penjualan mobil tersebut menjadi dua kelompok, satu kelompok bermesin diesel dan yang lainnya bermesin bensin. Ini dapat dihitung dengan $\binom{4}{3} = 4$ cara. Secara umum, kejadian menjual x mobil bermesin diesel dan 4 - x mobil bermesin bensin dapat terjadi dalam $\binom{4}{x}$ cara, di mana x bisa bernilai 0, 1, 2, 3, atau 4. Oleh karena itu, sebaran peluang f(x) = P(X = x) adalah

$f(x) = \frac{\binom{4}{x}}{16}$ untuk x = 0, 1, 2, 3, 4.

Dalam banyak kasus, kita perlu menghitung probabilitas bahwa nilai dari peubah acak X akan kurang dari atau sama dengan suatu bilangan riil x. Jika kita menyatakan F(x) = P(X ≤ x) untuk setiap bilangan riil x, maka F(x) disebut sebagai sebaran kumulatif atau fungsi sebaran dari peubah acak X.

6.4 Distribusi peluang kontinu

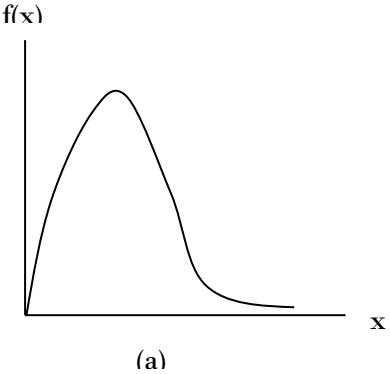
Sebuah peubah acak kontinu memiliki probabilitas nol pada setiap titik x. Oleh karena itu, sebaran probabilitasnya tidak dapat dijelaskan dalam bentuk tabel. Sehingga, yang diperhatikan adalah nilai dalam suatu interval (selang) daripada nilai pada titik tertentu dari peubah acak.

Kemudian, kita akan mendalami perhitungan probabilitas untuk berbagai interval dari peubah acak kontinu, seperti P(a < X < b), P(W > c), dan sebagainya. Penting untuk dicatat bahwa ketika X adalah peubah acak kontinu, ...

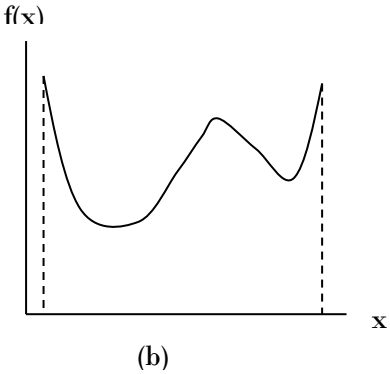
$$P(a < X \leq b) = P(a < X < b) + P(X = b)$$
$$= P(a < X < b)$$

Dalam konteks peubah acak yang kontinu, f(x) disebut sebagai fungsi kepadatan peluang, atau singkatnya fungsi kepadatan X. Karena X didefinisikan dalam ruang sampel yang kontinu, ada kemungkinan bahwa f(x) mungkin tidak kontinu pada beberapa titik yang berjumlah terbatas.

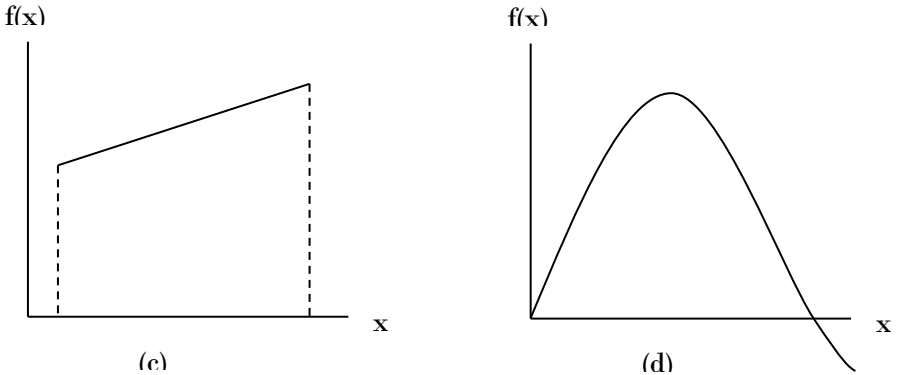
$$P(a < X \leq b) = \int_a^b f(x)dx$$



(a)



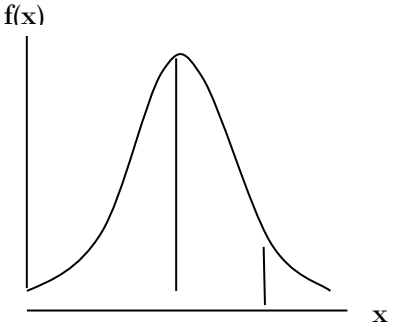
(b)



Gambar 6.1 Bentuk khas fungsi padat

DEFENISI. Fungsi f(x) adalah fungsi padat peluang peubah acak kontinu X, yang didefeniskan di atas himpunan semua bilangan real R, bila

1. $f(x) \geq 0$ untuk semua $x \in R$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$
3. $P(a < X < b) = \int_a^b f(x)dx$



Contoh.
Misalkan X mewakili galat suhu reaksi dalam derajat Celsius pada suatu percobaan laboratorium yang terkendali, dan X memiliki fungsi kepadatan peluang.

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

a. Mari kita perlihatkan bahwa kedua syarat dari definisi tersebut dipenuhi.
b. Sekarang, mari kita hitung $P(0 < x \leq 1)$.

Jawab.

a. $\int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{2} \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1$

b. $P(0 < X \leq 1) = \int_{0}^{1} \frac{x^2}{3} dx = \frac{x^3}{9}\Big|_{0}^{1} = \frac{1}{9}$

DEFINISI. Fungsi sebaran kumulatif (CDF) F(x) untuk suatu peubah acak kontinu X dengan fungsi kepadatan f(x) dapat dinyatakan sebagai berikut:

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

Sebagai hasil langsung dari definisi di atas, kita dapat menyusun kedua rumusan berikut:
$P(a < X < b) = F(b) - F(a)$
Dan

f(x) = $\frac{dF(x)}{dx}$
Jika derivatif fungsi ini tersedia.

Contoh
Mari kita cari F(x) dari fungsi kepadatan pada contoh di atas, dan kemudian kita akan menghitung P(0 < X ≤ 1).
Jawab
Untuk -1 < x < 2,
F(x) = $\int_{-\infty}^{x} f(t)dt = \int_{-1}^{x} \frac{t^2}{3} dt = \frac{t^3}{9} \Big|_{-1}^{x} = \frac{x^2+1}{9}$

Jadi,

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{x^3+1}{9}, & -1 < x < 2 \\ 1, & \text{Untuk x lainnya} \end{cases}$$

DEFINISI. Sebaran kumulatif F(X) suatu peubah acak diskret X dengan sebaran peluang f(x) dinyatakan oleh
F(X) = P(X ≤ x) = $\sum_{t \le x} f(t)$ untuk $-\infty < x < \infty$

Contoh

Hitunglah sebaran kumulatif peubah acak X dalam contoh f(x) = $\frac{\binom{4}{x}}{16}$ untuk x = 0, 1, 2, 3, 4. Dengan menggunakan F(X), perlihatkan bahwa f(2) = $\frac{3}{8}$
Jawab.
Dengan melakukan perhitungan secara langsung untuk sebaran peluang pada contoh f(x)= $\frac{\binom{4}{x}}{16}$ untuk x = 0, 1, 2, 3, 4. Diperoleh f(0) = $\frac{1}{16}$, f(1) = $\frac{1}{4}$, f(2) = $\frac{3}{8}$, f(3) = $\frac{1}{4}$, dan f(4) = $\frac{1}{16}$.

Jadi

$F(0) = f(0) = \frac{1}{16}$

$F(1) = f(0) + f(1) = \frac{5}{16}$

$F(2) = f(0) + f(1) + f(2) = \frac{11}{16}$

$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16}$

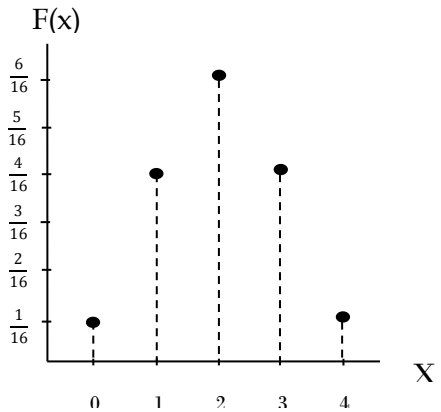
$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1$

Sehingga,

$$F(X) = \begin{cases} 0 & \text{bila} \\ \frac{1}{16} & \text{bila } 0 \leq \\ \frac{5}{16} & \text{bila } 1 \leq \\ \frac{11}{16} & \text{bila } 2 \leq \\ \frac{15}{16} & \text{bila } 3 \leq \\ 1 & \text{bila} \end{cases}$$

Sekarang

$f(2) = F(2) - F(1) = \frac{11}{16} - \frac{5}{16} = \frac{3}{8}.$



Rataan peubah acak

Jika dua mata uang dilemparkan 16 kali, dan kita menggunakan X untuk menggambarkan jumlah sisi depan yang muncul dalam setiap lemparan, maka X memiliki nilai yang berkisar antara 0, 1, hingga 2. Sebagai contoh, hasil dari serangkaian eksperimen tersebut mencatat bahwa tidak ada sisi muka muncul sebanyak 4 kali, satu sisi muka muncul sebanyak 7 kali, dan dua sisi muka muncul sebanyak 5 kali. Oleh karena itu, rata-rata jumlah sisi muka yang muncul dalam setiap lemparan dua koin adalah...

$$\frac{(0)(4)+(1)(7)+(2)(5)}{16}=1{,}06$$

Ini adalah perhitungan rata-rata dan tidak perlu menyebutkan hasil yang mungkin muncul dalam eksperimen tersebut. Sebagai contoh, rata-rata penghasilan seorang pedagang yang kemungkinannya kecil akan sama dengan penghasilannya dalam bulan tertentu. Kita dapat mengorganisir kembali perhitungan rata-rata jumlah muka yang muncul sebagai berikut...

$$(0)\left(\frac{4}{16}\right)+(1)\left(\frac{7}{16}\right)+(2)\left(\frac{5}{16}\right)=1{,}06$$

Definisi ini menjelaskan konsep nilai harapan atau rataan dari suatu peubah acak X yang memiliki sebaran peluang f(x). Nilai harapan atau rataan X adalah...

$$\mu=E(X)=\sum_{x}xf(x)$$

Bila X diskret, dan

$$\mu=E(X)=\int_{-\infty}^{\infty}xf(x)dx$$

Bila kontinu.

Contoh.
Mari kita hitung nilai harapan untuk jumlah Sosiolog dalam kepengurusan yang terdiri dari 3 orang yang dipilih secara acak dari 4 Sosiolog dan 3 psikolog..
Jawab.
kita akan menggunakan peubah acak X, yang akan mewakili jumlah Sosiolog dalam kepengurusan. Sebaran probabilitas X dapat dijelaskan sebagai berikut:

$f(X) = \frac{\binom{4}{x}\binom{3}{3-x}}{\binom{7}{3}}, x = 0, 1, 2, 3.$

Beberapa perhitungan sederhana menghasilkan f(0) = 1/35, f(1) = 12/35, f(2) = 18/35 dan f(3) = 4/35. Jadi

$$\mu = E(X) = (0)\left(\frac{1}{35}\right) + (1)\left(\frac{12}{35}\right) + (2)\left(\frac{18}{35}\right) + (3)\left(\frac{4}{35}\right) = \frac{12}{7} = 1{,}7$$

Maka, jika sebuah kepengurusan yang terdiri dari 3 orang dipilih secara acak berulang-ulang dari 4 Sosiolog dan 3 psikolog, rata-rata anggota kepengurusan diperkirakan sekitar 1,7.

Teorema . Misalkanlah X suatu peubah acak dengan sebaran peluang f(x). Rataan atau nilai harapan peubah acak g (X) adalah

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

Bila X diskret, dan

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

Bila X kontinu

Contoh .

Jumlah mobil, yang kita representasikan sebagai X, yang memasuki fasilitas pencucian mobil setiap hari antara jam 13:00 hingga 14:00 memiliki sebaran probabilitas sebagai berikut:

x	4	5	6	7	8	9
P(X = x)	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{6}$

Jika kita memakai g(X) = 2X - 1 untuk menggambarkan upah karyawan dalam ribuan rupiah yang dibayarkan oleh perusahaan pada jam tersebut, maka untuk menghitung rata-rata pendapatan karyawan pada jam tersebut...

Jawab.

Menurut teori, harapan pendapatan para karyawan...

$$E[g(X)] = E(2x-1) = \sum_{x=4}^{9}(2x-1)f(x)$$

$$= (7)\left(\frac{1}{12}\right) + (9)\left(\frac{1}{12}\right) + (11)\left(\frac{1}{4}\right) + (13)\left(\frac{1}{4}\right) + (15)\left(\frac{1}{6}\right) + (17)\left(\frac{1}{6}\right)$$

= Rp. 12,67.

Contoh.

Misalkan X sebagai peubah acak dengan fungsi kepadatan...

$$F(X) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{Untuk x lainnya} \end{cases}$$

Mari kita hitung nilai harapan dari g(x) = 4x + 3.

Jawab

$$E(4x+3) = \int_{-1}^{2} \frac{(4x+3)x^2}{3}\,dx$$

$$= \frac{1}{3}\int_{-1}^{2}(4x^3 + 3x^2)$$

= 8.

DEFENISI. Apabila X dan Y adalah dua peubah acak dengan sebaran peluang gabungan f(x, y), maka nilai harapan peubah acak g(X, Y) dapat dihitung dengan...

$$\mu_{g(X,Y)=E[g[X,Y]]=\sum_x \sum_x g(x,y)f(x,y)}$$

Bila X dan Y diskret, dan

$$\mu_{g(X,Y)=E[g[X,Y]]=\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dx\,dy}$$

Rampatan defenisi di atas untuk perhitungan harapan matematik dari fungsi dengan beberapa peubah acak telah jelas.

Contoh

Anggap X dan Y sebagai dua peubah acak dengan sebaran probabilitas bersama.

f(x, y)	x			Jumlah baris
	0	1	2	
0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
1	$\frac{3}{14}$	$\frac{3}{14}$		$\frac{3}{7}$
2	$\frac{1}{28}$			$\frac{1}{28}$
Jumlah lajur	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Hitunglah nilai harapan g (X, Y) = XY.

$E(XY) = \sum_{x=0}^{2} \sum_{x=0}^{2} xy\, f(x,y)$

= (0)(0) f(0,0) + (0)(1) f(0,1) + (0)(2) f(0,2) + (1)(0) f(1,0) + (1)(1) f(1,1) + (2)(0) f(2,0)

= f(1,1)

= $\frac{3}{14}$

Variansi dan konvariansi

Nilai harapan dari peubah acak X, yang juga disebut rata-rata, memiliki peranan penting dalam statistika karena mengindikasikan pusat dari sebaran probabilitasnya. Walaupun demikian, rataan sendiri tidak memberikan gambaran lengkap tentang seluruh karakteristik sebaran probabilitas..

Defenisi. Misalkan X peubah acak dengan sebaran peluang f(x) dan rataan μ. Variansi X adalah

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

Bila diskret dan

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

Bila X kontinu. Akar positif variansi, σ, disebut simpangan baku X.

Contoh

Anggaplah kita memiliki peubah acak X yang menggambarkan jumlah kendaraaan yang digunakan untuk keperluan dinas kantor setiap hari kerja. Sebaran probabilitas untuk kantor C adalah sebagai berikut:

x	1	2	3
f(x)	0,3	0,4	0,3

Dan untuk kantor D

x	1	2	3	4
f(x)	0,3	0,4	0,3	0,1

Buktikan bahwa varians dari sebaran probabilitas di kantor D lebih tinggi daripada varians di kantor C.

Jawab
Untuk kantor C, diperoleh
$\mu = E(X) = (1)(0{,}3) + (2)(0{,}4) + (3)(0{,}3) = 2{,}0$
Dan

$$\sigma^2 = \sum_{x=1}^{3} (x-2)^2 f(x)$$

$= (1-2)^2(0{,}3) + (2-2)^2(0{,}4) + (3-2)^2(0{,}3)$
$= 0{,}6$
Untuk kantor D, diperoleh
$\mu = E(X) = (1)(0{,}1) + (2)(0{,}3) + (3)(0{,}3) + (4)(0{,}1)$
$= 2{,}0$
Dan

$$\sigma^2 = \sum_{x=0}^{4} (x-2)^2 f(x)$$

$= (1-2)^2(0{,}1) + (2-2)^2(0{,}3) + (3-2)^2(0{,}3) + (4-2)^2(0{,}1)$
$= 1{,}6$

Dengan jelas, varians jumlah mobil yang digunakan untuk keperluan dinas lebih tinggi di kantor D daripada di kantor C.
Rumus
$\sigma^2 = \sum_x (x-\mu)^2 f(x) = \sum_x (x^2 - 2\mu x + \mu^2) f(x)$
$= \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x)$

Karena $\mu = \sum_x x f(x)$ menurut defenisi dan $\sum_x f(x) = 1$
Untuk sebaran peluang diskret, maka

$$\sigma^2 = \sum_x x^2 f(x) - \mu^2$$

$= E(X^2) - \mu^2$

Dalam situasi yang melibatkan variabel kontinu, langkah-langkah prosesnya serupa, hanya saja penjumlahan digantikan oleh integral.

Contoh
Pertimbangkan peubah acak X yang mewakili jumlah suku cadang yang cacat dari sebuah mesin ketika 3 suku cadang diambil dari rantai produksi dan diuji. Di bawah ini adalah sebaran probabilitas untuk X..

x	0	1	2	3
f(x)	0,51	0,38	0,10	0,01

Hitung σ^2 menggunakan Teorema 3.2

Jawab
Pertama – tama, hitung
$\mu = (0)(0{,}51) + (1)(0{,}38) + (2)(0{,}10) + (3)(0{,}01)$
= 0,61.
Sekarang
$E(X^2) = (0)(0{,}51) + (1)(0{,}38) + (4)(0{,}10) + (9)(0{,}01)$
= 0,87.

6.5 Distribusi Binomial

Setelah Anda menyelesaikan materi sebaran binomial, pelajari kembali konsep penting ini.

Untuk m kali eksperimen dari *y* sukses dan (m – *y*) gagal, diperoleh:
$P(X = y) = C_y^m \,.\, p^y \,.\, q^{m-y}$ atau b (*y*; m, p) = $C_y^m \,.\, p^y \,.\, q^{m-y}$

dengan y = 0, 1, 2, 3, ..., m

Bentuk ini disebut *sebaran Binomial*.

Teladan 1:

Buktikan bahwa jumlah nilai kemungkinan dari $C_x^n \,.\, p^x \,.\, q^{n-x}$ adalah sama dengan 1 !

Permainan Matematika

Dengan berbekal konsep di atas, sekarang Anda telah siap mengerjakan permainan matematika berikut. Jika tempat untuk menyelesaikan permainan tidak mencukupi, Anda boleh tambahkan di ruang kosong pada *Lembar Kerja* Anda.

Permainan 1

Dalam 10 lemparan koin, hitung probabilitas munculnya sisi muka (muka koin) sebanyak 5 kali.

Jawab:

Sebuah koin dilempar sebanyak 10 kali. Dalam hal ini, n = ...
Jika kita definisikan peristiwa hasil sukses sebagai kemunculan sisi muka (M), maka...
P (sukses) = P (M) = ... = ... , dan
P (gagal) = P (B) = ... = ...
Tampak bahwa p + q = 1 atau p = ... atau q = ...
Sehingga, b (5; 10, $\frac{1}{2}$) = $C^{\cdots}_{\cdots}$. (…)$^{\cdots}$. (…)$^{\cdots}$
= .....
= .....
= .....

Permainan 2

Dalam 12 lemparan dadu, hitung probabilitas munculnya angka 6 sebanyak 3 kali.

Jawab:

Sebuah dadu dilemparkan 12 kali. Berarti n = ...
Peristiwa sukses (s) adalah .....
Peristiwa gagal (g) adalah .....
P (sukses) = P (hasil 6) = ...
P (gagal) = P (hasil bukan 6) = ...
Ruang sampel = S = { ... , ... , ... , ... , ... , ... }

Maka probabilitas munculnya mata 6 sebanyak 3 kali adalah:
b (3; 12, $\frac{1}{6}$) = $C^{\cdots}_{\cdots}$. (…)$^{\cdots}$. (…)$^{\cdots}$
= .....

= .....
= .....

Uji Kompetensi

Kerjakan soal latihan berikut pada ruang kosong di *Lembar Kerja* Anda!

1. Diketahui bahwa 10% dari suatu kumpulan kaleng cat berisi cat warna hijau. Kemudian, secara acak diambil 30 kaleng. Berapa probabilitasnya bahwa:
   a. Semua 30 kaleng yang terambil berisi cat hijau.
   b. Salah satu dari 30 kaleng yang terambil berisi cat hijau.
   c. Paling sedikit satu dari 30 kaleng yang terambil berisi cat hijau.
   d. Paling banyak dua dari 30 kaleng yang terambil berisi cat hijau.
2. Diketahui bahwa probabilitas kesembuhan seorang pasien dari suatu penyakit langka adalah 0,4. Dalam populasi yang terdiri dari 15 orang yang terjangkit penyakit tersebut, kita ingin mencari probabilitas:
   a. Setidaknya 10 orang bertahan dan sembuh.
   b. Antara 3 hingga 8 orang sembuh.
   c. Tepat 5 orang yang sembuh.

Kesimpulan

Setelah mengerjakan soal-soal di atas, bagaimana menurut pendapat Anda? Apa yang dapat Anda simpulkan dari materi sebaran binomial ini?

6.6 Distribusi Poisson

Setelah Anda menyelesaikan materi sebaran poisson, pelajari kembali konsep penting ini.

Distribusi Poisson adalah sebaran probabilitas yang digunakan untuk peubah acak Poisson X, yang menggambarkan jumlah kejadian sukses yang terjadi dalam suatu periode waktu tertentu, dan dapat dinyatakan sebagai berikut:

$$P(X = x) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$$

atau ditulis dengan $f(x;\lambda) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$

Tampak bahwa untuk semua nilai x = 1, 2, 3, ...

Teladan 2:

Buktikan bahwa jumlah nilai kemungkinan dari $f(x;\lambda) = \frac{\lambda^x}{x!} \cdot e^{-\lambda}$ adalah sama dengan 1 !

Permainan Matematika

Dengan berbekal konsep di atas, sekarang Anda telah siap mengerjakan permainan matematika berikut. Jika tempat untuk menyelesaikan permainan tidak mencukupi, Anda boleh tambahkan di ruang kosong pada *Lembar Kerja* Anda.

Permainan 1

Misalkan rerata ada 1,4 orang buta huruf dari setiap 100 orang. Jika kita mengambil sampel berukuran 200 orang, maka kita ingin menentukan nilai kemungkinan bahwa tidak ada orang yang buta huruf dalam sampel tersebut.

Jawab:

Jika kita menggunakan X untuk mewakili jumlah orang buta huruf dari 200 orang tersebut, maka nilai λ (lambda) dapat dihitung sebagai λ = n * p.

= ...(...) = ...

Probabilitas bahwa tidak ada orang buta huruf dalam sampel tersebut adalah:

$$P(X = \cdots) = \frac{(\ldots^{\cdots})(\ldots^{\cdots})}{\ldots} = \cdots$$

Dengan demikian, nilai kemungkinan tidak adanya orang yang buta huruf dalam sampel tersebut adalah: 1 - (probabilitas X = 0) = ...

Permainan 2

Diasumsikan bahwa 1% dari sekumpulan orang berkacamata. Kemudian, secara acak dipilih 500 orang. Berdasarkan informasi ini, kita dapat menyelesaikan beberapa pertanyaan:

a. Probabilitas munculnya persis 2 orang yang berkacamata dari 500 orang yang dipilih adalah:
b. Rata-rata jumlah orang yang diharapkan berkacamata dari 500 orang tersebut adalah:
c. Heterogenitas atau keragaman jumlah orang yang berkacamata dari 500 orang tersebut adalah:

Jawab:

P = ...

n = ...

λ = n . p = ...

a. $P(X = \cdots) = \frac{(\ldots^{\cdots})}{\ldots}(\ldots^{\cdots}) = \frac{(\ldots^{\cdots})}{\ldots}(\ldots^{\cdots}) = \cdots$

b. E(X) = ...

Banyaknya orang yang diharapkan berkacamata = ...

c. Var(X) = ...

Varians banyaknya orang yang berkacamata = ...

Uji Kompetensi

Kerjakan soal latihan berikut pada ruang kosong di *Lembar Kerja* Anda!

1. Dengan menggunakan sebaran Poisson, tentukan:
   a. P(2 ; 1)
   b. P(3 ; ½)
   c. P(2 ; 0,7)
2. Dalam setiap kelompok 100 lembar kertas yang diproduksi oleh pabrik, diperkirakan terdapat 1 lembar yang rusak. Jika kita ingin menentukan probabilitas mendapatkan satu lembar kertas rusak dari 20 lembar kertas yang diambil secara acak, kita dapat menggunakan sebaran binomial. Dalam hal ini, probabilitas keberhasilan (p) adalah 1/100 (kemungkinan mendapatkan kertas rusak), dan ukuran sampel (n) adalah 20 lembar kertas. Probabilitasnya adalah:

3. Jika sekitar 2% dari total uang di Bank adalah palsu, maka probabilitas mendapatkan uang palsu dari 100 lembar uang yang diambil secara acak adalah sekitar 2%. Dalam kasus ini, kita tidak perlu menggunakan sebaran binomial karena hanya ada dua kemungkinan: uang palsu atau uang asli. Probabilitasnya adalah sekitar 2%.

Kesimpulan

Setelah mengerjakan soal-soal di atas, bagaimana pendapat Anda terhadap materi Sebaran Poisson? Coba kemukakan 2 simpulan Anda mengenai materi ini?

6.7 Distribusi Geometrik

Setelah Anda menyelesaikan materi sebaran geometrik, pelajari kembali konsep penting ini:

Sebaran probabilitas yang menggambarkan kejadian sukses pertama kali terjadi setelah sejumlah kejadian gagal disebut sebagai sebaran geometrik.
Sebaran geometrik ditulis:

$$g(x,p) = pq^{x-1}$$

dengan p = probabilitas terjadinya *sukses*
q = probabilitas terjadinya *gagal* = 1 – q

Dalam sebaran geometrik, terdapat nilai harapan (harapan matematis) dan varians yang dapat dihitung sebagai berikut:

$$E(X) = \frac{1}{p} \; dan \; Var(X) = \frac{q}{p^2}$$

Permainan Matematika

Dengan berbekal konsep di atas, sekarang Anda telah siap mengerjakan permainan matematika berikut. Jika tempat untuk menyelesaikan permainan tidak mencukupi, Anda boleh tambahkan di ruang kosong pada *Lembar Kerja* Anda.

Permainan 1

Dalam sebuah keranjang yang berisi apel, diperkirakan bahwa 5% dari apel tersebut berulat. Jika apel diambil satu per satu, diperiksa, dan kemudian dikembalikan ke dalam keranjang, kita dapat menghitung probabilitas terpilihnya apel berulat untuk pertama kalinya pada pemeriksaan yang ke-10 dengan menggunakan konsep sebaran geometrik.

Jawab:

Kita dapat mengasumsikan \(X\) sebagai peristiwa ketika apel berulat terpilih untuk pertama kalinya pada pemeriksaan yang ke-10.

x = ...

p = P (terpilihnya apel berulat) = ...

q = ... − ... = ...

g(10; 0,05) = ... (...)$^{...}$

= ...

= ...

Permainan 2

15% dari seluruh karcis di bioskop A adalah karcis bebas. Hitung peluang bahwa penonton yang merupakan penonton ke-100 adalah penonton pertama yang menggunakan karcis bebas.

Jawab:

Kejadian di atas adalah peristiwa sebaran geometrik, dengan:

p = ...

x = ...

g(100; 0,15) = ... (...)$^{...}$

= ...

= ...

Permainan 3

Tentukan harga harapan dan varians dari sebaran geometrik pada permainan 3 di atas!

Jawab:

p = ...

q = ... − ... = ...

$E(X) = \frac{\cdots}{\cdots} = \frac{\cdots}{\cdots}$

$Var(X) = \frac{\cdots}{\cdots} = \frac{\cdots}{\cdots} = \cdots$

Uji Kompetensi

Kerjakan soal latihan berikut pada ruang kosong di *Lembar Kerja* Anda!

1. Misalkan pabrik menghasilkan produk yang cacat sebanyak 2%. Tentukan peluang mendapatkan produk rusak pada saat mengeluarkan produk pertama dari jalur produksi.
2. Jika sebuah kartu diambil secara acak dari set kartu bridge, kemudian dikembalikan, hitung peluang untuk mendapatkan kartu As pada pengambilan kesepuluh.
3. Saat dua dadu dilempar bersama-sama, cari nilai ekspektasi dan varians untuk peristiwa pertama kali munculnya dua mata dadu yang sama.

Kesimpulan

Setelah mengerjakan soal-soal di atas, bagaimana menurut Anda mengenai materi sebaran geometrik? Apa yang dapat Anda kemukakan minimal 2 simpulan berdasarkan materi ini?

DAFTAR PUSTAKA

Bain, Lee J. dan Max Engelhardt, 1992. *Introduction to Probability and Mathematical Statistics, 2nd ed.* California: Duxbury Press.

Hogg, R.V. dan Allen T Craig, 1995. *Introduction to Mathematical Statistics, 5th ed.* New Jersey: Prentice-Hall.

BAB 7
UJI HIPOTESIS

Oleh Retna Kristiana

7.1 Pengertian Uji Hipotesis

Dalam rutinitas kegiatan sehari-hari, kita sering menemukan informasi yang dapat dianggap sebagai data. Tentunya informasi yang diperoleh harus diubah terlebih dahulu menjadi data yang mudah dibaca dan dianalisis. Statistika adalah cabang pengetahuan yang fokus pada metode pengolahan data(Rotondi, 2022). Untuk memperoleh data tersebut diperlukan penelitian melalui berbagai metode serta tahapan pengujian berbeda yang dilakukan oleh peneliti.

Sebelum memulai penelitian, langkah pertama adalah merumuskan hipotesis, yaitu pernyataan awal yang mengindikasikan apa yang akan diselidiki dalam penelitian (Talero-Sarmiento, 2019). Hipotesis berasal dari bahasa Yunani , dimana hupo berarti lemah atau kurang atau di bawah sedangkan thesis berarti teori, proposisi atau pernyataan yang disajikan sebagai bukti. Sehingga dapat diartikan sebagai Pernyataan yang masih lemah kebenarannya dan perlu dibuktikan atau dugaan yang sifatnya masih sementara. Pengujian Hipotesis adalah suatu prosedur yang dilakukan dengan tujuan memutuskan apakah menerima atau menolak hipotesis mengenai parameter populasi(Dorigo, 2022).

Uji hipotesis dalam statistika merupakan suatu anggapan atau pernyataan, yang mungkin benar atau tidak, mengenai satu populasi atau lebih. Hipotesis nol = H_0 : setiap hipotesis yang akan diuji dinyatakan dengan hipotesis nol. Pada pengujian hipotesis dihadapkan pada sekumpulan sampel untuk ditarik kesimpulannya dari hasil analisis sample yang dilakukan. Kesimpulan umum adalah kesimpulan demografis (Woo, 2023).

Jadi sampel hasil tangkapannya harus mewakili populasi (benar-benar representative populasi). Selalu ada dua tipe anggapan pada uji hipotesis untuk setiap masalah yang perlu diselesaikan. Hipotesis ini merupakan sesuatu yang perlu dilakukan karena akan menjadi gambaran temuan dalam penelitian.

Uji Hipotesis adalah metode pengambilan keputusan yang didasarkan dari analisis data, baik dari percobaan yang terkontrol, maupun dari observasi (tidak terkontrol)(L. Zhang, 2020). Dalam statistika, sebuah hasil dapat dikatakan signifikan secara statistik jika kejadian tersebut hampir tidak mungkin disebabkan oleh faktor yang kebetulan, sesuai dengan batas peluang yang sudah ditentukan sebelumnya.

Uji hipotesis kadang disebut juga "konfirmasi analisis data"(C. Zhang, 2022). Keputusan dari uji hipotesis biasanya berdasarkan uji hipotesis nol. Hal ini merupakan uji untuk menjawab pertanyaan yang mengasumsikan hipotesis nol adalah benar.

Hipotesis statistik adalah suatu anggapan atau pernyataan, yang mungkin benar atau tidak, mengenai satu populasi atau lebih. Hipotesis nol = H_0 : setiap hipotesis yang akan diuji dinyatakan dengan hipotesis nol atau hipotesis yang diartikan sebagai tidak adanya perbedaan antara ukuran populasi dan ukuran sampel(Kollo, 2020). Penolakan H_0 menjurus, pada penerimaan suatu hipotesis tandingan = H_1 atau lawannya hipotesis nol, adanya perbedaan data populasi dgn data sampel.

Galat jenis I : penolakan H_0 padahal hipotesis itu benar.
Galat jenis II : penerimaan H_0 padahal hipotesis itu salah.

Tindakan	H_0 benar	H_0 salah
Terima H_0	Keputusan benar	Galat jenis II
Tolak H_0	Galat jenis I	Keputusan benar

Kuasa suatu uji : peluang menolak H_0 bila suatu tandingan tertentu benar

Uji eka arah : uji hipotesis dengan wilayah penolakan H_0 ada di satu sisi saja
Uji dwi arah : Uji hipotesis dengan wilayah penolakan H_0 ada di dua sisi (kiri dan kanan) sebesar 0,05
Nilai -p: taraf (keberartian) terkecil sehingga nilai uji statistik yang diamati masih berarti (nyata). Karakteristik hipotesis yang baik merupakan dugaan terhadap keadaan variabel mandiri, perbandingan keadaan variabel pada berbagai sampel, dan merupakan dugaan tentang hubungan antara dua variabel atau lebih, dinyatakan dalam kalimat jelas, sehingga tidak menimbulkan berbagai penafsiran, dan apat diuji dengan data yang dikumpulkan dengan metode-metode ilmiah(Dorigo, 2022).

7.2 Langkah Pengujian Hipotesis

Dalam uji hipotesis, peneliti dapat menolak atau tidak menolak (menerima) hipotesis yang diajukan menganalisis data guna membuktikan/menguji hipotesis (Forcina, 2019). Kita akan menolak Ho apabila kenyataan yang ada berbeda secara meyakinkan atau tidak mendukung terhadap hipotesis yang diajukan. Demikian pula sebaliknya, kita akan menerima (tidak menolak) Ho, jika kenyataan yang ada (data) tidak berbeda dengan hipotesis yang diajukan. Dalam menerima/menolak hipotesis tidak akan selalu benar 100%, tetapi akan selalu terdapat kesalahan (kebenaran ilmiah tidak bersifat mutlak) terutama dalam inferensi sampel terhadap populasi(Eftimov, 2022).

Langkah pengujian hipotesis adalah sebagai berikut:

1. Penentuan hipotesis nol (H_0) dan hipotesis alternatif (H_1)
2. Penentuan significant level (α)
3. Menentukan statistik uji / kriteria uji yang digunakan
    a. Distribusi normal (distribusi t atau z)
    b. Distribusi x^2 (chi-square)
    c. Distribusi F (Analisis Of varians)

4. Pengambilan keputusan ***(H_0 diterima atau ditolak)***

7.3 Jenis Uji Hipotesis Statistik

Uji hipotesis adalah prosedur statistik yang digunakan untuk menguji klaim atau hipotesis tentang parameter populasi berdasarkan data sampel(Ganeshpurkar, 2018).

Terdapat beberapa jenis uji hipotesis, dan setiap jenis memiliki tujuan dan penggunaan yang berbeda. Berikut beberapa jenis uji hipotesis yang umum digunakan beserta pengertiannya:

Hipotesis dalam analisis statistik memiliki beberapa jenis disesuaikan dengan topik penelitian, diantaranya **:**

1. Uji Hipotesis Z digunakan ketika deviasi standar populasi diketahui dan untuk menguji rata-rata populasi(Balay, 2022). Contoh: apabila ingin mengetahui apakah rata-rata tinggi badan populasi adalah 170 cm.
2. Uji hipotesis T digunakan ketika deviasi standar populasi tidak diketahui atau ketika sampel(Winship, 2020). Contoh: apabila ingin menguji apakah terdapat perbedaan rata-rata berat badan antara dua kelompok.
3. Uji Chi-Kuadrat digunakan untuk menguji hubungan antara variabel kategorikal atau untuk menguji kesesuaian distribusi data dengan distribusi yang diharapkan(Bahraoui, 2018). Contoh: apabila ingin menguji apakah ada hubungan antara jenis kelamin (laki-laki atau perempuan) dan kecenderungan untuk memilih profesi tertentu.
4. Uji regresi digunakan untuk menguji hubungan antara satu atau lebih variabel independen dan variabel dependen dalam analisis regresi(Meilán-Vila, 2020). Contoh: Anda ingin menentukan apakah variabel X (misalnya, jumlah jam bekerja) mempengaruhi variabel Y (misalnya, kinerja karyawan).

5. Uji korelasi Pearson digunakan untuk mengukur hubungan linear antara dua variabel kontinu (Samsonova, 2020). Contoh: apabila ingin mengetahui sejauh mana ada hubungan antara usia dan pendapatan individu.
6. Uji Mann-Whitney U digunakan untuk membandingkan dua kelompok independen ketika data tidak terdistribusi normal(Arcuri, 2018). Contoh: apabila ingin mengetahui apakah terdapat perbedaan signifikan dalam waktu reaksi antara dua kelompok pengemudi.
7. Uji Kruskal-Wallis digunakan untuk membandingkan tiga atau lebih kelompok independen ketika data tidak terdistribusi normal(Zámková, 2020). Contoh: apabila ingin mengevaluasi perbedaan dalam tingkat kepuasan pelanggan di antara tiga merek produk yang berbeda.

Setiap jenis uji hipotesis ini digunakan untuk menjawab jenis pertanyaan penelitian yang berbeda dan memerlukan pendekatan statistik yang sesuai.

DAFTAR PUSTAKA

Arcuri, A. (2018). Evaluating search-based techniques with statistical tests. *Proceedings - International Conference on Software Engineering*, *Query date: 2023-10-20 09:45:35*, 21–21. https://doi.org/10.1145/3194718.3194732

Bahraoui, T. (2018). A Family of Goodness-of-Fit Tests for Copulas Based on Characteristic Functions. *Scandinavian Journal of Statistics*, *45*(2), 301–323. https://doi.org/10.1111/sjos.12300

Balay, İ. G. (2022). Estimation and hypothesis testing in the two-stage nested design under nonnormality. *Journal of Statistical Computation and Simulation*, *92*(5), 998–1014. https://doi.org/10.1080/00949655.2021.1982941

Dorigo, T. (2022). Statistics for Data Analysis. *Proceedings of Science*, *406*(Query date: 2023-10-20 08:55:25). https://api.elsevier.com/content/abstract/scopus_id/85143761239

Eftimov, T. (2022). Introduction to Statistical Analysis. *Natural Computing Series*, *Query date: 2023-10-20 16:00:55*, 23–31. https://doi.org/10.1007/978-3-030-96917-2_4

Forcina, A. (2019). Estimation and testing of multiplicative models for frequency data. *Metrika*, *82*(7), 807–822. https://doi.org/10.1007/s00184-019-00709-6

Ganeshpurkar, A. (2018). Concepts of Hypothesis Testing and Types of Errors. *Dosage Form Design Parameters*, *2*(Query date:

2023-10-20 09:10:44), 257–280. https://doi.org/10.1016/B978-0-12-814421-3.00007-5

Kollo, T. (2020). Covariance structure tests for t-distribution. *Recent Developments in Multivariate and Random Matrix Analysis: Festschrift in Honour of Dietrich von Rosen*, *Query date: 2023-10-20 16:00:55*, 199–217. https://doi.org/10.1007/978-3-030-56773-6_12

Meilán-Vila, A. (2020). A goodness-of-fit test for regression models with spatially correlated errors. *Test*, *29*(3), 728–749. https://doi.org/10.1007/s11749-019-00678-y

Rotondi, A. (2022). Basic Statistics: Hypothesis Testing. *UNITEXT - La Matematica per Il 3 Piu 2*, *139*(Query date: 2023-10-20 16:00:55), 259–318. https://doi.org/10.1007/978-3-031-09429-3_7

Samsonova, V. P. (2020). Common inaccuracies and errors in the application of statistical methods in soil science. *Bulleten' Pocvennogo Instituta Imeni V.V. Dokucaeva*, *102*, 164–182. https://doi.org/10.19047/0136-1694-2020-102-164-182

Talero-Sarmiento, L. H. (2019). Testing the Weak Form of Efficient Market Hypothesis and Causality Analysis in Colombian Food Supply Centers. *Apuntes Del Cenes*, *38*(67), 35–69. https://doi.org/10.19053/01203053.v38.n67.2019.8040

Winship, C. (2020). Interpreting t-Statistics Under Publication Bias: Rough Rules of Thumb. *Journal of Quantitative Criminology*, *36*(2), 329–346. https://doi.org/10.1007/s10940-018-9387-8

Woo, J. L. (2023). Board 29: Compiling Census Data and Atmospheric Repository Data to Infer Socio-Environmental Trends. *ASEE Annual Conference and Exposition, Conference Proceedings*, *Query date: 2023-10-20 08:35:58*. https://api.elsevier.com/content/abstract/scopus_id/85172071464

Zámková, M. (2020). Non-parametric anova methods applied on students' performance development in course of statistics. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *68*(1), 281–289. https://doi.org/10.11118/actaun202068010281

Zhang, C. (2022). Statistical damage simulation method for complete stress-strain path of rocks considering confining pressure effect and strength brittle drop. *Yantu Gongcheng Xuebao/Chinese Journal of Geotechnical Engineering*, *44*(5), 936–944. https://doi.org/10.11779/CJGE202205017

Zhang, L. (2020). The Research and Algorithm Design of Target Deception Hypothesis Testing Based on Position Residual. *Proceedings of 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology, ICCASIT 2020*, *Query date: 2023-10-20 08:40:57*, 182–185. https://doi.org/10.1109/ICCASIT50869.2020.9368681

BAB 8
ANALISIS KORELASI

Oleh Novita Aswan

8.1 Pendahuluan

Dalam dunia yang semakin terhubung dan kompleks ini, statistika telah menjadi alat yang tak tergantikan untuk memahami hubungan antara berbagai variabel dalam berbagai konteks. Salah satu konsep penting dalam statistika adalah analisis korelasi, yang memungkinkan kita untuk mengeksplorasi dan mengukur sejauh mana dua atau lebih variabel berkaitan satu sama lain. Dalam subbab ini, kita akan menjelajahi konsep dasar analisis korelasi, teknik-teknik yang digunakan untuk mengukur korelasi, serta interpretasi hasilnya.

Analisis korelasi memberikan wawasan yang berharga dalam berbagai disiplin ilmu, seperti ilmu sosial, ekonomi, kedokteran, dan ilmu alam. Dengan memahami sejauh mana dua variabel berkaitan, kita dapat membuat prediksi yang lebih akurat, mengidentifikasi faktor-faktor yang mempengaruhi suatu fenomena, dan mengambil keputusan yang lebih baik. Misalnya, dalam ilmu ekonomi, analisis korelasi dapat membantu kita memahami hubungan antara faktor-faktor ekonomi seperti pengangguran, inflasi, dan pertumbuhan ekonomi. Di bidang kesehatan, analisis korelasi dapat digunakan untuk mengidentifikasi hubungan antara gaya hidup dan risiko penyakit tertentu.(Hamzah, Awaluddin and Maimunah, 2016).

Subbab ini akan memandu pembaca melalui konsep-konsep dasar analisis korelasi, termasuk pengukuran korelasi seperti koefisien korelasi Pearson dan Spearman. Kami juga akan membahas bagaimana melakukan analisis korelasi dengan perangkat lunak statistik dan bagaimana menginterpretasi hasilnya

dengan benar. Selain itu, kami akan menjelaskan beberapa perhatian yang perlu diperhatikan saat menggunakan analisis korelasi, seperti asumsi-asumsi yang harus dipenuhi. Dengan memahami konsep-konsep dasar analisis korelasi dan teknik-teknik yang terkait, pembaca akan memiliki dasar yang kuat untuk mengambil keputusan yang berdasarkan bukti-bukti dalam berbagai bidang penelitian dan praktik.

8.2 Defenisi Analisis Korelasi

Analisis korelasi adalah teknik statistika yang digunakan untuk mengevaluasi hubungan atau ketergantungan antara dua atau lebih variabel dalam sebuah dataset. Tujuan utama dari analisis korelasi adalah untuk mengukur sejauh mana variabel-variabel ini berkaitan satu sama lain.(Priyono, 2021). Dengan kata lain, analisis korelasi membantu kita memahami apakah ada hubungan antara variabel tersebut, dan jika ada, seberapa kuat atau lemah hubungan tersebut. Kunci dalam Definisi Analisis Korelasi adalah (Roflin, Rohana and Riana, 2022):

1. **Hubungan Antar-Variabel**: Analisis korelasi fokus pada hubungan antara variabel. Variabel ini bisa berupa data numerik, seperti usia dan pendapatan, atau data kategorikal, seperti jenis kelamin dan preferensi produk.
2. **Ketergantungan:** Analisis korelasi membantu kita menentukan apakah perubahan dalam satu variabel berkaitan dengan perubahan dalam variabel lain. Ini mengimplikasikan adanya hubungan ketergantungan antara variabel-variabel tersebut.
3. **Pengukuran Hubungan**: Selain hanya mengidentifikasi apakah hubungan ada atau tidak, analisis korelasi juga memberikan pengukuran yang lebih kuantitatif tentang sejauh mana hubungan tersebut berlangsung. Hasil dari analisis ini dapat berupa angka yang mengindikasikan seberapa erat atau lemah korelasinya.

4. **Tujuan Analisis**: Analisis korelasi digunakan untuk berbagai tujuan, termasuk pemahaman lebih dalam tentang fenomena, pengambilan keputusan yang lebih baik, peramalan, atau penelitian ilmiah.

Dalam konteks analisis korelasi, kita mengukur ketergantungan antara variabel-variabel tersebut menggunakan koefisien korelasi. Terdapat beberapa jenis koefisien korelasi yang berbeda, seperti koefisien korelasi Pearson untuk hubungan linear dan koefisien korelasi Spearman untuk hubungan non-linear atau data yang tidak berdistribusi normal. (Cahyono, 2017). Dengan menggunakan analisis korelasi, kita dapat mengidentifikasi pola-pola penting dalam data, membuat prediksi yang lebih baik, dan memahami lebih dalam tentang bagaimana variabel-variabel berinteraksi satu sama lain dalam berbagai konteks, termasuk ilmu sosial, ekonomi, ilmu alam, dan bidang-bidang lainnya. Oleh karena itu, analisis korelasi merupakan alat yang sangat berharga dalam analisis data dan pengambilan keputusan.

8.3 Jenis-Jenis Analisis Korelasi

Analisis korelasi adalah alat statistik yang penting untuk memahami hubungan antara variabel-variabel dalam sebuah dataset. Terdapat beberapa jenis analisis korelasi yang digunakan tergantung pada sifat data dan hubungan yang ingin diukur. Berikut ini adalah beberapa jenis analisis korelasi yang umum:

8.3.1 Korelasi Pearson (Koefisien Korelasi Pearson)

Korelasi Pearson (korelasi produk-moment Pearson) adalah metode yang digunakan untuk mengukur tingkat hubungan linear antara dua variabel numerik. Ini adalah salah satu alat statistik yang paling umum digunakan untuk mengidentifikasi dan mengukur sejauh mana dua variabel bergerak bersama dalam suatu pola yang linier. (Priyono, 2021). Korelasi Pearson menghasilkan koefisien korelasi, yang biasanya disimbolkan sebagai *r.* Rumus menghitung

Koefisien Korelasi Pearson (r) adalah sebagai berikut (Roflin and Zulvia, 2021):

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

Dimana:

- X dan Y adalah data yang diamati.
- $\bar{X}$ dan $\bar{Y}$ adalah rata-rata dari X dan Y.

Rentang Nilai pada Koefisien korelasi Pearson berkisar antara -1 hingga 1 yang di interpretasi untuk menarik kesimpulan hipotesis sebagai berikut (Roflin, Rohana and Riana, 2022):

- Jika nilai $r = 1$ maka ini menunjukkan hubungan positif sempurna di antara kedua variabel, yang berarti bahwa ketika satu variabel naik, yang lainnya juga naik secara linier.
- Jika nilai $r = -1$ maka ini menunjukkan hubungan negatif sempurna di antara kedua variabel, yang berarti bahwa ketika satu variabel naik, yang lainnya turun secara linier.
- Jika nilai $r = 0$ maka ini menunjukkan bahwa tidak ada hubungan linear antara kedua variabel.

Adapun Kelebihan Korelasi Pearson ini adalah :

- Sangat intuitif dan mudah dipahami,
- Rentang nilai antara -1 hingga 1, sehingga mudah untuk menginterpretasikan tingkat hubungan;
- Dapat digunakan untuk mengukur hubungan antara variabel-variabel kontinu.

Sementara itu Keterbatasan Korelasi Pearson ini adalah:

- Hanya mengukur hubungan linear, sehingga tidak sensitif terhadap hubungan non-linear;

- Rentang nilai *r* bisa menjadi tidak representatif jika ada outlier dalam data.
- Tidak mengukur kausalitas; hanya mengidentifikasi hubungan statistik.

Korelasi Pearson banyak digunakan dalam berbagai bidang, termasuk ilmu sosial, ekonomi, sains, dan bisnis. Contohnya, dalam ekonomi, korelasi Pearson dapat digunakan untuk mengukur hubungan antara pendapatan dan pengeluaran rumah tangga. Dalam ilmu sosial, korelasi Pearson dapat digunakan untuk mengevaluasi hubungan antara variabel seperti pendidikan dan tingkat pengangguran. (Kadir, 2015). Penting untuk diingat bahwa korelasi Pearson hanya mengukur hubungan linier antara variabel. Jika Anda mencurigai adanya hubungan non-linear atau kausalitas antara variabel, maka analisis yang lebih mendalam mungkin diperlukan.

8.3.2 Korelasi Spearman (Koefisien Korelasi Spearman)

Korelasi Spearman, juga dikenal sebagai koefisien korelasi Spearman atau *rho (ρ),* adalah metode statistik yang digunakan untuk mengukur tingkat hubungan monoton antara dua variabel, terutama ketika hubungan antara kedua variabel tidak memiliki sifat linier. Dalam analisis korelasi Spearman, data diubah menjadi peringkat sebelum perhitungan. (Roflin and Zulvia, 2021). Rumus untuk menghitung koefisien korelasi Spearman (ρ) adalah sebagai berikut:

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

Di mana:

- *d* adalah perbedaan peringkat antara pasangan data.
- *n* adalah jumlah pasangan data.

Rentang nilai Korelasi Sprearman (ρ) berada antara -1 sampai dengan 1 yang merupakan Interpretasi untuk menentukan kesimpulan hipotesis yaitu (Roflin, Rohana and Riana, 2022):

- Jika $\rho=1$ maka ini menunjukkan hubungan monoton positif sempurna di antara kedua variabel, yang berarti bahwa ketika satu variabel naik, yang lainnya juga naik secara monotone.
- Jika $\rho=-1$ maka ini menunjukkan hubungan monoton negatif sempurna di antara kedua variabel, yang berarti bahwa ketika satu variabel naik, yang lainnya turun secara monotone.
- Jika $\rho=0$ maka ini menunjukkan bahwa tidak ada hubungan monotone antara kedua variabel.

Adapun Kelebihan Korelasi Spearman (ρ) adalah:

- Tidak terbatas pada hubungan linier; sensitif terhadap hubungan monoton, baik positif maupun negatif.
- Bisa digunakan untuk mengukur hubungan antara variabel ordinal atau data peringkat.
- Lebih tahan terhadap pengaruh outlier dibandingkan dengan korelasi Pearson.

Sementara itu Keterbatasan Korelasi Spearman (ρ) adalah:

- Tidak cocok untuk mengukur hubungan yang bukan monoton, seperti hubungan U-shaped atau kurva S.
- Kurang sensitif terhadap variasi di tengah rentang data.
- Tidak mengukur kausalitas; hanya mengidentifikasi hubungan statistik.

Korelasi Spearman sering digunakan dalam penelitian di berbagai bidang, terutama ketika hubungan antara variabel tidak linier atau data yang diukur dalam skala ordinal atau data peringkat. Contoh penggunaan korelasi Spearman termasuk penelitian psikologi untuk mengukur hubungan antara peringkat tes psikologis dan kinerja akademik, atau dalam ilmu ekonomi untuk

mengevaluasi hubungan antara peringkat kualitas hidup dan pendapatan. Ketika Anda memiliki keraguan tentang apakah hubungan antara dua variabel bersifat monoton, korelasi Spearman adalah pilihan yang baik untuk memahami tingkat hubungan mereka. Ini memberikan wawasan yang berharga dalam konteks data yang mungkin tidak sesuai dengan asumsi hubungan linier.(Priyono, 2021).

8.3.3 Korelasi Kendal Tau (Koefisien Kendal Tau)

Korelasi Kendal Tau, juga dikenal sebagai Koefisien Kendal Tau (τ), adalah metode statistik yang digunakan untuk mengukur tingkat hubungan monoton antara dua variabel. Metode ini sering digunakan ketika data yang diamati memiliki bentuk data peringkat atau data ordinal dan ketika Anda ingin mengukur tingkat kesamaan dalam urutan data. Rumus untuk menghitung Koefisien Kendal Tau (τ) adalah sebagai berikut(Roflin and Zulvia, 2021):

$$\tau = \frac{2}{n(n-1)} \sum \sum sgn\ (X_i - X_j)$$

Di mana:

- X_i dan X_j adalah pasangan data yang dibandingkan.
- *sgn(·)* adalah fungsi tanda, yang menghasilkan -1 jika *Xi* lebih kecil dari *Xj*, 1 jika *Xi* lebih besar dari *Xj*, dan 0 jika *Xi* sama dengan *Xj*.
- *n* adalah jumlah pasangan data.

Rentang nilai Koefisien Kendal Tau (τ) diinterpretasikan sebagai berikut (Roflin, Rohana and Riana, 2022):

- Jika maka τ=1 maka ini menunjukkan hubungan monoton positif sempurna di antara kedua variabel, yang berarti bahwa ketika satu variabel naik, yang lainnya juga naik secara monotone.
- Jika maka τ=−1 maka ini menunjukkan hubungan monoton negatif sempurna di antara kedua variabel, yang

berarti bahwa ketika satu variabel naik, yang lainnya turun secara monotone.

- Jika nilaia $\tau=0$ maka ini menunjukkan bahwa tidak ada hubungan monotone antara kedua variabel.

Adapun Kelebihan Korelasi Kendal Tau ada;ah sebagai berikut:

- Tidak terbatas pada hubungan linier; sensitif terhadap hubungan monoton, baik positif maupun negatif.
- Cocok untuk mengukur hubungan antara variabel ordinal atau data peringkat.
- Lebih tahan terhadap pengaruh outlier dibandingkan dengan korelasi Pearson.

Sementara itu Keterbatasan Korelasi Kendal Tau adalah sebagai berikut:

- Lebih sulit untuk dihitung secara manual dibandingkan dengan korelasi Pearson atau Spearman.
- Kurang sensitif terhadap variasi di tengah rentang data.

Korelasi Kendal Tau sering digunakan dalam penelitian di berbagai bidang ketika data yang diamati adalah dalam bentuk data peringkat atau data ordinal. Contohnya termasuk penelitian di bidang ekologi untuk mengukur hubungan antara peringkat spesies dalam ekosistem atau di bidang ekonomi untuk mengevaluasi hubungan antara peringkat perusahaan dalam industri tertentu. Korelasi Kendal Tau adalah alat yang berguna untuk memahami hubungan dalam konteks data yang mungkin tidak sesuai dengan asumsi hubungan linier. Ia memberikan ukuran korelasi yang kuat terhadap hubungan monoton dan sering digunakan ketika hubungan antara variabel harus diukur dalam tingkat ordinal atau peringkat.(Kadir, 2015)

8.3.4 Korelasi Point-Biserial:

Korelasi Point-Biserial adalah metode statistik yang digunakan untuk mengukur tingkat hubungan antara dua variabel, salah satunya adalah variabel biner (*dua kategori)* dan yang lainnya adalah variabel numerik (*dalam skala interval atau rasio*). Metode ini memberikan nilai yang mengindikasikan seberapa kuat hubungan antara variabel biner dan variabel numerik. Rumus untuk menghitung Koefisien Korelasi Point-Biserial (r_{pb}) adalah sebagai berikut (Cahyono, 2017):

$$r_{pb} = \frac{M_1 - M_0}{\sqrt{p(1-p)}}$$

Di mana:

- M_1 adalah rata-rata variabel numerik ketika variabel biner adalah 1.
- M_0 adalah rata-rata variabel numerik ketika variabel biner adalah 0.
- p adalah proporsi dari variabel biner yang bernilai 1.

Interpretasi dari nilai Koefisien Korelasi Point-Biserial (r_{pb}) adalah sebagai berikut (Roflin, Rohana and Riana, 2022):

- Jika nilai $r_{pb} > 0$ maka ini menunjukkan hubungan positif antara variabel biner dan variabel numerik, yang berarti bahwa ketika variabel biner adalah 1, variabel numerik cenderung lebih tinggi daripada ketika variabel biner adalah 0.
- Jika nilai $r_{pb} < 0$ maka Ini menunjukkan hubungan negatif antara variabel biner dan variabel numerik, yang berarti bahwa ketika variabel biner adalah 1, variabel numerik cenderung lebih rendah daripada ketika variabel biner adalah 0.
- Jika nilai $r_{pb} = 0$ maka Ini menunjukkan bahwa tidak ada hubungan antara variabel biner dan variabel numerik.

Adapun Kelebihan Korelasi Point-Biserial adalah sebagai berikut:

- Cocok untuk mengukur hubungan antara variabel biner (dichotomous) dan variabel numerik.
- Memberikan ukuran yang jelas tentang perbedaan rata-rata variabel numerik antara dua kategori variabel biner.

Sementara itu Keterbatasan Korelasi Point-Biserial adalah:

- Hanya cocok untuk mengukur hubungan antara satu variabel biner dan satu variabel numerik.
- Mengasumsikan hubungan antara dua variabel bersifat linier, yang mungkin tidak selalu benar.
- Tidak mengukur kausalitas; hanya mengidentifikasi hubungan statistik.

Korelasi Point-Biserial sering digunakan dalam berbagai disiplin ilmu, terutama dalam penelitian yang melibatkan variabel biner dan variabel numerik. Contohnya, dalam psikologi, korelasi Point-Biserial dapat digunakan untuk mengukur hubungan antara variabel biner seperti jenis kelamin (laki-laki atau perempuan) dan variabel numerik seperti skor tes psikologis. Dalam ilmu kesehatan, ini bisa digunakan untuk mengukur hubungan antara variabel biner seperti status vaksinasi (ya atau tidak) dan variabel numerik seperti tingkat antibodi dalam darah. Korelasi Point-Biserial memberikan pandangan yang jelas tentang bagaimana variabel biner dan variabel numerik berhubungan satu sama lain dan dapat membantu dalam mengidentifikasi faktor-faktor yang berkaitan dengan perbedaan dalam variabel numerik. (Cahyono, 2017).

8.3.5 Korelasi Phi (Koefisien Phi)

Korelasi Phi juga dikenal sebagai Koefisien Phi (φ), adalah metode statistik yang digunakan untuk mengukur tingkat hubungan antara dua variabel biner (dua kategori). Metode ini memberikan ukuran sejauh mana dua variabel biner berhubungan satu sama lain.

Rumus untuk menghitung Koefisien Phi (φ) adalah sebagai berikut (Djarwanto, 2003):

$$\varphi = \frac{(ad - bc)}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Dari rumus dijelaskan bahwa *a,b,c, dan d* adalah frekuensi kemunculan dalam tabel kontingensi 2x2, di mana:

- *a* adalah jumlah pasangan yang berada di kategori 1 pada kedua variabel.
- *b* adalah jumlah pasangan yang berada di kategori 1 pada variabel pertama tetapi di kategori 2 pada variabel kedua.
- *c* adalah jumlah pasangan yang berada di kategori 2 pada variabel pertama tetapi di kategori 1 pada variabel kedua.
- *d* adalah jumlah pasangan yang berada di kategori 2 pada kedua variabel.

Adapun Interpretasi dari nilai Koefisien Phi (φ) adalah sebagai berikut (Kadir, 2015) :

- jika nilai φ=1 maka Ini menunjukkan hubungan sempurna antara kedua variabel biner, yang berarti bahwa variabel-variabel tersebut selalu berada dalam kategori yang sama.
- Jika nilai φ=0 maka Ini menunjukkan bahwa tidak ada hubungan antara kedua variabel biner.
- Jika nilai φ<1 maka Ini menunjukkan adanya hubungan antara kedua variabel biner, tetapi tidak sempurna.

Sementara itu, Kelebihan Korelasi Phi adalah:

- Cocok untuk mengukur hubungan antara dua variabel biner (dichotomous).

- Memberikan ukuran yang jelas tentang sejauh mana dua variabel biner berhubungan satu sama lain.

Selanjutnya, Keterbatasan Korelasi Phi adalah:

- Hanya cocok untuk mengukur hubungan antara dua variabel biner.
- Tidak mengukur kekuatan atau arah hubungan; hanya mengidentifikasi hubungan statistik.

Korelasi Phi sering digunakan dalam berbagai disiplin ilmu ketika Anda ingin mengukur hubungan antara dua variabel biner (dua kategori). Contoh penggunaan Korelasi Phi termasuk dalam penelitian di bidang ilmu sosial untuk mengukur hubungan antara variabel biner seperti jenis kelamin (laki-laki atau perempuan) dan variabel biner lainnya seperti status perkawinan (menikah atau belum menikah). Korelasi Phi memberikan ukuran yang berguna untuk memahami sejauh mana dua variabel biner berhubungan satu sama lain dalam konteks data yang hanya memiliki dua kategori. Ini bisa membantu peneliti atau analis dalam mengidentifikasi hubungan statistik antara variabel biner yang relevan. (Roflin and Zulvia, 2021)

8.3.6 Korelasi Cramér's V

Korelasi Cramér's V (Koefisien Cramér's V) adalah metode statistik yang digunakan untuk mengukur tingkat hubungan antara dua variabel kategorikal dengan lebih dari dua kategori. Metode ini memberikan ukuran sejauh mana dua variabel kategorikal tersebut berhubungan satu sama lain. Rumus untuk menghitung Koefisien Cramér's V (V) adalah sebagai berikut (Roflin and Zulvia, 2021):

$$V = \sqrt{\frac{\chi^2}{n(\min(k-1, r-1))}}$$

Di mana:

- χ^2 adalah nilai statistik uji chi-squared.
- *n* adalah jumlah total data.
- *k* adalah jumlah kategori dalam variabel pertama.
- *r* adalah jumlah kategori dalam variabel kedua.

Interpretasi dari nilai Koefisien Cramér's V (V) adalah sebagai berikut (Roflin, Rohana and Riana, 2022):

- jika nilai V=0 maka Ini menunjukkan bahwa tidak ada hubungan antara dua variabel kategorikal.
- Jika nilai 0<V<1 maka Ini menunjukkan adanya hubungan antara dua variabel kategorikal, dengan semakin besar nilai V, semakin kuat hubungannya. Nilai mendekati 1 menunjukkan hubungan yang lebih kuat.

Adapun Kelebihan Korelasi Cramér's V adalah:

- Cocok untuk mengukur hubungan antara dua variabel kategorikal dengan lebih dari dua kategori.
- Memberikan ukuran yang jelas tentang sejauh mana dua variabel kategorikal berhubungan satu sama lain.

Sementara itu, Keterbatasan Korelasi Cramér's V adalah:

- Hanya cocok untuk mengukur hubungan antara dua variabel kategorikal.
- Tidak mengukur kekuatan atau arah hubungan; hanya mengidentifikasi hubungan statistik.
- Tidak memberikan informasi tentang kausalitas.

Korelasi Cramér's V sering digunakan dalam penelitian di berbagai bidang ketika Anda ingin mengukur hubungan antara dua variabel kategorikal yang memiliki lebih dari dua kategori. Contoh penggunaan Korelasi Cramér's V termasuk dalam penelitian di bidang ilmu sosial untuk mengukur hubungan antara variabel kategorikal seperti pendidikan (misalnya, tingkat pendidikan: SD,

SMP, SMA) dan variabel kategorikal lainnya seperti preferensi politik (misalnya, partai politik: A, B, C). Korelasi Cramér's V memberikan ukuran yang berguna untuk memahami sejauh mana dua variabel kategorikal berhubungan satu sama lain dalam konteks data dengan lebih dari dua kategori. Ini dapat membantu peneliti atau analis dalam mengidentifikasi hubungan statistik antara variabel kategorikal yang relevan dan memahami pola-pola yang mungkin terkait dengan data tersebut (Roflin and Zulvia, 2021)

8.4 Contoh Kasus Analisis Korelasi

Contoh kasus 1:

Seorang petani ingin mengetahui apakah ada hubungan antara curah hujan tahunan di wilayahnya dan hasil panen gandum selama lima tahun terakhir. Dia mencatat curah hujan tahunan (variabel X dalam milimeter) dan hasil panen gandum (variabel Y dalam ton) selama lima tahun terakhir. Berikut adalah data curah hujan tahunan (X) dan hasil panen gandum (Y) dalam lima tahun terakhir:

Tahun	Curah Hujan (mm)	Hasil Panen Gandum (ton)
2019	800	5.2
2020	950	5.8
2021	700	5.0
2022	1050	6.2
2023	850	5.6

Penyelesaian Analisis Korelasi Pearson:

Langkah-langkah analisis korelasi Pearson dalam konteks ini adalah sebagai berikut:

Langkah 1: Menyusun data dalam bentuk tabel kontingensi.

Tahun	X	Y
2019	800	5,2
2020	950	5,8
2021	700	5,0
2022	1050	6,2
2023	850	5,6
Σ	4350	27, 8

Langkah 2: Menghitung rata-rata ($\bar{X}\ dan\ \bar{Y}$) dan jumlah pasangan (n).

- $n = \ 5\ pasang\ data$
- $\bar{X} = \frac{1}{n}\sum_{i=1}^{5} X_i = \frac{1}{5}\,.(4350) = \frac{4350}{5} = 870$
- $\bar{Y} = \frac{1}{n}\sum_{i=1}^{5} Y_i = \frac{1}{5}\,.(27.8) = \frac{27.8}{5} = 5.56$

Langkah 3: Menghitung koefisien korelasi Pearson (r) menggunakan rumus:

$$r = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2}}$$

Menghitung setiap komponen dalam rumus dan kemudian menggabungkannya:

$\sum(X_i-\bar{X})(Y_i-\bar{Y})$ = (800-870)(5.2-27.8) + (950-870)(5.8-27.8) + (700-870)(5-27.8) + (1050-870)(6.2-27.8) + (850-870)(5.6-27.8) = - 2.958.

$\sum(X-\bar{X})^2$ = (800-870)^2 + (950-870)^2 + (700-870)^2 + (1050-870)^2 + (850-870)^2 = 73.000

$\sum(Y-\bar{Y})^2$ = (5.2-27.8)^2 + (5.8-27.8)^2 + (5-27.8)^2 + (6.2-27.8)^2 + (5.6-27.8)^2 = 2.469,68

Selanjutnya, kita menggabungkan semua nilai ini ke dalam rumus korelasi Pearson:

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}} = \frac{-2.958}{\sqrt{(73.000).(2.469{,}68)}}$$
$$= \frac{-2.958}{\sqrt{180.286.640}} = \frac{-2.958}{13.427{,}09} = -0{,}21$$

Maka diperoleh bahwa nilai *r* adalah:

$$r = -0{,}21$$

Berdasarkan hasil hitung nilai r maka diketahui bahwa Koefisien korelasi Pearson (r) adalah -0.21, yang menunjukkan adanya hubungan negatif kuat antara curah hujan tahunan dan hasil panen gandum. Ini berarti bahwa semakin tinggi curah hujan tahunan, semakin rendah hasil panen gandum, dan sebaliknya. Jika curah hujan naik, hasil panen cenderung turun secara linier.

Oleh karena itu, dapat diambil kesimpulan bahwa Dalam analisis ini, kami menggunakan korelasi Pearson untuk menentukan hubungan antara curah hujan tahunan dan hasil panen gandum selama lima tahun terakhir. Hasil analisis menunjukkan adanya hubungan negatif kuat antara kedua variabel tersebut, yang dapat membantu petani untuk membuat keputusan yang lebih baik terkait dengan pengelolaan pertanian mereka berdasarkan pola curah hujan dan hasil panen gandum.

Contoh Kasus 2:

Seorang petani ingin mengetahui apakah ada hubungan antara kehadiran hama tertentu (variabel biner: ada atau tidak ada) dan hasil panen tanaman jagung (variabel numerik dalam ton) di ladangnya selama beberapa musim tanam terakhir. Dia mencatat kehadiran hama (0: tidak ada, 1: ada) dan hasil panen jagung (variabel Y) selama lima musim tanam terakhir. Berikut adalah data kehadiran hama (X) dan hasil panen jagung (Y) selama lima musim tanam terakhir:

Musim Tanam	Kehadiran Hama (X)	Hasil Panen Jagung (ton)
Musim 1	1	4.2
Musim 2	0	5.0
Musim 3	1	4.1
Musim 4	1	3.8
Musim 5	0	5.2

Penyelesaian Analisis Korelasi Point-Biserial:

Langkah-langkah analisis korelasi Point-Biserial dalam konteks ini adalah sebagai berikut:

Langkah 1: Menyusun data dalam bentuk tabel kontingensi.

Musim Tanam	X	Y
Musim 1	1	4.2
Musim 2	0	5.0
Musim 3	1	4.1
Musim 4	1	3.8
Musim 5	0	5.2
Σ	3	22.3

Langkah 2: Menghitung rata-rata hasil panen jagung ($\bar{Y}_0$ dan $\bar{Y}_1$) untuk kedua kelompok (hama tidak ada dan hama ada) serta jumlah pasangan (n).

- $\bar{Y}_0$adalah rata-rata hasil panen jagung ketika hama tidak ada (X = 0).
- $\bar{Y}_1$ adalah rata-rata hasil panen jagung ketika hama ada (X = 1).
- n adalah jumlah musim tanam (jumlah pasangan data), yaitu 5.

Menghitung Y0¯Y0¯ dan Y1¯Y1¯:

- $\bar{Y}_0$ (hama tidak ada):

$$\bar{Y}_0 = \frac{\sum_{i=1}^{5} \bar{Y}_i}{jumlah\ musim\ tanam\ ketika\ tidak\ ada\ hama}$$
$$= \frac{5{,}0 + 5{,}2}{2} = 5{,}1\ ton$$

- $\bar{Y}_1$ (hama ada):

$$\bar{Y}_1 = \frac{\sum_{i=1}^{5} \bar{Y}_i}{jumlah\ musim\ tanam\ ketika\ ada\ hama}$$
$$= \frac{4{,}2 + 4{,}1 + 3{,}8}{2} = 4{,}367\ ton$$

Langkah 3: Menghitung koefisien korelasi Point-Biserial (r_{pb}) menggunakan rumus:

$$r_{pb} = \frac{M_1 - M_0}{\sqrt{p(1-p)}}$$

Menghitung r_{pb}:

- M_1 adalah $\bar{Y}_1$ (rata-rata hasil panen jagung ketika hama ada) = 4.367 ton
- M_0 adalah $\bar{Y}_0$ (rata-rata hasil panen jagung ketika hama tidak ada) = 5.1 ton
- p adalah proporsi musim tanam ketika hama ada (X=1) = 3/5= 0.6

Maka

$r_{pb} = \frac{4.367-5{,}1}{\sqrt{0{,}6(1-0{,}6)}}$ = -1.563

Interpretasi hasil:

Koefisien korelasi Point-Biserial (rpbrpb) adalah -1.563, yang menunjukkan adanya hubungan negatif kuat antara kehadiran

hama (variabel biner) dan hasil panen jagung (variabel numerik). Ini berarti bahwa ketika hama hadir, hasil panen jagung cenderung lebih rendah, dan sebaliknya, ketika hama tidak hadir, hasil panen jagung cenderung lebih tinggi.

Kesimpulan:

Dalam analisis ini, kami menggunakan korelasi Point-Biserial untuk menentukan hubungan antara kehadiran hama (variabel biner) dan hasil panen jagung (variabel numerik) selama beberapa musim tanam terakhir. Hasil analisis menunjukkan adanya hubungan negatif kuat antara kehadiran hama dan hasil panen jagung, yang dapat membantu petani untuk memahami dampak kehadiran hama pada hasil panen mereka.

DAFTAR PUSTAKA

Cahyono, T. (2017) *Statistik Uji Korelasi*. Purwokerto: Yayasan Sanitarian Banyumas.

Djarwanto (2003) *statistik non parametrik*. yogyakarta: BPFE.

Hamzah, L. M., Awaluddin, I. and Maimunah, W. (2016) *Pengantar Statistika Ekonomi*.

Kadir (2015) *Statistika Terapan*. Jakarta: PT. Raja Grafindo Persada.

Priyono (2021) *Analisis Regresi dan Korelasi untuk Penelitian Survey (Panduan Praktis oleh Data dan Interpretasi)*. Bogor: Guepedia.

Roflin, E., Rohana and Riana, F. (2022) *Analisis Korelasi dan Regresi*. Pekalongan: PT. Nasya Expanding Management.

Roflin, E. and Zulvia, F. E. (2021) *Kupas Tuntas Analisis Korelasi*. Pekalongan: PT. Nasya Expanding Management.

BAB 9
KORELASI BERGANDA

Oleh Joni Wilson Sitopu

9.1 Pendahuluan

Aspek yang paling penting dalam penelitian sosial adalah memisahkan dan memahami hubungan diantara variabel independen (bebas) dengan variabel dependen (terikat). Korelasi adalah hubungan antara dua atau lebih variabel. Sejauh mana hubungan diantara variabel-variabel penelitian bisa diukur dengan menggunakan korelasi Pearson. Koefisien korelasi Pearson bisa digunakan jika data dalam bentuk rasio dan kaitan antara variabel adalah linier (dalam Nazariah, et.al, 2022). Koefisien korelasi berganda menggeneralisasi koefisien korelasi standar. Ini digunakan dalam analisis regresi berganda untuk menilai kualitas prediksi variabel dependen. Ini sesuai dengan korelasi kuadrat antara nilai prediksi dan nilai sebenarnya dari variabel terikat. Dapat juga diartikan sebagai proporsi varians variabel terikat yang dijelaskan oleh variabel bebas. Jika variabel independen (digunakan untuk memprediksi variabel dependen) bersifat ortogonal berpasangan, maka koefisien korelasi berganda sama dengan jumlah koefisien korelasi kuadrat antara masing-masing variabel independen dan variabel dependen. Hubungan ini tidak berlaku jika variabel bebasnya tidak ortogonal. Signifikansi koefisien korelasi berganda dapat dinilai dengan rasio F. Besarnya koefisien korelasi berganda cenderung melebih-lebihkan besarnya hubungan korelasi populasi, namun perkiraan yang terlalu tinggi ini dapat dikoreksi. Sebenarnya kita harus menyebut koefisien ini sebagai koefisien korelasi berganda kuadrat, namun penggunaan saat ini sepertinya mengabaikan kata sifat "kuadrat", mungkin karena sebagian besar nilai kuadratnya yang dipertimbangkan.

9.2. Konsep Dasar Korelasi

Korelasi adalah metode statistik yang digunakan untuk menentukan apakah ada hubungan antar variable. Korelasi menyatakan derajat hubungan antara dua variabel tanpa memperhatikan variabel mana yang menjadi peubah. Karena itu hubugan korelasi belum dapat dikatakan sebagai hubungan sebab akibat. Koefisien korelasi merupakan ukuran seberapa besar hubungan antara dua variable atau lebih. Misapnya jika informasi tentang dua variabel seperti tinggi dan berat badan, pendapatan dan pengeluaran, permintaan dan penawaran, tersedia dan kita ingin mempelajari hubungan linier antara dua variabel, koefisien korelasi dapat memberikan besar atau derajat hubungan linier dengan arah apakah itu positif atau negatif. Namun dalam ilmu biologi, fisika, dan sosial, sering kali tersedia data mengenai lebih dari dua variabel dan nilai satu variabel tampaknya dipengaruhi oleh dua variabel atau lebih. Misalnya, kejahatan di suatu kota mungkin dipengaruhi oleh yang tidak berpendidkan, meningkatnya jumlah penduduk dan pengangguran di kota tersebut, dll. Produksi suatu tanaman mungkin bergantung pada jumlah curah hujan, kualitas benih, jumlah pupuk yang digunakan dan metode irigasi, dll. Demikian pula, prestasi mahasiswa dalam ujian di kampusmungkin bergantung pada IQ-nya, kualifikasi ibu, kualifikasi ayah, pendapatan orang tua, jumlah jam belajar, dll. Dalam mempelajari pengaruh antara dari dua variabel atau lebih pada satu variabel. Variable, korelasi ganda memberikan solusi untuk masalah ini. Koefisien korelasi yang memberikan derajat hubungan antara kedua variabel tersebut. Jika kita memiliki lebih dari dua variabel yang saling berhubungan dengan variabel lain, maka kita dapat mengetahui hubungan antara satu variabel dan himpunan yang lain. Hal ini merupakan kasus korelasi ganda.

Rumus yang dapat digunakan untuk menghitung harga R pada korelasi ganda yang digambarkan melalui paradigma di atas adalah:

$R = \sqrt{\frac{r_1^2 + r_1^2 - 2r_1 r_2 r_3}{1 - r_3^2}}$ (dalam Elfira Rahmadani, dkk., 2023)

Keterangan:

R= Nilai Koefisien Korelasi Ganda dan secara simultan terhadap Y

r_1= Nilai Koefisien Korelasi X_1 dan Y

r_2= Nilai Koefisien Korelasi X_2 dan Y

r_3= Nilai Koefisien Korelasi X_3 dan Y

X= Data Variabel Independen (Bebas)

Y= Data Variabel Dependen (Terikat)

9.2.1 Kovarians

Definisi 1:

Kovariansi (sampel) antara dua sampel $\{x_1, \ldots, x_n\}$ dan $\{y_1, \ldots, y_n\}$ adalah ukuran hubungan linier antara dua variabel x dan y berdasarkan sampel yang bersesuaian, dan ditentukan rumus,

$$cov(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

Kovariansi mirip dengan varians, hanya saja kovarians didefinisikan untuk dua variabel (x dan y di atas) sedangkan varians didefinisikan hanya untuk satu variabel.

Faktanya,

cov(x, x) = var(x).

Kovariansi dapat dianggap sebagai jumlah kecocokan dan ketidakcocokan di antara pasangan elemen data untuk x dan y: kecocokan terjadi ketika kedua elemen dalam pasangan tersebut berada pada sisi mean yang sama; ketidakcocokan terjadi ketika salah satu elemen dalam pasangan berada di atas rata-ratanya dan elemen lainnya berada di bawah rata-ratanya.

Kovariansinya positif jika kecocokannya lebih besar daripada ketidakcocokannya dan negatif jika ketidakcocokannya lebih besar

daripada kecocokannya. Besar kecilnya kovarians dalam nilai absolut menunjukkan intensitas hubungan linier antara x dan y: semakin kuat hubungan linier maka semakin besar nilai kovariansnya.

9.2.2 Korelasi sampel

Besar kecilnya kovarians dipengaruhi oleh skala elemen data, sehingga untuk menghilangkan faktor skala, koefisien korelasi digunakan sebagai metrik hubungan linier yang bebas skala.

Definisi 2:

Koefisien korelasi (sampel) antara dua sampel adalah ukuran hubungan linier tanpa skala dan diberikan rumus,

$$r = \frac{cov(x,y)}{s_x s_y}$$

Dimana r sebagai rxy untuk menampilkan kedua variabel secara eksplisit. Koefisien determinasi adalah r^2.

Varians dalam Ukuran Variabilitas, kovarians dapat dihitung sebagai berikut,

$$\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \frac{1}{n-1}(\sum_{i=1}^{n}x_i y_i - \bar{x}\sum_{i=1}^{x}y_i - \bar{y}\sum_{i=1}^{n}x_i + n\bar{x}\bar{y}) = \frac{1}{n-1}(\sum_{i=1}^{n}x_i y_i - n\bar{x}\bar{y})$$

Hasilnya, kita dapat menghitung koefisien korelasi sebagai berikut,

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

9.2.3 Kovarian dan korelasi populasi

Kovarians dan koefisien korelasi untuk sampel data. Kita dapat menentukan kovarians dan koefisien korelasi populasi, berdasarkan fungsi kepadatan probabilitasnya (pdf). Menurut Hogg, Robert V, Allen T. (dalam Joni Wilson Sitopu dan siswadi, 2022), Salah satu bentuk ekspektasi matematika lain adalah "fungsi

pembangkit momen (moment generating function)" yang merupakan ekspektasi matematika khusus, yang disingkat dengan "fpm"

Definisi 3:

Kovariansi populasi antara dua variabel acak x dan y untuk populasi dengan pdf diskrit atau kontinu ditentukan oleh

$cov(x, y) = E[x - \pi_x)(y - \mu_y)]$ (dalam Joni Wilson Sitopu, 2023)

dimana E[ ] adalah fungsi ekspektasi.

Definisi 4:

Koefisien korelasi (momen produk Pearson) untuk dua variabel acak x dan y untuk populasi dengan pdf diskrit atau kontinu adalah

$$\rho = \frac{cov(x, y)}{\sigma_x \sigma_y}$$

9.3 Koefisien Korelasi Berganda

Koefisien korelasi berganda (R) menghasilkan derajat hubungan liner maksimum yang dapat diperoleh antara dua atau lebih variabel bebas dan satu variabel terikat. (R tidak pernah diberi tanda + atau −. R^2 mewakili proporsi total varians dalam variabel terikat yang dapat diperhitungkan oleh variabel bebas.) Masing-masing variabel bebas diberi bobot optimal sehingga gabungannya akan mempunyai korelasi sebesar mungkin dengan variabel dependen. Karena penentuan bobot ini (koefisien beta), seperti statistik lainnya, selalu dipengaruhi (R selalu meningkat) oleh kesalahan pengambilan sampel, kelipatan R "dikecilkan" dengan tepat untuk mengoreksi bias yang disebabkan oleh kesalahan pengambilan sampel. Penyusutan R didasarkan pada jumlah variabel independen dan ukuran sampel. Jika jumlah variabel independen kecil (<10) dan ukuran sampel besar (>100), prosedur penyusutan mempunyai pengaruh yang dapat diabaikan. Selain itu, korelasi antar variabel independen yang masuk ke dalam perhitungan R dapat dikoreksi untuk atenuasi (kesalahan pengukuran), yang meningkatkan R. Selanjutnya, R dapat dikoreksi

untuk pembatasan rentang kemampuan dalam sampel tertentu jika variansnya pada variabel-variabel yang dimasukkan ke dalam R berbeda secara signifikan dari varians populasi, dengan asumsi bahwa variabel tersebut dapat diestimasi secara memadai. Koreksi korelasi untuk pembatasan rentang sering digunakan dalam penelitian yang didasarkan pada seleksi mahasiswa di perguruan tinggi, karena korelasi tersebut biasanya hanya mewakili setengah atas distribusi IQ pada populasi umum.

Definisi 5:

Dengan adanya variabel x, y, dan z, dapat kita definisikan koefisien korelasi berganda,

$$R_{z.xy} = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2r_{xz}r_{yz}r_{xy}}{1 - r_{xy}^2}}$$

dimana rxz, ryz, rxy sebagaimana didefinisikan dalam Definisi 2 adalah Konsep Dasar Korelasi. Dimana x dan y adalah sebagai variabel bebas dan z sebagai variabel terikat.

9.4 Sifat Koefisien Korelasi Berganda

9.4.1 Sifat Korelasi

Sifat 1

Jika r mendekati 1 (-1 ≤ r ≤1) maka x dan y berkorelasi positif. Jika r korelasi linier positif, maka nilai x yang tinggi berhubungan dengan nilai y yang tinggi, dan nilai x yang rendah berhubungan dengan nilai y yang rendah.

Sifat 2

Jika r mendekati -1 maka x dan y berkorelasi negatif. Jika r korelasi linier negatif maka nilai x yang tinggi berhubungan dengan nilai y yang rendah, dan nilai x yang rendah berhubungan dengan nilai y yang tinggi.

Jika r mendekati 0, maka terdapat sedikit hubungan linier antara x dan y. (-1 ≤ ρ ≤1)

9.4.2 Sifat Kovarian

Sifat 1

Untuk setiap konstanta a dan variabel acak x, y, z, hal berikut ini berlaku untuk definisi kovarians sampel dan populasi.

1. *cov(x, y) = cov(y, x)*
2. *cov(x, x) = var(x)*
3. *cov(a, y) = 0*
4. *cov(ax, y) = acov(x, y)*
5. *cov(x+z, y) = cov(x, y) + cov(z, y)*

Sifat 2

Berikut ini berlaku untuk definisi sampel dan populasi kovarians :

Jika x dan y saling bebas maka cov(x, y) = 0, sehingga koefisien korelasinya juga nol.

Kebalikan dari Sifat ini tidak benar, seperti yang ditunjukkan oleh contoh berikut.

Contoh,

Pertimbangkan variabel acak pada domain {-1,0,1} di mana P(x=-1) = P(x=1) = .5 dan P(x=0) = 0. : maka, tentukan variabel acak $y = x^2$.

Solusi,

diketahui x dan y tidak independen. Jika x diketahui, maka y dapat ditentukan. Sehingga mudah untuk menghitung bahwa korelasi antara x dan y adalah nol (lihat Gambar 1)

	A	B	C	D	E	F	G	H
3		x	y					
4	1	-1	1		cov(x,y)	0		=COVAR(B4:B6,C4:C6)
5	2	0	0					
6	3	1	1		corr(x,y)	0		=CORREL(B4:B6,C4:C6)

Gambar 9.1 – Contoh

Ternyata jika x dan y bivariat berdistribusi normal dan tidak berkorelasi (yaitu cov(x, y) = 0) maka x dan y saling bebas.

Sifat 3

Hal berikut ini berlaku untuk sampel dan populasi:

$$var(x - y) = var(x) + var(y) - 2cov(x, y)$$

Estimasi yang Tidak Bias

Ternyata r bukanlah estimasi ρ yang tidak bias. Estimasi ρ^2 yang umum digunakan dan relatif tidak bias diberikan oleh koefisien r^2_{adj} determinasi yang disesuaikan :

$$r^2_{adj} = 1 - \frac{(1 - r^2)(n - 1)}{n - 2}$$

Meskipun r^2_{adj} merupakan pendugaan koefisien determinasi populasi yang lebih baik, terutama untuk nilai n yang kecil, untuk nilai n yang besar mudah untuk melihat bahwa $r^2_{adj} = r^2$. Perhatikan bahwa $r^2_{adj} \leq r^2$, dan walaupun r^2_{adj} bisa negatif, hal ini relatif jarang terjadi.

Estimasi yang lebih tidak bias dari koefisien korelasi populasi yang terkait dengan data yang terdistribusi normal diberikan rumus,

$$\rho_{est} = r\left[1 + \frac{(1 - r^2)}{2(n - 3)}\right]$$

9.5 Koefisien Determinasi

Definisi koefisien determinasi berganda sebagai kuadrat dari koefisien korelasi berganda. Seringkali subskrip dihilangkan dan koefisien korelasi berganda dan koefisien determinasi berganda masing-masing ditulis sebagai R dan R^2. Definisi-definisi ini juga dapat diperluas ke lebih dari dua variabel independen. Dengan hanya satu variabel independen, koefisien korelasi berganda adalah r. Dimana, R bukanlah estimasi koefisien korelasi berganda populasi yang tidak bias, seperti pada sampel kecil. Versi R yang relatif tidak bias diberikan oleh R berikut ini.

Definisi 2: Jika R adalah Rz,xy seperti yang didefinisikan di atas (atau serupa untuk variabel lebih banyak) maka koefisien determinasi berganda yang disesuaikan adalah

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1}$$

dimana k = banyaknya variabel bebas dan n = banyaknya elemen data dalam sampel untuk z (yang harus sama dengan sampel untuk x dan y).

9.6 Fungsi Aplikasi SPSS

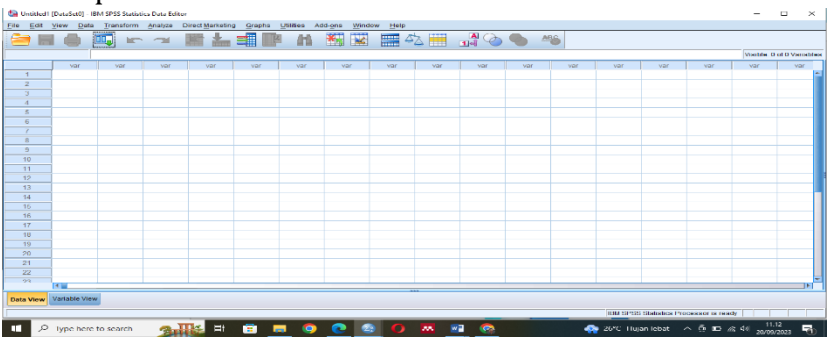
SPSS adalah program berbasis Windows yang dapat digunakan untuk melakukan entri dan analisis data serta membuat tabel dan grafik. SPSS mampu menangani data dalam jumlah besar dan dapat melakukan semua analisis yang tercakup dalam banyak teks. (dalam Joni Wilson Sitopu, dkk.,2023)

SPSS merupakan software aplikasi statistik yang dapat membantu pengolahan data dan pengujian hipotesis untuk berbagai uji dan analisis dalam statistika, seperti uji t, uji F, uji-uji non parametrik, analisis regresi kolerasi, dan analisis multivariat dan lain-lain.(dalam Joni Wilson Sitopu., dkk., 2021).

Langkah-langkah dengan SPSS 21 (dalam I Gede Purnawinadi, Etriana Meirista, dkk, 2023).

Langkah-langgkah mencari koefisien korelasi dengan SPSS

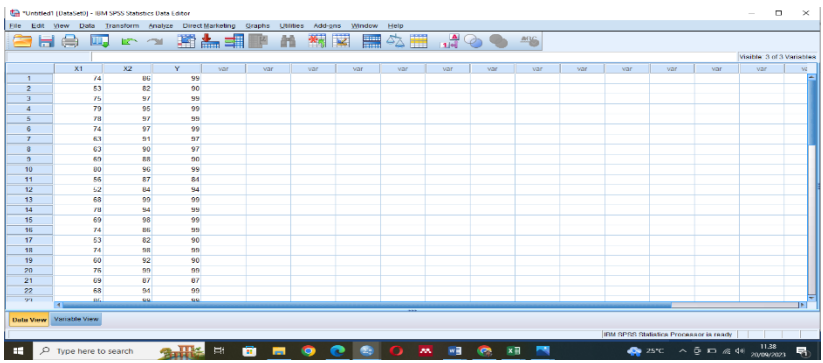
1. Buka aplikasi SPSS



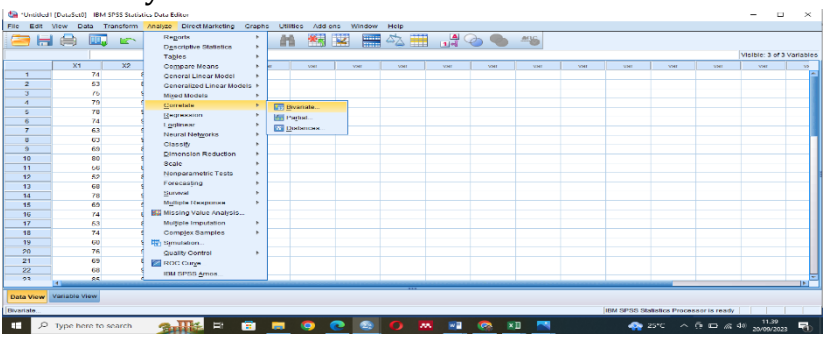
2. Masukan data X1,X2,Y

X1	X2	Y
74	86	99
53	82	90
75	97	99
79	95	99
78	97	99
74	97	99
63	91	97
63	90	97
69	88	90
80	96	99
56	87	84
52	84	94
68	99	99
78	94	99
69	98	99
74	86	99

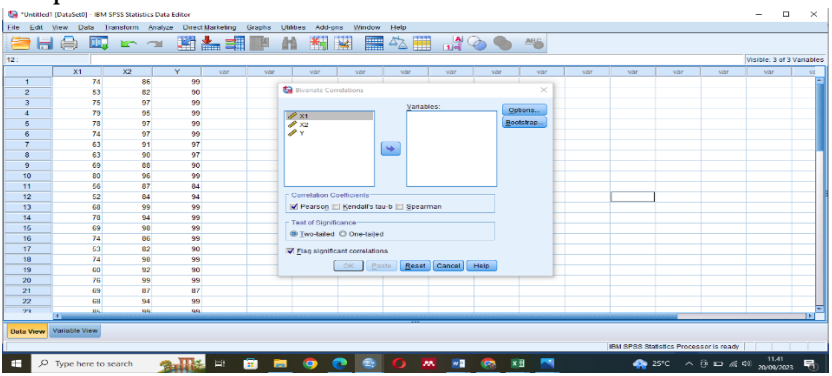
53	82	90
74	98	99
60	92	90
76	99	99
69	87	87
68	94	99
85	99	99
63	91	97
63	90	97
69	88	90
80	96	99
56	87	84
52	84	94
68	99	99
78	94	99
69	98	99
74	98	99
60	92	90



3. Klik Analyze-corralate-Bivariate



4. Diperoleh kotak



5. Pindahkan X1,X2,Y dengan mengklik tanda panah ke variabel
   Kemudian pilih pearson dan flag significant correlations.
   Lalu klik OK

Tabel 9.1 Correlations

	X1	X2	Y
Pearson Correlation	1	,704**	**0,672****
Sig. (2-tailed)		,000	,000
N	34	34	34
Pearson Correlation	,704**	1	,666**
Sig. (2-tailed)	,000		,000
N	34	34	34
Pearson Correlation	**0,672****	,666**	1
Sig. (2-tailed)	,000	,000	
N	34	34	34

**. Correlation is significant at the 0.01 level (2-tailed).

Sumber : Pengolahan data SPSS, 2023

6. Dari tabel 1 diperoleh koefisien korelasi (r) = 0,672.
7. Determininasi = R^2 x 100%= 67,2%.

DAFTAR PUSTAKA

Elfira Rahmadani., dkk., (2023). STATISTIKA PENDIDIKAN. Padang. PT Get Press.

Joni Wilson Sitopu., dkk., 2021). Pelatihan Pengolahan Data Statistik Dengan Aplikasi SPSS. *Jurnal Pengabdian Masyarakat*. Dedikasi Sains dan Teknologi

Joni Wilson Sitopu dan siswadi, (2022). Fungsi Pembangkit Momen dari Distribusi Probabilitas Diskrit. *FARABI. Jurnal Matematika dan Pendidikan Matematika*. UNIVA Medan.

Joni Wilson Sitopu, dkk, (2023). *Teori Peluang.* Padang. PT Get Press.

Joni Wilson Sitopu, (2023). *Bab 1 Konsep Teori Peluang*. Padang. PT Get Press.

Joni Wilson Sitopu, dkk, (2023). APLIKASI SPSS UNTUK ANALISIS DATA PENELITIAN KESEHATAN. Padang. PT Get Press.

Nazariah, Noviyanti, Dita Kurniawati, Roni Priyanda, Khairul Alim, Joni Wilson Sitopu, Joko Sabtohadi, Nurul Hidayah, P.A. (2022) 'Statistik Dasar', in statistik dasar. 1st edn. PT Get Press.

BAB 10
ANALISIS REGRESI

Oleh Laelatul Khikmah

10.1 Pendahuluan

Setelah Anda mempelajari terkait Uji Hipotesis dan analisis korelasi pada bab sebelumnya, maka pada Bab 10 ini kita akan mengenal model regresi linier sebagai alternatif pengambilan keputusan dan kesimpulan dengan pendekatan pemodelan.

Analisis regresi adalah metode dalam analisis statistik yang digunakan untuk menyelidiki pengaruh antar variabel. Sebagai contoh penilai real estat ingin mengetahui pengaruh karakteristik fisik bangunan yang dipilih dan besarnya pajak terhadap harga jual rumah. Kita mungkin ingin memeriksa apakah konsumsi rokok dipengaruhi oleh variabel sosial ekonomi dan demografi (umur, pendidikan, pendapatan, dan harga rokok). Pengaruh tersebut dibuat dalam bentuk persamaan atau model antara satu variabel dependen dan satu atau lebih variabel independen. Dalam konsumsi rokok misalnya, variabel dependen adalah konsumsi rokok (diukur dengan angka bungkus rokok dan variabel independennya adalah berbagai sosio ekonomi dan variabel demografis. Dalam contoh penilaian real estate, variabel dependennya adalah harga jual rumah dan variabel independennya adalah karakteristiknya bangunan dan pajak yang dibayarkan atas bangunan tersebut (Gordon, 2015). Dalam Bab ini kita akan lebih banyak belajar terkait analisis regresi linier sederhana.

10.2 Regresi Linier Sederhana

Regresi linier sederhana adalah metode statistik yang digunakan untuk menganalisis hubungan antara dua variabel, yaitu variabel dependen dan variabel independen. Tujuan utama dari

regresi linier sederhana adalah untuk memahami dan memodelkan bagaimana pengaruh variabel independen terhadap variabel dependen (Chatterjee and Simonoff, 2013).

Berikut adalah konsep dasar yang harus diketahui dan dipahami dari regresi linier sederhana:

1. Variabel Dependensi (Y): Variabel ini adalah variabel yang ingin kita prediksi atau modelkan. Variabel Y ini biasanya disebut juga sebagai variabel dependen, variabel respon, atau variabel terikat.
2. Variabel Independen (X): Variabel ini adalah variabel yang digunakan untuk memprediksi variabel dependen. Variabel X biasanya disebut juga sebagai variabel independen, variabel penjelas, atau variabel bebas.
3. Skala pengukuran yang digunakan pada variabel Y adalah skala pengukuran numerik. Sedangkan skala pengukuran pada variabel X adalah skala pengukuran numerik atau kategori.
4. Tujuan umum analisis regresi linier sederhana adalah untuk mengetahui pengaruh satu variabel X yang berskala numerik atau kategorik terhadap satu variabel Y yang berskala numerik.

Regresi linier sederhana dapat diterapkan dalam berbagai bidang dan kasus penggunaan untuk menganalisis dan memodelkan pengaruh variabel X terhadap variabel Y. Di bawah ini beberapa beberapa contoh penerapan regresi linier sederhana:

1. Kesehatan
   Regresi linier sederhana untuk memprediksi kunjungan pasien di rumah sakit berdasarkan jenis layanan dan umur pasien (Wiga Maulana Baihaqi, Melia Dianingrum, 2019), Penerapan Metode Regresi Linier Sederhana Untuk Prediksi

Persediaan Obat Jenis Tablet (Harsiti, Muttaqin and Srihartini, 2022).

2. Ekonomi

   Penerapan Data Mining untuk Prediksi Penjualan Produk Sepatu Terlaris Menggunakan Metode Regresi Linier Sederhana (Pohan, Dar and Irmayanti, 2022), Penerapan Metode Regresi Linier Sederhana dalam Memprediksi Jumlah Kebutuhan Ekspor Migas dan Non-Migas di Indonesia (Revaldi *et al.*, 2023).

3. Pertanian

   Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang (Padilah and Adam, 2019), Pengendalian Suhu dan Kelembapan Udara untuk Budidaya Microgreen Lobak menggunakan Metode Regresi Linier berbasis Arduino (Chan, Fitriyah and Widasari, 2023).

4. Pendidikan

   Analisis Regresi Linier Sederhana untuk Mengestimasi Pengaruh Kemampuan Self Regulated Learning terhadap Hasil Belajar Siswa Menggunakan Model Pembelajaran Rasi (Yustika, Sudarti and Handayani, 2022), Analysis of the Use of Google Classroom Applications for Simple Linear Regression Materials for Actuarial Study Program Students (Muchlian and Z, 2022).

5. Peramalan Cuaca

   Penentuan Batas Atas Normal dan Bawah Normal Curah Hujan Bulanan Setara Tercile dengan Koefisien Regresi Linier Sederhana (Muharsyah, 2014), Pemanfaatan Suhu Udara Dan Kelembapan Udara Dalam Persamaan Regresi

Untuk Simulasi Prediksi Total Hujan Bulanan Di Bandar Lampung (Swarinoto and Sugiyono, 2011).

10.2.1 Model Umum Regresi Linier Sederhana

Secara umum model regresi linier sederhana adalah:

$$Y = \alpha + \beta X + \varepsilon$$

dimana:

Y : adalah variabel dependen

α : adalah parameter regresi (intersep), dalam pemodelan regresi intersep merupakan nilai dugaan rataan y ketika x bernilai nol (jika x = 0 dalam selang pengamatan).

β : adalah parameter regresi (slope/ kemiringan), merupakan nilai dugaan perubahan rataan y (nilai harapan Y) jika x berubah satu satuan.

X : variabel independen

ε : error/ residual (Gordon, 2015).

10.2.2 Metode Kuadrat Terkecil

Dengan mengetahui adanya hubungan antar variabel, maka dapat dilakukan pendugaaan suatu variabel berdasarkan variabel lain melalui persamaan yang dibuat atas hubungan tersebut. Pendugaan parameter regresi linier sederhana adalah proses untuk memperkirakan nilai koefisien regresi (α dan β) dalam persamaan regresi linier sederhana. Ada 2 metode yang umum digunakan dalam menduga parameter regresi, yaitu metode kemungkinan maksimum atau *maximum likelihood method* dan Metode Kuadrat Terkecil atau *ordinary least square* sering disebut dengan OLS. Menduga persamaan regresi linier sederhana sama dengan menduga parameter-parameter regresi α dan β. Dua Metode tersebut memiliki sifat-sifat yang mampu menghasilkan penduga yang baik berdasarkan beberapa asumsi sehingga menjadikannya salah satu metode yang sering digunakan untuk mengestimasi parameter-parameter dalam persamaan regresi. Pendugaan

parameter α dan β yang sering digunakan pada analisis regresi linier sederhana menggunakan pendekatan Metode Kuadrat Terkecil atau sering disebut dengan metode *Ordinary Least Square* (OLS) (Chatterjee and Hadi, 2012).

Penduga bagi koefisien kemiringan garis α adalah:

$$\beta = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Penduga bagi intersep α adalah:

$$\alpha = \bar{y} - \beta\bar{x}$$

10.2.3 Asumsi dalam Regresi Linier Sederhana

Analisis regresi linier sederhana bergantung pada beberapa asumsi klasik yang harus dipenuhi untuk hasil yang akurat dan dapat diandalkan. Asumsi-asumsi ini adalah sebagai berikut:

1. Error/ residual berdistribusi normal, dengan nilai rata-rata nol dan ragam σ^2 atau disimbolkan :
$$\varepsilon_i \sim N(0, \sigma^2)$$
2. Tidak Terdapat Autokorelasi
Uji autokorelasi bertujuan untuk menguji apakah model regresi linier ada korelasi antara residual pengganggu pada periode t dengan residual pengganggu pada periode sebelumnya (t-1).
3. Asumsi Linearitas
Asumsi ini menyatakan bahwa hubungan antara variabel independen (X) dan variabel dependen (Y) adalah linier. Dalam regresi linier sederhana, ini berarti bahwa garis regresi adalah representasi yang tepat dari hubungan antara X dan Y.

4. Residual Bersifat Homokedastisitas

   Salah satu asumsi penting regresi linier sederhana adalah bahwa varian dari galat untuk variabel-variabel bebas yang diketahui bersifat konstan dengan istilah homokedastisitas.

10.2.4 Pengujian Hipotesis dalam Regresi Linier Sederhana

Pengujian hipotesis adalah salah satu komponen penting dalam analisis regresi linier sederhana. Ini digunakan untuk menguji signifikansi statistik dari koefisien regresi (α dan β) serta mengevaluasi sejauh mana model regresi cocok dengan data. Berikut adalah beberapa jenis pengujian hipotesis yang umum dalam regresi linier sederhana (Chatterjee and Hadi, 2012):

1. **Pengujian parameter α**

   Hipotesis:

   $H_0 : \alpha = 0$ (semua nilai Y dapat dijelaskan oleh x)

   $H_1 : \alpha \neq 0$ (ada nilai Y yg tidak dapat dijelaskan oleh x)

   Statistik uji yang digunakan adalah uji t.

$$t = \frac{a - \alpha}{s_a}$$

   dengan:

   a = intersep garis regresi

   α = intersep yg dihipotesiskan

   s_a = dugaan simp. baku intersep

   Kriteria tolak H_0, apabila nilai t > t_{tabel} atau p-value < taraf nyata (α)

2. **Pengujian parameter β**

   Hipotesis:

   $H_0 : \beta = 0$ (tidak ada pengaruh x terhadap Y)

   $H_1 : \beta \neq 0$ (ada pengaruh x terhadap Y)

   Statistik uji yang digunakan adalah uji t.

$$t = \frac{b - \beta}{s_b}$$

dengan:

b = koefisien kemiringan regresi

β = kemiringan yg dihipotesiskan

s_b = simpangan baku kemiringan

Kriteria tolak H_0, apabila nilai t > t_{tabel} atau p-value < taraf nyata (α)

10.2.5 Analisis Keragaman Regresi Linier Sederhana

Analisis keragaman adalah suatu metode untuk menguraikan keragaman total data menjadi komponen–komponen yang mengukur berbagai sumber keragaman. Analisis keragaman bertujuan untuk melihat apakah variabel independen X benar-benar memberikan sumbangan yang berarti bagi variabel dependen Y (Chatterjee and Simonoff, 2013).

Tabel 10.1 Tabel Analisis Keragaman Regresi Linier Sederhana

Sumber Keragaman	Db	JK	KT	F_{hitung}
Regresi	1	JKR	KTR	KTR/KTG
Galat	n-2	JKG	KTG	
Total	n-1	JKT		

Sumber: (Nugroho, 2008)

10.2.6 Koefisien Determinasi (R^2)

Kelayakan model regresi dapat diukur dengan menggunakan koefisien determinasi (R^2) yang didefinisikan sebagai berikut (Chatterjee and Simonoff, 2013):

$$R^2 = \frac{JKR}{JKT}$$

Koefisien determinasi mengukur seberapa besar proporsi keragaman pada Y yang dapat dijelaskan oleh X. Koefesien Determinasi (R^2) merupakan besaran tidak negatif yang kisaran nilainya $0 \leq R^2 \leq 1$. Koefesien Determinasi (R^2) semakian

mendekati 1 berarti model yang dipakai semakin sempurna karena kesalahan (error) yang tidak dapat dikendalikan semakin kecil, sedangkan R^2 yang bernilai nol berarti tidak ada hubungan antara peubah tak bebas dengan peubah bebas.

10.3 Studi Kasus

Sebuah penelitian dilakukan untuk mengetahui pengaruh jumlah objek pajak (bidang) terhadap realisasi pajak bumi dan bangunan (Rupiah) pada tahun 2016 di Kabupaten Kendal. Diperoleh data sebagai berikut:

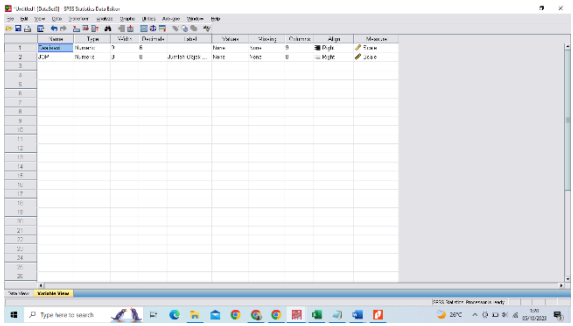
Tabel 10.2 Data

Realisasi (Y)	Jumlah Objek Pajak (X)
20,437537	1290
36,943647	2455
20,063949	1426
22,46589	1702
21,891546	1615
29,909558	1895
26,317196	2284
11,752992	1230
22,999254	1110
30,146747	3029
29,949571	2038
15,128165	728

Sumber: Dinas Pendapatan dan Pengelola Keuangan dan Aset Daerah (DPPKAD) Kabupaten Kendal

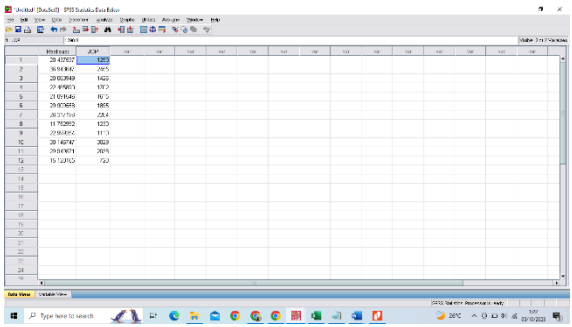
Untuk dapat menyelesaikan permasalahan diatas, akan dilakukan analisis regresi linier sederhana dengan menggunakan software SPSS sebagai berikut:

1. Definiskan variabel yang akan digunakan pada lembar kerja SPSS (***Variable View***), seperti pada gambar berikut ini:



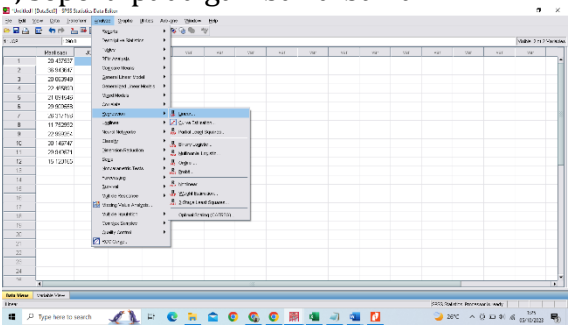
Gambar 10.1 Tampilan *Variable View* SPSS
(Sumber : SPSS 17)

2. Input data yang akan digunakan pada lembar kerja SPSS (***Data View***), seperti tampak pada gambar dibawah ini:



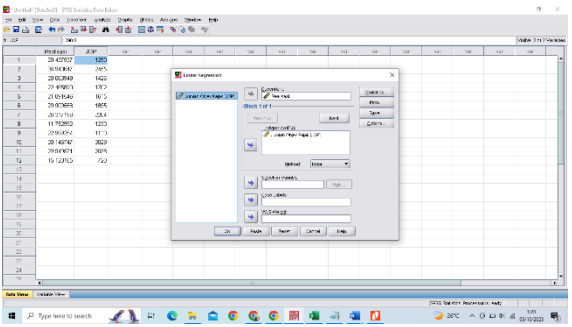
Gambar 10.2 Tampilan *Data View* SPSS
(Sumber : SPSS 17)

3. Klik Menu **Analyze**, pilih **Regression**, lalu pilih **Linier Regression**, seperti pada gambar dibawah ini:



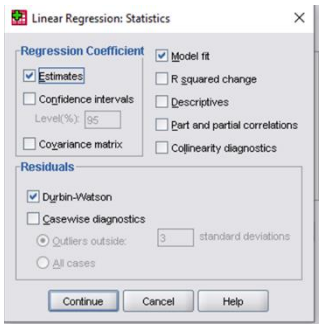
Gambar 10.3 Tampilan Menu Regression
(Sumber : SPSS 17)

4. Maka akan tampil kotak dialog Linear Regression. Isikan Realisasi pada kotak Dependent, dan Jumlah Objek Pajak pada kotak Independent. Seperti terlihat pada gambar dibawah ini:



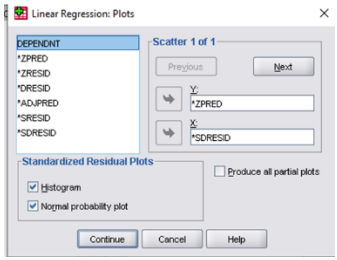
Gambar 10.4 Tampilan Kotak Dialog Linear Regression
(Sumber : SPSS 17)

5. Pilih **Statistics**, lalu beri tanda centang (√) pada pilihan Durbin-Watson, seperti terlihat pada gambar dibawah ini:



Gambar 10.5 Tampilan Menu *Statistics*
(Sumber : SPSS 17)

6. Pilih menu **Plots**, isikan "ZPRED" pada Y dan "SDRESID" pada X, seperti gambar dibawah ini:



Gambar 10.6 Tampilan Menu *Plots*
(Sumber : SPSS 17)

7. Lalu klik OK, maka kita akan peroleh hasil dari analisis regresi linier sederhana yang kita lakukan.

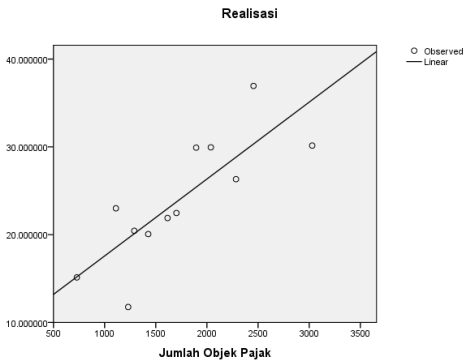
Hasil Analisis

1. Model Regresi yang diperoleh berdasarkan kasus diatas adalah:

$$Y = 8{,}800 + 0{,}009X + e$$

2. Pengujian Asumsi

a. Asumsi Linieritas

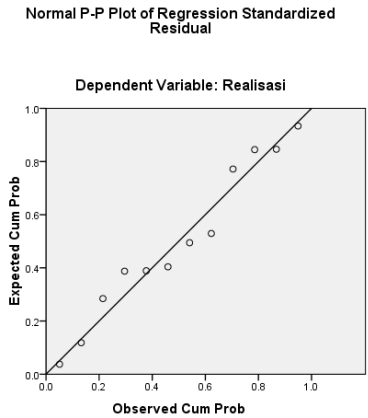


Gambar 10.7 Asumsi Linieritas
(Sumber : Output SPSS 17)

Berdasarkan Gambar diatas, dapat diketahui bahwa hubungan antara variabel Realisasi dan Jumlah Objek Pajak adalah linier. Sehingga analisis regresi linier sederhana dapat diterapkan pada studi kasus ini.

b. Asumsi Normalitas

Untuk melakukan pengujian asumsi Normalitas, dapat dilakukan melalui 2 tahap yaitu dengan menggunakan grafik atau dengan pengujian Kolmogorov-Smirnov dari nilai residual yang dihasilkan dari model. Secara Grafik dapat dilihat pada Gambar dibawah ini:



Gambar 10.8 Asumsi Normalitas
(Sumber : Output SPSS 17)

Berdasakan Gambar diatas, dapat terlihat bahwa data menyebar di sekitar garis linier regresi, sehingga dapat dikatakan bahwa residual yang dihasilkan mengikuti distribusi Normal atau asumsi normalitas terpenuhi. Untuk lebih meyakinkan maka dapat dianalisis dengan hasil Kolmogorov-Smirnov dibawah ini:

Tabel 10.3 Tabel Hasil Kolmogorov-Smirnov

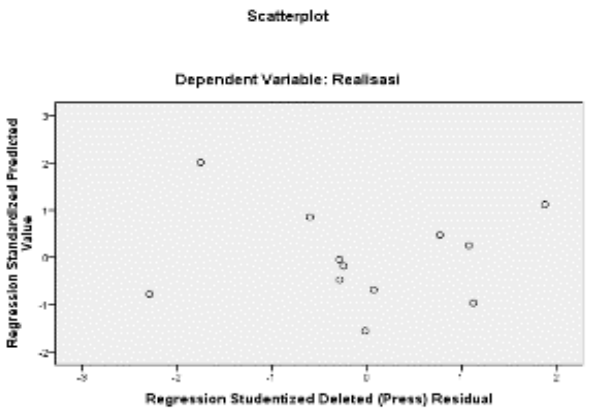
Keterangan	Nilai
Kolmogorov-Smirnov Z	0,471
Asymp. Sig. (2-tailed)	0,980

Sumber: Output SPSS 17

Berdasarkan Tabel diatas, dapat terlihat bahwa nilai p-value yang dihasikan sebesar 0,980 > taraf nyata 5% (0,05). Sehingga dapat dikatakan residual yang

dihasilkan mengikuti distribusi Normal atau asumsi normalitas terpenuhi.

c. Residual Bersifat Homokedastisitas



Gambar 10.9 Residual Bersifat Homokedastisitas
(Sumber : Output SPSS 17)

Berdasarkan Gambar diatas dapat dilihat bahwa data menyebar sehingga dapat dikatakan bahwa residual bersifat homoskedastisitas.

d. Asumsi Autokorelasi

3. Pengujian Hipotesis

Tabel 10.4 Hasil Uji t

Keterangan	Koefisien	t	Sig.
Konstanta	8,808	2,326	0,042
Jumlah Objek Pajak	0,009	4,260	0,002

Sumber: Output SPSS 17

Dari Tabel diatas dapat diketahui bahwa:

- Nilai intersep dan slope siginifikan pada taraf nyata 5% (0,05). Hal ini dapat dilihat bahwa nilai p-value pada intersep maupun slope kurang dari taraf nyata 5% (0,05).
- Nilai 8,808 menyatakan bahwa dugaaan rataan Realisasi PBB sebesar 8,808 Rupiah saat Jumlah Objek Pajak bernilai 0 bidang. Atau apabila seseorang tidak memiliki Objek Pajak tertentu dalam ukuran bidang, maka besarnya Realisasi PBB yang harus dibayarkan adalah Rp. 8,808,00.
- Nilai 0,009 emngidentifikasikan bahwa apabila Jumlah Objek Pajak naik 1 bidang, maka dugaan rataan Realisasi PBB akan naik sebesar 0,009 Rupiah.

4. Analisis Keragaman (uji F)

Tabel 10.5 Hasil Uji F

Model	Sum of Square	df	Mean Square	F	Sig.
Regresi	351,824	1	351,824	18,147	0,002
Residual	193,869	10	19,387		
Total	545,693	11			

Sumber: Output SPSS 17

Dari Tabel diatas dapat diketahui bahwa nilai p-value yang dihasilkan sebesar 0,002 atau kurang dari taraf nyata 5% (0,05). Sehingga dapat dikatakan bahwa variabel Jumlah Objek Pajak benar-benar memberikan sumbangan yang berarti bagi variabel Realisasi PBB.

5. Koefisien Determinasi

Tabel 10.6 Hasil Koefisien Determinasi

Model	R	R square	Adjusted R square
1	0,803	0,645	0,609

Sumber: Output SPSS 17

Berdasarkan Tabel diatas, dapat diketahui bahwa besarnya R^2 adalah 0,645 atau 64,5%. Artinya bahwa keragaman variabel Realisasi PBB mampu dijelaskan oleh variabel variabel Jumlah Objek Pajak. Selebihnya 35,5% dijelaskan oleh variabel lain diluar Model. Atau dapat juga dianalisis sebagai keragaman variabel Realisasi PBB dipengaruhi oleh variabel variabel Jumlah Objek Pajak. Selebihnya 35,5% dipengaruhi oleh variabel lain diluar Model.

DAFTAR PUSTAKA

Chan, R.M., Fitriyah, H. and Widasari, E.R. (2023) 'Pengendalian Suhu dan Kelembapan Udara untuk Budidaya Microgreen Lobak menggunakan Metode Regresi Linier berbasis Arduino', 7(5), pp. 2534–2541.

Chatterjee, S. and Hadi, A.S. (2012) *Regression Analysis by Example*. Wiley-Interscience.

Chatterjee, S. and Simonoff, J.S. (2013) *Handbook of Regression Analysis*, *Handbook of Regression Analysis*. Available at: https://doi.org/10.1002/9781118532843.

Gordon, R.A. (2015) *Regression analysis for the social sciences: Second edition*, *Regression Analysis for the Social Sciences: Second Edition*. Available at: https://doi.org/10.4324/9781315748788.

Harsiti, Muttaqin, Z. and Srihartini, E. (2022) 'Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet', *JSiI (Jurnal Sistem Informasi)*, 9(1), pp. 12–16. Available at: https://doi.org/10.30656/jsii.v9i1.4426.

Muchlian, M. and Z, Y.R. (2022) 'Analysis of the Use of Google Classroom Applications for Simple Linear Regression Materials for Actuarial Study Program Students', *Mathline : Jurnal Matematika dan Pendidikan Matematika*, 7(2), pp. 208–218. Available at: https://doi.org/10.31943/mathline.v7i2.249.

Muharsyah, R. (2014) 'Penentuan Batas Atas Normal dan Bawah Normal Curah Hujan Bulanan Setara Tercile dengan Koefisien Regresi Linier Sederhana', *Jurnal Meteorologi Dan Geofisika*, 15(1), pp. 59–69.

Nugroho, S. (2008) *Dasar – Dasar Metode Statistika*. Bengkulu: PT. Gramedia Widiasarana Indonesia.

Padilah, T.N. and Adam, R.I. (2019) 'Analisis Regresi Linier Berganda Dalam Estimasi Produktivitas Tanaman Padi Di Kabupaten Karawang', *FIBONACCI: Jurnal Pendidikan Matematika dan Matematika*, 5(2), p. 117. Available at: https://doi.org/10.24853/fbc.5.2.117-128.

Pohan, D.A., Dar, M.H. and Irmayanti (2022) 'Penerapan Data Mining untuk Prediksi Penjualan Produk Sepatu Terlaris Menggunakan Metode Regresi Linier Sederhana', *Jurnal Nasional Informatika dan Teknologi Jaringan*, 2, pp. 2–6. Available at: https://doi.org/10.30743/infotekjar.v6i2.4795.

Revaldi, K.A. *et al.* (2023) 'Penerapan Metode Regresi Linier Sederhana dalam Memprediksi Jumlah Kebutuhan Ekspor Migas dan Non-Migas di Indonesia', *Seminar Nasional Teknik Elektro, Sistem Informasi, dan Teknik Informatika*, (November 2022), pp. 268–275. Available at: https://ejurnal.itats.ac.id/snestikdanhttps://snestik.itats.ac.id.

Swarinoto, Y.S. and Sugiyono, S. (2011) 'Pemanfaatan Suhu Udara Dan Kelembapan Udara Dalam Persamaan Regresi Untuk Simulasi Prediksi Total Hujan Bulanan Di Bandar Lampung', *Jurnal Meteorologi dan Geofisika*, 12(3), pp. 271–281. Available at: https://doi.org/10.31172/jmg.v12i3.109.

Wiga Maulana Baihaqi, Melia Dianingrum, K. aswin N. ramadhan (2019) 'Regresi linier sederhana untuk memprediksi kunjungan pasien di rumah sakit berdasarkan jenis layanan dan umur pasien', *Simetris*, 10(2), pp. 671–680.

Yustika, D., Sudarti and Handayani, R.D. (2022) 'Analisis Regresi Linier Sederhana untuk Mengestimasi Pengaruh Kemampuan Self Regulated Learning terhadap Hasil Belajar Siswa Menggunakan Model Pembelajaran Rasi', *Jurnal Pendidikan MIPA*, 12(September), pp. 682–689.

BAB 11
ANALISIS REGRESI LINIER BERGANDA

Oleh Siti Ulfa Nabila

11.1 Pendahuluan

Istilah Regresi sudah dikenal sejak tahun 1886 oleh Francis Galton dimana ditemukan kecenderungan orang tua bertubuh tinggi akan memiliki anak dengan tubuh yang tinggi dan juga sebaliknya (Ghozali, 2005). Konsep tersebut dikaji oleh banyak peneliti seperti Karl Pearson dan A. Lee (Suliyanto, 2011).

Analisis regresi adalah metodologi statistik yang sangat berharga dalam upaya memahami korelasi antara beragam variabel dan dalam usaha meramalkan nilai suatu variabel berdasarkan variabel lainnya. Dalam ranah analisis regresi, kami mendefinisikan dua jenis variabel yang sangat penting:

1. Variabel terikat, yang merupakan variabel yang menjadi subjek prediksi utama.
2. Variabel bebas atau prediktor adalah variabel yang diasumsikan memengaruhi variabel terikat.

Melalui pendekatan analisis regresi yang cermat, kami dapat mengungkap pola dan relasi yang rumit antara variabel-variabel tersebut. Hal ini berperan penting dalam pembentukan model yang kuat untuk tujuan meramalkan atau menjelaskan hubungan yang kompleks di antara variabel-variabel tersebut.

11.2 Analisis Regresi Linear Berganda

Analisis regresi adalah metode statistika yang digunakan untuk menentukan hubungan antara variabel terikat, dengan satu atau lebih variabel bebas. Tujuannya adalah untuk memahami

sejauh mana variasi dalam variabel dependen dapat dijelaskan oleh variasi dalam variabel independen. Konsep ini diperkenalkan oleh Gujarati pada tahun 2006 dan telah menjadi alat penting dalam penelitian statistika.

Rumus umum regresi linear berganda,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i,$$
$$i = 1,2,3, \ldots, n \text{ dan } j = 1,2,3, \ldots, k$$

Y_i = variabel tak bebas yang akan diduga
X_{ji} = variabel bebas
β_0 = nilai konstanta (*intercept)* dari model
β_j = nilai koefisien regresi
ε_i = galat pengamatan dari model

Persamaan linear dari pendugaan garis regresi linear ditulis dalam bentuk:

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki}$$
$$i = 1,2, \ldots, n \text{ dan } j = 1,2,3, \ldots, k$$

$\hat{y}_i$ = nilai dugaan dari variabel tak bebas *Y*
x_{ji} = variabel bebas
b_0 = nilai dugaan konstanta (*intercept)* dari model
b_k = nilai dugaan koefisien regresi

Apabila dibatasi pada kasus dua variabel bebas X_1 dan X_2. Maka bentuk model regresi menjadi

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \varepsilon_i$$

Nilai dugaan b_0, b_1, dan b_2 dapat diperoleh dengan memecahkan persamaan linear simultan

$$nb_0 + b_1 \sum_{i=1}^{n} x_{1i} + b_2 \sum_{i=1}^{n} x_{2i} = \sum_{i=1}^{n} y_i$$
$$b_0 \sum_{i=1}^{n} x_{1i} + b_1 \sum_{i=1}^{n} x_{1i}^2 + b_2 \sum_{i=1}^{n} x_{1i} x_{2i} = \sum_{i=1}^{n} x_{1i} y_i$$

$$b_0 \sum_{i=1}^{n} x_{2i} + b_1 \sum_{i=1}^{n} x_{1i}x_{2i} + b_2 \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} \mathrm{x}_{2i}y_i$$

Hasil penyelesain dari sistem persamaan linear tersebut didapatkan pendugaan untuk nilai b_1 dan b_2 yaitu

$$b_0 = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2$$

Model regresi linear berganda dalam bentuk matriks,

$$Y = X\beta + \varepsilon$$

Y = vektor variabel terikat
X = matriks variabel bebas
β = vektor parameter yang akan diduga
ε = vektor galat

11.3 Asumsi Klasik

Menurut Draper dan Smith (1992) di dalam model regresi terdapat beberapa asumsi yang harus dipenuhi untuk mencapai kesimpulan yang benar dari parameter β_0 dan β_k.

Asumsi-asumsi tersebut meliputi:

1. Nilai ε_i bebas satu dengan yang lainnya $(\varepsilon_i\ , \varepsilon_j) = 0$. Artinya ε_i merupakan variabel acak dengan nilai tengah nol dan σ^2 yang tidak diketahui.

   $$E(\varepsilon_i) = 0, V(\varepsilon_i) = \sigma^2$$

   ε_i dan ε_j tidak berkolerasi dengan $i \neq j$, sehingga $covv(\varepsilon_i\ , \varepsilon_j) = 0$. Jadi, ε_i merupakan variabel acak normal, dengan σ^2 dengan $\varepsilon_i \sim N\ (0, \sigma^2)$.

2. Nilai tengah dari variabel Y merupakan fungsi linear dari variabel X. Jika dihubungkan titik-titik dari nilai tengah yang berbeda, maka akan diperoleh garis lurus.

   $$\mu(y|x) = \beta_0 + \beta_k X_{ki}.$$

3. Ragam galat homogen (homoskedasitas) berarti bahwa tingkat variasi kesalahan di antara data atau pengukuran adalah konstan, disimbolkan sebagai $Var(\varepsilon_i) = \sigma^2$. Artinya, galat memiliki tingkat variasi yang seragam di seluruh data.

4. Asumsi keempat adalah bahwa kesalahan (galat) ε_i dalam model memiliki distribusi normal dengan rata-rata nol dan ragam tertentu, disimbolkan sebagai $\varepsilon_i \sim N(0, \sigma^2)$. Ini berarti bahwa galat dianggap memiliki pola distribusi yang mirip dengan distribusi normal dengan nilai rata-rata nol dan variabilitas tertentu.

5. Multikolinearitas diperkenalkan oleh Ragnar Frisch (1934), merupakan istilah dalam statistik yang menggambarkan hubungan linear antara variabel bebas dalam regresi linear berganda (Myers, 1990). Multikolinearitas dapat mengindikasikan adanya masalah dalam analisis regresi berganda, dan kita dapat mengidentifikasinya dengan melihat beberapa statistik penting. Salah satu indikator penting adalah "toleransi" atau "Tolerance" dan "Variance Inflation Factor" (VIF) yang terkait dengan setiap variabel bebas dalam model regresi (Gujarati, 1995).
    - Toleransi (tolerance) adalah ukuran yang menunjukkan sejauh mana variabel bebas tidak berkorelasi atau tidak terkait satu sama lain dalam analisis regresi. Nilai toleransi yang rendah menunjukkan bahwa variabel bebas tersebut mungkin memiliki hubungan kuat dengan variabel bebas lainnya, dan ini dapat menyebabkan masalah multikolinearitas.
    - Variance Inflation Factor (VIF), adalah pengukuran yang mengidentifikasi sejauh mana variabilitas (variance) dari koefisien regresi meningkat karena adanya multikolinearitas. Nilai VIF yang tinggi menunjukkan bahwa variabel bebas tersebut

memiliki tingkat multikolinearitas yang signifikan dengan variabel bebas lainnya dalam model.

Pengaruh multikolinearitas dapat dilihat dengan memeriksa nilai (*tolerance*) dan (*Variance Inflation Factor*) dari setiap variabel bebas terhadap variabel terikatnya (Gujarati, 1995). Model Regresi yang baik memiliki VIF<10.

Proses pengujian multikolinearitas dilakukan melalui langkah-langkah berikut:

- Hitung nilai VIF untuk variabel X_1.
- Lakukan regresi variabel bebas yang lain terhadap variabel X_1.
- Dapatkan nilai koefisien determinasi R_j^2 dari regresi tersebut.
- Hitung nilai TOL dengan rumus $TOL = (1 - R_j^2)$.
- Terakhir, hitung nilai VIF dengan rumus $= \frac{1}{TOL}$.

Dengan cara ini, kita dapat mengidentifikasi apakah ada gejala multikolinearitas dalam model regresi.

11.4 Proses Analisis Regresi Berganda

Setelah memahami pentingnya analisis regresi berganda dalam mengungkap hubungan antara variabel-dependen dan variabel-independen yang kompleks, mari kita lanjutkan untuk menjelajahi langkah-langkah kunci dalam proses analisis ini.

Proses analisis regresi berganda melibatkan beberapa tahapan yang berurutan, dimulai dari pengumpulan data hingga interpretasi hasil. Setiap langkah memiliki peran penting dalam memastikan bahwa analisis yang dilakukan memiliki dasar yang kuat dan hasil yang bermakna. Mari kita mulai dengan langkah pertama, yaitu pengumpulan data, yang merupakan fondasi dari seluruh proses analisis regresi berganda.

1. Pengumpulan Data

Pengumpulan data adalah langkah pertama dalam analisis regresi berganda. Data yang dikumpulkan harus relevan dengan variabel dependen dan independen yang akan digunakan dalam analisis. Data ini dapat diperoleh dari berbagai sumber, seperti survei, pengamatan lapangan, atau basis data yang ada.

2. Spesifikasi Model

Setelah data dikumpulkan, langkah berikutnya adalah merancang model regresi yang tepat. Ini melibatkan pemilihan variabel independen yang akan dimasukkan ke dalam model. Pemilihan variabel harus didasarkan pada pengetahuan domain dan teori yang relevan. Misalnya, jika Anda ingin menganalisis faktor-faktor yang mempengaruhi pendapatan individu, variabel independen dapat mencakup pendidikan, pengalaman kerja, dan lokasi geografis.

3. Estimasi Parameter

Estimasi parameter regresi dilakukan menggunakan metode MKT. Metode ini mencari nilai-nilai parameter $(\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$ yang meminimalkan jumlah kuadrat kesalahan (galat) antara prediksi model dan data yang diamati. Proses ini menghasilkan penduga yang optimal untuk koefisien regresi.

4. Uji Hipotesis

Setelah model diestimasi, langkah selanjutnya adalah melakukan uji hipotesis untuk mengevaluasi signifikansi parameter regresi dan asumsi-asumsi klasik. Beberapa uji hipotesis yang umum digunakan dalam analisis regresi berganda meliputi uji t-statistik untuk masing-masing koefisien regresi, uji F-statistik untuk keseluruhan model, dan menguji asumsi-asumsi klasik (Wooldridge, 2016).

5. Evaluasi Model

Evaluasi model adalah langkah penting dalam analisis regresi berganda. Beberapa statistik evaluasi yang umum digunakan termasuk:

- R-squared (R^2): Mengukur sejauh mana variabilitas dalam variabel dependen dapat dijelaskan oleh variabel independen dalam model.

- Mean Squared Error (MSE): Mengukur seberapa baik model cocok dengan data observasi, dengan nilai yang lebih rendah menunjukkan model yang lebih baik.
- Uji AIC (Akaike Information Criterion) atau BIC (Bayesian Information Criterion): Mengukur kualitas model berdasarkan kompromi antara kesesuaian data dan kompleksitas model.

6. Interpretasi Hasil

Langkah terakhir dalam analisis regresi berganda adalah menginterpretasikan hasil. Ini mencakup menjelaskan bagaimana variabel independen mempengaruhi variabel dependen berdasarkan koefisien regresi yang diestimasi. Interpretasi harus dilakukan dengan hati-hati dan dengan mempertimbangkan konteks masalah yang sedang dianalisis.

11.5 Metode Kuadrat Terkecil

Persamaan umum nilai pendugaan dengan metode kuadrat terkecil ditulis

$$\hat{\beta}^{MKT} = \arg\min_{\beta} \|y - X\beta\|_2^2 = \arg\min_{\beta} \left\| y - \sum_{j=1}^{p} \beta_j \tilde{x}_j \right\|_2^2$$

Jumlah kuadrat galat e_i^2 dapat ditulis dalam bentuk

$$\sum_{i=1}^{n} e_i^2 = e_i^T e_i = [e_1 \quad e_2 \quad \cdots \quad e_i] \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_i \end{bmatrix}$$
$$= e_1^2 + e_2^2 + \ldots + e_i^2$$

Dari persamaan umum regresi linear berganda, dapat dituliskan bahwa

$$Q(\hat{\beta}_j) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \cdots - \beta_k x_{ki})^2$$

Dengan mengacu pada persamaan umum regresi linear berganda dalam bentuk matriks, kita dapat menyimpulkan sebagai berikut :

$$e = Y - X\beta$$

Oleh Karena itu, perkaliaan matriks galat menjadi :

$$e_i^T e_i = (Y - X\beta)^T (Y - X\beta)$$

$$e_i^T e_i = Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta$$

(karena $X^T \beta^T Y = Y^T X\beta$)

$$e_i^T e_i = Y^T Y - 2Y^T X\beta + \beta^T X^T X\beta$$

Untuk mencari nilai β dilakukan dengan meminimumkan jumlah kuadrat galat, yaitu menemukan turunan dari $Q(\beta_j)$ secara parsial terhadap β_j ; $j = 1,2,\ldots,k$ lalu disamakan dengan nol.

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i2} & \dots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i1} & \sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i2}X_{i1} & \dots & \sum_{i=1}^{n} X_{ik}X_{i1} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i1}X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & \dots & \sum_{i=1}^{n} X_{ik}X_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{i1}X_{ik} & \sum_{i=1}^{n} X_{i2}X_{ik} & \dots & \sum_{i=1}^{n} X_{ik}^2 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_{11} & X_{21} & \dots & X_{n1} \\ X_{12} & X_{22} & \dots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{ik} & X_{2k} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix}$$

Secara umum dapat dinotasikan dalam bentuk matriks menjadi

$$[X^T X]\hat{\beta} = [X^T Y]$$

$$(X^T X)^{-1} X^T X \hat{\beta} = (X^T X)^{-1} X^T Y$$

$$I\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Sehingga diperoleh estimator,

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Karakteristik dari metode estimasi kuadrat terkecil.

1. $\hat{\beta}$ linear jika $\hat{\beta}$ merupakan fungsi linear dari β

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

$$= (X^T X)^{-1} X^T (X\beta + \varepsilon)$$

$$= (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon$$
$$= I\beta + (X^T X)^{-1} X^T \varepsilon$$

2. $\hat{\beta}$ tak bias
   $\hat{\beta}$ adalah penduga tak bias jika $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$

$$E(\hat{\beta}) = E((X^T X)^{-1} X^T Y)$$
$$= E((X^T X)^{-1} X^T X\beta + \varepsilon)$$
$$= E((X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon)$$
$$= (X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T E(\varepsilon)$$
$$= (X^T X)^{-1} X^T X\beta$$
$$= \beta$$

Sehingga $\hat{\beta}$ merupakan penduga tak bias dari $\boldsymbol{\beta}$

3. $\widehat{\boldsymbol{\beta}}$ memiliki variansi minimum

$$var(\hat{\beta}) = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T]$$
$$= E[(\,(X^T X)^{-1} X^T Y - \beta)((X^T X)^{-1} X^T Y - \beta)^{\mathrm{T}}]$$
$$= E[\,(\,(X^T X)^{-1} X^T (X\beta + \varepsilon) - \beta)\,((X^T X)^{-1} X^T (X\beta + \varepsilon) - \beta)^{\mathrm{T}}]$$
$$= E[((X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon) - \beta)((X^T X)^{-1} X^T X\beta + (X^T X)^{-1} X^T \varepsilon) - \beta)^{\mathrm{T}}]$$
$$= \mathrm{E}\,[(\beta + (X^T X)^{-1} X^T \varepsilon - \beta)\,(\beta + (X^T X)^{-1} X^T \varepsilon - \beta)^{\mathrm{T}}]$$
$$= \mathrm{E}\,[((X^T X)^{-1} X^T \varepsilon)\,((X^T X)^{-1} X^T \varepsilon)^{\mathrm{T}}]$$
$$= \mathrm{E}\,[((X^T X)^{-1}\, X^T \varepsilon^T \varepsilon\, X\, (X^T X)^{-1}]$$
$$= \mathrm{E}\,[((X^T X)^{-1}\, X^T\, E(\varepsilon\varepsilon^T)\, X\, (X^T X)^{-1}]$$
$$= (X^T X)^{-1}\, X^T X\, (X^T X)^{-1}\, E[\varepsilon\varepsilon^T]$$

$= (X^T X)^{-1}\sigma^2$

Varians dari penduga $\widehat{\boldsymbol{\beta}}$, merupakan yang paling rendah di antara semua penduga linear tak bias, dan telah dibuktikan dalam teorema Gauss-Markov. Untuk menegaskan bahwa $Var(\widehat{\boldsymbol{\beta}})$ adalah yang paling minimal, kita akan melakukan asumsi bahwa ada penduga lain yang bersifat linear dan tak bias. Setelah itu, kita akan membuktikan bahwa varians dari penduga lain tersebut lebih besar dari $Var(\widehat{\boldsymbol{\beta}})$.

Misalkan $\widehat{\boldsymbol{\beta}}^*$ linier dan tak bias bagi $\boldsymbol{\beta}$. Asumsikan bahwa :

$$\hat{\beta}^* = ((X^T X)^{-1}X^T + Z)y$$

dimana **Z** adalah matriks konstanta $(k \times n)$ yang diketahui

$$\begin{aligned}
&\hat{\beta}^* = ((X^T X)^{-1}X^T + Z)y \\
&= ((X^T X)^{-1}X^T + Z)(X\beta + \varepsilon) \\
&= \left((X^T X)^{-1}X^T X\beta + (X^T X)^{-1}X^T\varepsilon + Z(X\beta + \varepsilon)\right) \\
&= ((X^T X)^{-1}X^T X\beta + (X^T X)^{-1}X^T\varepsilon + ZX\beta + Z\varepsilon) \\
&= (I\beta + (X^T X)^{-1}X^T\varepsilon + ZX\beta + Z\varepsilon)
\end{aligned}$$

sehingga

$$\begin{aligned}
E(\hat{\beta}^*) &= E(\beta + (X^T X)^{-1}X^T\varepsilon + ZX\beta + Z\varepsilon) \\
&= \beta + ZX\beta \qquad (karena\ E(\varepsilon) = 0) \\
&= (1 + ZX)\beta
\end{aligned}$$

Agar $\hat{\beta}^*$ estimasi tak bias dari β maka $ZX = 0$, sehingga :

$$Var(\hat{\beta}^*) = E\left((\hat{\beta}^* - \beta)(\hat{\beta}^* - \beta)^T\right)$$

dengan $(\hat{\beta}^* - \beta) = (X^T X)^{-1}X^T\varepsilon + ZX\beta + Z\varepsilon$. Dan diasumsikan bahwa $E(\varepsilon\varepsilon^T) = \sigma_u^2 I_u$. Karena $ZX\beta = 0$ sehingga

$$\begin{aligned}
&Var(\hat{\beta}^*) = E\left(((X^T X)^{-1}X^T\varepsilon + Z\varepsilon)((X^T X)^{-1}X^T\varepsilon + Z\varepsilon)^T\right) \\
&= E\left(((X^T X)^{-1}X^T\varepsilon + Z\varepsilon)(\varepsilon^T Z^T + \varepsilon^T X(X^T X)^{-1})\right)
\end{aligned}$$

$$= E\left(((X^TX)^{-1}X^T + Z)\varepsilon\varepsilon^T(Z^T + X(X^TX)^{-1})\right)$$
$$= \left(((X^TX)^{-1}X^T + Z)\sigma_u^2 I_u(Z^T + X(X^TX)^{-1})\right)$$
$$= \sigma^2\left(((X^TX)^{-1}X^T + Z)(Z^T + X(X^TX)^{-1})\right)$$
$$= \sigma^2((X^TX)^{-1}X^TZ^T + (X^TX)^{-1}X^TX(X^TX)^{-1} + ZZ^T + ZX(X^TX)^{-1})$$
$$= \sigma^2((X^TX)^{-1}X^TZ^T + I(X^TX)^{-1} + ZZ^T + ZX(X^TX)^{-1})$$

Karena $ZX = X^TZ^T = 0$ maka

$$Var(\hat{\beta}^*) = \sigma^2((X^TX)^{-1} + ZZ^T)$$

Matriks ZZ^T *definite positive*, dengan elemen diagonalnya berbentuk kuadrat. Sehingga, varians dari setiap elemen dalam vektor $\hat{\beta}^*$ selalu lebih besar, atau sama dengan varians elemen $\hat{\beta}$ yang sesuai telah terbukti. Penduga kuadrat terkecil yang memenuhi sifat linear, tak bias, dan varians minimum ini dikenal dengan penduga Linear Terbaik dan Tak Bias (BLUE).

11.6 Ukuran Pemusatan dan Penskalaan (*centering and scaling)*

Pemusatan dan penskalaan data adalah cara untuk membuat data lebih mudah dibandingkan. Ini membantu kita saat kita ingin membandingkan hal-hal yang berbeda dalam dataset kita.

- **Pemusatan** adalah ketika kita mengambil setiap angka dalam data kita dan mengurangkan angka itu dengan angka rata-rata dari semua angka dalam data. Ini membuat angka-angka dalam data kita lebih fokus pada perbedaannya daripada nilai sebenarnya.
- **Penskalaan** adalah ketika kita mengevaluasi setiap angka dalam data kita dengan cara menghitung seberapa jauh angka itu berbeda dari nilai rata-rata, diukur dalam satuan deviasi standar. Ini membantu kita memahami betapa "istimewanya" sebuah angka dalam data kita

Ini adalah teknik statistik yang berguna ketika kita ingin membandingkan data yang berbeda dalam skala dan variasi. Teknik ini membantu kita membuat data lebih mudah dimengerti dan diinterpretasikan (Kutner, *et al.*, 2005).

Model regresi linear berganda setelah dibakukan menjadi

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

Berikut ini merupakan pembakuan variabel terikat Y dan variabel bebas $X_1, X_2, \ldots, X_k$

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \; dengan \; S_Y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$$

$$X_j^* = \frac{X_{ij} - \bar{X}_j}{S_{X_j}} \; dengan \; S_{X_j} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{ij} - \bar{X}_j\right)^2}{n-1}}$$

untuk $i = 1,2, \ldots, n \; dan \; j = 1,2, \ldots, k$

$\bar{Y}$ = rata-rata dari Y

$\bar{X}_j$ = rata-rata dari pengamatan X_j

S_Y = standar deviasi dari Y

S_{X_j} = standar deviasi dari X_j

Model regresi berganda terstandarisasi merupakan tranformasi dari model regresi berganda

$$Y_i^* = \beta_1^* X_{1i}^* + \beta_2^* X_{2i}^* + \cdots + \beta_k^* X_{ki}^* + \varepsilon_i^*$$

Model di atas merupakan bentuk *standardized regression model*. Hubungan antara kedua parameter dari dua model yang berbeda tersebut dijabarkan sebagai berikut (Kutner, *et al*., 2005).

$$\beta_j = \left(\frac{S_Y}{S_{X_j}}\right)\beta_j^* \ , \qquad j = 1,2, \dots , k$$
$$\beta_0 = \bar{Y} - \beta_1\bar{X}_1 - \beta_2\bar{X}_2 - \cdots - \beta_k\bar{X}_k$$
$$\beta_0 = \bar{Y} - \sum_{j=1}^{k} \beta_j\bar{X}_j$$

Prosedur ini merupakan prosedur penskalaan. Dari persamaan di atas dapat dijabarkan dalam bentuk

$$Y_i = \beta_0 + \beta_1(X_{1i} - \bar{X}_1) + \beta_1\bar{X}_1 + \beta_2(X_{2i} - \bar{X}_2) + \beta_2\bar{X}_2 + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \beta_k\bar{X}_k + \varepsilon_i$$
$$Y_i = (\beta_0 + \beta_1\bar{X}_1 + \cdots + \beta_k\bar{X}_k) + \beta_1(X_{1i} - \bar{X}_1) + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \varepsilon_i$$

Maka berlaku :

$$\bar{Y} = \beta_0 + \beta_1\bar{X}_1 + \beta_2\bar{X}_2 + \cdots + \beta_k\bar{X}_k$$

Sehingga

$$Y_i - \bar{Y} = (\beta_0 + \beta_1\bar{X}_1 + \cdots + \beta_k\bar{X}_k) + (\beta_1(X_{1i} - \bar{X}_1) + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \varepsilon_i) - (\beta_0 + \beta_1\bar{X}_1 + \beta_2\bar{X}_2 + \cdots + \beta_k\bar{X}_k)$$
$$Y_i - \bar{Y} = \beta_1(X_{1i} - \bar{X}_1) + \beta_2(X_{2i} - \bar{X}_2) + \cdots + \beta_k(X_{ki} - \bar{X}_i) + \varepsilon_i$$
$$y_i = Y_i - \bar{Y}$$
$$x_{1i} = X_{1i} - \bar{X}_1$$
$$x_{2i} = X_{2i} - \bar{X}_2$$
$$x_{ki} = X_{ki} - \bar{X}_k$$

Maka didapat model regresi baru yaitu:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$$

Langkah-langkah untuk mengubah model awal menjadi model akhir dikenal sebagai prosedur pemusatan. Dalam prosedur ini, intercept β_0 dihilangkan, yang menyederhanakan perhitungan dalam pencarian model regresi. Keseluruhan langkah-langkah di atas disebut sebagai prosedur pemusatan dan penskalaan

Contoh :
Hitung persamaan regresi linier berganda dari data dibawah ini.

X1	X2	Y
45	16	29
42	14	24
44	15	27
45	13	25
43	13	26
46	14	28
44	16	30
45	16	28
44	15	28
43	15	27

Penyelesaian

n	X_1	X_2	Y	X_1Y	X_2Y	X_1X_2	X_1^2	X_2^2	Y^2
1	45	16	29	1305	464	720	2025	256	841
2	42	14	24	1008	336	588	1764	196	576
3	44	15	27	1188	405	660	1936	225	729
4	45	13	25	1125	325	585	2025	169	625
5	43	13	26	1118	338	559	1849	169	676
6	46	14	28	1288	392	644	2116	196	784
7	44	16	30	1320	480	704	1936	256	900
8	45	16	28	1260	448	720	2025	256	784
9	44	15	28	1232	420	660	1936	225	784
10	43	15	27	1161	405	645	1849	225	729
$\overline{X}$	$\overline{X_1}$ = 44,1	$\overline{X_2}$ = 14,7	$\overline{Y}$ = 27,2						
Total	441	147	272	12005	4013	6485	19461	2173	7428

Perhitungan :

$$\sum_{i=1}^{n} y^2 = \sum_{i=1}^{n} Y^2 - n(\overline{Y})^2 = 7428 - 10\,.\,(27{,}2)^2 = 29{,}6$$

$$\sum_{i=1}^{n} x_1^2 = \sum_{i=1}^{n} X_1^2 - n(\bar{X}_1)^2 = 19461 - 10\,.\,(44{,}1)^2 = 12{,}9$$

$$\sum_{i=1}^{n} x_2^2 = \sum_{i=1}^{n} X_2^2 - n(\bar{X}_2)^2 = 2173 - 10\,.\,(14{,}7)^2 = 12{,}1$$

$$\sum_{i=1}^{n} x_1 y = \sum_{i=1}^{n} X_1 Y - n\,.\,\bar{X}_1\,.\,\bar{Y} = 12005 - 10\,.\,44{,}1.\,27{,}2 = 9{,}8$$

$$\sum_{i=1}^{n} x_2 y = \sum_{i=1}^{n} X_2 Y - n\,.\,\bar{X}_2\,.\,\bar{Y} = 4013 - 10\,.\,14{,}7.\,27{,}2 = 14{,}6$$

$$\sum_{i=1}^{n} x_1 x_2 = \sum_{i=1}^{n} X_1 X_2 - n.\,\bar{X}_1.\,\bar{X}_2 = 6485 - 10\,.\,44{,}1\,.\,14{,}7 = 2{,}3$$

$$b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2} = \frac{(12{,}1)(9{,}8) - (2{,}3)(14{,}6)}{(12{,}9)(12{,}1) - (2{,}3)^2} = 0{,}564$$

$$b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2} = \frac{(12{,}9)(14{,}6) - (2{,}3)(9{,}8)}{(12{,}9)(12{,}1) - (2{,}3)^2} = 1{,}099$$

$$a = \bar{Y} - b_1\overline{X_1} - b_2\overline{X_2} = 27{,}2 - (0{,}564)(44{,}1) - (1{,}099)(14{,}7) = -13{,}828$$

Sehingga diperoleh model regresi

$$\hat{Y} = -13{,}828 + 0{,}564X_1 + 1{,}099X_2$$

DAFTAR PUSTAKA

Gujarati, Damodar. 2006. Dasar-Dasar Ekonometrika. Erlangga: Jakarta

Kutner, M.H., Nachtsheim, C.J, dan J. Neter, J. 2004. Applied Linear Regression Models. 4th ed. New York: McGraw-Hill Campanies, Inc.

Usman, M. dan Warsono. 2000. Teori Model Linear dan Aplikasinya. C.V. Darmajaya, Bandar Lampung.

Rencher, C.A, 2008, LInear Model in Statistics, New Jersey: John Wiley and Sons.

Draper, N.R. and Smith, H. 1992. Analisis Regresi Terapan. Ed. Ke-2.

Myers, R.H. 1990. Clasical and Modern Regression With Application. PWSKENT publishing Company, Boston.

Sumodiningrat, G. 1998. Ekonometrika Pengantar. BPFE, Yogyakarta.

Kutner, M.H., et al. 2005. Applied linear Statistic Model. Ed. Ke-5. Mc-Grawhill, New York.

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

BAB 12
ANALISIS RUNTUN WAKTU

Oleh Hanna Hilyati Aulia

12.1 Pendahuluan

Deret berkala (*time series*) atau runtun waktu adalah serangkaian pengamatan terhadap peristiwa, kejadian atau variabel yang diambil dari waktu ke waktu, dicatat secara teliti menurut urut-urutan waktu terjadinya, kemudian disusun sebagai data statistik. Dari suatu rutut waktu akan dapat diketahui pola perkembangan suatu peristiwa, kejadian atau variabel. Jika perkembangan suatu peristiwa mengikuti suatu pola yang teratur, maka berdasarkan pola perkembangan tersebut akan dapat diramalkan peristiwa yang akan terjadi dimasa yang akan datang. (Robinson, 2009) Secara konvensional, analisis deret berkala selalu didasarkan pada anggapan bahwa nilai deret berkala merupakan hasil perkalian (multiplikatif) dari trend sekuler, variasi musim, variasi siklikal, dan variasi random. (Aswi and Sukarna, 2006; Hornik, 2009)Namun demikian, data deret berkala juga dapat merupakan hasil penjumlahan atau kombinasi antara perkalian dan penjumlahan dalam seribu satu cara dari komponen-komponennya.

Analis runtun waktu merupakan prosedur analisis yang dapat digunakan untuk mengetahui gerak perubahan atau perkembangan nilai suatu variabel sebagai akibat dari perubahan waktu. Dalam analisis ekonomi dan lingkungan bisnis biasanya analisa deret berkala digunakan untuk meramal (*forecasting*) nilai suatu variabel pada masa lalu dan masa yang akan datang berdasarkan pada kecenderungan dari perubahan nilai variabel tersebut.

12.2 Jenis Variasi Runtun Waktu

Metode tradisional analisis rangkaian waktu terutama berkaitan dengan penguraian variasi dalam rangkaian menjadi tren, variasi musiman, perubahan siklus lainnya, dan fluktuasi 'tidak teratur' yang tersisa. Pendekatan ini tidak selalu merupakan yang terbaik namun sangat berharga ketika variasinya didominasi oleh tren dan/atau musiman. Namun, perlu dicatat bahwa dekomposisi tersebut umumnya tidak unik kecuali ada asumsi tertentu yang dibuat. Jadi semacam pemodelan, baik eksplisit maupun implisit, mungkin terlibat dalam teknik deskriptif ini, dan ini menunjukkan batas yang kabur antara teknik deskriptif dan inferensial. (Chatfield, 1990) Berbagai sumber variasi sekarang akan dijelaskan secara lebih rinci pada tabel berikut.

Trend	Merupakan gerakan jangka panjang yang menunjukkan arah perkembangan baik arah menaik maupun menurun. Data deret berkala diperoleh dari rata-rata perubahan dari waktu ke waktu dengan nilai yang cukup rata (smooth). Trend arah menaik disebut trend positif dan trend dengan arah menurun disebut trend negatif.
Variasi Siklus	Adalah gerakan jangka panjang di sekitar garis trend gelombang perekonomian mengalami siklus yang terulang setiap jangka waktu tertentu (satu tahun, tiga tahun, atau lima tahun). Contoh gerakan siklus terjadinya kemunduran (resesi)
Variasi Musim	Adalah gerakan yang mempunyai pola tetap dari waktu ke waktu. Perubahan waktu berhubungan dengan perubahan musim-musim tertentu. Satuan yang digunakan dapat berupa bulan, triwulan, semester. Contoh : Produksi pertama dipengaruhi oleh musim seperti seperti

	padi triwulan pertama produksi meningkat, triwulan kedua dan ketiga menurun.
Variasi Tidak Teratur	Adalah gerakan yang sporadis, perubahannya tidak beraturan baik dari sisi waktu maupun lama siklusnya. Misalnya gerakan perekonomian yang disebabkan perang dan bencana alam. Data berkala terdiri dari uraian komponen-komponen yang menyebabkan gerakan yang tercermin pada fluktuasi. Data berkala dapat digunakan untuk pembuatan garis trend.

12.3 Runtun Waktu Stationer

Secara garis besar suatu deret waktu dikatakan stasioner jika tidak ada perubahan mean yang sistematis (tidak ada tren), jika tidak ada perubahan varians yang sistematis, dan jika variasi periodik yang ketat telah dihilangkan. Sebagian besar teori probabilitas deret waktu berkaitan dengan deret waktu yang stasioner, dan karena alasan ini analisis deret waktu sering kali mengharuskan seseorang untuk mengubah deret waktu yang tidak stasioner menjadi deret waktu yang stasioner agar dapat menggunakan teori ini. Misalnya, dengan menghilangkan tren dan variasi musiman dari sekumpulan data dan kemudian mencoba memodelkan variasi dalam residu melalui proses stokastik stasioner. Namun, perlu juga ditekankan bahwa komponen nonstasioner, seperti tren, terkadang lebih menarik dibandingkan residu stasioner. (Chatfield, 1990)

12.4 Plot Waktu

Langkah pertama dan terpenting dalam analisis deret waktu adalah memplot observasi terhadap waktu. Grafik yang dibuat harus menampilkan fitur-fitur penting dari rangkaian seperti tren, musiman, outlier, dan diskontinuitas. Plot untuk mendeskripsikan dan merumuskan model dari data. Pilihan skala, ukuran titik potong, dan cara titik-titik diplot (misalnya sebagai garis kontinu atau

sebagai titik-titik terpisah) secara substansial dapat mempengaruhi 'tampilan' plot, sehingga analis harus berhati-hati dan mengambil keputusan. Selain itu, aturan umum dalam menggambar grafik yang 'baik' harus diikuti: judul yang jelas harus diberikan, satuan pengukuran harus dinyatakan, dan sumbu harus diberi label dengan jelas.

12.5 Transformasi Data

Pembuatan plot data mungkin perlu mempertimbangkan transformasi data, misalnya dengan mengambil logaritma atau akar kuadrat. Tiga alasan utama dilakukannya transformasi adalah sebagai berikut.

a. Untuk menstabilkan varians

   Jika terdapat tren pada deret dan variansinya meningkat dengan mean, maka data dapat ditransformasikan. Khususnya jika standar deviasi proporsional dengan mean maka digunakan transformasi logaritma.

b. Untuk membuat efek musiman menjadi penjumlahan

   Jika terdapat tren pada deret dan ukuran dari efek musiman meningkat sepanjang mean, maka data dapat ditransformasi agar efek musiman berupa konstan dari tahun ke tahun.

c. Untuk membuat data terdistribusi normal

   Dalam pemodelan dan peramalan biasanya mengansumsikan bahwa data berdistribusi normal. Pada praktiknya, tidak selalu menjadi masalah, adanya skewness pada plot waktu akan sulit dihilangkan dan membutuhkan distribusi kesalahan (error) yang lain.

Transformasi logaritma dan akar kuadrat adalah kasus khusus dari jenis transformasi yang disebut transformasi Box-Cox.

Misalkan terdapat deret berala $\{x_t\}$ dan parameter transformasi λ, maka deret transformasinya adalah

$$y_t = \begin{cases} \frac{x_t^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ log(x_t) & \lambda = 0 \end{cases}$$

saat $\lambda \neq 0$ adalah transformasi pangkat dan saat $\lambda = 0$ adalah konstan untuk membuat deret y_t kontinu terhadap λ.

12.6 Analisis Deret yang Memuat Tren

Berikut diberikan model tren yang paling sederhana, untuk waktu observasi t, variabel acak X_t tuliskan

$$X_t = \alpha + \beta t + \varepsilon_t$$

dimana α, β adalah konstan dan ε_t menyatakan suku error dengan mean nol.

Tren pada persamaan di atas merupakan fungsi deterministik waktu dan terkadang disebut tren linier global. Hal ini umumnya tidak realistis, dan sekarang ada lebih banyak penekanan pada tren linier lokal di mana parameter α dan β dalam persamaan dibiarkan berkembang seiring waktu. Alternatifnya, trennya mungkin berbentuk non-linier seperti pertumbuhan kuadrat. Pertumbuhan eksponensial bisa sangat sulit untuk ditangani, bahkan jika logaritma digunakan untuk mengubah tren ke bentuk linier.

Analisis rangkaian waktu yang menunjukkan tren bergantung pada apakah seseorang ingin (a) mengukur tren dan/atau (b) menghilangkan tren untuk menganalisis fluktuasi lokal. Hal ini juga bergantung pada apakah data menunjukkan pola musiman. Dengan data musiman, ada baiknya untuk memulai dengan menghitung rata-rata tahunan berturut-turut karena ini akan memberikan gambaran sederhana tentang tren yang mendasarinya. Pendekatan seperti ini terkadang cukup memadai, terutama jika trennya cukup kecil, namun terkadang diperlukan pendekatan yang lebih canggih dan teknik-teknik berikut dapat dipertimbangkan.

12.6.1 *Curve Fitting*

Metode tradisional untuk menangani data non-musiman yang berisi tren, khususnya data tahunan, adalah dengan menyesuaikan fungsi sederhana seperti kurva polinomial (linier, kuadrat, dll.), kurva Gompertz, atau kurva logistik (misalnya lihat Levenbach dan Reuters, 1976; Meade, 1984). Kurva Gompertz diberikan oleh

$$logx_t = a + br^t$$

dimana a, b, r adalah parameter dengan $0 < r < 1$, sedangkan kurva logistik diberikan oleh

$$x_t = \frac{\alpha}{l + be^{-ct}}$$

Kedua kurva ini berbentuk S dan mendekati nilai asimtotik sebagai $t \to \infty$. Menyesuaikan kurva dengan data dapat menghasilkan persamaan simultan non-linier. Untuk semua kurva jenis ini, fungsi yang dipasang memberikan ukuran tren, dan residu memberikan perkiraan fluktuasi lokal, dengan residu adalah perbedaan antara pengamatan dan nilai yang sesuai dari kurva yang dipasang.

12.6.2 *Filtering*

Metode kedua untuk tren adalah dengan filter linier yang mengubah satu deret waktu $\{x_t\}$ ke $\{y_t\}$ dengan operasi linier

$$y_t = \sum_{r=-q}^{+s} a_r x_{t+r}$$

dimana $\{a_r\}$ adalah himpunan bobot. Untuk memuluskan fluktuasi lokal dan memperkirakan mean lokal, kita harus memilih bobot dengan jelas sehingga $\sum \alpha_r = 1$, dan kemudian operasi ini sering disebut sebagai rata-rata bergerak (*moving average*).

12.6.3 *Differencing*

Metode ini merupakan bentuk khusus dari filtering yang berguna untuk menghilangkan trend dengan membuat beda deret waktu tertentu hingga menjadi stationer. Metode ini merupakan bagian integral dari prosedur yang dianjurkan oleh Box dan Jenkins (1970). Untuk data non-musiman, pembedaan orde pertama biasanya cukup untuk mencapai stasioneritas nyata, sehingga deret baru $\{y_1 \ldots, y_{N-1}\}$dibentuk dari deret asal $\{x_1, \ldots, x_N\}$ oleh

$$y_t = x_{t+1} - x_t = \nabla x_{t+1}$$

Diferensiasi orde pertama banyak digunakan. Kadang-kadang diperlukan pembedaan orde kedua dengan menggunakan operator ∇^2, dimana

$$\nabla^2 x_{t+2} = \nabla x_{t+2} - \nabla x_{t+1} = x_{t+2} - 2x_{t+1} + x_t.$$

12.7 Analisis Deret yang Memuat Variasi Musiman

Tiga model musiman yang sering digunakan antara lain

A. $X_t = m_t + S_t + \varepsilon_t$
B. $X_t = m_t S_t + \varepsilon_t$
C. $X_t = m_t S_t \varepsilon_t$

dimana m_t adalah tingkat rata-rata desasionalisasi pada waktu t, S_f adalah efek musiman pada waktu t dan ε_t adalah kesalahan acak.

Model A mendeskripsikan kasus aditif, sedangkan model B dan C keduanya melibatkan musiman multiplikatif. Dalam model C, istilah kesalahannya juga bersifat perkalian, dan transformasi logaritmik akan mengubahnya menjadi model aditif (linier) yang mungkin lebih mudah ditangani. Plot waktu harus diperiksa untuk melihat model mana yang mungkin memberikan gambaran yang lebih baik. Indeks musiman $\{S_t\}$ biasanya diasumsikan berubah perlahan sepanjang waktu sehingga $S_t \simeq S_{t-s}$, dimana s adalah jumlah observasi per tahun. Indeks biasanya dinormalisasi sehingga berjumlah nol dalam kasus penjumlahan, atau rata-rata menjadi satu dalam kasus perkalian. Kesulitan muncul dalam praktiknya jika

istilah musiman dan/atau kesalahan tidak benar-benar perkalian atau penjumlahan. Sebagai contoh, efek musiman dapat meningkat seiring dengan tingkat rata-rata namun tidak dengan kecepatan yang begitu cepat sehingga efek tersebut berada di antara bersifat multiplikatif atau aditif.

Analisis deret waktu yang menunjukkan variasi musiman bergantung pada apakah seseorang ingin (a) mengukur pengaruh musiman dan/atau (b) menghilangkan kemusiman. Untuk rangkaian yang menunjukkan sedikit tren, biasanya cukup memperkirakan pengaruh musiman untuk periode tertentu (misalnya Januari) dengan mencari rata-rata setiap observasi bulan Januari dikurangi rata-rata tahunan terkait dalam kasus aditif, atau observasi bulan Januari dibagi dengan rata-rata tahunan dalam kasus perkalian.

12.8 Proses Stokastik

Proses stokastik dapat digambarkan sebagai fenomena statistik yang berkembang dalam waktu sesuai dengan hukum probabilistik. Contoh yang umum adalah panjang antrian, ukuran koloni bakteri, dan suhu udara pada hari-hari berturut-turut di suatu lokasi tertentu. Kata 'stochastic', yang berasal dari bahasa Yunani, berarti berkaitan dengan kebetulan, dan banyak penulis menggunakan proses acak sebagai sinonim untuk proses stokastik.

Secara matematis, proses stokastik dapat didefinisikan sebagai kumpulan variabel acak yang diurutkan dalam waktu dan ditentukan pada serangkaian titik waktu yang mungkin kontinu atau diskrit. Kita akan menyatakan variabel acak pada waktu t dengan $X(t)$ jika waktu kontinu ($biasanya\ -\infty < t\ < \infty$), dan dengan X_t jika waktu diskrit (biasanya $t == 0, \pm 1, \pm 2, \ldots$).

Sebagian besar masalah statistik berkaitan dengan memperkirakan sifat-sifat suatu populasi dari suatu sampel. Dalam analisis deret waktu terdapat situasi yang agak berbeda, meskipun dimungkinkan untuk memvariasikan panjang deret waktu yang diamati - sampel - biasanya tidak mungkin melakukan lebih dari satu pengamatan pada waktu tertentu. Jadi kita hanya mempunyai

satu hasil dari proses dan satu observasi pada variabel acak pada waktu t. Meskipun demikian, kita dapat menganggap rangkaian waktu yang diamati hanya sebagai salah satu contoh rangkaian waktu tak terhingga yang mungkin telah diamati. Kumpulan deret waktu yang tak terhingga ini terkadang disebut ansambel. Setiap anggota ansambel merupakan kemungkinan realisasi proses stokastik. Deret waktu yang diamati dapat berupa

dianggap sebagai satu realisasi tertentu, dan akan dilambangkan dengan $x(t)$ untuk $(0 \leq t \leq T)$ jika pengamatannya kontinu, dan dengan x,untuk $t = 1, \dots, N$ jika pengamatannya diskrit.

Karena hanya terdapat populasi nosional, analisis deret waktu pada dasarnya berkaitan dengan evaluasi properti model probabilitas yang menghasilkan deret waktu yang diamati.

Salah satu cara untuk mendeskripsikan proses stokastik adalah dengan menentukan distribusi probabilitas gabungan $X(t_1), \dots, X(t_n)$ untuk himpunan waktu $t_1, \dots, t_n$ dan nilai n apa pun. Namun hal ini agak rumit dan biasanya tidak dilakukan dalam praktik. Cara yang lebih sederhana dan berguna untuk mendeskripsikan proses stokastik adalah dengan memberikan momen-momen dari proses tersebut, khususnya momen pertama dan kedua, yang disebut fungsi mean, varians, dan autokovarians. Ini sekarang akan didefinisikan untuk waktu berkelanjutan, dengan definisi serupa berlaku dalam waktu diskrit.

fungsi mean didefinisikan oleh

$$\mu(t) = E[X(t)]$$

fungsi varians didefinisikan oleh

$$\sigma^2(t) = Var[X(t)]$$

fungsi varians saja tidak cukup untuk menspesifikasi momen kedua dari deret variabel acak. Oleh karena itu, kita harus mendefinisikan fungsi autokovariansi

$$\gamma(t_1, t_2) = E\{[X(t_1) - \mu(t_1)\,][X(t_2) - \mu(t_2)]\}$$

12.9 Proses Stasioner

Salah satu kelompok proses stokastik yang penting adalah proses stasioner. Suatu deret waktu dikatakan stasioner ketat (*strictly stationer*) jika distribusi gabungan $X(t_1), \dots, X(t_n)$ sama dengan distribusi gabungan $X(t_1 + \tau), \dots, X(t_n + \tau)$ untuk semua $t_1, \dots, t_n, \tau$. Dengan kata lain, menggeser asal waktu sebesar τ tidak berpengaruh pada distribusi gabungan, yang karenanya harus bergantung hanya pada interval antara $t_1, \dots, t_n$. Definisi di atas berlaku untuk nilai n apa pun.

dengan kata lain, jika $n = 1$, maka stationer ketat menyatakan bahwa distribusi $X(t)$ sama untuk setiap t, sehingga dua momen bertama terbatas, yaitu

$$\mu(t) = \mu$$
$$\sigma^2(t) = \sigma$$

keduanya merupakan konstan dan tidak bergantung pada t.

Lebih lanjut, jika $n = 2$, distribusi gabungan dari $X(t_1)$ dan $X(t_2)$ hanya bergantung pada $(t_2 - t_1)$, yang disebut *lag*. Kemudian, fungsi autokovariansi juga bergantung pada $(t_2 - t_1)$ yang ditulis sebagai

$$\gamma(\tau) = E\{[X(t) - \mu\,][X(t + \tau) - \mu]\} = Cov[X(t), X(t + \tau)]$$

fungsi tersebut disebut koefisien autokovariansi pada lag τ. Selanjutnya, fungsi autokovariansi disebut $acv.f$.

Besaran koefisien autokovarians bergantung pada satuan pengukuran $X(t)$. Jadi, untuk tujuan interpretasi, akan berguna untuk membakukan $acv.f$ untuk menghasilkan fungsi yang disebut fungsi autokorelasi, yang diberikan oleh

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)}$$

yang mengukur korelasi antara $X(t)$ dan $X(t + \tau)$.

12.9.1 Stasioner orde kedua

Dalam praktiknya, sering kali berguna untuk mendefinisikan stasioneritas dengan cara yang tidak terlalu dibatasi dibandingkan yang dijelaskan di atas. Suatu proses disebut stasioner orde kedua (atau stasioner lemah) jika rata-ratanya konstan dan $acv.f$nya hanya bergantung pada lag, sehingga

$$E[X(t)] = \mu$$

dan

$$Cov[X(t), X(t+\tau)] = \gamma(\tau)$$

Tidak ada asumsi yang dibuat tentang momen yang lebih tinggi daripada momen orde kedua. Dengan memisalkann $\tau = 0$, perhatikan bahwa asumsi di atas tentang $acv.f$ menyiratkan bahwa varians dan meannya adalah konstan. Perhatikan juga bahwa varians dan mean harus berhingga.

Definisi stasioneritas yang lebih lemah ini umumnya akan digunakan mulai sekarang, karena banyak sifat proses stasioner yang hanya bergantung pada struktur proses sebagaimana ditentukan oleh momen pertama dan kedua. Salah satu kelas proses yang penting dimana hal ini khususnya berlaku adalah kelas proses normal dimana distribusi gabungan $X(t_1), \dots, X(t_n)$ adalah multivariat normal untuk semua $t_1, \dots, t_n$. Distribusi normal multivariat sepenuhnya dicirikan oleh momen pertama dan kedua, dan karenanya oleh $\mu(t)$ dan $\gamma(t_1, t_2)$, sehingga stasioneritas orde kedua menyiratkan stasioneritas ketat untuk proses normal.

12.10 Fungsi Autokorelasi

Koefisien autokorelasi dari deret waktu yang diamati merupakan kumpulan statistik penting untuk menggambarkan deret waktu. Demikian pula fungsi autokorelasi $ac.f$ dari proses stokastik stasioner merupakan alat penting untuk menilai sifat-sifatnya. Bagian ini menyelidiki sifat umum $ac.f$

Misalkan proses stokastik stasioner $X(t)$ mempunyai mean μ, varians σ^2, $acv.f$ $\gamma(\tau)$, dan $ac.f$ $\rho(\tau)$ maka

$$\rho(\tau) = \frac{\gamma(\tau)}{\gamma(0)} = \frac{\gamma(\tau)}{\sigma^2}$$

Perhatikan bahwa $\rho(0) = 1$.

Ciri-ciri fungsi autokorelasi

1. fungsi ac.f adalah fungsi genap $\rho(\tau) = \rho(-\tau)$
2. $\rho(\tau) \leq 1$
3. Kurangnya keunikan. Meskipun proses stokastik tertentu memiliki struktur kovarians yang unik, hal sebaliknya tidak berlaku secara umum. Biasanya dimungkinkan untuk menemukan banyak proses normal dan tidak normal dengan ac.f. dan ini menciptakan kesulitan lebih lanjut dalam menafsirkan contoh ac.f.

Beberapa contoh proses stokastik

1. Proses Random Murni

Suatu proses waktu diskrit disebut proses acak murni jika proses tersebut terdiri dari barisan variabel acak $\{Zt\}$ yang saling bebas dan terdistribusi secara identik. Dari definisi tersebut dapat disimpulkan bahwa proses tersebut memiliki mean dan varians yang konstan dan

$\gamma(k) = Cov(Z_t, Z_{t+k}) = 0$ untuk $k = \pm 1,2, \ldots$

karena mean dan $acv.f$ tidak bergantung oleh waktu. Fungsi $ac.f$ diberikan oleh

$$\rho(k) = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1, \pm 2, \ldots \end{cases}$$

poses random murni biasanya disebut *white noise.*

2. *Random Walk*

Misalkan $\{Z_t\}$ diskrit, proses random murni dengan mean μ dan variansi σ_z^2. Proses $\{X_t\}$ dikatakan *random walk* jika

$$X_t = X_{t-1} + Z_t$$

Proses dimulai untuk $t = 0$, selanjutnya

$$X_1 = Z_1$$

dan

$$X_t = \sum_{i=1}^{t} Z_t$$

sehingga diperoleh bahwan $E(X_t) = t\mu$ dan $Var(X_t) = t\sigma_z^2$. Karena mean dan varians berubah sepanjang waktu, maka proses tidak stasioner. Namun, perhatikan bahwa persamaan beda dari suku awal *random walk* adalah

$$\nabla X_t = X_t - X_{t-1} = Z_t$$

membentuk proses random murni yang stasioner.

3. Proses Rerata Bergerak/*Moving Average* (MA)

Misalkan $\{Z_t\}$ diskrit, proses random murni dengan mean μ dan variansi σ_z^2. Proses $\{X_t\}$ dikatakan proses *moving average* (MA) dengan orde q jika

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}$$

dengan $\{\beta_i\}$ merupakan konstanta. Nilai Z biasanya diskalakan sehinga $\beta_0 = 1$. Dengan demikian diperoleh

$$E(X_t) = 0$$

$$Var(X_t) = \sigma_z^2 \sum_{i=0}^{q} \beta_i^2$$

karena Z independen, diperoleh

$$\gamma(k) = Cov(X_t, X_{t+k})$$

$$= Cov\left(\beta_0 Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q}\right)$$

$$= \begin{cases} 0 & k > q \\ \sigma_z^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & k = 0, 1, \ldots, q \\ \gamma(-k) & k < 0 \end{cases}$$

karena

$$Cov(Z_s, Z_t) = \begin{cases} \sigma_z^2 \; s = t \\ 0 \;\; s \neq t \end{cases}$$

Karena $\gamma(k)$ tidak bergantung pada t dan meannya konstan, maka prosesnya stasioner orde kedua untuk semua nilai $\{\beta_i\}$. Lebih lanjut, jika Z terdistribusi normal, maka X juga terdistribusi normal, dan kita mempunyai proses normal yang stasioner.

Ac.f dari proses MA(q) diberikan oleh

$$\rho(k) = \begin{cases} 1 & k = 0 \\ \frac{\sum_{i=0}^{q-k} \beta_i \beta_{i+k}}{\sum_{i=0}^{q} {\beta_i}^2} & k = 1, \ldots, q \\ 0 & k > q \\ \rho(-k) & k < 0 \end{cases}$$

DAFTAR PUSTAKA

Aswi and Sukarna (2006) 'Analisis Deret Waktu Analisis Deret Waktu', (January), p. 303.

Chatfield, C. (1990) 'The Analysis of Time Series.', *Biometrics*, 46(2), p. 550. Available at: https://doi.org/10.2307/2531477.

Hornik, K. (2009) *Use R ! Robert Gentleman*, *Media*. Available at: http://www.springerlink.com/index/10.1007/978-0-387-88698-5.

Robinson, G.M. (2009) *Time Series Analysis*, *International Encyclopedia of Human Geography*. Available at: https://doi.org/10.1016/B978-008044910-4.00546-0.

BIODATA PENULIS

Hetty Elfina, S.Pd., M.Pd
Dosen Program Studi Teknik Mesin
Fakultas Teknik dan Ilmu Komputer
Universitas Pembinaan Masyarakat Indonesia
Medan

Penulis lahir di kota Sibolga pada tanggal 07 Januari 1990. Penulis merupakan dosen tetap pada Program Studi Teknik Mesin, Fakultas Teknik dan Ilmu Komputer, Universitas Pembinaan Masyarakat Indonesia. Penulis menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika di Universitas Negeri Medan tahun 2013 dan melanjutkan S2 pada tahun yang sama di Jurusan Pendidikan Matematika di Universitas Negeri Medan dan lulus tahun 2015. Penulis telah menulis 4 buku, yaitu Kalkulus, Statistika Pendidikan, Teori Bilangan, dan Metodologi Penelitian.

BIODATA PENULIS

Yenny Anggreini Sarumaha, S.Pd., M.Sc.
Dosen Program Studi Pendidikan Matematika
Fakultas Keguruan dan Ilmu Pendidikan

Penulis lahir di Padang tanggal 22 Januari 1988. Penulis adalah dosen tetap pada Program Studi Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan, Universitas Cokroaminoto Yogyakarta. Penulis menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang dan melanjutkan S2 pada Jurusan Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sriwijaya.

BIODATA PENULIS

Suryanti, S.Pd., M.Pd.
Dosen Program Studi Pendidikan Matematika
Fakultas Keguruan dan Ilmu Pendidikan
Universitas Pancasakti

Penulis lahir di Ujung Pandang tanggal 12 Juli 1987. Penulis adalah dosen tetap pada Program Studi Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan, Universitas Pancasakti. Menyelesaikan pendidikan S1 pada Jurusan Pendidikan Matematika dan melanjutkan S2 pada Jurusan Matematika.

BIODATA PENULIS

Novrianti, S.Si., M.Si., Ph.D.
Dosen Program Studi Teknik Mesin
Sekolah Tinggi Teknologi Nasional (STITEKNAS) Jambi

Penulis lahir di Jambi, 27 November 1989. Penulis merupakan dosen LLDIKTI X yang ditempatkan di STITEKNAS Jambi sejak April 2015. Menyelesaikan pendidikan S1 dan S2 di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas. Kemudian pada tahun 2018 melanjutkan pendidikan S3 di Gifu *University*, Jepang. Kelompok keahlian penulis adalah Matematika Terapan, dan sedang melanjutkan penelitian dalam berbagai bidang terkait.

Disamping sebagai seorang Dosen, penulis juga merupakan guru lepas (*freelance tutor*) pada lembaga bimbingan belajar Nurul Fikri di Padang (2011 – 2015), dan melanjutkan di Nurul Fikri Jambi (2015 – sekarang).

BIODATA PENULIS

Dr. Joni Wilson Sitopu, M.Pd.
Dosen Universitas Simalungun

Dr. Joni Wilson Sitopu, M.Pd, lahir di Medan, Pendidikan S1 Prodi Matematika FMIPA USU Medan, S2 Prodi Pendidikan Matematika UNIMED Medan dan S3 Prodi Ilmu Matematika FMIPA USU Medan. Dosen di FKIP, SPS Prodi S2 PW, IPS dan Manajemen, serta Kepala Lembaga Akreditasi Internal (LAI) Universitas Simalungun. Penulis aktif mengikuti sebagai pemakalah pada Seminar Nasional dan internasional, aktif mengikuti Seminar Nasional dan internasional dan aktif menulis Buku, Jurnal OJS, Sinta, dan Scopus.